CHAPTER 1

# Science and Sentience: Some Questions Regarding the Scientific Investigation of Consciousness

Jonathan D. Cohen
*Carnegie Mellon University*
*University of Pittsburgh School of Medicine*
Jonathan W. Schooler
*University of Pittsburgh*

The topic of consciousness has occupied philosophers and metaphysicians for centuries. What distinguishes the scientific approach to this topic, however, is a methodology—a set of tools that permit the empirical and theoretical study of a phenomenon in an objective, reproducible way. The relation between this methodology and the phenomenon of consciousness has a restless and controversial history. The topic of consciousness was central to psychology at its inception as a scientific enterprise. However, it rapidly fell into disfavor, in part because the method that seemed most appropriate for studying it (introspection) was abandoned as a scientific method. During the 1960s there was a brief reawakening of the topic. This was probably due in part to sociological factors, but almost certainly it had something to do with the advent of a set of empirical and theoretical methods, now known as cognitive science, that provided a scientific way of asking and answering questions about phenomena such as memory, language, and attention—phenomena clearly related to consciousness.

The reason why the topic of consciousness fell back out of favor is not entirely clear. Perhaps, again, it had to do with sociological factors. Possibly its association with the excesses and pop psychology that flourished as part of the 1960s counterculture revolution cost the topic its scientific credibility. Or, perhaps it was that early cognitive scientists knew, consciously or unconsciously, that additional groundwork was necessary before this topic could be productively addressed. Instead of taking consciousness head on, scientists chose to see how far the new tools of cognitive science could take them, each

in their own direction. Thus, it seemed possible to make solid scientific headway on phenomena such as strategic versus automatic processing, allocation of attention, working memory, subliminal processing, implicit memory, and differences between information processes in the awake and dreaming states, all without directly addressing the vexing questions that surround consciousness qua consciousness. These lines of research share a common assumption of a particular quality, state, or process—awakeness, awareness, alertness—that seems closely related to consciousness, but do not address the question directly. Newell (1990) made this point with characteristic clarity in his book about the Soar architecture:

> SOAR provides a theory of awareness. Soar is aware of something if its deliberate behavior can be made to depend on it. In this sense, awareness is operationally defined and fundamental, and it is a much-used notion throughout cognitive psychology. But consciousness can be taken to imply more than awareness . . . namely, the phenomenally subjective. It can mean the process, mechanism, state (or whatever) that establishes when and what an honest human would claim to be conscious of, both concurrently or retrospectively. Soar does not touch the phenomena of consciousness, thus designated. Neither does much else in cognitive psychology. That it seems out of reach at the moment might justify pushing the issue into the future, until additional regularities and phenomena accumulate, but the challenge remains. (p. 433)

Accordingly, researchers may have been reluctant to directly take on the issue of consciousness because it is such a highly subjective phenomenon. Another aspect of the challenge was captured by James (1890/1952), when he articulated the difficulty of examining consciousness directly:

> As a snow-flake crystal caught in the warm hand is no longer a crystal but a drop, so, instead of catching the feeling of relation moving to its term, we find we have caught some substantive thing, usually the last word we were pronouncing, statically taken, and with its function, tendency, and particular meaning in the sentence quite evaporated. The attempt at introspective analysis in these cases is in fact like seizing a spinning top to catch its motion, or trying to turn up the gas quickly enough to see how the darkness looks. (p. 158)

Thus, consciousness is not only subjective, but also ephemeral. The question then becomes: How can we come to examine and understand the snowflake without melting it in the process?

We have no illusions that this volume will resolve all of the difficulties associated with the scientific investigation of consciousness, although we hope it will make progress on a few. Most importantly, however, the goal is to explore the extent to which consciousness can be the target of direct scientific inquiry, that is, to get on the table some of the relevant scientific work and to consider the degree to which this research can help inform our understand-

ing of consciousness. Put simply: Is it now possible to study consciousness directly in a scientific way?

With this goal in mind, we open this volume by raising a few questions and issues that any scientific account of consciousness must ultimately address. Some are addressed head-on in this volume, whereas others remain as challenges for the future.

## HOW CAN WE IDENTIFY CONSCIOUSNESS?

How do we identify consciousness, either within ourselves or in others? What are our criteria? One way to begin is to consider the extremes: that is, by identifying processes or phenomena we consider to be nonconscious, and those we take to be clearly indicative of consciousness. There are at least two dimensions or contrasts that we can focus on in this respect. Consider, for example, the differentiation of reportable and unreportable experiences. At the one extreme, we have unreportable cognitive processes such as implicit memory and learning and subliminal perception. At the other extreme, we have reportable and some highly self-reflective processes such as explicit memory and metacognition. Automaticity is another dimension that seems relevant—that is, the distinction between automatic versus attentionally dependent processes with completely automatized processes, such as typing or driving a car at one extreme, and tasks heavily demanding attention (e.g., playing chess) at the other.

How does consciousness relate to the dimensions of reportability and attention? It seems, at least intuitively, that an important criterion we use in identifying consciousness in others is the ability to report the experience of consciousness. And this, in turn, seems to hinge on the ability for self-reflection. Specifically, to what extent are self-reflection (cf. Hobson, chapter 19; Kihlstrom, chapter 24; Johnson & Reeder, chapter 13) and the reportability of experience (cf. Reber, chapter 8; Dulany, chapter 10; Merikle & Joordens, chapter 6) useful or criterial for identifying consciousness? Related to the question of reportability is the question about the relation between language and consciousness. If reportability is a critical component of consciousness or of our ability to identify it, then what is the relation between language of any type and consciousness (see Reber, chapter 8, and Dulany, chapter 10, for contrasting discussions of this issue)? If there are potential limitations to the use of reportability as a criterion for consciousness, then what other types of criteria might we use to determine or infer when conscious processes are operating (see Jacoby, Yonelinas, & Jennings, chapter 2; Merikle & Joordens, chapter 6; and Greenwald & Draine, chapter 5, for possible alternatives to a simple reportability criterion for consciousness)? At-

tention also seems to be related to consciousness in some important way, but exactly how? Is attention "the gateway to consciousness," as has often been assumed, or is its relation to consciousness more complex (see Shiffrin, chapter 3, for one response to this question)? Finally, how do attention, reportability, and self-reflection relate to one another?

These questions all focus on the identification of consciousness. It seems unlikely that we will make much scientific progress in understanding a phenomenon until we can reliably identify it. Assuming this is possible, several additional questions arise.

## WHAT IS THE RELATIONSHIP BETWEEN CONSCIOUS AND UNCONSCIOUS PROCESSES?

As suggested earlier, one way of approaching the identification of consciousness is to focus on its distinction from unconsciousness. That is, how do conscious and unconscious processes differ? Do, as a number of contributors suggest (e.g., Jacoby et al., chapter 2; Rajaram & Roediger, chapter 11; Merikle & Joordens, chapter 6; Wegner, chapter 14; Dulany, chapter 10), conscious and unconscious processes involve qualitatively different operations? If conscious and unconscious processes do involve distinct operations, then how do these systems interact? Are they truly independent processes (cf. Jacoby et al.), or do they impact one another (cf. Wegner)?

The suggestion that conscious and unconscious processes may be qualitatively different raises important questions about the nature of unconscious processing. How cognitively sophisticated are unconscious processes? Are unconscious processes capable of deriving and representing complex rules (cf. Lewicki, Czyzewska, & Hill, chapter 9; Reber, chapter 8), or must complex symbolic thought be limited exclusively to the domain of consciousness (cf. Dulany, chapter 10)? Determining the limitations of unconscious processes may help to reveal the attributes that make consciousness unique (cf. Baars, Fehling, LaPolla, & McGovern, chapter 22).

## WHAT ARE THE MECHANISMS UNDERLYING CONSCIOUSNESS?

Ultimately, any satisfactory scientific account of consciousness will require an articulation of the mechanisms that underlie consciousness. Traditional cognitive psychological and computational approaches have suggested that higher cognitive processes—what many refer to as "controlled" processes—rely on a central executive, and this would seem to be a natural candidate for the seat of consciousness. Dennett (1991) referred to this as the Cartesian

theater, that is, the arena in which "it all comes together and consciousness happens" (p. 39).

This approach has intuitive appeal and helps explain some of the central observations about conscious experience, including its limited capacity and its apparently sequential or serial nature. Nevertheless, neurobiological and neuropsychological data, as well as recent research within cognitive science and psychology, have suggested an alternative view—a more distributed view of neural and cognitive processing. Further, there are now a number of theories of consciousness to go with this approach. For example, Minsky's (1985) Society of Minds and Dennett's Multiple Drafts theory, as well as two presented in this volume (Baars, Fehling, LaPolla, & McGovern's Global Workspace hypothesis; and Kinsbourne's distributed systems view), have the appeal of fitting better with what we know about neural organization. However, what they seem to lack, at least up to this juncture, is the ability to account for the phenomenological unity of consciousness. How do cooperation, seriality, sequentiality, and the focality of consciousness actually come about within a distributed system? This has yet to be fully specified or demonstrated. It also raises a closely related question: Should consciousness be treated as a unitary construct? Specifically, does it represent the operation of a delineable set of identifiable processes (as Hobson, chapter 19, and Mandler, chapter 26, seem to suggest), a particular type of interaction among some or all processes (as suggested by Farah, O'Reilly, & Vecera, chapter 17; Kinsbourne, chapter 16), or is it an emergent and irreducible property of the processing system as a whole? In particular, does consciousness, as a construct, promise to add anything to our understanding of neurobiological and/or psychological processes that a complete account of each component processing system, and the interaction among them, would not otherwise provide? This question leads naturally to questions about the conditions under which consciousness may arise.

## WHERE AND WHEN DOES CONSCIOUSNESS ARISE?

Suppose we could fully specify the mechanisms underlying consciousness—either computationally, psychologically, or even neurally. It is then tempting to ask whether or not these would be sufficient to support consciousness. For example, could we ever consider a computer to be conscious? Two related questions should also be considered. First, are any nonhuman animals conscious (if so, at what phylogenetic level)? Second, are infants conscious (if not, at what age does consciousness emerge)?

We think questions about what can be conscious are critical if only because they are so inescapably asked, and once asked, can become vexing. It is important, therefore, to define the scope of this volume with respect to such

questions. We think these can lead in one of two directions. One raises the issue of dualism, that is, whether a physical mechanism is sufficient to produce consciousness, or whether some additional ingredient (e.g., some ectoplasm or spirit) is necessary for consciousness to occur. This volume does not address the issue of dualism, not because we believe it is an illegitimate, uninteresting, or even unimportant issue, but because we believe it is not an issue that science can profitably address. In our view, scientific questions are by their very nature directed at the physical, observable world, and therefore their scope should be limited to this domain. We realize this is subject to philosophical debate, and the very terms *physical* and *observable* are themselves subject to interpretation. However, as a matter of procedural course, we believe the scientific community can agree about what is and is not observable and that, by definition, the elusive twin in the dualist view—the ghost in the machine—is neither physical nor observable. The challenge for us is to see how far we can get in our understanding of consciousness without invoking dualism. Referring back to James' observation, our goal as scientists is the description and explanation of the snowflake without melting it.

Keeping this in mind, we can more adequately consider the following question: If we could fully specify the mechanisms and processes underlying consciousness in humans, then would reproducing or identifying these in a nonhuman be sufficient to produce (or attribute) consciousness in (to) that system? We would like to suggest that this question might be answerable only by a Cartesian version of the Turing test: If an entity that we assume a priori has consciousness—that is, a human being—interacts with the system under question, and cannot distinguish between its possession of consciousness and that of another human being, then the mechanisms we have specified are indeed sufficient to produce consciousness, at least as far as it is identifiable and studiable in a scientific fashion.

Once we have a system that passes the Cartesian Turing test, would we now be in the position to derive a complete set of criteria for consciousness? That is, could we now just figure out which components of the system are critical for allowing it to pass the test? Here, we would like to offer the following monition. Even if we have successfully identified consciousness in an entity, it does not follow that there is a distinct, specifiable set of criteria that precisely delimit the phenomenon. By analogy, we can say unequivocally that at noon it is day and at midnight it is night. But at what point does the night become the day? Our point is that it may be less useful to identify a rigid set of criteria for consciousness than it is to identify the variable or variables that form the continuum or dimensions along which consciousness lies. In other words, we would like to suggest that one key to success in understanding consciousness will lie in our ability to identify the relevant dimensions of neurobiological, psychological, and computational processing that define a continuum

between conscious and unconscious processes (cf. Farah et al., chapter 17; Hobson, chapter 19; Johnson & Reeder, chapter 13; Kinsbourne, chapter 16; Reber, chapter 8).

In conceptualizing consciousness as involving the interplay of continuous dimensions, we must be cautious to avoid some potential traps that can ensnare discussions of psychologically contintuous dimensions. In particular, positing continuous dimensions to consciousness does not necessarily rule out the possibility that the extremes along certain dimensions may rely on qualitatively different processes. For example, our perception of the transition from day to night may seem quite continuous, even though neurobiologically it entails a gradual shift in the relative contributions of two different, and qualitiatively distinct processing mechanisms: rods and cones. Thus, qualitatively different mechanisms may work together to instantiate processing along a continuous dimension. In short, it is quite possible that consciousness lies along a continuum while conscious and unconscious processes draw on distinct processes.

## A MULTIDISCIPLINARY APPROACH

Ultimately, answers to the questions raised here require the contributions of scientists addressing the issue of consciousness at a variety of different levels of analysis. We need to consider research addressing the low levels of consciousness associated with domains such as automaticity, subliminal perception, and implicit learning, and research addressing the high levels of consciousness associated with domains such as attention and metacognition. We also need to consider a variety of different research approaches, including the experimental, clinical, neurobiological, neuropsychological, philosophical, theoretical, and computational. By considering the contributions of scientists examining aspects of consciousness at different levels and from the vantage of multiple disciplines we believe it will indeed be possible to address questions about consciousness in a direct and scientific fashion. We hope this volume will contribute productively to this effort.

## ACKNOWLEDGMENTS

The writing of this chapter was supported by grants to both authors from the National Institute of Mental Health. Steve Fiore provided helpful comments on an earlier draft.

## REFERENCES

Dennett, D. (1991). *Consciousness explained.* Boston: Little, Brown.

James, W. (1952). *Great books of the western world: The principles of psychology.* Chicago: Encyclopedia Britannica. (Original work published 1890)

Minsky, M. (1985). *The society of mind.* New York: Simon & Schuster.

Newell, A. (1990). *Unified theories of cognition.* Cambridge, MA: Harvard University Press.