Self-Awareness

Its Nature and Development

Edited by
Michel Ferrari
Robert J. Sternberg

Self-Awareness

ITS NATURE AND DEVELOPMENT

Edited by

MICHEL FERRARI
ROBERT J. STERNBERG

THE GUILFORD PRESS
New York London

A Division of Guilford Publications, Inc.
72 Spring Street, New York, NY 10012
http://www.guilford.com

Printed in the United States of America

This book is printed on acid-free paper.

Last digit is print number: 9 8 7 6 5 4 3 2 1

Library of Congress Cataloging-in-Publication Data

Self-awareness : its nature and development / edited by Michel
Ferrari, Robert J. Sternberg.
p. cm.
Includes bibliographical references and index.
ISBN 1-57230-317-4
1. Self-perception. I. Ferrari, M. D. II. Sternberg, Robert J.
BF697.5.S43.S434 1998
126—dc21 97-52300
CIP

Contents

SELF-AWARENESS

PART ONE

DEFINITIONS AND UNDERPINNINGS OF SELF-AWARENESS

CHAPTER ONE

The Mind and Education

✧

JOHN R. SEARLE

I have been asked to discuss my work in the philosophy of mind and cognitive science with a view to its relevance to education. I must say at the outset that I am not sure how much relevance my views have for the actual practice of educating young minds. However, if I am right in my criticisms of certain prevailing views about the mind, we may at least avoid making certain mistakes in that practice, mistakes that derive from holding false philosophical and psychological theories about the mind. In this talk, I am going to summarize some of my more controversial views and then make some speculations about their possible relevance to education. For the sake of brevity and clarity, I will state my views as a set of theses.

FIVE THESES ABOUT THE MIND

Thesis 1. The essence of the mental is consciousness. Descartes almost ruined this sentence forever when he said it (or at least its Latin and French versions) to mean something else. What Descartes meant was that there are two kinds of substances in the world, mental and physical, and each has an Aristotelian essence, an essential trait that makes the substance the kind of substance it is. The essential trait of mental substance is that it is always and forever conscious. It is impossible for a mind to exist without being conscious.

In contrast to Descartes, I construe mental states as either conscious or unconscious, and the unconscious ones are understandable as mental because they are the sorts of things that are at least in principle accessible to consciousness. So when I say that the essence of the mental is consciousness, what I mean by that is that our ordinary notion of an unconscious

mental state is understandable as mental only in terms of its accessibility to consciousness. *All mental states are either actually or potentially conscious*. A person might not be able to bring his or her unconscious belief to consciousness because of some impediment. These impediments range all the way from brain lesions, to the tip-of-the-tongue phenomenon, to Freudian style repression, but we have no notion of a belief state except in terms of conscious mental states. Our paradigms of mental states are conscious states such as thinking that Bill Clinton is President of the United States, or trying to solve a mathematical problem, or worrying about the size of the deficit, or wondering whether our income tax will be audited this year.

When I say consciousness is the essential trait of the mind, I am presenting that thesis in contrast to a number of other views that say things such as that the essential thing about the mind is that it is behavior, or functional adaption, or perhaps worst of all, that the mind is just a computer program. I will come back to this view later.

It is usually said that "consciousness" is especially difficult to define. I don't really think that "consciousness" is hard to define, if what we are seeking is a commonsense definition rather than the sort of analytic definitions that typically come at the end of a scientific analysis. If we just want a commonsense definition that will identify the phenomenon in question, then we can say that consciousness consists in subjective states of awareness or sentience that begin when one wakes from a dreamless sleep and continue until one goes to sleep again, or falls into a coma, or dies, or otherwise becomes, as one would say, "unconscious."

So construed, consciousness is a biological phenomenon like any other. On my view, we need to stop thinking of the mind as something mysterious and ethereal, and recognize it as an ordinary biological phenomenon, on all fours with digestion, mytosis, meiosis, or the secretion of bile.

Thesis 2. Consciousness is caused by processes in the brain and is a higher-level feature of the brain system. This thesis is an answer to the traditional "mind–body problem." The problem of the relation of consciousness to physical processes, such as brain processes, has obsessed philosophers for the past 350 years. I think, in fact, that if this problem did not have such a particularly sordid history, we would not find it very difficult. I believe that the mind–body problem has a rather simple solution given what we know about how the world actually works.

Mental states, and by that I mean all of our mental states—thoughts, pains, tickles, itches, feelings, worries, anxieties, angst, love, hate, and so on, all of our mental states—are caused by lower-level neuronal processes in the brain and are themselves higher-level features of the brain. Formally speaking, these types of relationships are very common in nature. For ex-

ample, the liquidity of the water in this glass or the solidity of the wood in this table are both caused by the behavior of lower-level elements of the system, in these cases, by the behavior of the molecules. But at the same time, neither liquidity nor solidity are separate entities. They are, rather, the states that the systems composed of the molecules happen to be in. It is similarly the case with the mind and the brain. Consciousness is caused by the behavior of the lower-level elements, neurons and synapses, but it is not thereby a separate substance or entity. Consciousness is just the state the brain happens to be in at certain times, those times when the neurons and synapses are behaving appropriately. The solidity of the table and the liquidity of the water are not separate things distinct from the molecules that make up the water or the table. Analogously, we should not think of consciousness as a separate thing, distinct from the neuronal systems that make up the brain. Rather, consciousness and other mental phenomena are just states that the brain happens to be in at any particular point, in the same sense that the water in this glass is in a liquid state at present, or the wood in this wooden table is in a solid state.

There are immense empirical mysteries about how the brain works in fact, but I believe there is no deep mystery about the overall relationship between the mind and the brain.

Thesis 3. If we construe the mind as a biological phenomenon, the most important functional feature of the mind is intentionality. Intentionality is simply that feature of the mind by which it is capable of representing objects and states of affairs in the world other than itself. The word "intentionality" is an unfortunate word for English speakers because it seems to imply some special connection with intending, in the sense in which I intend to go to the movies, or intend to get a haircut sometime in the next week. But "intentionality" in the technical sense does not imply intending. Intending is just one kind of intentionality along with perceiving, believing, desiring, hoping, fearing, wondering whether, wishing that, falling in love, or being disappointed. In short, any intentional state that is *about* something other than itself is said to be intentional. In this sense, pains, tickles, itches are not intentional, nor are undirected feelings of elation and anxiety. In this respect, they are unlike intentional states such as believing, desiring, hoping, fearing, and so on.

Intentional states and processes function to relate us to our environment by way of representations that act on the environment and are in turn acted on by the environment. Typical intentional phenomena that direct us to act on the environment are desires, intentions, and intentional actions. Typical intentional states that are caused by the environment and represent the environment are beliefs, memories, and perceptual experiences.

It should be obvious from all this that there is a close connection be-

tween intentionality and consciousness, but they are not coextensive. Most conscious states are intentional, but not all of them. If I have an undirected feeling of anxiety, or simply feel an unpleasant sensation in my elbow, these conscious states are not intentional. Furthermore, I have many intentional states that at any given point in my life are not conscious. The man who is sound asleep, for example, can literally be said to believe that Bill Clinton is President of the United States. Thus, not all intentional states are conscious; and not all conscious states are intentional. However, this point needs to be reemphasized: Any unconscious intentional state is understood as an intentional mental state only in terms of its accessibility to consciousness—only in terms of its potentially being conscious.

I guess it is obvious to everyone here that education is for the most part about intentionality. Our aim is to produce knowledge, and knowledge consists in having beliefs that meet certain conditions; that is, they must be true and they must be justified in certain ways.

Thesis 4. The Background: All intentional states, conscious or unconscious, only function given a Background of abilities, capacities, tendencies, dispositions, and general know-how that do not consist in more intentional states. All of our mental life is conditioned by a set of presuppositions that are not part of our conscious awareness. It is a mistake to think of these as a set of unconscious beliefs that we could bring to awareness at will. Rather, it seems to me best to think of them as a set of abilities that we have for coping with the environment and with other people. For the most part, the Background is not a matter of "knowing that" but of "knowing how."

Any educator has the following sort of experience: Some things you say are systematically misunderstood, because the people to whom you are speaking have a different Background. They just bring a different mode of sensibility to bear on the problem.

The most obvious proof of the thesis of the Background is to consider ordinary sentences. If you consider sentences containing the word, "cut," for example, as in "Bill cut the grass," or "Sally cut the cake," we understand the word "cut" as having the same meaning, yet it has a different application. What counts as cutting the grass is quite different from what counts as cutting the cake, not because the meaning of the word is different, but rather because we have a set of abilities and practices derived from our biology and our culture, and these are what I am calling "Background."

Background abilities are not simply cultural; many of them are biological. If you read in a book the sentence "Bill ate the hamburger," you know that he ate it by putting it in his mouth and not by, let's say, stuffing it in his ear or ingesting it through capillary action in his skin. All of that is

pretheoretical, preintentional Background. Much education is about shaping the Background. I will return to this point later.

Thesis 5. It is a mistake to suppose that the mind is a computer program. Perhaps the biggest single mistake we make about the mind today is to suppose that the brain is a digital computer and the mind is simply the program of that computer. This mistake is the latest in a long series of reductionist attempts to get rid of consciousness and intentionality and replace them with something else. If you are as old as I am, you will remember certain earlier versions of this reductionist tendencies, such as behaviorism.

Behaviorism says that the essential thing about the mind is publicly observable behavior. I think that view is obviously false, as we all know from our own experiences. There is no essential connection between our mental states and our behavior, though of course, our mental life evolved to enable us to cope better with our environment, and that coping is done through our behavior. But you can have mental states without exhibiting behavior, and you could, at least in principle, build systems that would exhibit the behavior but not have the mental states. I think behaviorism is pretty much dead, but occasionally there are efforts to revive it.

The more interesting and currently more popular view is that mental states are essentially computational, that a computer, if it had the right program and the right inputs and outputs, would have to have the same mental states we have, and indeed, that our mental states consist entirely in computer programs that are capable of mediating the relations between input and output. I will not spend much time on this view, because I believe it can be refuted decisively in a couple of minutes. Briefly, here is the refutation:

What is a computer program? A computer program is a set of rules for formal symbol manipulation. These are usually defined in terms of zeros and ones. A computer manipulates zeros and ones. In fact, on Turing's original definition of a Turing machine, a Turing machine can perform exactly four operations: It can move its head one square to the left, or move it one square to the right; it can erase a zero and print a one, or erase a one and print a zero. That is all a computer does. It does those things very rapidly—several million operations per second, and it does them in accordance with rules of the form: Under condition C, perform act A; if C, then A. It is one of the great intellectual achievements of the 20th century that we are able to do so much with such a simple apparatus, but equally it is one of the great intellectual mistakes of the latter part of the 20th century to suppose that that is what is going on in our minds. Now, we know that the mind could not be just an implemented computer program. This view could not be right for the following simple reason: The operation of the

computer is defined purely in terms of meaningless symbols such as the zeros and ones. But what goes on in our minds is more than the manipulation of meaningless symbols; rather, our mental states actually have a mental content. They have something more than just zeros and ones. If I wish I had a cold beer or am planning a trip to Seattle, I actually have thoughts about Seattle or beer, and not just about zeros or ones.

I gave a simple proof of this over 15 years ago. If it were true that a computer program were sufficient for mental states, then just imagine that you are carrying out the computer program for some mental state that you do not in fact have. I do not understand Chinese, so I imagine myself carrying out a computer program that simulates the understanding of Chinese. I imagine I receive a series of inputs in Chinese. I look up in the rule book, that is to say, the program, I am supposed to manipulate these symbols, and I give a set of outputs in Chinese—I give back Chinese symbols to those people who have been giving me Chinese symbols. Unknown to me, it might be the case that the program is enabling me to give the correct Chinese answers to the Chinese questions that are presented to me, but I know nothing of that. I am simply doing what a computer does, manipulating meaningless symbols. Now, the point of this whole parable can be stated quite simply: If I do not understand Chinese on the basis of implementing a digital computer program for the understanding of Chinese, then neither does any other digital computer solely on that basis, because that is all a computer has. It has a set of rules for manipulating formal symbols. There is a much simpler summary of this argument—it can be stated in two slogans of four words each. Syntax is not semantics; and simulation is not duplication. The syntax of the implemented computer program is not sufficient for the semantics of ordinary human minds; and simulating a process is not the same as duplicating it.

Intellectual historians will regard it as one of the most amazing features of the 20th century that the computational theory of the mind, though it is preposterous on its face, survived as long as it did.

I think, incidentally, that these two mistakes are important for education in that behaviorist theories of education and computational theories of education have both had deleterious consequences.

The upshot of my discussion so far can be stated quite simply: What is going on inside our skulls consists exactly of two sorts of things. First, there are brain processes with all their enormous complexity, and these have many levels of description ranging all the way from quarks and muons up through synapses, neurons, and cell assemblies to large neuronal circuits and even gross molar organs such as the hypocampus and the thalamus. Second, there is consciousness with all of its color and variety. But that is it. There is not, in addition, a whole lot of mental models, or the language of thought, or computer programs; there is just a brain, and

sometimes it is conscious. Consciousness functions against a biological and cultural Background of nonconscious capacities that are realized in the brain.

CONSEQUENCES FOR EDUCATION

What are the practical applications to education, if any, of what I just told you? Well, as I said at the beginning, I am not at all sure that there are any practical applications, and my own instinct about education is to think that if we have been doing it more or less successfully, both systematically and unsystematically, for the past 2,500 years, we must have worked out some effective strategies for doing it, and we ought to take our own knowhow as superior to anything some theoretician might tell us. But, anyway, there are perhaps a few lessons, so here goes.

One of the most surprising things about educational theories in the 20th century is how certain common expectations were disappointed. When radio was first invented, it was assumed it would revolutionize education. We had similar expectations for the movies and then for television. Instead of having ordinary fellows trying to educate you, you could be educated by the world's leading superstars, because you would be able to see them on television and nowadays see them on videotape. Education, in short, was going to be revolutionized by the electronic media. Why did it not work? Every study that I have seen shows that the retention rate from material acquired while watching television is vastly lower than material acquired while listening to a live lecture in the classroom. Why? What is so magical about actual classrooms inhabited by living people? Well, no doubt, the answer is complex, but there is a rather simple aspect that I want to call to your attention. If you have to get up, put on your overcoat, go to the classroom, sit in the classroom, look at an actual human being and take notes, you have a much greater intensity of commitment and involvement than you do if you are sitting at home in front of a television set. Live education, in short, involves a degree of *conscious* involvement and commitment that is not present in the passive reception of visual and auditory images. Furthermore, not only does it take a greater commitment on the part of a student, but also there is actually a self-awareness of the student's interaction with the teacher. When I am a student in a classroom, the teacher is not only there to be an object of my scrutiny, but also I am aware that the teacher is looking at me.

It may be that the most educationally effective of the electronic media will be those interactive forms of electronic technologies that involve active participation by the student.

The points I am making here are not especially confined to education.

I think that you are aware that you are much more affected by a drama if you see it live on stage or even if you have to get up and go to a movie house to see it, than if you see the same thing at home on television. At home, when you see it on television, you are busy carrying on a conversation with your spouse, your mind is on other things, and at various points you will get up and go to the refrigerator to get a beer, and so on. So the first lesson of what I have been saying is not very surprising. *The key to successful education is total commitment of the consciousness.* To the extent that you succeed in doing that, you will have an impact on the brain of the recipient; to the extent that you fail, you will not. So let's turn and talk about brains a bit.

The most important advice anyone can give about the relevance of brain research to education is not to take this research too seriously. We just do not know enough about how the brain works to use it as guide to education. However, as far as we know, the brain is losing neurons throughout most of our lives. This is not necessarily a bad thing. As far as we now know anything about how it works, it seems best to think of the brain more as a selectional rather than as an instructional mechanism. What I mean by that is that, in general, learning does not take place through reinforcing existing circuitry in an undifferentiated fashion. Rather, especially among the young, learning is often a matter of selecting certain circuitry and allowing other circuits to become inactive or simply to die. It is very important to reinforce the right circuitry at the right age. This is obvious to us in the case of physical skills. The child who does not learn ice-skating, musicianship, or skiing at an early age will never become a world-class performer in these activities. If you start as an adult, it is too late. Any classroom teacher knows that this is most obvious in the case of the pronunciation of foreign languages. Children exposed to a foreign language prior to adolescence will be able to speak that language without any trace of an accent. The older you get, the harder it is. This is not because of middle-aged laziness or lack of attention; it is simply that the circuitry is no longer able to accommodate the perception and the production of the appropriate phonetic shapes. There are some wonderful experiments that show this in the case of the contrast between Japanese children and Japanese adults, where the perception of the distinction between the English "r" and the "l" is concerned. The adults simply cannot hear a difference. The children have no problem hearing the difference. And just as they have no problem hearing the difference, they have no problem producing the difference, in contrast to their elders.

The moral of this is obvious: Teach the child as much as you can as early as you can. One of the tragedies of American education is the idea that bright children should not proceed faster than the rest of the class. The aim of education is to develop each child to the maximum of his or her po-

tential, and though that is an ideal we will seldom realize in practice, all the same, we should not adhere to policies that go dead against that ideal. In his autobiography, Santayana says that the only really well-educated people he knows were educated privately at home by tutors and not in schools, neither public nor private schools. One of the advantages of private tutoring is that the student can go as fast and as far as he or she is able and as fast and as far as the tutor can teach.

Now I want to say a bit about the Background. Any teacher knows that much of what is important in what you teach is not the explicit content of the information that the child learns, but the attitude, the cast of mind, the ways of responding to knowledge and information, and indeed the whole mode of sensibility that the successful teacher can only convey by exemplifying it and by exposing the student to works that exemplify it. The French have a saying that translates as: You forget the information you learned, but the education remains. I take it that the point of that remark is that education is not just a matter of acquiring information, but the effect of acquiring the information is to change pretty much every aspect of your life, in ways that are typically not represented in the information you acquire.

Please do not misunderstand the point I am making. I am not denigrating the educational value of acquiring large amounts of sheer information. You will not learn the history of the United States without memorizing a great many names and dates, and you will not learn how to speak French without memorizing the meaning of a lot of French words. If anything, there is not enough memorization of large amounts of information going on in the schools at present. The point is that the information gives you a much deeper understanding of American history, and learning French—or any other foreign language—well, will change your whole way of responding to language and experience. But that understanding and those ways of responding are not additional items of information. They are matters of what I call the Background.

The real disaster produced by television is not that children waste so many hours watching it, but that it alters their Background. It alters their whole mode sensibility as they get a conception of how adults are normally supposed to behave in responding to each other. I do not need to tell this audience that the model of human behavior presented in most popular television programs is violent, brutish, and stupid.

CHAPTER TWO

Consciousness and Self-Awareness

THOMAS NATSOULAS

Before embarking on a scientific discussion of the topic of consciousness, a psychologist does well to consider the several ordinary meanings of the word "consciousness" (Dewey, 1906; Natsoulas, 1983; cf. Husserl, 1900/1970, pp. 535–536). Attention to these nontechnical concepts should reduce the risk of difficult-to-avoid referential displacements, namely, from consciousness to phenomena that are more amenable to investigation by means of the methodologies and theoretical approaches currently in favor among psychologists (cf. Koch, 1980, p. 45; Miller, 1990, p. 7; Wittgenstein, 1947/1980, p. 180e).

One might suppose that in the years since the beginning of the cognitive revolution, the disciplinary factors largely responsible for such referential displacements have been mitigated in major part. However, note this recent astute observation:

> But the basic issue, I think, is never really addressed. . . . I have the same feeling about consciousness because whenever consciousness comes up—and it usually comes up in meetings of philosophers and psychologists—it's the same kind of phenomenon, where people are prepared to give a theory of consciousness but what they give you is a theory of something else. And the relationship between consciousness and that other thing is never made explicit. (Pylyshyn, 1990, p. 201)

I too have observed similar intellectual behavior among present-day psychologists: a continued, albeit anachronistic, championing of operational definitions as providing the true path for achieving the conceptual ad-

vances that all of us in this area of psychological science seek; and an advocacy of empirical research as itself constituting a way by which to bootstrap ourselves somehow—how exactly is not made clear—to a more faithful and complete conceptual grasp of the phenomena that fall commonsensically under the name "consciousness."

Quite germane to the purposes of the present chapter is the ordinary meaning of "consciousness" that is listed second in the *Oxford English Dictionary* (1989; henceforth, the *OED*). The *OED* describes as follows the kind of psychological state picked out when one properly uses "consciousness" in its second sense. At one time, the phrase "consciousness to oneself" was also used to make the same reference: "Internal knowledge or conviction; knowledge as to which one has the testimony within oneself; esp. of one's own innocence, guilt, deficiencies, etc." As the present chapter proceeds, it will be seen repeatedly that consciousness in the second *OED* sense involves reference to oneself as such in every case; and self-awareness is a necessary part of any instance of this kind of consciousness—as is not the case for every other kind.

The preceding dictionary definition expresses the oldest ordinary sense that continues to find expression with the word "consciousness." Use of the word in a different early sense—which the *OED* defines simply as "joint or mutual knowledge"—is now obsolete (Natsoulas, 1991a).[1] Both of these senses date from the 17th century, which is when "consciousness" first came into the English language. Later, four additional meanings of the word were added. I soon consider two of these four, together with the second sense. The two remaining senses of "consciousness" have no role to play in the present chapter. However, here is how the *OED* expresses them: "The totality of impressions, thoughts and feelings, which make up a person's conscious being." "The state of being conscious, regarded as the normal condition of healthy waking life." For detailed discussion of all six *OED* concepts, one may want to consult two previous publications of mine (Natsoulas, 1983, 1986–1987).

The late entry of "consciousness" into the English language does not signify that, prior to that time, ancestors of ours who knew only English could not refer to what we can refer to with the word. They were in firm possession of, among other linguistic means, the cognate word "conscience." From the 14th century onward, "conscience" had to serve, among its other functions, some of the functions that "consciousness" was later to assume. Indeed, one might wonder: Given the availability of "conscience" to users of English, why should there have emerged a word "consciousness" at all? This interesting question lies beyond the scope of the present chapter. However, one might begin to answer it along the following lines: After its poetical introduction into the English language in imitation of Latin authors (see Lewis, 1967), "consciousness" caught on and its

number of uses expanded because "conscience" could thereby become more specialized and less ambiguous (cf. Engelberg, 1972).

What more exactly is consciousness in the *OED*'s second sense? What is the character of the kind of psychological state to which the dictionary is adverting with its second definition? The definition leaves one wanting to know more, especially if one is a psychologist who is seeking to address consciousness itself, rather than something else that may be related to it. The *OED*'s definition fails to provide adequate information to enable the discernment of instances of the indicated kind of consciousness. If one attempts to decipher the second meaning on the basis of the explicit definition alone, questions such as the following come to mind, and they are not answered:

1. Does "having the testimony within oneself" amount simply to being in a position to give testimony about something that one has witnessed? What are the limitations, if any, on the possible objects of consciousness in the *OED*'s second sense? Does one possess consciousness in this sense of everything to which one can attest on a firsthand basis?

2. Can a happening or state of affairs external to one, even independent of one, be an object of one's consciousness in the second sense? Can a crime, for example, be an object of one's consciousness in the second sense whether or not one has participated in it, so long as one actually witnessed it? If it can, is the crime an object of one's consciousness only whenever it comes back to mind? Or does it suffice that, possessing the knowledge, one is in a position to remember the crime?

3. How is consciousness in the second sense "internal"? Is there a counterpart phenomenon that qualifies as external to oneself yet is not "internal" to someone else? Does the *OED* describe consciousness in the second sense as internal because the object of this consciousness, whenever one instantiates the latter, is something that no one else can apprehend and give testimony about on a firsthand basis, something that is private to oneself in principle? Is the knowing, or awareness, that defines the *OED*'s second kind of consciousness in some way directed round upon the mind itself?

4. Or is consciousness in the second sense internal to oneself merely because, in the particular instance, one happens to be or to have been the only witness to the particular matter that one now knows or believes? Is it simply that the knowledge or conviction to which the *OED*'s definition refers is personal, at most secret, rather than held in common?

It is such questions regarding consciousness in the *OED*'s second sense that the present chapter seeks to address. I shall call the psychological referent of the second sense "consciousness$_2$," which has been my term for it in previous publications as well (e.g., Natsoulas, 1983, 1986–1987, 1991b). With the help of numerical subscripts, I have been distinguishing

the six kinds of consciousness that the *OED* describes in its series of entries under the word. Comparisons between kinds of consciousness will again be useful as I proceed with making explicit what consciousness$_2$ is. Quite soon, the distinction between consciousness$_2$ and consciousness$_3$ enters the picture that I am drawing. Later, I also show how consciousness$_4$ is involved, always, in consciousness$_2$.

Moreover, as I proceed, I bring out the relation between each of three kinds of consciousness and self-awareness. In summary:

1. A case of consciousness$_3$ may or may not be a case of self-awareness; consciousness$_3$ can take place although one is not self-aware at the time.
2. Consciousness$_4$ is a form of self-awareness; it is a way in which one is occurrently aware of certain occurrent parts of oneself, typically, of them as being such.
3. Every instance of consciousness$_2$ involves, as a feature of itself, self-awareness in several ways.

My comments on the difference between consciousness$_2$ and consciousness$_3$, and on the relation between consciousness$_2$ and consciousness$_4$, serve a further purpose, in addition to helping to spell out what consciousness$_2$ is. They show that I am not engaged in advocacy for a particular sense of "consciousness." I am not suggesting that, for purposes of their research, psychologists should adopt the concept of consciousness$_2$ to the exclusion of all other technical or nontechnical concepts of consciousness. Nor am I engaged in the process—which has become traditional in psychology—of constructing a technical concept to be expressed by means of a familiar, ordinary word.

So too, the reader should not draw from my bringing out, perforce, the ethical dimension of consciousness$_2$, that I consider the *OED*'s other referents of "consciousness" to be less worthy as topics of psychological investigation. Actually, I believe even consciousness$_1$ is an intriguing and worthwhile topic, notwithstanding that the word "consciousness" is no longer used to speak of the interpersonal cognitive relation that consciousness$_1$ is (Natsoulas, 1991a).[2]

Thus, I simply hold that how we commonly use ordinary language reflects differentiations of proven utility over a considerable period of time. These differentiations may well be more refined than the distinctions that psychologists, at this early point in the history of their science, are finding useful or attractive. It will become evident as I proceed that the concept of consciousness$_2$ has reference to phenomena that should be of special interest to those psychologists, and others, who are currently rediscovering consciousness and want to advance their understanding of self-awareness.

THE CHARACTER OF CONSCIOUSNESS$_2$

More Than Being Conscious$_3$

Whenever someone is, as the *OED* puts it, "mentally conscious or aware *of* anything," that is, whenever there takes place an occurrent awareness of anything at all, consciousness$_3$ is therein instantiated by definition (Natsoulas, 1992b). Regarding consciousness$_3$, John Dewey (1906) accurately wrote: "'Conscious' means aware: 'consciousness' the state of being aware. This is a wide, colourless use; there is no discrimination as to contents, as to what there is awareness of,—whether mental or physical, personal or impersonal, etc." (p. 40). A case of consciousness$_3$ may fail to qualify as a case of each of the five other kinds of consciousness that are described under "consciousness" in the *OED*.

Note also that a psychological phenomenon does not need to occur in any special way for it to qualify as a case of consciousness$_3$. Its being a case of consciousness$_3$ does not depend on how an occurrent "awareness-of" comes to take place. Paranormal and strange futuristic means would be all right, if only they existed. In contrast, consciousness$_2$ requires witnessing or having witnessed evidence bearing on what, along certain lines, one is now occurrently aware of regarding oneself. Consciousness$_2$ always includes reference to oneself, and involves self-awareness in several ways, as I explain further on. An instance of consciousness$_3$, however, may be an instance of self-awareness, but often it is not. It is often an awareness of something else, which can be something completely unrelated to oneself. As the *OED* states, instances of consciousness$_3$ can be "of anything."

Mead

I must acknowledge that, according to George Herbert Mead's (1934) influential and enlightening conception of mind and self, every instance of consciousness$_3$ that one produces, however simple that instance may be, includes a reference to oneself (see Natsoulas, 1985). Whenever one is occurrently aware of anything at all, one is therein indicating it (and, often, characteristics of it) to oneself as though one were another person. Although Mead does introduce, in addition, a kind of experience entirely lacking the feature of self-reference, such experience is claimed by Mead not to possess any cognitive aspect. Assuming this noncognitive kind of experience does in fact take place,[3] it would not fall under the category of consciousness$_3$. Mead's proposed involvement of self-awareness in every case of consciousness$_3$ implies that all cognitions are founded on other cognitions, which in turn require self-awareness, and so on (see Natsoulas, 1991b). In any case, Mead's understanding of consciousness$_3$ is not the

common one—which amounts simply to being aware of anything at all, and does not necessarily include oneself as the object of awareness.[4]

Being Conscious to Oneself

In its entry for consciousness$_2$, the *OED* calls our attention to a corresponding entry under the adjective "conscious." We are told to see "*conscious to oneself* (*of* anything, *that*, etc.)." The parallel meaning turns out to be very much the same. To be conscious to oneself of something in particular is, with regard to the latter: "Having the witness of one's own judgement or feelings, having the witness within oneself, knowing within oneself, inwardly sensible or aware." Whenever one has consciousness of . . . or that . . . in the second *OED* sense, one is conscious to oneself of . . . or that . . . , or, as I shall also say, conscious$_2$ of . . . or that. . . . Both of the latter, "conscious to oneself" and "conscious$_2$," will serve here as adjectival descriptors of someone who is, was, or will be instantiating consciousness$_2$.

Is there an external kind of knowing which, in the previous definition of "conscious to oneself," the *OED* is implicitly contrasting with the internal kind that is consciousness$_2$? The previous definition of "conscious to oneself" implies that the witnessing could be external to oneself; the testimony could be within someone else, instead of being within oneself, although it is not external to oneself in those cases picked out by the proper use of "conscious to oneself." Indeed, one may partake of the witness of someone else's judgments or feelings—if he or she reports about them to one. However, so to partake is not equivalent to being conscious to oneself of what the other person has witnessed. In the very process of being conscious to oneself (i.e., in being conscious$_2$ of something), the witness cannot be someone else; it is oneself alone—one knows within oneself.

Thus, the previous definition of "conscious to oneself" suggests not only (1) the exclusivity of the personal knowing involved in consciousness$_2$, but also (2) to be conscious to oneself of . . . , or that . . . , is to know something that one may want to keep to oneself, something about which one would keep one's own counsel. This is consistent with the original source of the words "consciousness" and "conscious," as well as the word "conscience." Near the start of its entries for "conscience," the *OED* helpfully identifies the Latin source as follows:

> L. *conscientia* privity of knowledge (with another), knowledge within oneself, consciousness, conscience, f. *conscient-* [present participle] of *conscire*, f. *con-* together + *scire* to know; thus *conscire alii* to know along with another, to be privy with another to a matter, thence, *conscire sibii* to know with oneself only, to know within one's own mind.

"Consciring"

C. S. Lewis (1967) coined the word "consciring" to refer to knowing together, or sharing the knowledge of something with someone, that is, to the state of affairs that the Latin *conscire* was primarily employed to indicate. Lewis explained that his new English word would have a use only in cases of shared knowledge between two or a few people in a secret together with respect to that knowledge. Also, just as the Latin allowed one to be one's own *conscius* or *conscia*, that is, to be the one who knows along with oneself, so Lewis generalized the meaning of his word "consciring" to the reflexive, or intrapersonal, case. Lewis justified his latter move by means of a number of quotations from literature. And, drawing upon common knowledge, he eloquently stated,

> Man might be defined as a reflexive animal. A person cannot help thinking and speaking of himself as, and even feeling himself to be (for certain purposes), two people, one of whom can act upon and observe the other. Thus he pities, loves, admires, hates, despises, rebukes, comforts, examines, masters or is mastered by, "himself." Above all he can be to himself in the relation that I have called consciring. He is privy to his own acts, is his own *conscius* or accomplice. And of course this shadowy inner accomplice has all the same properties as an external one; he too is a witness against you, a potential blackmailer, one who inflicts shame and fear. (p. 187)

What more specifically is involved in the makeup of the psychological state, relation, or activity that is Lewis's consciring with oneself? Clearly from the aforementioned statement, consciring is not to be equated with self-observation, self-awareness, or self-referential behavior, nor with emotion, thought, or action that is directed on oneself. Therefore, are intrapersonal consciring and the *OED*'s consciousness$_2$ the same? In a deeper sense, what is it for a person to be his own *conscius* or her own *conscia*? What is it for someone to have the testimony or witness within himself or herself? Having learned to distinguish particular cases of consciousness$_2$ from similar though different psychological phenomena, we will still want to know: In what does consciousness$_2$ consist? What more exactly is going on when one is conscious$_2$?

Relation to Consciousness qua Guilty Awareness of Wrongdoing

The *OED*'s definitions for "conscious to oneself" (quoted at the beginning of the preceding subsection) and "consciousness" in its second sense (quoted near the beginning of the present chapter) are, I believe, too broadly

drawn. Surely, the compilers of the *OED* did not intend to include under the second concept so large a variety of psychological phenomena. The two definitions—on their own (i.e., ignoring etymology and examples of usage)—allow for too many possible kinds of items that one might attest to, or be in a position to attest to, and thereby be conscious_2 of . . . or that. . . .

There are at least two grounds for thinking that the definitions are too broad for capturing the true second sense of consciousness. I develop the first ground in this subsection, the second in the subsection right after the next one. The first ground pertains to a further meaning of "conscious," which the *OED*'s compilers considered to be very closely related to the meaning of "conscious to oneself."[5] The second ground is provided by the *OED*'s totality of quotations taken from literature for the purpose of illustrating either the use of "consciousness" in its second sense or the corresponding uses of "conscious to oneself" and "conscious."

The entry under "conscious" immediately after the one for "conscious to oneself" is comprised of two subentries. The first subentry attributes to "conscious" on its own the same meaning as for "conscious to oneself." The second subentry gives the following, distinct definition of "conscious": "Having guilty knowledge (*of* anything); *absol.* inwardly sensible of wrong-doing, guilty." This use of "conscious" is no longer current; yet the connection between being conscious_2 and being conscious in this old sense is no less pertinent.

Guilty awareness of wrongdoing is not a feature of every case of consciousness_2. However, it is such a feature often enough for "conscious" to have meant, in certain contexts, one's feeling guilty concerning something about which, being conscious_2 of it, one is in a position to attest. Compare the *OED*'s several illustrative quotations for the old sense of "conscious"; here are two of them: (1) "She being conscious, did of her own accord . . . make confession of her wickedness;" (2) "*conscious*, inwardly guilty, privy to ones self of any fault or errour." The reference of "conscious" in this sense was not simply to the fact that someone instantiated consciousness_2. Feelings of guilt played a necessary role in this further way in which an individual could be conscious.

Both quotations clearly implicate consciousness_2. It was not by hearsay that "she" knew of her wickedness, faults, or errors; rather, she had within herself the witness to the personal facts about which she felt guilty. However, to have guilty awareness of wrongdoing is more than to be conscious_2 that one has done something wrong. Guilty feelings are not essential to being conscious_2. This can be seen from the *OED*'s relevant definitions, and especially from the examples of usage accompanying them.

Note too that any case of guilty awareness of wrongdoing is also a case of being conscious_3 of something. Yet the respective concepts are not as closely related as the pair under discussion. The uniquely close relation

between being conscious$_2$ and having guilty awareness of wrongdoing strongly suggests that what one can be conscious$_2$ of is not just anything at all; nor can it be just anything at all about oneself. As Lewis (1967, pp. 188–189) points out, one would fail to grasp some of what was going on if one construed the statement "I am conscious to myself of many failings" as though its author was making reference to a psychological state that did not involve intrapersonal consciring, that is, that did not involve his being "privy to himself, in his own secret" as regards many of those failings.

Distinct from Consciousness$_4$

The *OED* provides a total of 17 illustrative quotations that are proposed to be instances of exercising the concept of consciousness$_2$. This total does not include, of course, the quotations illustrating the use of "conscious" to refer to guilty awareness of wrongdoing. Nevertheless, many of the 17 statements imply that the individual referred to is not merely in a position to give testimony; also, he or she feels guilty regarding the witnessed state of affairs—not for having witnessed it, of course, but for what that state of affairs is or indicates about him or her. Moreover, except for one case out of the 17, the author is speaking of someone's being aware of something personal to him or to her that possesses ethical significance as defined broadly (to be explained).

Before I develop the latter point, let me address the single aberrant example. This example turns out to be useful for the present purpose because it has reference to a basic kind of consciousness that is distinct from consciousness$_2$. It is the kind that I have been calling "consciousness$_4$" following the *OED* (e.g., Natsoulas, 1983, 1986–1987, 1994a). Also, I have been calling consciousness$_4$ "inner awareness," because consciousness$_4$ is the direct, noninferential awareness that we all have of some of our own mental-occurrence instances. Although there is disagreement among theorists concerning the process by which inner awareness is accomplished (e.g., Brentano, 1911/1973; Dulany, 1991, p. 105; Humphrey, 1987; Natsoulas, 1993b; Rosenthal, 1993; Woodruff Smith, 1989), it is the rare psychologist who will deny that we possess the ability to be directly aware of some of our own mental-occurrence instances (e.g., Hebb, 1972, 1982; cf. James, 1890/1950, pp. 304–305).[6]

The stray *OED* example of usage is none other than John Locke's (1706/1975) application of "conscious to oneself" in the first sentence of this passage from *An Essay Concerning Human Understanding*:

> If they say, That a man is always conscious to himself of thinking; I ask How they know it? Consciousness is the perception of what passes in a

> Man's own mind. Can another man perceive, that I am conscious of any thing, when I perceive it not my self? No Man's knowledge here, can go beyond his Experience. (p. 115)

Clearly, Locke is not speaking here of consciousness$_2$. His foregoing use of "conscious to oneself" should have been included in a different *OED* entry. In the previous passage, Locke is using both "consciousness" and "conscious to himself" in a very important psychological sense. He is referring to a kind of consciousness that is more basic than consciousness$_2$. He is clearly referring to inner awareness.[7]

To say, as I just did, that consciousness$_4$ is more basic than consciousness$_2$ is not to side, so to speak, with one kind of consciousness against another, in the fashion of some psychologists. Among much else, the *OED*'s six kinds of consciousness are *all of them* our subject matter, and *in no part* to be denied. If consciousness$_4$ is the more basic of the two, then it is essential to study consciousness$_4$ in order fully to understand consciousness$_2$. However, study of consciousness$_4$ cannot substitute for study of consciousness$_2$, except on pain of omitting psychological phenomena of great human importance.

But how did Locke's statement find its way into the *OED* entry for "conscious to oneself"? The form of words must have been responsible; as was, no doubt, the fact that every case of consciousness$_2$ instantiates the kind of "inwardness" that Locke was adverting to in the previously quoted passage. The latter fact will be seen in the final subsection of the present chapter, which bears the title The Self-Awareness Involved.

Locke's kind of consciousness is described later in the *OED*'s two lists of entries. The second sentence of the quoted passage from Locke serves as one of the illustrative quotations in the first half of the *OED*'s fourth entry under "consciousness." Indeed, the compilers of the *OED* could have adopted "Consciousness is the perception of what passes in a Man's own mind" for their definition of "consciousness" in the fourth sense.[8]

One's being conscious$_4$, or having inner awareness (cf. Brentano, 1911/1973) of one's thinking now going on, or of any other kind of mental activity or mental state or event now taking place within one, is a distinct kind of consciousness. It is distinct with two qualifications:

1. Consciousness$_4$ is part of what essentially occurs whenever one of certain of the other five *OED* kinds of consciousness happens to be instantiated.
2. Every instance of consciousness$_4$ is also an instance of consciousness$_3$—which is, as already identified, occurrent awareness of anything at all by whatever means, including however it may be that, whatever the process by which, we have inner awareness.

The familiar distinction between conscious and nonconscious mental-occurrence instances amounts to whether or not a particular mental-occurrence instance is an object of inner awareness. Also, we now commonly speak of conscious mental occurrences and nonconscious mental occurrences, meaning, respectively, mental occurrences that can and cannot be the objects of inner awareness; that is, to qualify for the epithet "conscious," a mental occurrence need not be an object of inner awareness on every occasion of its taking place, but it must be capable of being such an object.[9]

Like consciousness$_3$, consciousness$_4$ is instantiated with very great frequency in everyday life (Natsoulas, 1993a; contrast Delius, 1981). And consciousness$_4$ certainly does not require that one be instantiating consciousness$_2$ at the time. Typically, consciousness$_2$ is, as it were, a more reflective process; that is, consciousness$_4$ occurs most often merely in passing, as one orients oneself with reference to what is taking place in one's "stream of consciousness" (James, 1890/1950, 1892/1984).[10] However, as already mentioned and as will be seen in some detail, in every case of being conscious$_2$, inner awareness of mental-occurrence instances does take place as a feature of consciousness$_2$ itself.

At the start of the preceding paragraph, I contradicted a conviction that would seem to be held by many psychologists. Consciousness$_4$ is supposed to take place under special circumstances, namely, when one has a task or a problem requiring that one determine what exactly one is now experiencing. Otherwise, one's awarenesses are directed on the environment and on oneself as inhabiting the environment, but not on the stream of consciousness itself. In this erroneous view, although the stream of consciousness always has the potential to be consulted (i.e., it is open to one's consciousness$_4$), there is normally no need to perform such a consultation.

I disagree with this view frequent among psychologists. They so hold, I believe, because they construe consciousness$_4$ as a process of reflecting on one's mental life, that is, as a matter of turning one's thoughts back upon one's mental life. They fail to realize that, in the absence of inner awareness, we would be ignorant as we go along of our seeing, feeling, thinking, or intending whatever we may be seeing, feeling, thinking, or intending (Natsoulas, 1993a). Absent all inner awareness, we would be, with respect to all of our mental life, like people who possessed only "blindsight" over the entirety of their field of view (Weiskrantz, 1993; cf. Natsoulas, 1996). We would be "mindblind." All of our self-awarenesses would result either from perceiving our bodies and behaviors by means of our senses or from apprehending ourselves as objects of thought.[11] The stream of consciousness could not be, in any part, its own direct object (cf. James, 1890/1950, pp. 304–305).

The Possible Objects of Consciousness$_2$

With the exception of that stray statement quoted from Locke, the *OED*'s 17 illustrative quotations for "consciousness" in the second sense, for "conscious to oneself," and for "conscious" in the same sense characterize one or more persons, respectively, as being conscious$_2$ of his, her, or their own

> wants; ignorance; guilt; wrongdoing; immeasurable superiority to others; well-spent life; great weakness; great defectiveness; useless medicines; having done everything possible to warn the nation; offence; highest worth; errors of omission; defects and vices; unfitness; or having engaged in adultery.

Whoever was said to be conscious$_2$ of any one of most of these items probably felt guilty about instantiating it. However, the exceptions to the latter concomitance are useful in helping us to determine the range of items of which one can be said to be conscious$_2$. The following five are among the exceptions:

1. Happy in the consciousness$_2$ of a well-spent life, one is likely to have applied moral standards to one's past actions and activities in coming to this consciousness. In judging one's life to have been well spent, one may also have applied other kinds of standards, in addition to or instead of the moral. In any case, one's favorable judgment implies an absence of feelings of guilt at least concerning the general way in which one has spent one's life.
2. If one were conscious$_2$ of being immeasurably superior to other people and drew support from being so—just as the respective illustrative quotation states about someone called Bentley—one would not feel guilty about this state of affairs, which one judged to be the case, although one would likely keep one's judgment to oneself.
3. In *A Journal of the Plague Year*, Daniel Defoe (1722/1960) states, "Abundance of quacks too died, who had the folly to trust to their own medicines, which they must needs be conscious to themselves were good for nothing" (p. 43). These quacks could not but be conscious$_2$ of the uselessness of the medicines that they were dispensing. Thus, they could not but be conscious to themselves of their own powerlessness against the plague. Nevertheless, Defoe goes on to say, they failed to feel guilty concerning their fraudulent practice. And, curiously, they did not act in such a way as would have avoided the punishment that they deserved. Defoe would seem to be suggesting that these quacks were deeply engaged in self-deception. They placed faith in medicines of which they were conscious to themselves could not protect them from illness, or restore them to health if they contracted the disease, as was very likely.
4. John Milton (1667/1935, II, 428) describes Satan as conscious of

his "highest worth" when Satan announces to his fellows that he alone will take on the great task that the rest of them feared. Satan's "monarchal pride" is mentioned in this connection, but no feelings of guilt are implicitly or explicitly ascribed to him.

5. In a letter, Edmund Burke (1779/1963) wrote, "I am low and dejected at times in a way not to be described. The publick Calamities affect me; and would much more, if I were not conscious to myself of having done every thing in my power to warn the Nation of the Evils that were bringing upon them" (p. 125). Although the government's actions and inactions were endangering the nation's safety from its enemies, Burke did not have to feel guilty, at least, for having remained silent regarding the situation. He was conscious to himself of having done all that he could in this respect.

Now, one's being immeasurably superior to other people, for example, or one's possessing the highest worth, is not obviously the kind of personal characteristic that one might feel guilty about. The potential for guilty feelings concerning the personal characteristic being judged would not seem to delimit the potential objects of consciousness$_2$. If every object of consciousness$_2$ is not to be distinguished by accompanying feelings of guilt regarding one's instantiating it, what do all the possible objects of consciousness$_2$ have in common?

I mentioned earlier that a large number of them clearly do have ethical significance. However, this statement cannot be generalized without a special interpretation of the ethical that includes intellectual powers and achievements within that category. For example, one can be conscious to oneself of being ignorant regarding how parts of one's body specifically function, or conscious to oneself of one's superior ability in treating the sick.

James's Spiritual Self

William James's discussion of "the spiritual self" in *The Principles of Psychology* (1890/1950) can help to explain what all of the possible objects of consciousness$_2$ have in common. It would seem that everything of which one can be conscious$_2$ has direct relevance to those features of a person that James considered to be parts of his or her spiritual self. The possible objects of consciousness$_2$ are either psychological characteristics of the kind constituting the spiritual self, or they are behaviors or mental occurrences that indicate one's having a particular one or more such characteristics.[12]

James (1890/1950) defined the spiritual self as

> a man's inner or subjective being, his psychic faculties or dispositions. . . . These psychic dispositions are the most enduring and intimate part of the

> self, that which we most verily seem to be. We take a purer self-satisfaction when we think of our ability to argue and discriminate, of our moral sensibility and conscience, of our indomitable will, than when we survey any of our other possessions. Only when these are altered is a man said to be *alienatus a se*. (p. 296)

We gather better what James's spiritual self consists of in its entirety when James explains that "spiritual self-seeking" corresponds to "every impulse towards psychic progress, whether intellectual, moral, or spiritual in the narrow sense of the term [i.e., religious]" (p. 309). Thus, a person's spiritual self, when it is considered abstractly, is equivalent to all of the psychological powers, abilities, traits, dispositions, and tendencies that make up the intellectual, moral, and religious dimensions of his or her personality.

When, instead, the spiritual self is theoretically considered by James as something concrete, then it amounts simply to the stream of thought or mental life. In contrast, the aforementioned features, which are said to make up the spiritual self, do so abstractly; that is, although they can be judged or inferred about by the person whose spiritual self it is (and by others as well), they cannot be directly apprehended themselves, whether by inner awareness or in any observational sort of way.

James states that the spiritual self may be theoretically considered both abstractly and concretely. This does not mean that someone's spiritual self amounts merely to how that individual thinks it is in his or her own case, that is, to a kind of conception that one keeps on developing of an important part of oneself. Certainly, one can think about one's spiritual self and come to conclusions concerning each of the features that comprise it. But these features are one thing and how we construe them is another thing, not to be confused with them. The features of the spiritual self are each something objective about the individual and susceptible to being gotten wrong by the individual and others. That certain of an individual's powers or tendencies largely manifest themselves by affecting his or her stream of consciousness does not make them any less objective.

James's inclusion of the intellectual as part of the spiritual self is, of course, intriguing. For we commonly hold that the intellect and its development may serve a variety of purposes, including morally neutral ones, spiritually irrelevant ones, or worse. I want to be sure to emphasize that James considered the intellectual to be part of the spiritual self only insofar as intellectual ends are being pursued for themselves alone, without concern for any material or social advantage—not even God's approval—that may result from the intellect's advancement. According to James and others of his time, the desire to improve one's intellectual functioning derives from the same group of basic instinctive impulses as do moral and spiritual purposes.

The Self-Awareness Involved

Certain kinds of judgments are a part of being conscious$_2$. However, judgments do not comprise all that any instance of consciousness$_2$ essentially involves. Another essential part of being conscious$_2$ is the *basis* of the specific judgments involved. Elsewhere, I have indicated what this basis is in the following words:

> I newly learn or remind myself, on a firsthand basis (not from hearsay), about the kind of person I am in one or another specific respect. . . . I newly learn or remind myself of this, from having witnessed relevant actions I performed or experiences I had, and by now bringing this evidence to bear on how I conceive of myself, in terms of a trait or ability I therefore consider myself to possess, on perhaps other grounds as well. (Natsoulas, 1991b, p. 344)

In any instance of consciousness$_2$ that is based on evidence from the past, several kinds of self-awareness must take place, as described next; one cannot be conscious$_2$ in their absence.

Witnessing of Oneself

Consciousness$_2$ cannot be, as already noted, about just anything at all. Rather, one is, or has been at some point, aware firsthand, on the spot, at the time of its occurrence, of a certain relevant piece of one's own behavior or a certain relevant part of one's stream of consciousness. This behavior or the particular segment of one's stream must be suitable for serving as evidence about one's intellectual, moral, or religious powers, abilities, traits, dispositions, or tendencies. The evidence cannot pertain instead to merely social or material facts about oneself. Evidence pertaining only to one's health, physical attributes, kinship, wealth, social status, sociability, and the like, would not be relevant, because consciousness$_2$ is always about one or more aspects of one's spiritual self. Of course, in evaluating evidence concerning an aspect of one's spiritual self, one may take into account such factors as the aforementioned. One may acknowledge one or more of these factors as determinants of one's intellectual performance for example. However, it is not these factors that one has consciousness$_2$ of or about; rather, it is the relevant outcomes to which they may have contributed, such as a habitual way of thinking about the world or treating other people.

Appropriation to Oneself

The self-witnessing just mentioned cannot be of an alienated sort. In James's term, the self-witnessing must "appropriate" to oneself that which

is witnessed. One must not only be aware firsthand of the particular piece of one's behavior or segment of one's stream of consciousness that would serve as evidence, but also one must be aware of that piece or segment as being one's own. In this additional sense, the self-witnessing must be "personal." It is possible for one to have, instead, inner awareness wherein one's stream of consciousness seems to have been taken over by another agent or, at least, seems not to be one's own (Natsoulas, 1979). Also, one may observe someone behaving (e.g., as a result of an arrangement of mirrors or by means of video equipment) and not realize it is oneself whom one is observing. Reed (1972) brings out that actions, as well as temporal sections of one's stream of consciousness, may suffer "a loss of personal attribution," either transiently, as a result of situational stress in the case of normal people, or more consistently, as part of a schizophrenic syndrome.

Retrowareness of Oneself

Instances of consciousness$_2$ very frequently take place at a temporal distance from the specific self-witnessing that furnishes evidence for them. For example, something of yourself that you witnessed at a certain, perhaps early point in your life may repeatedly serve you as evidence for judgments regarding the kind of person you are. Thus, remembering proper is usually a part of being conscious$_2$, and all instances of such remembering, whether or not they take place as part of consciousness$_2$, necessarily involve occurrent awareness now of a past happening or state of affairs. In being conscious$_2$, one now remembers something that one previously did or underwent; that is, consciousness$_2$ usually involves a kind of self-awareness that is a "retrowareness" (Natsoulas, 1986). For example, I am apprehending in thought right now, not for the first time, how I behaved (i.e., a particular action of mine) at the party my parents gave to celebrate the 12th anniversity of my birth.

Inner Awareness

For an occurrent awareness to qualify as an instance of remembering proper, the awareness must be a retrowareness of a special kind. It does not suffice that one was originally aware firsthand of the object of the particular retrowareness. In an act of remembering something in particular, one must be occurrently aware now of oneself as now apprehending that which one had earlier apprehended. Thus, necessarily involved at the point of remembering is inner awareness (consciousness$_4$) of one's present self-retrowareness.

Consciousness$_4$ Extended Backward

And there is involved, as well, a kind of retrowareness that resembles inner awareness. There occurs a present retrowareness of one's past experience. By past experience, I mean, for example, one's perceiving, emoting over, thinking about, planning, remembering, or expecting something or other in particular that was taking place, had taken place, or was going to take place. Thus, in those many cases of consciousness$_2$ wherein one remembers witnessing the relevant evidence, there occurs both (1) inner awareness of present components of one's stream of consciousness that are retrowarenesses of the evidence witnessed and (2) inner awareness of present components of one's stream of consciousness that are retrowarenesses of some of the mental acts that took place as part of one's witnessing the evidence.

Self-Thoughts and Self-Judgments

There is more to consciousness$_2$ than remembering the right sort of item. One is conscious$_2$ not simply because one is remembering one's act of witnessing something about oneself, however important the latter feature of oneself may happen to be and whatever it may be that one remembers. Only through putting what one remembers to use does it become evidence. It becomes evidence by being brought to bear on something that one already believes or something that one may come to believe. To be conscious$_2$, one must put what one remembers to use as evidence regarding the kind of person that one is with respect to the spiritual sphere as James broadly defined it.[13] In addition to one's undergoing retrowareness of the particular past events as being occurrent parts of oneself, one must have thoughts regarding one or more characteristics that may belong to the intellectual, moral, or religious dimensions of one's personality, and one must make judgments regarding how the remembered evidence bears on whether those characteristics do so belong.

The necessary emphasis on remembering as a feature of consciousness$_2$ based on past evidence should not serve as a distraction from the fact that consciousness$_2$ from present evidence also takes place. It can occur on the spot, that is, while one is still engaged in witnessing the relevant behavior or segment of one's stream of consciousness. Spelling out the self-awareness involved in contemporaneous consciousness$_2$, as I have done for consciousness$_2$ from past evidence, would be a belaboring of the closely analogous features. Indeed, the main points of the present subsection on the self-awareness involved in consciousness$_2$ can now be, in effect, very briefly reviewed simply by noting how the present analysis extends quite naturally to contemporaneous consciousness$_2$ as well. Accordingly, when-

ever any instance of consciousness$_2$ takes place at the time of the occurrence of the evidence on which it is based, there is essentially involved self-awareness in, at the least, all of the following forms:

- One witnesses potential evidence about oneself.
- One has inner awareness of this witnessing.
- One has occurrent awareness in thought of one or more features of one's character or personality.
- One brings self-witnessed evidence to bear in judging of this feature or these features.

NOTES

1. However, see the Cambridge physiologist Horace B. Barlow's (1987) use of the obsolete interpersonal sense of "conscious" to make his case for a certain scientific understanding of consciousness (cf. Barlow, 1980). For some discussion of Barlow's view in relation to the *OED*'s first sense of "consciousness," see Natsoulas (1991a).

2. Nor do we still use "conscious" in the corresponding sense that the *OED* defines as follows: "Knowing, or sharing the knowledge of anything with another; privy to anything with another."

3. Compare the following statement from James's *The Principles of Psychology* (1890/1950):

> No one ever had a simple sensation by itself. Consciousness from our natal day is always of a teeming multiplicity of objects and relations, and what we call simple sensations are results of discriminative attention, pushed often to a very high degree. It is astonishing what havoc is wrought in psychology by admitting at the outset apparently innocent suppositions, that nevertheless contain a flaw. (p. 224; see also p. 478)

James held that all basic durational components of the stream of consciousness—therefore, all mental-occurrence instances—possess the property of intentionality (cf. Brentano, 1911/1973); that is, all are "intellections," all have a cognitive aspect—which does not mean that they are not, each of them, feelings as well according to James.

4. In the present chapter, I abbreviate and simplify what needs to be said regarding consciousness$_3$ (see Natsoulas, 1992b, 1995). Also, I do not discuss views, such as James's (1890/1950), to the effect that every basic durational component of one's stream of consciousness intrinsically involves awareness of one's body or of certain occurrences in one's body. This would mean that every instance of consciousness$_3$ is a kind of self-awareness by involving awareness of a part of oneself (see Natsoulas, 1996–1997).

5. The closeness of meaning is of unusual degree, as will be seen. There are a number of entries listed under "conscious" in the *OED*. But the remaining mean-

ings, beyond those already mentioned in the text, need not be brought into the present discussion of consciousness$_2$.

6. A particular theoretical disagreement that is relevant to the relation between consciousness and self-awareness is the one between (1) those who hold that inner awareness requires the ascription or appropriation of its objects to oneself (e.g., Kihlstrom, 1987; Rosenthal, 1986), and (2) those who allow for impersonal or anonymous inner awareness (James, 1890/1950; Woodruff Smith, 1989). For objections to Rosenthal's view, see Natsoulas (1992a). For discussion of this issue with special reference to James, see Natsoulas (1996–1997).

7. Margaret Atherton (1983/1992) does not agree. I have argued against her position elsewhere (Natsoulas, 1994b). It serves no present purpose, however, to summarize or to consider her understanding of Locke again here.

8. Except for one consideration: Insofar as we know our own mental life firsthand, do we really know it by the efficacy of a perception-like process, by, as it were, "inner spection," or Locke's power of reflection? This is a controversial thesis (see Brentano, 1911/1973, pp. 429–434; Woodruff Smith, 1989, pp. 83–88).

9. Thus, all the basic durational components of James's (1890/1950) stream of consciousness are conscious mental occurrences, although inner awareness of every one of them does not occur. But any one of them that happens not to be an object of inner awareness could have been such an object. All of the basic durational components of the stream are open to inner awareness, as they come into existence one after another. However, James's notion of the specious present renders his understanding of inner awareness somewhat more complex than I have intimated here; see Natsoulas (1992–1993).

10. Compare the phenomenologist Aaron Gurwitsch (1964): "Our mental activity is always [?] accompanied by awareness of facts and data belonging to the following three orders of existence: 1. *The stream of our conscious life;* 2. *our embodied existence;* 3. *the perceptual world*" (p. 415).

11. But what could such self-awareness amount to? Lacking inner awareness, we would have to infer the occurrence of an instance of self-awareness from something about ourselves that we could observe. But we could not have inner awareness of any observations and would have to infer their occurrence as well—and so on.

12. James also discussed other constituents of the self, none of which are mentioned in the text of the present chapter, namely, the material self, the social self, and the pure ego (whose existence James rejected). Their exclusion from the present chapter, however, does not signify a judgment against their importance for understanding self-awareness. Rather, as should be clear, James's notion of the spiritual self serves here as a likely key to comprehending consciousness$_2$.

13. Putting evidence to use is indeed what one must do in being conscious$_2$ of . . . or that . . . ; but this need involve no more than drawing a conclusion concerning oneself, partially or entirely on the basis of that evidence. Thus, I do not mean that potential evidence must be worked over, although this may be necessary in some cases if it is to contribute to one's conclusion. What may serve as evidence can point in more than a single direction regarding the kind of person one is, or who is responsible for what happened.

REFERENCES

Atherton, M. (1992). Locke's theory of personal identity. In V. Chapell (Ed.), *John Locke* (pp. 85–105). New York: Garland. (Original work published 1983)

Barlow, H. B. (1980). Nature's joke: A conjecture on the biological role of consciousness. In B. D. Josephson & V. S. Ramachandran (Eds.), *Consciousness and the physical world* (pp. 81–90). Oxford, England: Pergamon.

Barlow, H. B. (1987). The biological role of consciousness. In C. Blakemore & C. Greenfield (Eds.), *Mindwaves* (pp. 361–376). Oxford, England: Blackwell.

Brentano, F. (1973). *Psychology from an empirical viewpoint*. New York: Humanities Press. (Second German edition published 1911)

Burke, E. (1963). *The correspondence of Edmund Burke* (Vol. 4). Cambridge, England: Cambridge University Press. (Letter cited written 1779)

Defoe, D. (1960). *A journal of the plague year*. New York: Signet/New American Library. (Original work published 1722)

Delius, H. (1981). *Self-awareness*. Munich: Beck.

Dewey, J. (1906). The terms "conscious" and "consciousness." *Journal of Philosophy, Psychology and Scientific Method, 3*, 39–51.

Dulany, D. E. (1991). Conscious representation and thought systems. In R. S. Wyer, Jr., & T. K. Srull (Eds.), *Advances in social cognition* (Vol. 4, pp. 97–119). Hillsdale, NJ: Erlbaum.

Engelberg, E. (1972). *The unknown distance*. Cambridge, MA: Harvard University Press.

Gurwitsch, A. (1964). *The field of consciousness*. Pittsburgh: Duquesne University Press.

Hebb, D. O. (1972). *A textbook of psychology* (3rd ed.). Philadelphia: Saunders.

Hebb, D. O. (1982). Elaboration of Hebb's cell assembly theory. In J. Orbach (Ed.), *Neuropsychology after Lashley* (pp. 483–496). Hillsdale, NJ: Erlbaum.

Humphrey, N. K. (1987). *The uses of consciousness* (Fifty-seventh James Arthur Lecture on the Evolution of the Human Brain). New York: American Museum of Natural History.

Husserl, E. (1970). *Logical investigations* (2 vols.). New York: Humanities Press. (Original work published 1900)

James, W. (1950). *The principles of psychology* (2 vols.). New York: Dover. (Original work published 1890)

James, W. (1984). *Psychology*. Cambridge, MA: Harvard University Press. (Original work published 1892)

Kihlstrom, J. F. (1987). The cognitive unconscious. *Science, 237*, 1445–1452.

Koch, S. (1980). Psychology and its human clientele: Beneficiaries or victims. In R. A. Kasschau & F. S. Kessel (Eds.), *Psychology and society* (pp. 30–53). New York: Holt, Rinehart & Winston.

Lewis, C. S. (1967). *Studies in words* (2nd ed.). Cambridge, England: Cambridge University Press.

Locke, J. (1975). *An essay concerning human understanding*. Oxford, England: Clarendon. (Fifth edition published 1706)

Mead, G. H. (1934). *Mind, self, and society*. Chicago: University of Chicago Press.

Miller, G. A. (1990). The place of language in a scientific psychology. *Psychological Science*, *1*, 7–14.

Milton, J. (1935). *Paradise lost*. New York: Odyssey. (Original work published 1667)

Natsoulas, T. (1979). The unity of consciousness. *Behaviorism*, *7*, 45–63.

Natsoulas, T. (1983). Concepts of consciousness. *Journal of Mind and Behavior*, *4*, 13–59.

Natsoulas, T. (1985). George Herbert Mead's conception of consciousness. *Journal for the Theory of Social Behaviour*, *15*, 60–75.

Natsoulas, T. (1986). Consciousness and memory. *Journal of Mind and Behavior*, *7*, 463–501.

Natsoulas, T. (1986-1987). The six basic concepts of consciousness and William James's stream of thought. *Imagination, Cognition and Personality*, *6*, 289–319.

Natsoulas, T. (1991a). The concept of consciousness$_1$: The interpersonal meaning. *Journal for the Theory of Social Behaviour*, *21*, 63–89.

Natsoulas, T. (1991b). The concept of consciousness$_2$: The personal meaning. *Journal for the Theory of Social Behaviour*, *21*, 339–367.

Natsoulas, T. (1992a). Appendage theory—pro and con. *Journal of Mind and Behavior, 13*, 371–396.

Natsoulas, T. (1992b). The concept of consciousness$_3$: The awareness meaning. *Journal for the Theory of Social Behaviour*, *22*, 199–225.

Natsoulas, T. (1992-1993). The stream of consciousness: II. William James's specious present. *Imagination, Cognition and Personality, 12*, 367–385.

Natsoulas, T. (1993a). The importance of being conscious. *Journal of Mind and Behavior, 14*, 317–340.

Natsoulas, T. (1993b). What is wrong with appendage theory of consciousness. *Philosophical Psychology*, *6*, 137–154.

Natsoulas, T. (1994a). The concept of consciousness$_4$: The reflective meaning. *Journal for the Theory of Social Behaviour*, *24*, 373–400.

Natsoulas, T. (1994b). The concept of consciousness$_5$: The unitive meaning. *Journal for the Theory of Social Behaviour*, *24*, 401–424.

Natsoulas, T. (1995). Consciousness$_3$ and Gibson's concept of awareness. *Journal of Mind and Behavior, 16*, 305–328.

Natsoulas, T. (1996). Blindsight and consciousness. *American Journal of Psychology, 110*, 1–34.

Natsoulas, T. (1996–1997). The stream of consciousness: XII. Consciousness and self-awareness. *Imagination, Cognition, and Personality, 16*, 161–180.

Pylyshyn, Z. (1990). Roundtable discussion. In P. P. Hanson (Ed.), *Information, language and cognition* (pp. 198–216). Vancouver: University of British Columbia Press.

Reed, G. (1972). *The psychology of anomalous experience*. London: Hutchinson.

Rosenthal, D. M. (1986). Two concepts of consciousness. *Philosophical Studies, 49*, 329–359.

Rosenthal, D. M. (1993). Higher-order thoughts and the appendage theory. *Philosophical Psychology*, *6*, 155–166.

The Oxford English dictionary. (1989). (2nd ed., 20 vols.). Oxford, England: Clarendon.

Weiskrantz, L. (1993). Unconscious vision. *The Sciences*, *32*(5), 23–28.

Wittgenstein, L. (1980). *Remarks on the philosophy of psychology* (Vol. 1). Oxford, England: Blackwell. (Original German typescript dated 1947)

Woodruff Smith, D. (1989). *The circle of acquaintance*. Dordrecht, The Netherlands: Kluwer.

PART TWO

THE EVOLUTION OF SELF-AWARENESS

CHAPTER THREE

When Self Met Other

✧

DANIEL J. POVINELLI
CHRISTOPHER G. PRINCE

"Starting from what I know of the operations of my own individual mind, and the activities which in my own organism they prompt, I proceed by analogy to infer from the observable activities of other organisms what are the mental operations that underlie them." So wrote John George Romanes (1882, pp. 1–2), outlining his method for gaining scientific leverage on the problem of animal minds. In the wake of Darwin's (1871/1982) publication of *The Descent of Man*, Romanes took up the challenge of investigating the evolution of mind with a fervor. Darwin had marshaled an impressive array of observations to suggest that not only were humans descended from other species in bodily structures, but also in mental structures. For Romanes, the agenda for a new science was clear. Just as anatomists "aim at a scientific comparison of the bodily structures of organisms," he observed, "so [comparative psychology] aims at a similar comparison of their mental states" (Romanes, 1883, p. 5). But Romanes faced a problem. Although his analogy between comparative anatomy and a new science of comparative psychology was powerful, it began to break down when it came to the substances to be compared. Anatomists had access to dead bodies, but the stuff of psychology was not so easily available for examination on the laboratory bench. Recognizing this, Romanes offered an interim solution by turning to the only source of material available—the spontaneous behavior of animals. But even here, he was far behind the anatomist in that there was no existing corpus of data. Romanes knew he might be roundly chastised for doing so—and he was—but ultimately he was forced to rely on anecdotes as his database. "If the present work is read without reference to its ultimate object of supplying facts for

the subsequent deduction of principles," Romanes apologized, "it may well seem but a small improvement upon the works of the anecdote-mongers" (1882, p. vii).

Although the foundations of Romanes' approach collided with an age-old philosophical problem far more profound than that of inferring mental states in other species—the uncertainty of making inferences even about the minds of our fellow humans—in a very real sense, the core of his method was unassailable. After all, we are not just bodies cohabiting the same physical locations. We are thinking, feeling beings, negotiating a social world teeming with each others' desires, goals, intentions, and emotions. We are linked not just in space, but also in mind. Each of us possesses an intense desire to have our lives understood by others. Indeed, this desire to understand and be understood, to be part of a group—whether it be the village, school, family, or kibbutz—would appear to be as universal a trait of the human species as any.[1] This inextricable connection between self and other even occurs in our most silent moments. Someone looks around a corner, furrows his or her brow, and we effortlessly attribute a visual experience of having just seen something. In this sense, Romanes' solution to the problem of understanding animal minds was simple: If we can make reasonably accurate inferences about what other humans are thinking and feeling by just observing them, why can we not do the same with members of other species? If the method of introspective analogy can bridge the gap that separates the mind of one human from another, why cannot it not likewise span the distance that separates one species from another?

In this chapter, we return to Romanes' problem by examining the evolution of the psychological connection between self and other that appears so characteristic of our species. We explore three related questions. First, when did the ability to conceive of the self and others as mentalistic agents evolve? That is, when and in what lineages did organisms first evolve the capacities to reason about themselves and others in terms of mental states and events such as thinking, knowing, believing, attending, desiring, and perceiving? The data we review suggest that humans may be largely unique in being able to reason about such internal states. Second, we ask whether the evolution of self–other psychology occurred in synchrony, or whether an understanding of self as a mental agent preceded a comparable understanding of others. A main issue we address is whether the fusion of self and other understanding we observe in humans can be dissociated in other species. For example, the data we review are consistent with the hypothesis that although chimpanzees may possess at least a limited objective self-concept, they may lack the ability to conceive of others (and perhaps even themselves) as mentalistic agents. Finally, we address the thorny problem of how it is that nonhuman primates can share with us so many behavioral

patterns that in humans are clearly associated with a mentalistic understanding of others, yet not possess such understanding themselves. Our conclusion is that humans may have evolved these abilities not because they endowed us with scores of novel behaviors per se, but rather because they allowed for the complex reorganization and redeployment of existing behavioral patterns. Thus, although humans may indeed possess unique behavioral capacities (e.g., active pedagogy), the initial utility of interpreting behavior in terms of internal, nonobservable mental states may have been at an organizational level—at a higher level of abstraction than any specific set of behaviors. If this general idea is correct, it may mean that traditional efforts to find a coherence between an understanding of self and others as mental agents, and some set of naturally occurring behaviors, will remain largely futile.

In order to set the stage for examining these ideas, we review several aspects of current theorizing and research concerning self and social understanding. First, by examining certain aspects of the spontaneous social behavior of primates in nature, we explore how the traditional approach to understanding the evolution of social intelligence has led to the idea that many, if not most, nonhuman primates possess some understanding of the mental states of themselves and others. Second, we explore the evidence concerning the evolution of self-conception and explain why these data have been interpreted as suggesting a qualitative psychological difference between the great apes and humans on the one hand, and most other forms of life on the other. Third, we examine experimental data concerning the development of one aspect of human infants and children's understanding of their own and others' minds, and parallel research with chimpanzees. Our interpretation of these data is that despite their striking similarity to us at the level of specific behavioral patterns, not even chimpanzees possess a theory-of-mind system comparable to that which develops in humans during late infancy and early childhood.

SELF, SOCIALITY, AND INTELLIGENCE

Evolving an Intelligence of Others

Alison Jolly (1966) authored the first careful statement of the idea that the truly remarkable features of primate intelligence had evolved in the context of coping with each other—a kind of social intelligence of others. Returning from an early field survey of the prosimian primates of Madagascar, Jolly was left with the impression that primate intelligence was far from a unitary concept. She reflected that intelligence about objects and physical events appeared to have evolved independently from intelligence deployed

in the social realm. She minimally differentiated between intelligences "toward objects, including food; toward other active species, including predators; and toward fellow members of one's own species" (p. 504). Some of her own previous laboratory research had suggested that prosimians lagged far behind other primates both quantitatively and qualitatively on standardized tests of intelligence (Jolly, 1964a, 1964b; see also Andrew, 1962). In contrast, the sophistication of their social behaviors seemed comparable to what had been previously observed in various monkey species, including the characteristically long chains of social interactions.

Given that prosimians appeared able to generate the same level of social complexity as the anthropoid primates, but were far less proficient at traditional object-oriented laboratory tests of intelligence, it occurred to Jolly that prosimians might share a different form of intelligence with other primates—a social intelligence. As she explained:

> The social use of intelligence is of crucial importance to all social primates. As the young develop, they depend on the troop for protection and for instruction in their role in life. Since their dependence on the troop both demands social learning and makes it possible, social integration and intelligence probably evolved together, reinforcing each other in an ever-increasing spiral. And although it is very likely that the learned social relations of monkeys are in fact more complex than those of lemurs, our present techniques of description emphasize the similarity between lemur and monkey social interactions. (1966, p. 504)

Jolly's speculations amounted to suggesting a possible dichotomy or compartmentalization between these two types of intelligence. Indeed, experimental evidence from rhesus monkeys reared in social isolation was already hinting that there might be some validity to this claim by demonstrating that despite their inability to learn a variety of species-typical patterns of social behaviors, they performed like normal monkeys on most traditional, object-oriented intelligence tests (Harlow, 1965; Harlow, Schlitz, & Harlow, 1968).

Nicholas Humphrey (1976) independently reached a similar conclusion in the mid-1970s. Like Jolly, he noted that most primates, on the surface, appeared to possess a veritable "surplus" of intelligence: a seemingly unneeded cornucopia of intellectual abilities unrelated to the demands of their way of life in nature. Given his view of nature as an uncompromising optimizer, Humphrey quickly reached the conclusion that the idea that such a surplus truly existed "was most likely to be wrong" (p. 303). The solution, he reasoned, might lie in the function that intellect played in the social arena:

> [The] social primates are required by the very nature of the system they create and maintain to be calculating beings; they must be able to calcu-

ISBN 1-57230-317-4