CONSCIOUSNESS

R. J. AUMANN
Center for the Study of Rationality, and Departments of Mathematics and Economics, The Hebrew University of Jerusalem, 91904, Jerusalem, Israel

1. Introduction

Consciousness is the last great frontier of science. Sixty years ago, life was not understood; it was a mystery. With the discovery of DNA, that mystery was solved; today, we more or less understand, at least in principle, how life "works." But we do not at all understand how consciousness "works."

We start by defining our terms; already there, there are difficulties. By "consciousness," we mean, in the first instance, the ability to *experience*. To see, hear, smell, feel, taste, desire, enjoy, suffer, like, dislike, love, hate, fear, become excited by an idea, be saddened by a loss. We do *not* mean to "sense." A machine—a "sensor"—also senses. Machines read; record sounds; detect odors, touches, and flavors; win at chess. They can even be programmed to exhibit a frownie, or emit "downbeat" sounds when something isn't right. But presumably they do not *experience*.

What, exactly, does "experience" mean? Ah, that is where the difficulty lies. The word cannot be defined in technical terms that do not refer to the concept itself. "Experience" cannot be defined in terms that a machine can understand. If you yourself are not conscious—have never experienced something—then you will not understand what the term means; just as little as congenitally deaf people understand what music means. To be sure, they can understand about the vibrations of taut strings, about air waves, about the workings of musical instruments, and even about musical notation and rhythm; but they can never understand what music *is*. For that, one must *hear* it. Similarly, someone who is not conscious cannot understand experience.

Ernst Mayr (1982) distinguishes between two fundamental questions in biology: "how" and "why." "How" refers to mechanism, "why" to function. The question "how do we digest food?" is answered by describing the process, involving saliva, chewing, swallowing, processing in the stomach and intestines, absorption into the bloodstream, and so on; and, disposing of wastes. The question "why do we digest food?" is answered by saying that food must be digested in order to provide vital ingredients for the functioning of the body, particularly energy.

To these questions, we add a third, which logically comes before the other two: "what." This refers to the descriptive aspect of biology, and of science in general. The answer to the question "what is digestion?" is that it is the process whereby food

is transformed to a state that the body can use. "What" questions also include observational, descriptive, and classifying matters, and also methodological or conceptual matters, like "what is a species?"

The remainder of this essay is divided into three sections: "What," "Why," and "How." In the first, we discuss what consciousness is, and how it differs radically from other scientific phenomena. In the second, we discuss a possible function of consciousness, from the evolutionary viewpoint; and the third discusses the mechanism. Unfortunately, the "How" section is particularly short: we really have nothing to say about this, other than to describe the difficulties.

2. What

We have defined consciousness as the ability to experience. This puts the phenomenon into a completely different category from other scientific phenomena, in several ways.

(i) Unlike almost every—indeed every—other scientific phenomenon, consciousness is completely *subjective*. No verifiable outside characteristics of consciousness are known. No matter how complex an organism's behavior is, a computer could conceivably be programmed to mimic that behavior. An individual can observe consciousness only in himself.

Specifically, *I* can observe consciousness in myself only; I cannot be certain that anybody else really is conscious. To be sure, since other people appear roughly similar to me, and act similarly, I may surmise that they, too, are conscious; but I'm not *certain*. Each individual can be certain of consciousness *only* in himself, where he directly observes it.

(ii) Whereas the phenomenon of consciousness is highly subjective, it is, paradoxically, the *only* phenomenon of which the observer is absolutely certain. All other phenomena and observations could conceivably be attributed to hallucinations, dreams, and/or mental illness. But also hallucinations, dreams, and the ravings of a madman are experiences; in each case, the observer is *sure* that he is experiencing—is "conscious"—and he is right (we include dreams under the heading of consciousness).

(iii) Some people express perplexity as to the nature of the problem. They do not see anything mysterious about consciousness, and do not understand in what way it is different from other neurological functions like, say, the regulation of breathing. Asked whether a computer could in principle be conscious, they answer, "why not?"

We are dumbfounded by this reaction, and can only conjecture that these people are themselves not conscious. To me, it is evident that no combination of silicon chips and wires could conceivably "experience" in the sense that I do. Consciousness involves something beyond the merely physical and mechanical.

(iv) It seems only slightly less evident that no combination of off-the-shelf chemicals could experience in the sense that I do. But here, we are entering a gray area. In an e-mail message dated March 22, 2005, the U.S. National Academy of

Sciences announced that "a worldwide ban on human reproductive cloning—a technique that attempts to produce a child—is justified." If it is being banned, it is presumably within reach. Thus, the day seems not far off when it will be possible to synthesize a human being. Will such a golem be conscious?

Each of the possible answers—"yes" and "no"—is problematic. "Yes" is problematic because a combination of chemicals is in principle no different from a combination of silicon chips and wires, which we intuitively feel cannot experience. But "no" is also problematic, because there is no reason to believe that a golem that is identical, molecule for molecule, with a live human being would not in all respects—including consciousness—be like that human being.

(v) Are animals conscious? On the face of it, there is no reason to suppose that they cannot be. But as stated above (in (i)), it is not even clear that all humans—other than me—are conscious. By analogy with me, I can surmise that other human beings *are* conscious; but the analogy is less compelling in the case of animals. The further one gets from human beings on the evolutionary scale, the less compelling the analogy. So the short answer is, "possibly; we don't know."

(vi) Consciousness may have levels. For example, dreaming is certainly an expression of consciousness, but perhaps at a lower level than waking consciousness. Newborn children, and the mentally impaired, may be conscious at a lower level. In the opposite direction, people taking certain drugs sometimes report a "heightened state of consciousness."

Here again, we are at a loss, because we cannot really imagine what it is like to be, say, a newborn child. We personally have never taken drugs, so cannot make a judgment in the opposite direction either. We are stuck in our own conceptual prison: consciousness is about subjective experience, so it is difficult to imagine levels of consciousness other than our own.

Though it may have various levels, its existence at any level already poses the conceptual problems discussed here.

(vii) Conscious experience appears to be associated with certain physiological-neurological processes in the brain: Prof. Rafael Malach of the Weizmann Institute of Science has reported that it is associated with the simultaneous firing of many neurons in a well-defined group of neurons. (Based on this observation, Malach makes the fascinating suggestion that a *group* of individuals acting in concert may also be conscious.) This, however, does not explain how consciousness works—just as little as noticing that human reproduction is associated with sexual intercourse explains how reproduction works.

(viii) Up to now we have discussed only the "input" component of consciousness: experience. There are also two other vital components. One is the "processing" component: thought, including intention. The other is the "output" component: volition—consciously choosing to do something, and doing it. True, a person could be conscious, but have no power to take any action—as when asleep, or as a result of a totally debilitating stroke. But under most normal circumstances, thought and volition are intimately associated with consciousness. Indeed, as we

shall presently see, it is the combination of all three elements—experience, thought, and volition—that enables consciousness to perform its function.

3. Why

We next address the issue of "why": what is the *function* of consciousness, from, say, the evolutionary viewpoint?

The answer we propose is based on the two related notions of *decentralizing* and *decoupling*—roughly speaking, splitting a difficult or complex task into several easier or less complex tasks, often with the aid of an auxiliary *"driving force."* Here are some examples:

(i) **Tearing a Manhattan telephone book in two**—perhaps the granddaddy of all decoupled processes. Taken as is, it is very difficult. But if one first separates the book into a number of thinner parts, then one can easily tear each of these parts in two, thus accomplishing the task. Here the difficult task is tearing the whole book; the easier tasks are tearing each of the thinner parts; and the decoupler is separating the entire book into the thinner parts.

(ii) **Operating an economy.** An economy can be centrally planned, as in a Kibbutz (Israeli cooperative village). A central planner decides how much of each good will be allocated to each individual, how much—and where—each individual will work, and so on. In theory, the entire economy of a whole country could be planned in this way. That is the conception behind socialist economies, like that of the former Soviet Union.

Centrally planning an entire economy is enormously complex and difficult. To start with, the informational requirements—finding out what each person wants to consume and what he is capable of producing—seem utterly beyond reach. But even if that could somehow be achieved, the problem of getting the people to do what you want them to do, and the sheer complexity of the task, makes the efficient central planning of an entire economy practically unachievable. Indeed, socialist economies like that of the Soviet Union achieve levels of efficiency that are far below those of advanced "free" economies.

What enables free economies to work more efficiently is that they are, by and large, decentralized. Within limits, each person seeks to acquire the goods and services that he wants, and to work at the tasks that he wants and that he is able to perform. There is no problem in getting the people to do what you want them to do, as each individual makes his own choices. And, the complexity of the task is greatly reduced, as the task of operating the economy is transferred from a central planner to individuals, each planning only the segment of the economy that interests him—usually, his own consumption and production. Instead of one planner (or planning agencies) planning many billions of transactions, we have several million planners planning thousands of transactions each. The total size of the task is perhaps roughly the same, but the task is *decoupled* into millions of individual tasks—and so is much more easily accomplished.

The driving force that makes the whole process work is *individual motivation*. Each individual is motivated to seek for himself the best "deal" that he can get—the goods and the work that he likes most—and it is this that operates the entire economy. That is Adam Smith's "invisible hand."

(iii) **Money and prices.** Primitive decentralized markets work by barter: two or more people get together and exchange goods or services, to the mutual benefit of all parties to the transaction; the process may be repeated as often as desired. One may hope that the final outcome is *optimal*, in the sense that no traders could have done better by trading with each other (more precisely, that no group of traders could have improved the welfare of all its members by trading within the group only. In economic theory, such outcomes are called *core* outcomes). In practice, in reasonably large markets—or even in fairly small ones—identifying the bartering groups, and deciding on the barters to be implemented, is so complex, involved, and fraught with uncertainty that barter is unlikely to achieve an outcome that is anywhere close to optimal.

Enter the institutions of money and prices. Rather than bartering, each trader sells his goods at market prices; with the proceeds he buys the goods he desires, again at market prices. Then if the prices are "right," the market "clears": the demand for each good exactly matches the supply. Moreover, the resulting reallocation of goods is optimal (in the above sense); and in large markets, *all* optimal outcomes are achievable in this way.

Thus, the unsophisticated but highly complex, involved, and uncertain bartering process is replaced by the price mechanism, which, though a lot more sophisticated than barter, is far easier to execute. In barter, each trader needs some knowledge about the preferences of the traders with whom he trades, and also about at least some of the others, so that he will know with whom to trade. But the price mechanism requires only that each trader buy what *he* likes, given *his* budget (the proceeds from the sale, at market prices, of the goods that he brought to the market).

Here the complex task is achieving optimality; the simpler tasks are for each trader to decide what he wants, given his budget; and the decoupler is the price mechanism. And again, the driving force is *individual motivation*. Each individual is motivated to "maximize over his budget set"—sell and buy, at the given market prices, so as to be left with those goods and services that he most prefers.

An interesting aspect of this decoupling is that historically, it has emerged by itself in *every* reasonably advanced society, without being imposed by any planning entity. Not only has the price mechanism emerged by itself, but the market prices themselves also usually emerge by themselves—determined by supply and demand—without being imposed by planning entities. And when planning entities do enter the process of price determination, as in the former Soviet Union or with rent control in various cities, they often wreak havoc, causing shortages and other distortions.

Whereas this example is related to the previous one (Example ii), they are not the same. In the previous example, the point was decentralizing the economy, letting each individual see to himself. In the current example, the point is the formation of a

price system. Logically, the examples are independent; a centrally planned economy can have a price system, and a decentralized economy can work on barter.

(iv) **Chess.** The ideal way to play chess is to plan the entire course of play from the beginning, taking into account anything the other player might do. In practice, this is beyond the power of any man or machine. Instead, the players "evaluate" the situation at each move, using numerical indices for the various pieces; e.g., 8 for the queen, 5 for a rook, 3 for a bishop or knight, 1 for a pawn. They also take account of the general characteristics of the position: development of the pieces, castling, passed pawns, and so on. Using such criteria, each player "looks ahead" for a few moves, trying to maximize his valuation of the position at the end of that period, and taking into account that the other player is doing likewise. Weaker players often do not look ahead more than a move or two, and even that only partially. Stronger players may look as much as five or six moves ahead, and sometimes even more; but they, too, do not examine all possibilities—all "branches of the tree"—in the process. Human players do not use a precise numerical valuation, but take a generalized view. Chess-playing computer programs basically do use a precise numerical valuation function; though even there, the "depth" of the look-ahead varies, with some branches of the tree being examined more thoroughly than others.

Here the complex task is planning the whole game beforehand; and the simpler, decoupled tasks are playing move by move, with a more or less modest look-ahead. The overall, macroscopic driving force is, of course, the desire to win; but "microscopically"—at each separate move—the driving force is to maximize the valuation several moves ahead.

In Examples (ii) and (iii) above, the decoupling is achieved by spreading the task over many individuals, each with his own motivation. In contrast, in this case the decoupling is achieved by spreading the task over time. There is a single overall motivation—winning—which is expressed at each move by looking ahead a few moves and maximizing valuation.

(v) **Solomon's judgment.** Rather than rendering his judgment (1 Kings 3, 16–28), King Solomon could have entered into a complex factual investigation of the women's claims. He "decoupled" the process by motivating them unwittingly to reveal the truth themselves. Here the complex task is determining which woman is the live baby's mother; the simpler tasks are for the women to express their preferences given the judgment; and the driving force is the women's motherly love.

(vi) **Fair division.** This may be achieved by cumbersome methods of direct measurement. Alternatively, the process may be decoupled by the method of "one cuts, the other chooses," which motivates the parties themselves to divide fairly.

(vii) **Evolution.** Suppose a Creator had wanted to create the living world as we know it. He could have designed each individual organism separately, together with the appropriate interactive adaptations. This would have been enormously complex. Alternatively, He could have decoupled the process by means of evolution—survival of the fittest—which runs by itself, automatically, with no need for "hands on" direction. Here the complex task is creating the world; the decoupled alternative

tasks are for each organism—or even each gene—to adapt to its surroundings; and the driving force is survival of the fittest.

These examples should give the reader some idea of what we mean by decoupling. What we now suggest is that consciousness serves as a decoupler that allows human beings to perform tasks that otherwise would be unachievable because of their complexity. Let us illustrate.

(a) **Food.** The body needs food for energy and other vital purposes. The process of supplying food may be divided into two parts: (1) before it enters the mouth, and (2) afterwards. Part 1 consists of triggering the individual to seek food, and acquiring, preparing, and eating it. Part 2 consists of digesting it, i.e., transforming it into a state that the body can use (see Section 1, above).

Both parts are extremely complex, but Part 1 is by far the more so. We must choose the food to buy, shop for it, store it, clean it, cook it, and serve it. We must also choose and buy closets and a refrigerator to store it; sinks to wash it; stoves, ovens, gas tops, microwaves, pots, pans, cake forms, pie dishes, utensils, mixers, vegetable peelers and so on to prepare it; plates, bowls, platters, and utensils, to serve it. We must earn money to buy the food and all the auxiliary items we have mentioned, and those we have not mentioned. Earning money usually involves various skills—not the least of them social—which must be acquired.

There is another fundamental difference between the two parts. Part 2 is "hard-wired": it works "automatically," by mechanical, chemical, and electrical means, with no conscious, voluntary component. Part 1 is precisely the opposite: all conscious, all volitional. To start with, when the body runs short of the required nutrients, it must be prompted to eat. Once this is done, the food must be acquired, prepared, and eaten. Conceivably, parts of the process *could* be hard-wired: a gauge could tell the brain when the stomach is empty; the eye and brain could identify food, then send signals to the hand automatically to take it and put it in the mouth. But it seems unlikely that acquiring the food and preparing it, and earning the necessary money, could all be hard-wired.

How, then, does the process work? What drives it?

The answer is simple, even obvious: hunger. And, the other side of the same coin: enjoyment of food.

Hunger does *not* mean an empty stomach. An empty stomach by itself will not prompt us to eat. We need the pain or discomfort of hunger, and/or the pleasure of food, to make us eat. Pain and pleasure depend on consciousness. If you are not conscious, you cannot experience pain or pleasure. So consciousness is an important component of the mechanism that supplies us with nutrients. Moreover, it decouples the nutritive process into the two parts we have described: until the food reaches the mouth, and afterwards.

Indeed, it does much, much more: the entire process of acquiring, preparing, and serving food is decoupled into many small steps. Each step is conscious, with a well-defined goal; it is *motivated.* Going to market, picking the items to buy, standing in line at the check-out counter, bringing the items to your car, unloading, putting in the refrigerator, all the myriad tasks involved in cooking, all the myriad tasks involved in earning the money with which to buy the food, all the myriad tasks

involved in all stages of the process—all together, and each one separately—are driven by hunger, through the medium of consciousness; they are *motivated.*

More precisely, the experiences of hunger and food enjoyment are the overall, macroscopic driving forces, like the desire to win in chess (Example iv above). The thought process, which is the second component of consciousness (see Section 2, Item viii), translates this into many small tasks—making money, baking a cake, and so on—each with its own driving force, like valuation-maximizing in chess. And then volition—the third component of consciousness—comes into play, enabling the individual actually to carry out these individual tasks.

It is important to note that while "hard-wired" processes like digestion may be highly complex, they are fairly repetitive. Digestion works on the same materials, in the same way, every day; there are few, if any, surprises. The processes of food acquisition and preparation are much less repetitive; they require a good deal of instantaneous adaptation to various different environments, environments that may be unfamiliar. Consciousness is particularly important for motivating and carrying out these non-repetitive tasks.

Finally, we remark that the mechanisms provided by nature to facilitate eating—hunger and the enjoyment of food—may sometimes "misfire." It is well known that severely undernourished people whose hunger leads them to overeat may well die as a consequence; there are documented cases of people who survived the concentration camps during the Holocaust, only tragically to die in this way. Even better known is that the enjoyment of food may cause people to overeat or to eat foods that are not nutritious. While not immediately fatal like with the concentration camps, this may nevertheless be detrimental to health; or at least, serve no useful adaptive purpose.

(b) **Sex.** Biologists identify two basic drives in living organisms: nourishment and reproduction. What we have said about food applies, *mutatis mutandis*, also to sex. It is hunger and the enjoyment of food—*not* the need for nourishment!—that makes people eat. Similarly, it is *not* the desire for offspring that makes them have sex; it is the sex drive—the enjoyment of sex. Enjoyment is a function of consciousness. You cannot enjoy if you are not conscious. Many people do consciously want offspring, but that is not why they have sex.

As with nourishment, the process of reproduction has several distinct parts: (1) before the semen enters the woman's body; (2) conception and pregnancy; and (3) childhood, when the offspring cannot fend for itself. The first two parts are decoupled by the sex drive. The second part, though highly complex, is hard-wired; once sex has taken place, there is no conscious intervention until birth. After birth, parental love takes over; like hunger and the sex drive, this depends on consciousness. As with food, the first and third parts of the process—which are not hard-wired—split up into a myriad of distinct small steps, starting with dating and earning the required money; each is conscious, each has a well-defined goal, each is motivated.

Like with food, the hard-wired part of the process is fairly repetitive, the other parts not.

Unlike food, sex is programmed to be profligate. In a single season, a single flowering tree produces many hundreds of thousands of blossoms, and billions of grains of pollen. It is doubtful if even one of these comes to fruition. In each ejaculation there are hundreds of millions of sperms, at most one of which is used. Many sexual episodes lead nowhere, many relationships lead nowhere, many dates lead nowhere. The sex drive leads to many activities that have no chance of producing offspring: sex with birth control, sex after the reproductive age, homosexuality, masturbation, oral sex, bestiality, pornography, and so on. Love and sex play dominant roles in advertising, literature, film, music, painting, photography, dance, almost all cultural activity. In the case of food, we used the term "misfire" to describe situations in which food does not provide nourishment. In the case of sex, the corresponding situations are so ubiquitous that they must be considered a part of the process.

(c) **Pain.** It has long been recognized that pain calls attention to something in the body having gone wrong, so that it can be attended to. Again, pain depends on consciousness; if you are not conscious, you cannot experience pain. Machines cannot suffer.

As before, we have here a decoupling process. Pain motivates the individual to seek medical or surgical treatment. Unlike with food and sex, though, the effect of treatment is not entirely automatic; repeated intervention may be required. Pain decouples the treatment process into many distinct small steps, motivating the individual at each step to do what is required in order to alleviate his condition. It, too, may "misfire," as when a medical or surgical treatment causes more pain than what the patient can suffer.

To summarize: consciousness enables the decoupling of highly complex, non-repetitive tasks into many simpler tasks, mainly through the element of motivation.

4. How

This is the shortest of our sections: we have little to contribute on this score, other than to say that the neurological phenomena that have been observed to be associated with consciousness do not explain how it works; the "how" remains a deep mystery.

One last remark is worth making. "How" questions are usually answered by analogy with something else, with which we are familiar. For example, the workings of the circulatory system are explained by analogy with plumbing. But consciousness is unique; there is nothing else in the world that is even remotely like it. Since there is nothing like it, what analogy can we use to explain it? And if not by analogy, how else can it be "explained"?

5. Summary

Consciousness is the last great frontier of science. This chapter discusses what it is, how it differs fundamentally from other scientific phenomena, what adaptive function it serves, and the difficulties in trying to explain how it works. The emphasis is on the adaptive function.

6. References

Ernst Mayr, E., (1982) *The Growth of Biological Thought*, Belknap Press, Cambridge, Massachusetts.