

האוניברסיטה העברית בירושלים

THE HEBREW UNIVERSITY OF JERUSALEM

WHY CONSCIOUSNESS?

By

ROBERT J. AUMANN

Discussion Paper # 740 (May 2021)

מרכז פדרמן לחקר הרציונליות

THE FEDERMANN CENTER FOR
THE STUDY OF RATIONALITY

Feldman Building, Edmond J. Safra Campus,
Jerusalem 91904, Israel

PHONE: [972]-2-6584135 FAX: [972]-2-6513681

E-MAIL: ratio@math.huji.ac.il

URL: <http://www.ratio.huji.ac.il/>

Why Consciousness?

Robert J. Aumann*

April 26, 2021

Abstract

Emotions—specifically desire and the objects of desire, like enjoyment and satisfaction—drive much of what we do; indeed they drive *all* we do that is not recurrent. They are thus indispensable to human life. Inter alia, emotions enable the operation of incentives—like hunger for eating—that motivate us to perform tasks that are vital to our lives. We suggest that the adaptive function of consciousness is to enable emotions to operate.

*Department of Mathematics and Federmann Center for the Study of Rationality, The Hebrew University of Jerusalem; raumann@math.huji.ac.il

1. What, how, why

Biologists ask two kinds of questions: (i) How does it work? And (ii) why did it evolve—what is its purpose, its adaptive¹ function? To these we add a third, which perhaps should come before the other two: (iii) What is it? What are we talking about?

An example is digestion. **What** is it? It is the transformation of food to a state the body can use. **How** does it work? The answer involves chewing and saliva, the operations of the stomach and intestines, waste disposal, and many other processes. **Why** are we endowed with this capability—what is its purpose? To provide energy, enable growth, and deliver ingredients for vital bodily functions and maintenance (e.g., calcium for bones).

Another example is hearing. Already there it's not entirely easy to give a straightforward answer to the “what” question; we return to that below. The “how” question is answered by referring to air waves, the ear and its components, the relevant parts of the brain, and neural connections thereto. As for “why:” Hearing is used for communication, for becoming aware of dangers and opportunities, and for other purposes, like experiencing music—which is also adaptive, though perhaps less clearly so.

This essay concerns consciousness—perhaps the last great enigma of science.² Almost all the attention that this phenomenon has received has been directed at the “how” question. Here we address the “why” question. Needless to say, before explaining “why”, one must answer the “what” question. We also discuss “how”, though only briefly; it is not our main concern here.

2. *What is consciousness?*

Consciousness has many manifestations, in at least four categories. First, we experience the outer world: see, hear, smell, taste, touch and are touched; feel hot or cold, wet or dry, rough or smooth texture; experience long or short durations, heavy or light loads, Second, we experience our inner worlds, our emotions: we desire, enjoy, suffer, like, dislike, love, hate, fear, admire and experience admiration, despise and experience contempt; are excited, indifferent, angry, calm, satisfied, dissatisfied, glad, sad, interested, bored, pleased, pained, comfortable,

¹In the sense of adapting to the environment, evolutionary advantage.

²See Section 3 below.

uncomfortable, Third, we think, analyze, and remember consciously. Fourth, we intend, plan, carry out intentions and plans, act volitionally.³

There may also be other categories; some of the manifestations may belong to more than one category; and there may be aspects of consciousness that we have missed altogether.

Several remarks are in order. First, when we say experience, we do *not* mean sense. Many of the above phenomena, like sight, are often called senses; but machines also “sense”—detect and record activity of one kind or another and respond in some specific way. But machines do not *experience*.

We also think and remember consciously; and while machines may perform tasks with functions reminiscent of thinking and remembering, they do not think or remember consciously, just as they do not experience.

If not sensing, then what *do* we mean by “experiencing”? Ah, that is the difficulty. The word cannot be defined in terms that do not refer to the concept itself. If you yourself are not conscious—have never experienced something—then you cannot understand what the word means; just as little as congenitally deaf people can understand what music is. 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.

Second, we started this section by listing various manifestations of consciousness. But though consciousness has many manifestations, it is a single phenomenon. Like electricity—a single phenomenon with many manifestations.

Third, unlike every other natural phenomenon, a person’s consciousness is completely subjective—can be verified only by that person. No matter how complex an organism’s behavior is, a computer could conceivably be programmed to mimic that behavior.

Specifically, *I* can verify consciousness in myself only; I cannot be certain that anyone 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 can’t be *certain*. Each individual can be certain of consciousness only in himself, where he directly observes it.

³Including the operation of limbs and other body parts.

3. *How* does consciousness work?

The short answer is that we simply don't know.

A longer answer is that while we do have some idea of the detailed workings—the “neural correlates of consciousness”—we have no idea at all of the big picture, the underlying phenomenon, the so-called “Hard Problem.”⁴ For an incomplete analogy, imagine a group of astute seventeenth century scientists arriving in the twenty-first century by time machine, and being given a pc to examine. Eventually, they get to “understand” some of the hard- and software; but having no inkling of the phenomenon of electricity, they do not fathom how the device works, and specifically, how it is that it works if and only if a pair of wires emanating from it is stuck into a pair of holes in the wall.

It has been suggested that consciousness is transcendental—cannot be explained in ordinary scientific, “physical” terms. We do not espouse this view. Though it is tempting to ascribe transcendental to phenomena that one has tried without success to understand, one must overcome the temptation. When I was young, transcendental was ascribed to life; since the discovery of DNA in the early nineteen-fifties, the phenomenon of life is in principle understood (though many specifics remain unexplained). Similarly, one may expect that sooner or later, we will get to understand how consciousness works.

4. *Why* are we conscious?

4.1. Emotions

We come now to the main object of this study—the “why” question. What purpose does consciousness serve? What (evolutionary) advantage does it confer?

Most scientific studies of consciousness and its neural correlates are centered on the first category of conscious phenomena: experiencing the outer world—seeing, hearing, smelling, tasting, and so on. This category is the more amenable to ordinary scientific methods, inter alia objective observation and measurement. But by the same token, the tasks it accomplishes are more similar to those that can be accomplished by inanimate machines: As pointed out above, though machines cannot experience, they can and do “sense” sights, sounds, odors, tastes, contact, temperature, duration, weight, and so on. Similarly for the third category of

⁴A term coined by the philosopher David Chalmers in his book *The Conscious Mind: In Search of a Fundamental Theory*, Oxford: Oxford University Press, 1996.

conscious phenomena: Machines engage in activities analogous to thinking, analyzing, and remembering. And in part, also for the fourth category: Though indeed machines cannot intend, they can and do carry out intentions.

Contrast this with the second category of conscious phenomena—emotions, experiencing our inner worlds—of which machines are incapable.⁵ Machines cannot desire; they cannot feel pain or discomfort or hatred; they cannot enjoy or love. AI—artificial intelligence—has made, and is making, tremendous strides; but AE—artificial emotion—has not gotten off the ground, indeed has not even been proposed. In the sequel we will argue that emotions are the basic driving force of human activity; it is they (with the support of the other three categories of conscious phenomena) that enable human life.

To start with, we discuss incentives, which we will show are indispensable for human life.

4.1.1. Incentives

Recall that an “incentive” is a condition that makes people want to act—or refrain from acting—in a specified way, for reasons not directly related to the action in question.⁶ For example, the threat of imprisonment is an incentive to refrain from crime; executive bonuses are incentives for managers to manage their companies profitably; wages are incentives for workers to do the work that the employer specifies.

To operate, incentives require consciousness, three times over. First, they are based on desire—they make people **want** to act in a specified way; and desire is an emotion, a function of consciousness—you can’t want if you’re not conscious. Second, the desire invariably involves the prospect of a conscious emotional experience like enjoyment, fulfillment, or suffering. Finally, the actions engendered by the incentives are volitional, and volition is a function of consciousness.

Living organisms have many needs. Most basic are nourishment, reproduction, and self-preservation. Conceivably, these needs could be met without consciousness; as they are, for example, in plants. But in human beings, and perhaps also in some animals, each of the three basic needs is in fact met by processes in which incentives—and so consciousness—play a fundamental role. We examine the three in turn.

⁵At least as of now.

⁶After some hesitation, we decided that this last requirement is indeed part of the usual meaning of “incentive.” For example, one would not say that to have bread is an incentive to buy bread.

Nourishment Nourishment—food—is needed for energy, growth, and providing vital ingredients for bodily functions.

But that is not what makes people eat. People eat because they are hungry, and/or because the food is tasty. Hunger is uncomfortable; eating tasty food is pleasant. People want to eat in order to alleviate the discomfort of hunger, and/or to enjoy the pleasure of consuming tasty food. They then take purposeful action to satisfy that desire: They eat.

Note that physiologically, hunger is not in principle a necessary concomitant of an empty stomach. One can conceive of organisms that consume food unconsciously, like the oxygen that we breathe. Rather, hunger evolved specifically to motivate eating.⁷ Similarly, the tastiness of food is not a necessary concomitant of its nutritive value; it, too, evolved specifically to motivate eating.

Indeed, both hunger and food enjoyment may “misfire”—motivate eating when eating is not adaptive, even counteradaptive. For example, in the obese. Or, at the opposite end of the spectrum, in people whose stomachs have shrunk because they have eaten too little for too long.⁸

The bottom line: Hunger and food enjoyment are the prime incentives to eat.

Reproduction Sex is required for reproduction, the driving force of evolution. But mostly, people do not have sex because they want children. People have sex because they enjoy it.

Here again, physiologically, enjoyment is in principle not a necessary concomitant of sex; it evolved specifically to motivate having sex.

Indeed, like hunger, the sex drive may “misfire”—motivate having sex when it is not adaptive. As with sex with birth control, sex after the reproductive age, homosexuality and lesbianism, oral sex, bestiality, and masturbation; and sex may even be counteradaptive, as with sexually transmitted diseases.

The bottom line: Enjoyment is the prime incentive to have sex.

Self-preservation An incentive for self-preservation is the phenomenon of pain. Pain calls attention to a problem in the body, something that needs fixing—and then motivates us to fix it.

⁷That is, the emotion of hunger evolved because it motivates individuals possessing it to eat, and so increases their fitness.

⁸As happened tragically to Holocaust survivors whose stomachs burst after eating all they wanted upon being released from concentration camps.

Again, physiologically, pain is not a necessary concomitant of problems in the body. Indeed, some people cannot feel pain; such people are in deep trouble, and usually do not survive long. Also, there are problems in the body—like most cancers—that do not cause pain until it is too late to address them. Thus, like hunger and the sex drive, pain may “misfire”—motivate trying to fix a problem in the body when it is too late.

The bottom line: pain provides an incentive to fix problems in the body.

More generally, self-preservation (including self-defense and collective defense) is based on the desire to live—to experience life—and both desire and experience are functions of consciousness.

This, too, may and does “misfire.” Living is adaptive only until the end of the reproductive age (it being understood that “reproduction” includes raising offspring to maturity). Beyond that, there is a transition stage during which people may still be useful to their offspring; but at some point, they become more of a burden than a help. Nevertheless, the desire to experience life continues, and society accedes to that desire.

4.1.2. Beyond the basic needs

Above, we adduced specific incentives that motivate meeting the three basic needs: hunger and food enjoyment for nourishment, the sex drive for reproduction, pain and the desire to live for self-preservation. We will now argue that emotions (like desire, happiness, satisfaction, disappointment, pleasure, interest, boredom) play a far larger role in meeting these basic needs, and in fact all our needs.

To start with, nourishment is not only a matter of eating; one must also obtain the food to eat. Specifically, one must choose the food to buy; shop for, store, clean, cook, and serve it; choose and buy ovens, closets, refrigerators, sinks, stoves, microwaves, pots, pans, utensils, . . . ; earn money to buy the food and the foregoing; acquire the skills required to earn the money. And so on.

In the short term, it is indeed hunger and food enjoyment that motivate eating. But we do not think of hunger and food enjoyment when we study in schools or universities to acquire the skills required for earning the money required to buy the food and buy the appurtenances required to prepare the food. Indeed, while acquiring the skills, we do not even think directly of earning the money. We want to get good grades in examinations, are happy or satisfied when we do, disappointed when we do not. We want to study subjects that interest us. And when buying sinks, stoves, closets, refrigerators, we are not hungry. We have some

notion of what we want, seek it out, and buy it when we're satisfied or pleased. Perhaps hunger lurks somewhere in the backs of our minds, but nowhere near the front.

Like getting food, reproducing is complicated. One must choose a date, get her or him to go out with you, choose what to do on the date, buy clothes, give presents, marry or otherwise choose a partner of the opposite gender, bring up children—part of the reproductive process!—earn the money to pay for the foregoing, acquire the skills required to earn the money. And so on.

Also self-preservation—health and defense—is in practice highly complex, and consciousness plays a vital role in the process.

Examinations offer an incentive to study hard. Indirectly, doing well in examinations enables us to obtain nourishment, sex, health, self-defense. But the direct motivation is getting a good grade—not hunger, not enjoying food or sex, not avoiding pain. When shopping for a dress, we want something attractive and not too costly. When deciding on a career, we look for something interesting, enjoyable, something we do well. We want to get job promotions not only because they usually imply a rise in salary, but also because they nourish our ego. When going on a date, we look for an interesting movie, an enjoyable concert, a memorable theater performance; we are not directly motivated by the sex drive.

Indeed, just about everything we do is driven by consciousness. First, we **want** to achieve a certain goal; desire is an emotion, belonging to the second category of consciousness. Second, we want to achieve the goal **because** it will lead to some emotion such as pleasure or satisfaction, or because we are angry or sad. Third, the inner experience (emotion) is often⁹ triggered by an outer (first category) experience—seeing a beautiful flower, hearing a moving song, avoiding foul odors or repellent tastes, being aroused by an intimate touch.

In brief, we **want** to achieve the goal; decide what to do in order to achieve it, taking into account what is involved (easy or hard, pleasant or unpleasant, ...); and **intentionally** do it. Desire, experience (outer and inner), and volition—all are functions of consciousness, and they are what enables us to meet the three basic needs, and indeed all our needs.

⁹But not always. For example, if we hear that a loved one has died, we are sad because of the knowledge that has been conveyed, not because of the conscious experience of hearing.

4.2. Alternatives to consciousness?

4.2.1. First Proposal

Though we do not know how consciousness works, presumably it is via the nervous system. One question is, what is the added value of consciousness? Why can't the self-same neurological processes that produce consciousness—and so make us do what we do—operate without consciousness?

We answer this question with an imaginary dialogue.

Q. What is artificial heat good for?

A. Inter alia, for cooking. Frying an egg, for example.

Q. How is artificial heat produced?

A. With an electric stove top, for example.

Q. How does the electric stove top work?

A. Frankly, I myself don't know. It's a highly complex process, involving electricity, which I don't quite understand, the effect of electricity on certain materials, and a whole bunch of other stuff. Though I don't understand it, I see that it works; it makes the heat, which I *do* understand.

Q. Well then, why is the heat really needed? You could use the same process, with the same electricity, the same materials, and the same other stuff that you don't understand, and fry the egg that way. Why do you need the heat?

A. That's a strange question. First, it's the heat that fries the egg, not the process of creating it. No matter how you create it, without the heat itself, the egg won't fry. Second, if you use the same process, you'll get the same result—heat—so the egg *will* fry.

Consciousness is similar. First, as explained above, it itself—not the neurological processes that produce it—makes us act. And second, if you use the same neurological processes, you'll get the same result—consciousness.

4.2.2. Second Proposal

Another question is, is consciousness really necessary? Could not our needs be met in some other way, not involving consciousness—for example by some complex neurological process resembling deep learning?

The answer is that indeed it might be possible to meet the three basic human needs—as well as other, less basic needs—in a different way, not involving consciousness. The biological world is highly diverse, provides many different mechanisms for meeting the needs of organisms. Among them is consciousness, but

there is no reason to think that other systems could not work. Here we explain the function of this specific mechanism; we don't claim exclusivity for it.

4.2.3. Background

Underlying the above two proposals is the feeling that consciousness is somehow problematic and mysterious, and that it is desirable to dispense with it when studying evolutionary biology. Indeed, there are first-rate scientists who are so uncomfortable with consciousness that they deny its existence. Needless to say, if it did not exist, there would be no call to explain its purpose.

But it does exist.¹⁰ Though we do not understand how it works, it is a fact of life. And its adaptive function—like that of all physiological features—calls for elucidation.

5. Summary

The fundamental function of consciousness is enabling the operation of emotions. Emotions—specifically desire and the objects of desire—drive much of what we do; indeed they drive *all* we do that is not recurrent¹¹. Consciousness is thus indispensable to human life as we know it.

¹⁰Indeed it is the only scientific phenomenon of whose existence we may be absolutely certain—all others 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, and he is right.

¹¹like breathing or digesting.

HEBREW UNIVERSITY OF JERUSALEM
מרכז פדרמן לחקר הרציונליות
THE FEDERMANN CENTER FOR THE STUDY OF RATIONALITY

List of Recent Discussion Papers

These and earlier papers can be found on our website: www.ratio.huji.ac.il

706. **Abraham Neyman , Elon Kohlberg**, Cooperative Strategic Games (February 2017)
707. **Ohad Dan, Drorith Hochner- Celnikier, Amy Solnica, Yonatan Loewenstein**, Association Of Catastrophic Neonatal Outcomes With Increased Rate Of Subsequent Cesarean Deliveries (April 2017)
708. **Bezalel Peleg, Shmuel Zamir**, Sequential Aggregation Of Judgments (May 2017)
709. **Johannes Müller-Trede, Shoham Choshen-Hillel, Meir Barneron and Ilan Yaniv**, The Wisdom of Crowds in Matters of Taste (May 2017)
710. **Elon Kohlberg and Abraham Neyman**, Games of Threats (September 2017)
711. **Sergiu Hart**, Repeat Voting: Two-Vote May Lead More People To Vote (October 2017)
712. **Sergiu Hart and Philip J. Reny**, The Better Half of Selling Separately (December 2017)
713. **Shevy Waner, Uzi Motro, Yael Lubin, and Ally R. Harari**, Male mate choice in a sexually cannibalistic widow spider (February 2018)
714. **Maya Bar-Hillel, Tom Noah, and Shane Frederick**, Learning psychology from riddles: The case of stumpers (February 2018)
715. **Bezalel Peleg and Ron Holzman**, Representations of Political Power Structures by Strategically Stable Game Forms: A Survey (February 2018)
716. **Kristoffer Arnsfelt Hansen, Rasmus Ibsen-Jensen, and Abraham Neyman**, The Big Match with a Clock and a Bit of Memory (February 2018)
717. **Bezalel Peleg and Hans Peters**, Self-implementation of social choice correspondences in strong equilibrium (April 2018)
718. **Abraham Neyman**, Additive valuations of streams of payoffs that obey the time-value of money principle: characterization and robust optimization (April 2018)
719. **Bezalel Peleg and Shmuel Zamir**, Judgements aggregation by a sequential majority procedure (June 2018)

720. **Lotem Elber-Dorozko and Yonatan Loewenstein**, Striatal Action-Value Neurons Reconsidered (July 2018)
721. **Orli Bobek, Adiv Gal, David Saltz, and Uzi Motro**, Effect of Nest-Site Microclimatic Conditions on Nesting Success in the Lesser Kestrel (*Falco Naumanni*) (July 2018)
722. **Barry O'Neill**, Two-Party Agreements as Circular Sets (September 2018)
723. **Adiv Gal, David Saltz, and Uzi Motro**, Effect of Supplemental Feeding on Nesting Success in the Lesser Kestrel (*Falco Naumanni*) (September 2018)
724. **Maya Bar-Hillel**, The unbearable lightness of self-induced mind corruption (November 2018)
725. **Tomer Siedner**, Optimal pricing by a risk-averse seller (May 2019)
726. **Maya Bar-Hillel and Cass R. Sunstein**, Baffling bathrooms: On navigability and choice architecture (June 2019)
727. **Maya Bar-Hillel and Jacob Lavee**, Lay attitudes toward involuntary organ procurement from death-row prisoners: no, but (June 2019)
728. **Maya Bar-Hillel**, Why didn't I see it earlier? (July 2019)
729. **Maya Bar-Hillel, Tom Noah, and Shane Frederick**, Solving stumpers, CRT and CRAT: Are the abilities related? (October 2019)
730. **Sergiu Hart and Yosef Rinott**, Posterior probabilities: dominance and optimism (November 2019)
731. **Sergiu Hart and Dean P. Forester**, Forecast-hedging and calibration (November 2019)
732. **Maya Bar-Hillel**, The base-rate fallacy in probability judgments (December 2019)
733. **Yigal Attali and Maya Bar-Hillel**, The false allure of fast lures (February 2020)
734. **Uri Zak**, Female Chess Players Do Underperform When Playing Against Men: Commentary on Stafford (2018) (March 2020)
735. **Daniel Kahneman and Maya Bar-Hillel**, Laplace and Cognitive Illusions (June 2020)
736. **Sergiu Hart and Yosef Rinott**, Posterior Probabilities: Nonmonotonicity, Log-Concavity, and Turán's Inequality (July 2020)
737. **Maya Bar-Hillel**, An annotated compendium of stumpers (July 2020).
738. **Alex Gershkov and Eyal Winter**, Exploitative Priority Service (August 2020).
739. **Alexander Kravtsov and Eyal Winter**, An Axiomatic Approach to Sensors' Trust Measurements (August 2020).
740. **Robert J. Aumann**, Why Consciousness? (May 2021)