THE "EQUATION" OF EXPERIENCE

dbruiger [at] gmail [dot] com

1. Consciousness

Those who would journey together are wise to agree upon a point of departure. Since human beings agree upon very little, this is easier said than done. Perhaps the sheer fact that *there is something rather than nothing* is the one thing to which all readers will unconditionally assent. Beyond that, the *nature* of this something and how we come to experience it have always been open to debate.

In order to underline that fundamental basis for agreement, let us call this sheer presence of something-rather-than-nothing *consciousness*. Already we moderns, through the prism of mind, disperse this experiential totality into subjective and objective aspects, calling the latter 'reality'. Kant called the former *phenomenon* and the latter *noumenon*. But the undifferentiated totality itself is the one thing that transcends and underlies everything else, which is perhaps why consciousness remains unexplained scientifically and God remains a seductive idea.

No doubt there *is* a real "external" world, Kant's noumenon. But, the nature of our experience is not an open window upon it. Consciousness is more like a virtual reality, guided in real time by the physical world, which in modern times has been defined by science. Even our notions of a spiritual existence beyond the physical (literally, meta-physical) are unavoidably colored by science. Contemporary religion reacts to the view of the world presented by modern science and the seemingly reduced position of humankind within its vision of nature. "Materialism" is the dominant ethos of our time, paradoxically empowering humanity, and endorsed by the successes of technology. But our collective unconscious may know in its bones that the scientific program is hubris. In common with religion, it reduces reality to ideal terms proposed by mere humans, whether theorists or theologians.

2. The autonomy of nature

The root of the word *physics* means *nature*. The word *nature*, in turn, means *born of itself*, self-generating. Nature, the physical world, is autonomous. Unlike our inventions, we did not make it. Its very autonomy means that there is always more than meets the eye; it is more than what we think it is. Nature was here before us and contains *us*, not the other way around. It can only provisionally and partially be contained within scientific theories. Nature cannot be exhaustively known and will always surprise us. The universe may or may not be infinite in extent. But the principle reason it eludes final comprehension is that knowledge of things we did not make ourselves can never be completely certain. Knowledge of a *theory* can be perfect, since the theory

was humanly created. Since we did not make nature, it is unavoidably ambiguous, indefinite, and mysterious from the limited perspective of the human subject. While there is progress in science, the scientific picture of nature can never be perfect or complete. There will never be a truly final theory—unless that means the one our species conceives in its final breath.

3. Science and religion

However, this is hardly the understanding modern science inherited from antiquity or its origins in medieval religious thought. For, according to our Christian heritage, the natural world is *not* self-generating or mind-independent. Rather it is an invention of the mind of God. The early scientists were creationists, who believed they could have an insider's understanding of nature because the human mind mirrored the mind of God. We could understand nature just as we understand our own inventions. And, through understanding natural creations, we could understand the mind of their Designer.

It might seem that science has outgrown its religious origins. However, there are significant vestiges in modern times of the prejudice against the autonomy of nature. A world that is created by divine will has only the derived reality of a created artifact. It is not mind-independent, since it depends directly on the mind of God! The natural philosophers of the Enlightenment held that God imposed order on natural chaos, just as rulers imposed it in the human realm. This identification of human with divine will reflected the heritage of patriarchal religion, which identified God with the masculine psyche. Nature's design was the important thing, rather than nature itself, as though it could be documented in a blueprint. The rational structure of nature could be contemplated and discussed, as from one engineer to another, upstaging nature's manifestly irrational aspects. It was only in the 20th century that physicists began to pay serious attention to the latter.

Religion has always had a materialist and realist side tempering and compromising its idealism. The Reformation was a revolt against the material excesses and hardened doctrines of the medieval Church. Every splinter group since was motivated by a return to purity, inspired by some direct revelation, only to eventually become doctrinaire in turn, and in many cases ironically materialist. Conversely, science has an idealist side informing its official materialism and realism. Mutually dependent, theory and experiment leapfrog through time. Revolution occurs when theory grows overblown, self-contained and beyond the reach of experiment. It is no coincidence that the original Scientific Revolution coincided with the Reformation, both rebelling against the inbred doctrines of medieval scholasticism. Humanism, especially in the arts, was another facet of the general cultural revolution in Europe. The Church had been the mainstay of the arts, and the tides of religion had a direct impact on the history of art. As in science and religion, realism and idealism in art have seesawed through time. Similar cycles can be identified in politics, economics, and every aspect of human affairs.

4. Subject and object

Realism and idealism are dialectical aspects of something that encompasses them both. I will refer throughout to what I call the "Equation of Experience," which simply states that any possible experience or thought you can have involves a contribution from inside you and one from outside you. Subject and object are inseparably entwined, as are matter and idea. This accounts for misunderstandings even among philosophers and scientists.

Of course, in speaking of *subject* and *object*, *matter* and *idea*—and even of *experience*—one has already departed any indisputable ground. We have now entered a discussion of *what* exists and how to characterize it, about which there are many opinions, with polarized extremes that come and go from generation to generation. For, human beings love dispute and are natural extremists. Some people go so far as to deny the existence of a subject, while others deny the existence of a material world. Some claim there is no such thing as conscious experience, or a self who experiences; or claim that these have no place in scientific discussions, which pertain only to the physical world. Others hold that mind is primary and self-evident, since obviously we are conscious; whatever the real world is, its *appearance* is a construction of the mind.

The "equation" of experience states that subject and object are two factors that *co-determine* our moment-to-moment experience, not to mention our behavior. Though this is metaphor, it might be expressed as a literal mathematical function: E= f(s, o). As with all such functions in two variables, it cannot be "solved" without a second equation in the same variables. In life, however, there is no second equation! The only way to understand how our experience changes with *one* of the factors is to *exclude* the other—either by holding it constant or by ignoring it. Thus, science tries to describe real events by standardizing the observer, whose individual subjectivity is to be left out of the picture. This is the basis of scientific realism, method and protocol. If the conditions of observation were left to the idiosyncrasies of the individual observer or experimenter, there could be no reliable results. Conversely, if we wish to study how subjective experiences vary in a population, we could standardize the objects they perceive for comparison. In a sense, this is what words do informally.

Strict materialists may object to taking "experience" as primary, rather than the physical world. They might misread the Equation as a closet idealism that holds everything reducible to the stream of consciousness. In a sense that is right, since conscious experience—and not the physical world—is what is *given* to us directly. However, the Equation certainly does not deny the existence of the physical world. It only claims that any possible knowledge or experience of it is as much a product of the mind as of objective reality. To make that clear, the relation can be written in an alternative form: O = f(e, s). This makes the world (the 'object') primary. In other words, whatever one takes the objective world to be co-varies with what is presented in experience *and* with the structure of one's mind (the 'subject'). If one wishes to be objective, one cannot simply take experience at face value, but must stand back from it to question how the mind influences perception and thought. This is the function of *self*-consciousness. Hence, one should pay particular attention to how *others* view one's mental idiosyncrasies! Hence also the creative role of skepticism, which plays a significant part in the evolution of culture.

Finally, the equation can also be written S = f(o, e). If we wish to understand the subject's mentality, we must consider it not only in objective terms (of brain processes, for example) but also from the point of view of the subject's own experience. This is not only the key to empathy but also to explaining consciousness, as we shall see.

If mathematical-looking expressions are not your cup of tea, just remember these points in a nutshell: Your experience depends jointly on what happens "out there" and on what happens "in here." Your notion of objective reality *naturally* seems self-evident and to follow from your personal experience; but it also depends on the objective particulars of *your* nature, about which others are well positioned to have an opinion. On the other hand, if you wish to understand someone else's mind, consider their own point of view as well as what *you* think makes them tick. Whatever our differences, there is basis for agreement insofar as we are united in facing the raw fact of being here together in this mysterious existence.

To understand that all our experiences, thoughts, feelings, and behaviors depend on these two factors—subject and object—can be very helpful. First of all, it explains how we can so easily disagree. Everyone knows how people can give entirely different accounts of the "same" situation. If *only* objective events determined our experience, our descriptions of an event could only differ in minor ways, such as the literal view from where we happen to stand. And if *only* our particular subjective make-up determined our experience, then actual events would hardly matter; everyone would be living in a personalized dream. But experience is a product of both together, which makes it real but inherently ambiguous. Unfortunately, we are not well constituted to tolerate ambiguity, which explains a second thing: human beings lean toward polarized views. It is just too confusing and demanding to deal with both variables at the same time—which could leave us precariously indecisive. While it may be natural to trust our perceptions as literal truth, it is just as natural for the other fellow, who perceives differently, to disagree!

We mortals can hardly conceive the whole of a situation. We focus on one aspect to the exclusion of others. Our ideas are always only partial truths. Even an outright lie contains a kernel of truth, and even the most self-evident principle must be applied in context. Language conditions thought in such a way that it is unavoidably partial, in both senses of the word. Yet, the human tendency is to claim it as eternal and absolute. This is a recipe for disaster, because it sets the stage for the dialectical opposite to eventually emerge. This is why there are political and religious rebellions, scientific revolutions, changes in fashion, stock market trends—and wars. Moreover, where neither of two competing forces can permanently gain the upper hand, they will be locked in a *cycle* in which they alternate in ascendency. Night follows day as day follows night. Yet the cycle does not repeat perfectly. Daylight lengthens toward summer, recedes toward winter. In every revolution of the earth there are minute changes: the "day" gets longer as the earth's rotation slows down; the month gets longer as the moon slowly recedes from the earth. Just so, history does not repeat exactly, but more in a pattern like a spiral. The perpetual interaction of opposing factors drives culture through a progression that is hardly linear. History breathes through a dialectic of opposites. Even science has its fads.

5. Realness

The most obvious example of how we "naturally" trust our perceptions is the very fact that we experience the world as *real* and *out there*. Most people today acknowledge that conscious experience is produced by the brain, which happens to be sealed within the skull. Yet, experience does not seem to unfold inside the head but in a space outside the body. Unlike our dreams, this world external to the skull appears real to us. For reasons we will soon explore, we are natural born realists. On the other hand, we have imagination and thought, which in some sense do seem to take place within an interior space—an inner realm that is not the same as the external world. The elements of this realm are not real objects but "ideas," which in many cases are idealized versions of external things. It is appropriate to call them objects of thought, since they can be manipulated in the mind's eye in much the way that external objects can be manipulated under the literal eye's view. Thus, in a sense, we are natural born idealists as well. The interplay of these two aspects of human nature reflects the two factors of the equation of experience. The 'object' factor represents the real external world, and thus has a special relation to various forms of realism. The 'subject' factor represents the creative participation of the mind in producing experience and thought, with a special relation to idealism expressed in various ways.

However, if experience is indeed produced by the brain, then obviously the brain does something that results in the experience of a world outside the skull. Events within the skull are projected outside it as events in the world. This ability of the brain is not restricted to projection in space (which is largely a visual experience). It applies to anything the brain considers "real," which can include ideas.

The brain is an organ of the body, which is an organism that has come to exist through natural selection. The world appears "real" because it holds over the organism the power of life and death. The organism is totally dependent on its environment. If it did not take this dependency seriously it would have been eliminated long ago by natural selection. Indeed, it would never have come to exist. Treating the world as real is *how* the organism takes it seriously. *Realness* is therefore both a property of the external world itself and also a quality with which the brain must imbue its experience of the world in order to survive. It is both objective and subjective.

Because of this primary commitment to external reality, the brain tends to consider real whatever it is inclined to take seriously. Often enough, this includes its own fabrications. *Reification* is treating an idea as a literal thing. Because we are programmed to defer to the power that the external world holds over us as physical organisms, when we want to invest some notion with the authority of that power, we give it the status of a real thing. This is especially true when we try to convince others of the validity of our ideas, wishes, or opinions. Rather than claim responsibility for a personal point of view, we *dis*claim responsibility by asserting that we are simply pointing to a reality "out there" that has nothing to do with us. When reality is attributed to something, someone is doing the attributing for some reason of their own.

The ambiguous nature of experience leads not only to disagreement between individuals but also to cultural divisions as well. Think of differences between the arts, the sciences, and religion. These involve different emphases on the subject or the object factor. Science focuses on the object, by standardizing the subject. Art focuses rather on the subject, who intentionally redefines the object. One aspect of religion focuses on God as the ultimate object, but another aspect of religion focuses on the needs and spiritual evolution of the subject, who finds in God a metaphor for the ideal human being and how to live. One aspect of science focuses on experiment and observation; another aspect focuses on mathematical theory. One aspect of religion attempts to manipulate the gods through prayer and sacrifice; another aspect attempts to master the self through spiritual practices and discipline. Science manipulates matter through technology, as art does through technique. Both involve self-discipline.

Opposing political philosophies and parties tend to focus on complementary aspects of governance and economics, and usually the interests of different groups. While each may appeal to "reality" to vindicate its cause, their interests are partial (in both senses of the word) and cannot objectively represent the overall long-term welfare of the community or state as a whole. In politics too, one attends poorly to both the subject and object factors together. If one looks only at statistics and economic theories, for example, it may be difficult to empathize with people's subjective experiences. If one takes only a "bleeding heart" approach, it is difficult to accommodate everyone's conflicting needs. The true resolution of conflict is not a matter of arbitrary compromise, but of finding a higher ground on which there no longer appears to be a conflict.

6. Agency

One is said to *have* a mind, and to *have* a body, as though this "one" were an agent distinct from either. This suggests an inner witness, *interior to any possible experience.* This self or 'I' is not one's body, one's mind, nor any other *thing* one can be aware of. It is not what is seen but is the seat from which seeing occurs. This self cannot be *materially* real, because then it would be something perceivable, an object of awareness rather than the subject. (If I can perceive it, it is not interior to what is perceived; therefore it is not 'I'.) Hence, the brain, which most modern people believe is the material agent responsible for perception, is not the self. Likewise, while *I* can feel pain *in* my body, it does not seem to be the body which feels. The body then seems like some form of prosthetic device which the self inhabits in order to be in the world. And perhaps "the mind" is no more than a subtle version of the same thing—some software that comes with the space suit. Such intuitions have influenced everything from religion to science fiction and left their mark on western philosophy as well.

A spirit or soul is often imagined as a vaporous entity, like a ghost. It is not quite material but not quite immaterial either. This quasi-materialism may result from observation of the natural world, where visible things appear to be material but there also seem to be invisible forces at work. Our ideas about causality are based on experience in the material world, so that when an

unseen cause appears to be at work it is imagined as material, even if invisible. Animals and people are familiar agents, but there may be uncanny invisible agents as well. Yet, for these to be causal, they must be able to influence the visible world. While such forces ring of animism, it is no exaggeration to say that forces conceived in physics are no less ambiguous and uncanny.

The idea that mind—or the self, or the soul—is some sort of quasi-material entity gives the misleading appearance that there are two sorts of *stuff* in the world: matter and mind. This leads to many fruitless questions, such as how they interact or share the space of the world. This dualism cannot be a duality of competing *nouns*, so to speak, but rather of nouns versus verbs—a basic intuition enshrined in language. Mind, then, is not (only) a thing that can be acted upon, but is above all the source of action.

This concept of *agency* is compatible with modern ideas about neural or computational systems. An agent is a material entity when it is embodied as a natural creature. We can also now imagine autonomous computers as agents. But their material nature alone does not explain the activity of these systems. Brains and computers are composed of matter, which of itself is essentially passive. Something more must account for their intentional behavior. (For the human being, this was once thought to be the soul; for the animal, some vital force; for the computer, it is the programmer.) Yet, little ground is gained simply by naming the agent involved. For, we have not yet clarified what *intention* is, or what *agency* is, or by virtue of what a system can be an agent or have intentions.

The brain is clearly an integral part of the body, with a special role. It is not literally a machine. Yet, the metaphorical importance of the computer is that the operation of the brain might be understood in terms of the human intentions, actions, and logic involved in creating software. Of course, genetic engineering notwithstanding, the "software" of natural organisms—including the human one—is not programmed by human programmers. Rather, organisms program themselves, or are programmed by nature. Can we understand the intentionality behind this programming in terms of *human* intentions, while not limited to those terms? For this to be so we must broaden the concept of intentionality to allow that other sorts of entities than human beings have intentions. Certainly, that is our intuition concerning animals, especially those closest to us on the tree of life. Can we accord them agency that is not merely a projection of our own? Ultimately, we need to explain what agency is in a way that applies universally to all organisms, even to hypothetical aliens from outer space and intelligent machines that might qualify as artificial organisms. We need an abstract concept of intention independent of human intention as commonly understood. We could then say that an agent is an entity with its own intentions, and agency is the exercise of them.

7. Intentional and causal description

In that case, two kinds of description are available to us: intentional description and causal description. The latter is the standard fare of science. However, while causal forces animate physical systems, intentions animate agents. It is not sufficient to think of agents only in physical

or causal terms. To be sure, mind is a feature of the physical organism. Yet, a purely physical (causal) description does not account for the organism's behavior, which is intentional; let alone can it account for consciousness. To account for either the behavior or the experience of an agent, we must include intentional description, even though the intentional operations of the system are carried out through physical operations.

To illustrate these two kinds of description, imagine an electronic circuit. The physical circuit consists of material components such as wires, resistors, and transistors. These are connected physically with solder. Their *logical* connection is illustrated schematically in a circuit diagram. This diagram, however, does not depict the *functioning* of the circuit or its purpose. Even to read the logic of the diagram presupposes an implicit understanding of such matters, which resides in the human agent, not the circuit itself.

Or, imagine instead a steam engine. It too consists of physical parts that interact through various physical pushes and pulls. However, a detailed "cutaway" illustration of the machine does not by itself reveal the machine's functioning or purpose. It invites the human observer familiar with such machines, and who already knows something of the purposes behind them, to speculate on how the thing works—that is, to second-guess the intentions and reasoning of its designers. An intentional description of it presumes such understanding of design and purpose.

This presents a problem when it comes to understanding organisms and their behavior, or indeed anything not designed by human beings. This is one reason why even biological science tends to restrict itself to causal description. For, we can only second-guess the functioning of organisms—or, for that matter, of any physical system we did not design! Yet, to some extent, even causal description faces this dilemma, since it proposes to describe the interaction of parts, as we identify them, as though these were as clearly defined in natural systems as they are in artifacts. In seeking to understand the functioning of natural organisms, the best we can do is try to put ourselves in the creature's place, to try to fathom *its* intentions and the reasons behind them. The risk is that this exercise may simply read human intentions and reasoning into the behavior of the creature. The benefit is that it is the only way to understand the behavior of creatures above a level of complexity that defies mere causal description: the level of agency. For the benefit to outweigh the risk requires that the human observer's perspective be clearly distinguished from that of the organism.

But we have not yet defined 'intention.' Before proceeding further, therefore, let us clarify what is meant, lest it be understood by default merely in human terms. I propose that an intention—or *intentional connection*—is simply any internal connection made by an agent rather than by outside forces. This is admittedly a circular definition: an intention is a connection made by an agent, and an agent is an entity that makes such connections. Yet, the concept contrasts usefully with a *causal connection* between observed external events, which is asserted to exist by an outside observer. Moreover, any observer is a material agent that may also be observed, by another agent who may assert causal or intentional connections. Causal connections are asserted by outside observers, whereas intentional connects are made by the observed system itself. This does not prevent the observer from speculating about the intentional connections within an agent, as well as about causal connections. But such speculations are the observer's own intentionally

made internal connections, not necessarily those of the system observed. Making this distinction is crucial in the study of organisms, including human beings, where the intentionality of the observer must not be confused with that of the subject creature.

8. How to explain conscious experience

If the operation of physical systems can be explained in terms of the pushes and pulls of physical causes, the operation of agents (specifically organisms) can only be explained in terms of intentions and reasons. This is not to deny that an agent is necessarily a physical system. Intentional connections within physical systems must also be causal connections. The physical system underlying the organism's agency can be explained in causal terms; but that will not account for the actions of the agent or the behavior of any but the simplest organism. In particular, it will not account for the brain's operations resulting in conscious experience. As Leibniz early observed, nowhere in the "machinery" of the brain will you come across a "perception," such as the color blue, the smell of a rose, or the sting of pain. While hardware can shape the characteristics of experience, it cannot *explain* sensation, feeling, thought, or other conscious experience. Software can only be described in intentional terms. Yet, where in the physical brain do we encounter intentions, any more than perceptions? We can identify causal connections in the brain (synapses), but how do we identify intentional connections, the software?

The brain, like the computer, is not only a physical system. It is above all a *logical* system. By this I do not mean that it necessarily follows the formal rules of proper reasoning as understood by logicians. A "logic" is a broader concept, which simply means following *some* rules in an orderly way, for some reason. In the case of an organism, these are determined by its bodily needs, its evolutionary history, and its personal interactions with the real environment, which includes other organisms. In the case of a computer, it follows whatever rules the programmer specifies.

While it may be impossible to see how physical causes can produce conscious experience, it may be easier to see how the organism's *intentions* can produce it. I propose that conscious experience *is* how an agent keeps track of its own internal operations, inputs, and actions. Experience is a sort of real-time bookkeeping and inventory. The function of consciousness is to monitor the organism's relationship to the external world, including aspects of its own internal processing. The organism does this for its own purposes, by making internal connections which are at once intentional and causal. These may be observed by an outside observer to consist of physical connections such as nerve synapses.

Human observers, at least if they are scientists, usually confer with each other to justify their attributions of causality. In other words, causality is proposed from a third-person point of view. In contrast, the organism is under no obligation to justify to human observers its internal connections. They are simply the connections that work in the long term of evolutionary success,

justified by the fact that the creature exists! The external observer can only speculate about the "reasons" for these connections, which may not correspond to human reasons.

The human phenomenon this resembles is language. The symbols of language have no inherent meaning; they are *assigned* meaning by those who use them. To a person who does not speak that language, it sounds like gibberish and looks like meaningless squiggles on the page. That person may note the patterns in which the symbols are used and make inferences, eventually coming to understand and use the language. Indeed, this is how children learn their native tongue and how codes are decrypted. The observer may also study how use of the symbols changes over time and arrive at an understanding of how certain expressions came into usage. Yet, at any given moment, the meaning of a word is not necessary in any sense, but is simply a matter of convention. It has its meaning by common consent, because a group of language users have *made* similar connections between the symbol and what it represents to them.

The essential point is that intentional connections are simply predicated, made by *fiat.* In the case of language, which is interpersonal by definition, these "decrees" are made collectively. Within the organism, however, the "language" is purely internal; yet it functions in the same way as a coherent set of conventions established for the sake of communication. The connection is not made because it is meaningful; rather, it acquires meaning by being made.

I propose that conscious experience arises in the same way as meaning in language. We are not aware *of* the internal connections our brains make; rather, we are aware *through* them in the same way that we are aware of the meanings of words through convention and use. There is a language of the senses, through which intentional connections (embodied in neural events) become conscious experience. Just as words heard or read give rise to mental images and arouse emotions, so neural connections give rise to the conscious experiences of sensory perception, thoughts, and feelings. The brain assigns meaning to its own internal representations, evoking sensory experience in the way that words evoke mental images.

Perhaps it seems that the analogy gains us no ground, since it is as great a mystery (or perhaps the same one) how words evoke mental images as how intentional connections evoke sensory experiences. More than analogy is involved; for, meaning emerges in language and conscious experience alike through a process of assertion: by fiat. Sensory experiences such as colors, smells, pains, and other sensations are meanings read into coded impulses of sensory input, just as a telegrapher can read Morse code. Thus pain, for example, *stands* for something in much the way that words or dashes and dots do. Namely, it stands for tissue damage and appropriate response implied for the organism.

The meaning of pain, as an internal communication implying behavior, may seem clear enough. But what of the meanings of colors, for example, which do not obviously entail a behavioral response? If meanings in the language of the senses are conventional and arbitrary, as they are in human languages, then why are ripe apples experienced as red and grass as green, rather than vice-versa or some other way? What is it about the experience we call greenness that commends it to represent foliage in the vocabulary of the senses? And what is it about redness that commends it to represent fruit that must stand out against a green background? This is rather like asking why a particular meaning is denoted in a given language by a particular word, written

and pronounced its specific way, rather than by some other symbol. For the native language user, the association of the word with what it represents seems natural and unquestioned, though of course it is actually a social convention. Given a symbol system, s*ome* symbol must be chosen; however arbitrary, through usage it will inevitably come to seem *imbued* with the meaning it conveys. Hence, it is backwards to ask why grass appears green; rather, greenness is what it is because of the association with grass. Given consciousness as a symbol system, greenness is the way we visually experience the totality of associations related primarily to chlorophyll.

One can acknowledge the arbitrariness of words because human languages use different terms for what is presumably the same conscious experience. The experience itself, however, cannot be compared interpersonally. There is no way to determine whether two people have the same private experience of what they both agree to publicly call green. The *sensation* of greenness, unlike the word *green* (or *vert* or *grün*), is not merely a linguistic convention, interchangeable with other symbols, and subject to social change. Rather, it is a convention of neuro-logical organization, with the force of long genetic precedent. While the words of a natural language are transient and specific to culture, the meanings of sensory experiences are more universal and enduring, backed by the relative stability of biology. Indeed, human visual cognition adapts to distorting colored lenses or filters in such a way that color experience (of verdant foliage, for example) is eventually restored to normalcy. The sensation of greenness is what it is, and different than the sensation of redness, precisely because of the stable real-world things it refers to in our common evolutionary history, from which it cannot be arbitrarily dissociated.[1]

The qualities of sensory experiences, such as specific colors, tones, or smells, are thus not something gratuitously added to the information they represent. Nor are they "caused" by it, any more than words are caused by the things they represent. Rather, they are a *version* of that information, which an internal agent presents to itself in consciousness as a kind of synopsis. Sensations, in other words, are an in-house presentation of information that is also often publicly available. If the experienced quality of greenness (for example) seems to convey privileged or ineffable information beyond that involved in the scientific analysis of light, this is because it also bears information about the brain's internal language, its relationships to the world, its priorities and evolutionary history. That is information about the subject rather than the object, or rather about their specific relationship. One could compare this to the personal connotations words acquire, in contrast to explicit dictionary definitions.

Information represents a *discrimination*, which is an action a conscious observer can perform and which even a system that is *not* conscious might also perform. Thus, laboratory equipment can discriminate between wavelengths of light or sound, or detect chemical odors. Using such equipment, observers can arrive at a similar conclusion about the world as they might "directly" through sensation. The same information can be gained in different ways. Apart from

---

[1] This is why there can be no "inverted spectrum," which is the "apparent possibility of two people sharing their color vocabulary and discriminations, although the colors one sees... are systematically different from the colors the other person sees." [Wikipedia: inverted spectrum].

precision, the difference is point of view: the senses provide information from the point of view of the perceiving subject, for that subject's use; laboratory instruments provide information to multiple observers (who are also perceiving subjects), *for their* purposes. By convention we call the former information subjective, because it is private and relies on one's natural equipment; and we call the latter objective, because it is public and relies on mechanical devices. Of course, laboratory instruments have been designed to extend or enhance natural perception, to provide information not detectable with the natural senses, or to measure it more precisely. Yet, at the end of either process there is a subject who interprets the information. Sensory experience is the subject's interpretation of information provided by the natural senses.

9. Is consciousness an illusion?

By definition, *fiat* is active. The agent does not merely receive information passively, to evaluate whether a proposition is true. In effect, it *declares* what is true. If fiat is the basis of consciousness, it would seem that what we experience is a sort of fiction, guided by fact, which the mind may treat as literally true or real. This does not mean that it is false or *merely* a story. (Much less does it imply that the external world does not exist.) It simply means that the mind's representation of the world involves the subject's own creative effort. Perception is not a direct window on the world. While there *is* a real world, what we experience is not that world itself but a representation of it, like a map, which can be "true" only in the sense that it guides us effectively.

Of course, the map can be wrong, even when not lethally so. Hallucination is an example of how the brain simply asserts something to be there when it is not. Yet, this sort of fabrication is actually the normal basis of perception! An example is the visual blind spot. Where the nerves of the retina are gathered to exit the eyeball, there are no receptors, so there is no sensory input in this small area of the retina. Yet it is not experienced as a blank or hole in the visual field. Instead, the brain fills in the gap, or rather ignores it, so that the visual world appears continuous. While the presence of visual input there is fabricated, this illusion correctly represents real continuity in the world. To be true to reality, the brain must tell itself a white lie. It must make the world *look* continuous because it *is* continuous, even when the most direct evidence is lacking. And this leap is an act of simply *deciding* to see it that way.

There are many examples of such perceptual leaps. We are not normally aware of them precisely because they are a matter of disregarding appearance (or its lack) in favor of truth. For example, to experience motion pictures as continuous is only possible because the brain disregards the gap between individual frames when they succeed each other fast enough. In a sense this is hallucination, because continuity is experienced that is literally not there. But, again, the illusion serves the larger truth that something really is continuous.

This characterizes normal visual perception, not only anomalies revealed in laboratory experiments. After all, the retinal field is digitized, as are the other sensory surfaces. While there are myriad visual receptors, they are discrete, separate, and limited in number. There is minute

space between them where light falls undetected. (In fact, the light is also digitized as quanta.) The brain ignores these discontinuities, creating the illusion of continuity. It continually reinvents its representation of reality.

What is the nature of that representation? We have only metaphors at our disposal. *Painting* is a classic metaphor to understand the brain's creative effort involved in seeing. *Fiction* refers primarily to literature. *Film* is another way to present fictional narratives. A more recent metaphor is *virtual reality*, which is an interactive fiction created by computer. These are all *representations*, in different media. Using this metaphor, normal perception is a virtual reality created by the brain for its own use, largely to represent the external world. Literal virtual reality is for the entertainment of a human user, who dons an apparatus to provide an artificial sensory input that is nevertheless processed through the normal sensory channels. The user can remove the device and return to "direct" sensory experience. Imagine, however, that you could *not* remove it. And imagine also that you could program it, change the representation, create the virtual reality yourself as you went along. Unless you could somehow bring this private show into synch with external reality, you would probably not survive very long. If the show "in here" were not reliable as a guide to whatever is "out there," you would soon run into serious, probably fatal trouble.

Let us take this fantasy a step further and say you were sealed, from birth, inside a sophisticated windowless vehicle from which there is no exit. (There are, of course, appropriate life support systems.) Imagine further that your contact with the outside consists *only* of data supplied by various sensing instruments that are part of this vehicle. You would also have various controls that do something (you don't know what), which indirectly changes the input of data. An outside observer would say that your attempts to control the behavior of the vehicle in external space cause the input to be updated. However, at this point you are merely randomly pulling levers to see how that affects the readings of instruments. Through trial and error, you could learn that when you pull *that* lever, *this* pattern of input from the instruments results. The outside observer would say that you are slowly learning to "fly by instrument." To an outside observer, real things are happening to this vehicle as you try to master control of it, some of which could be disastrous or lethal. Inside the vehicle, however, you can only speculate on what is "really" happening "out there." You form a representation of it, inferred from these correlations between levers and instrument readings. And yet, you must bring this representation into line with what really is out there if you are to survive.

In such a situation, you are not in a position to compare your representation to the external world through any direct experience outside the vehicle. This changes the meaning of "representation" considerably, for it's more a *theory* of the external world than a portrait. If the theory works well enough that you are not destroyed, then you have an "accurate" picture of the outside world. Your *experience* of the world *is* this picture, which you get purely through navigating by instrument on the basis of your theory.

It is not so easy to decide whether your virtual-reality representation is "illusory," or what that even means. The concept of illusion depends on the possibility to perceive things as they "really" are. In our metaphor, that would require getting outside the sealed chamber of the

"vehicle," which is not possible. This vehicle is obviously the body, and the sealed chamber is the skull. The brain has only the input of electro-chemical impulses from afferent nerves, and the output of efferent nerves, with which to guide the body's movements in "external space." On the basis of correlations between these, it develops a theory representing what is "out there." Your experience of the external world *is* that representation. The very notions of 'body' and 'space' are simply part of it!

10. Why consciousness is necessary

What would happen if we put a sophisticated computer inside the sealed chamber instead of a hypothetical human guinea pig? It is certainly plausible that such a machine could solve the problems of navigation and control. But would it, like the brain, be conscious? Would it *need* to be conscious in order to do its job? But to turn the question on its head: why exactly is the *brain* conscious?

Let us backtrack a moment to ask what a *machine* is and how even a sophisticated computer differs from an organism and its brain. First of all, a machine is a human artifact: there are no naturally occurring machines, and (so far) we know of no machines built by alien civilizations.[2] A machine is *made,* whereas natural things are *found.* To think of *nature* in mechanistic terms is therefore no more than a metaphor. Moreover, a machine (or any artifact) consists of a finite number of well-defined parts. Natural systems, by contrast, are *not* well defined and might be indefinitely complex.

To view the brain as a machine is to assume that it has finite well-defined parts that we can clearly identify, which may or may not be so. For, like any aspect of the external world, we have no direct access to the complete reality of the brain, only partial access through our cognitive processes. Nature often turns out to be more complex than theory predicts (which is why there continue to be discoveries!). *Ideas* are finite and machine-like, but nature is not. The very point in question is always how well these ideas resemble the real thing. We cannot assume, therefore that the brain is a machine, operates like one, or corresponds perfectly to theoretical models of its workings.

The brain is part of an organism, which has its own purposes and priorities, whereas a machine (so far, at least) dutifully carries out the purposes of its human designers and users, according to *their* priorities. Furthermore, unlike the machines we are familiar with, an organism is not only self-maintaining and self-reproducing; it is also *self-defining*. Since the brain is part of an organism, in order for a machine to truly take the brain's place in our thought experiment, the machine would effectively have to *become* an organism, even if not made of flesh and blood. It would have to be in the *relationship* of embodiment with its environment. That would be quite

---

[2] It might be argued that some machines are now built by other machines. Yet, the intentionality involved remains human. There are not yet any machines with their own intentionality.

unlike our ordinary notion of machine, which is an artifact defined by us and not by itself, and without any relationship of its own with an environment.

Even so, it remains unclear why an *organism* should be conscious—and *which* organisms? It may also be unclear exactly what we mean by 'conscious'. Some people believe that all living matter is "sentient"; and some people attribute some form of consciousness even to rocks! In the case of human beings, it is clear that a lot of our behavior does not require conscious attention and can be carried out effectively without involving what we have been calling "experience." For example, we have little conscious access to, control over, or experience of autonomic functions of the body. They just happen automatically. We can also drive cars without paying much attention, if the route is familiar enough. Routine behaviors in general seem to require little conscious effort. But there are other situations that do seem to require attention—especially situations of novelty. This suggests a specific role for conscious experience: to make real-time sensory input available to higher centers in order to deal with situations that cannot be handled automatically by established algorithms. Consciousness thus plays a separate role from non-conscious behavioral responses, and presumably involves different or additional neural processes.

For example, the conscious experience of initiating motor activity (willing) comes only *after* the neural events that have caused the activity. However, this conscious experience (with its memory) serves as the basis for choosing *future* action, or action in a larger context. Consciousness serves as a registration system for information to be tagged for future retrieval and use. The conscious experience indicates *acknowledgment*, after the fact, of the particular non-conscious processing underlying it and what it responds to. It serves a distinct purpose, with a different associated behavior.

Let's say you experience a tickle in the throat and immediately you begin to cough. What is the relation between the unpleasant tickle *feeling* and the *behavior* of coughing? The experience does not *cause* the behavior in the scientific sense of causation. For, modern science considers only *physical* processes to be causally effective, and those have already taken place as neural events, slightly preceding the conscious experience. Yet the tickle *sensation* does play a functional role by *representing* a state of your organism. The tickling is a *sign* indicating irritation, a state of affairs upon which you (an agent) might act voluntarily, independently of the cough reflex. While the coughing *behavior* can be an unconscious reflex, the *sensation* serves further to inform you about the very occurrence of the reflex, the input that causes it, and the condition to which it responds. The job of this inner agent ("you") is to monitor the state of the organism, the world, and the activities of various sub-agencies. It has executive powers to suppress the cough or to plan some other action to deal with the irritation, such as to get a drink of water. This function cannot itself be automated like the coughing reflex, because reflexes can deal only with precise situations for which they are adapted, while the conscious sensation facilitates behavior with a higher degree of freedom and longer-range purpose. Swallowing may involve a reflex, but drinking a glass of water involves complex actions that beyond a reflex.

This inner agent might be likened to the CEO of a corporation, who is responsible for decisions at the highest level, based on "reports" provided by various subordinates. The job of

this executive is to monitor and trouble-shoot the overall operation of the system, to plan ahead on various time scales, and to take charge in situations where established protocols are inadequate and too mechanical or crude. The literal CEO relies upon charts, graphs, and other iconic representations and visual aids, in addition to verbal and written reports, to summarize complex information supplied by various departments. As in a corporation or government, sub-agencies passing information up must package it for decisive action. This means that what the conscious self perceives is ideally unambiguous even when wrong! The very nature of visual cognition, for example, is to identify clearly what is seen, to know how to act upon it. Yet, what renders something clear or certain is ultimately no more than the decision that it is so. Moreover, there should be but *one* boss to decide; it would not be functional to have multiple seats of consciousness competing in the organism. Sub-agencies might lack crucial resources required for the executive job and so be limited to non-conscious processing or to options to be consciously considered.

Yet, one may still ask why even this executive function could not be performed *non-consciously* or by a good enough machine. Could a superintelligent computer take the place of the CEO? If it did, would it necessarily be conscious?

One thing that could set the executive function apart from others is that its protocols are created on the spot, in real time, not simply drawing on pre-programmed algorithms. The executive can *direct* attention where it is not necessarily *demanded*, on a time scale permitting planning, reason, and reflection. When immediate attention *is* demanded, it may be preceded by a pre-programmed quick first response, such as a reflex. Yet, beyond that, the demand itself puts the executive on notice for subsequent short-term and long-term planning. Nevertheless, given the deep influence of computation on all aspects of modern life, it still makes sense to ask whether all these same operations of the organism could be performed by algorithms—that is, non-consciously by a sophisticated machine. But to put question the other way around, could it turn out that the ability of a machine to perform exactly the same as a conscious organism would necessarily render it conscious, whether or not its "algorithms" could be identified by human observers?

11. Formalism

Here we must digress to inquire what it means for operations to be identifiable or "the same." While birds and airplanes both "fly," they achieve flight in very different ways. A baseball player and a pitching machine both perform the operation of "pitching," but the device only crudely approximates the person, no matter how accurately it hurls the ball. The characteristic and deceptive "chunking" involved in language and thought alike, whereby a rose is a rose is a rose, gives the false impression that operations are identical when they are only analogous. (This is what makes perfect simulation seem plausible.) When one "operation" seems to resemble another, however, they are both implicitly being compared to a common formalism, which has been abstracted as the essence of that behavior ("flying" or "pitching"). Taken to the extreme, it

is assumed that all aspects of an organism's behavior as a physical system can be reduced to such formalisms: algorithms, which are the equivalent of written instructions. The algorithm, program, or formalism is the bottleneck through which the whole being of the object, system, or behavior must pass in order to be simulated.

Two operations are "the same," therefore, when they both embody a common formalism. This can work perfectly well for two artifacts, such as a propeller aircraft and a working model airplane: these are structured alike and fly on the same principles. They are in fact two alternative constructions from the same design, scaled differently. It is a fallacy, however, to think that the being of a *natural* object, such as a bird or a brain, is exhausted in a formalism abstracted from it. For, abstraction seeks the essence, intentionally leaving out what could turn out to be crucial detail. The formalism is then mistakenly thought of as its blueprint or essence, in the same sense that the aeronautical engineer's design is the blueprint capturing the essence of the flying machine. The machine is designed to serve human purposes. The brain, however, is a found object, not an invention constructed from design. The theoretical model of the brain—the formalism, program or blueprint—is imposed after the fact, through an analysis that can never be guaranteed complete or perfectly accurate. The fallacy involved in the reverse engineering of natural systems is the belief that it is possible to perfectly replicate a natural thing or phenomenon by first codifying its structure and behavior and then constructing an artifact from that design. The artifact *will* instantiate the design, of course. But it will *not* replicate the natural object, any more than an airplane replicates a bird.

Even if we aren't trying to replicate the brain, but only trying to get a machine to accomplish its tasks, the same argument applies. For, it is a fallacy to assume that the tasks set by the programmer or designer are the tasks set by the subject's brain. Obviously, computers or robots can perform many human tasks, and often better. But whether we are talking about doing arithmetic, playing chess, driving a car, or performing domestic duties, these are operations defined by people, not by machines. They may be formalized in such a way that a machine can perform them to human satisfaction. But this says nothing about how the natural brain addresses such tasks. It may seem reasonable that actions a person performs "automatically" (i.e., unconsciously) could be accomplished by a machine because they are "mere algorithms." But then it is equally reasonable to think that actions that humans can only perform consciously would require something analogously different for the computer—at least a "higher" algorithm.

The modern assumption is that the brain functions like a computer, using identifiable algorithms. In fact, we understand no better how artificial neural nets solve tasks than how natural ones do. Determinism is supposed to dispense with free will—and also with conscious experience as a cause of behavior. However, that argument works two ways. *If* there is a one-to-one correlation between brain processes and algorithms, then for a computer to truly perform the *identical* algorithms involved in human consciousness might imply that the computer too must be conscious!

12. Why consciousness is hard to explain scientifically

The scientific explanation of consciousness remains a major unsolved mystery, second only, perhaps, to the question of why there is anything at all. The challenge for any scientific theory of consciousness is to fit within the materialist framework. Of course, it might *not* fit, either because consciousness is not material (or not produced by the brain); or because the framework itself is too constricting (the approach we will pursue here).

The problem of consciousness is peculiar because it is difficult even to properly formulate. Part of that difficulty lies in the fact that philosophy of mind does not have a precise, well-defined vocabulary. Mental terminology is especially ambiguous and confusing. *Awareness*, for example, can refer either to a subjective experience or to the behavior of taking cognizance of something. For that matter, *experience* bears a dual meaning, as either a momentary subjective state (such as experiencing pain) or a history of events lived through (such as 'work experience' or a 'traumatic experience'). These ambiguities lead to redundancy, for the sake of clarity, in such expressions as *conscious experience,* where a single term ought to suffice. Other ambiguous psychological terms include: mind, thought, attention, perception, cognition, phenomenon, sensation, representation, intentionality, sense-data, qualia, introspection, disposition, introspection, conscious and unconscious. The last two terms, for example, can refer either to a state or to a compartment of the mind.

Language makes little distinction between objects of thought and physical objects. An object can be a material thing; but can also be anything that receives attention, intention, or action. It is even a part of speech. Language conditions us to believe that anything that can be named or labeled must "exist" in some fashion—if not in the external world, then in the mind. We are all too ready to make nouns out of other parts of speech.

The language for analysis of mind is frequently awkward. It is filled with mixed metaphors, category errors, logical inconsistency, and circularity. This results in a lot of misunderstanding and talking at cross-purposes. Language can only define things in terms of other known things, or describe them as *like* other things that remain undefined. It is fundamentally challenging to define mental states in a way that does not circularly refer to other mental states. Attempting to define even simple psychological terms can end in proliferating mental terminology.

Many proposed solutions to the problem of consciousness fail because they actually address *behavior* rather than experience (i.e., phenomenology). However difficult the problems solved, they often turn out not to be the problem posed by conscious experience. On the other hand, not everyone agrees that there even is a problem to solve. Mind remains the elephant in the room for science, and the diverse approaches to it suggest the story of the blind men who each describe a different part of the creature and can reach no consensus.

However, communication is only part of what makes the problem posed by consciousness difficult for science to engage. In part, the structure of science itself is responsible for this failure. Right from the beginning, the focus was exclusively on properties of the physical world, especially those that can be described mathematically. Historically, what was meant by *physical*

excludes the subject's *experience* by definition. Subjectivity and consciousness do not fit within the program to describe the world objectively. In terms of the Equation of Experience, the strategy of science is to disregard the subject in order to study the object. The price to pay, however, is that we lack a strictly scientific theory of consciousness.

In particular, the Scientific Revolution rejected those aspects of sensory experience that do not unequivocally reflect properties of the external world, especially those aspects that do not lend themselves to mathematical treatment. These were known as "secondary qualities," which include color, sound, taste, smell, and tactile feel. In other words, science focused mainly on visual properties such as shape and location in space, which could readily be measured. As soon as there were reliable clocks, the challenge was to describe how the position and motion of objects changed with time. The visual sense became identified with objectivity (since it best reveals these "primary qualities" of objects), and the other senses were dismissed as merely "subjective." This left unsolved the problem of how to explain secondary qualities (such as color) in a world that was defined without them. Everything, including secondary qualities, was to be reduced to a description in space and time. But how to describe color and odor, for example, in such terms? Molecular and chemical theory might provide a description of how receptors in the retina transduce the energy of light into nerve impulses, which are the movement of molecules, or how the receptors in the nose transduce information about molecules in the air. But such a description does not seem to reveal anything about the subject's *experience* of color or odor.

Secondary qualities are ambiguous insofar as they depend on both the external world and the perceiving organism. However, according to the principle of co-determination we have called the Equation of Experience, even so-called primary qualities must be a function of both subject and object. Vision, after all, is but another sense of the organism, not an open window on the world. For human beings and many creatures, vision includes color perception, so that subjective quality re-enters science by the back door. The question of whether color is a property residing in the world or in the sense organ and brain is no different than the question of whether odor resides in smelly substances or in the nose. Clearly, the answer in each case is: both. Detecting odors is an ability of the organism to discriminate the objective presence of airborne chemicals. The subjective experience of a particular smell *refers* to an objective reality. Similarly, color experience refers to objective properties (wavelength, reflectance, ecological context, etc.). Hence, color, like odor, can alternatively be understood as subjective experience, as a sensitive capacity of the nervous system, or as a property of the external world. And that is only half of the story, since the *experience* of color or odor refers also to the internal communication of the organism—its "language of the senses."

The primary qualities are supposed to represent observer-independent properties of the world: how nature is carved at its real joints. However, someone is required to do the carving, so there remains a subjective factor involved in so-called objectivity. Furthermore, not all physics concepts are based on the visual sense. Concepts of mass and force refer not only to vision but also to touch and proprioception. Scientific instrumentation is designed to free measurement from the human senses. Yet, we interpret such measurements in terms of our bodily experience as

organisms. The scientist cannot escape being an organism who participates in the measurement process. Indeed, the very ideal of "objectivity" is a survival strategy of the human organism.

Here is another way to understand why science does not deal effectively with consciousness: science describes the world from a third-person point of view. Our conscious experience, however, is a non-verbal description of the world from a first-person point of view. How can one describe this first-person perspective itself from a third-person perspective? Let's say I tell you I have a toothache. As a fellow human being, you know what that feels like; but as a scientist you may believe "feelings" are irrelevant to discussions about the physical world. If you happen to be my dentist, you can examine my mouth and see that there is swelling and discoloration, signs of infection and tooth decay. You also note signs of my discomfort, my winces and grimaces. As a human organism, you know perfectly well what these things feel like. (It is this "what it is like" that constitutes your own first-person experience and memories thereof.) However, as a scientist and professional, you must focus on what you observe (for the most part, visually—which is also, nonetheless, *your* first-person experience). You describe in detail all my symptoms and behavior to your dental assistant, who also knows perfectly well what I must be experiencing. Yet, the description itself can avoid referring to my subjective experience, which is incidental to the outcome of the procedures underway. Your reading of the symptoms (including my verbal statements, which are a form of behavior) and your skill as a dentist enable you to restore my mouth to health. I tell you afterward I no longer feel pain, and this confirms your success. But scientifically this is not strictly necessary to your diagnosis and treatment. You can relieve my pain without ever admitting there is such a thing!

However, science is not merely a disinterested study of the external world. It may seem so at the most abstract theoretical level. Yet, such high-level research is usually funded with the implicit understanding that there may eventually be some practical payoff. The social importance of science lies ultimately in its ability to relieve human suffering and enhance human experience. Ironically, it fulfills this role so well precisely by ignoring human experience. Science works by implicitly presuming conscious experience and human values, while pretending that an "objective" description of the world is independent of them.

13. Value and embodiment

What is "value" and where does it come from? The answer is simple: value comes from the interests of the body. Machines do not have interests, and so do not have values, except those given them by human beings. Machines have no stake in the outcomes of their actions or events in the world. A living organism, to the contrary, only exists *because* it takes these stakes very seriously. By definition, an organism must maintain itself to exist, and must reproduce successfully if its species is to exist. It therefore must evaluate stimuli with regard to these goals. Only those organisms exist that are appropriately programmed to prefer some outcomes and

events to others.[3] Obviously, a creature must eat other organisms, and not be eaten. It must maintain itself within a zone of conditions in which it can function well, and (if it reproduces sexually) must find and attract mates. All these provisions imply values that the creature must implicitly embrace. If it is complex enough to have a brain, then this brain will operate on the basis of these values as fundamental axioms.

In other words, value derives from the embodiment of organisms competing in the game of survival. Embodiment does not simply mean that the system is physical. It means that it is the kind of system an organism is: one that has come into being through natural selection and continues to exist by adhering to mandates that are products of natural selection. Embodiment is a relationship to the world. Feeling is a conscious experience of valuation. That is clear enough in the case of emotions and pleasure and pain. Not all valuation is consciously experienced, yet valuation is the basis of any organism's non-conscious processing. However, all conscious experience involves valuation, including sensory experience, thought, imagination, etc. Value necessarily permeates everything about an organism's relationship to the world upon which it crucially depends.

By this definition, machines are not (yet) embodied, even when they are physical. They do not have values because they are not organisms dependent on an environment. I do not doubt that machines *could* be organisms, self-defining and embodied in the above sense. Yet, aside from the sheer god-like power required to bring about such things, that would not be in the interests of human beings. Quite the contrary, artificial organism implies the menace of a parallel artificial ecology interpenetrating the natural one and competing with it for resources. Those resources would likely include us! Artificial organisms would be no more under human control than natural ones are now, and perhaps far less controllable if they are less vulnerable physically than their organic counterparts. The evolution of an artificial ecology would be as unpredictable as the evolution of the natural one has been. Human beings have painfully achieved a relative mastery over their natural environment. At best, they would have to begin that struggle all over again in an unnatural environment. At worst, they would be wiped out.

14. The role and consequences of self-consciousness

The experience we know as *awareness* is how an input is represented in a conscious intentional system. It serves also to register that input in memory, tagged for later retrieval. *Self-consciousness*, in contrast, is awareness of being aware. Only a being aware of being aware is (or has) a "self," in the usual sense of knowing that you exist. An executive agent isn't necessarily aware of itself, but its abilities are augmented if it is. Only such a self-conscious agent can ascribe to itself a point of view that is distinct from what is viewed. It can bracket its experience, considered to be a function of its own mental processes, rather than take it at face value as a

[3] If human beings have defied this rule to some extent, it is because cultural practices of the species make protective allowances for the aberrations of individuals.

direct view upon the world. It can thus own responsibility for its part in the construction of that view. Self-conscious experience then has an ambivalent character. As when watching the news, one is at once looking at the TV monitor in the room and at the remote but real scene transmitted. While attention is naturally drawn to one aspect or the other, self-consciousness insures we can choose focus according to context.

Self-consciousness entails awareness of one's role in producing, and thus in choosing, what one experiences. It is awareness of the 'subject' variable in the Equation of Experience. To some extent, this awareness confers the ability to change the experience, if not the world. That means greater flexibility and range of response, as well as a potentially more objective perception of the world. The concept of truth is born of utility, but grows beyond it because the self-conscious mind is inherently self-transcending and open-ended. In addition, self-consciousness implies an inner domain with which one identifies. In this domain, imagination is the research and development laboratory for the invention of new ideas. Recombining and transforming already-known elements, perhaps borrowed from the external world, is the basis of both grammatical language and of our powers to manipulate the world to our advantage. In this way, we can change experience indirectly as well as directly. The natural world is redesigned by human subjectivity, at first in imagination and thought, and eventually in deed, through technology.

Paradoxically, the evolutionary advantage of subjectivity is that it can lead to greater objectivity and control. To some extent, an instinctual or compulsive behavior can be brought under conscious control by bracketing the associated perception as an element of an inner domain. This reminds us, for example, that what may seem an objective situation may be no more than a transitory feeling. In this way, one can step outside the behavioral implications of perception to gain a more detached perspective. The normally overarching certainty of perception—seeing situations as clear-cut and events as external and *real*—is tempered by recognizing one's own contribution to that perception. One is therefore in a better position to look before leaping, even literally.

The possibility of emotional detachment and greater objectivity may well have facilitated cooperative behavior in a highly socialized species. One function of self-consciousness is to qualify and relativize the mind's tendency to perceive in absolutes—in objective, external, certain terms. However, this function itself must be qualified in turn: too much detachment is as dangerous as too little. Freedom from the organism's biological programming can be more dangerous than the programming itself. Freedom from social programming can also put one at risk. For, we can scarcely live as individuals without the support of others.

Thus, self-consciousness is a double-edged sword. It is common to think of it as an awkward state of social embarrassment. Self-consciousness in that sense involves a split of attention—part directed to external demands, part to how one should respond to them and how one will be viewed by others. With self-consciousness in the reflexive sense, there is also a split. Attention is divided between inside and outside. Consciousness often requires certainty, while self-consciousness invokes doubt and reflection. Moment to moment awareness of one's role in things can inhibit spontaneity even while it provides additional information and an enhanced

point of view. To the degree that the purpose of self-consciousness is to question one's experience, it can even be paralyzing. Yet, often the questioning pays off, allowing one to stand back and take a wiser perspective. In any case, one can assess whether the situation warrants the digression of thought or demands immediate action.

The awareness of self gives rise to an inner, subjective world apart from the external world and the life of the body. Indeed, one finds "oneself" *inside* the body, perhaps inside the head. It may seem that one's consciousness is the true inhabitant, the body a mere dwelling or vehicle: *you*, not your body, are who you *really* are. This arrogation of identity can pit one against the world and especially that part of the world known intimately as one's body.

Self-consciousness is undeniably useful to a social creature, to qualify the absoluteness of perception and the compulsiveness of behavior. This renders action more circumspect, which can also make it more cunning. To see oneself as an actor in the scene can facilitate either benevolent or selfish intentions. Yet, the evolutionary import of subjective consciousness goes further. The inner realm of *idea* constitutes a distinct domain, parallel to the physical world. This distinction plants the seed of separation from nature and gives rise to the dualism traditionally permeating modern thought. Within this inner world develop the templates for the cultural and technological worlds that substitute for nature.

When a creature evaluates a stimulus, implicitly in terms of its well-being, it is capable of feeling pain along with pleasure. Yet, only self-conscious beings can be said to *suffer*, which requires knowledge of one's context as well as the direct sensing of one's condition. The natural state of any organism is perforce one of limitation, mortality, and participation in an evolutionary contest. This state dictates the creature's perception, behavior, and very being. For a self-conscious organism, there is suffering in the awareness of these constraints, in the longing for possibilities it can conceive beyond such limitations. The very fact of being able to see the natural context of one's life implies a ground on which to stand outside it, at least in thought. But it also permits a painful feeling of being trapped within that context or within the individual circumstances of one's life. Self-conscious mind is perpetually in revolt against its perceived entrapment in the system of nature and its own identification with the body and survival. We are the creature with one foot in each of two worlds.

15. Consciousness versus intelligence

The concept of intelligence is as vague as that of consciousness. Let us informally define intelligence as a general ability to pursue goals and solve problems. As amply illustrated by computers, such an ability does not in itself entail conscious experience. Rather, consciousness is a special manifestation of intelligence.

Humanity is now faced with the looming possibility of super-intelligent machines: computers that can play chess better than people, cars that drive themselves, apps with which you can have a conversation, etc. These examples may be only the beginning of a new era of technology, in which machines can do everything that humans can do—and more and better and

faster. What will be the place of life and consciousness in a world dominated by super-intelligent machines? What is the relationship between intelligence and conscious experience? What are the limits of *non-conscious* superintelligence?

As early as the 1950s, scientists began to think seriously about the practical uses of self-replicating machines. Sent forth to colonize planets for automated mining, for example, these could overcome the limitations of biology in outer space. If life has not been able to permeate the universe, perhaps machines can! Technology now makes such ideas seem feasible. It also suggests that a human brain, for example, could be systematically replaced by a non-organic version, one nerve at a time. Many futurists now speak of an imminent technological "singularity"—an irrevocable runaway advancement of artificial intelligence. The idea of self-programming, self-improving, self-reproducing AI, operating at the speed of light instead of the speed of neural impulses, seems plausible. Such a development could rapidly outstrip human capabilities and displace humanity, raising many anxious questions—technical, philosophical, social, political, and moral. On many fronts, we are facing possible consequences of astounding new developments that could quickly change the form and definition of human being itself. As well as many science fiction films on this topic, there is a growing academic literature. Here, however, we will consider only the relationship between superintelligence and consciousness.

Natural organic life is a product of natural selection. What has been selected *for* in organisms is the ability to survive long enough to reproduce. Advantage, not truth, is the natural function of intelligence. In that sense, every existing creature is intelligent by definition. Yet humans have imposed their own culturally relative values and purposes on the concept of intelligence, which is basically human-centered. The idea of artificial *general* intelligence (AGI) is an idealized extension of certain capabilities valued by modern people.

The natural competition among organisms led to an arms race of mutual adaptations, resulting in a trend toward more complex forms with greater "intelligence." These new forms do not displace those lower in the pyramid of life, but add new layers. Generally, the number of individuals decreases with the size and complexity of the organism. Single-celled and simple organisms continue to constitute the great mass of the "pyramid," which supports the upper layers in the food chain. However, the dependence is mutual: viruses, bacteria, and worms feed on the corpses of mammals, for example. Perhaps the model for intelligence should not be the competitive edge of certain individuals or species, but the success of the biosphere as a whole. To the extent we are a threat to that, humans can hardly be considered intelligent.

Artificial intelligence was originally conceived and pursued as a product of rational design, as a tool for human use—not produced through some form of unnatural selection. However, the mechanistic worldview suggests that organisms can be copied, artificially engineered, or artificially evolved. Artificial organisms could display the equivalent of natural intelligence or vastly improved versions of it. Thus, there are two paths for AI to pursue. One is the development of increasingly sophisticated *tools* that remain under human control. The other follows the ancient dream of playing god by creating life, *tool-users* that will almost certainly *not* remain within human control. However, a significant question is whether these two paths can even remain distinct. It seems a "logical" next step to design tools that can design themselves,

setting them on an accelerating path to the sort of autonomy that characterizes organisms. It may seem desirable to improve on the natural design of organisms through genetic engineering, for example. Such thinking could result in artificial organisms that "improve" themselves many times faster than natural ones, eluding human control.

If natural evolution produced consciousness, possibly artificial evolution would also, for the similar reasons. But this possibility would depend on whether the real conditions of natural evolution are actually recreated artificially. So far, no computer is embodied like an organism. This does not mean merely that it is not connected to physical apparatus it can control, for we already have robots, computers that control industrial plants, and AIs connected to the internet. But, to be truly autonomous, an AI must have its *own* "body" to look after, in which it has a compelling vested interest. What it values would reflect that self-interest, its own survival mandate. Some theorists talk about programming values into AI, mainly to insure it respects human beings. However, these are not the values of the artificial organism itself, acquired through unnatural selection. As with natural organisms, the AI's own values would be its primary commitment, which would accord no special favor to human beings.

Certainly, we can imagine superintelligence *without* consciousness—AI tools. No one believes that chess-playing computers or self-driving cars are conscious. But then, human beings can drive cars without conscious attention, at least for periods. Furthermore, we can imagine lowly artificial creatures lacking consciousness; robotics has reached the stage of creating artifacts that look and behave very much like insects, for example. Yet, we might believe that insects are "sentient," because so much of their basic creature behavior is like ours: responding to stimuli, eating, moving about, evading destruction or seeking prey. If artificial insects (or even artificial bacteria) had the ability to self-reproduce, we would have reason to fear them regardless of any sentience they might have, just as we fear plagues.

We know, because we build them, that robot insects are simulacra, *not* organisms.
Is it inevitable that superintelligence would become conscious? This depends on how intelligent superintelligence can become without becoming an organism. There is little doubt that general intelligence, as people define it, will continue to increase in machines, even exponentially. How general can it be, and how far can it develop, without becoming truly embodied? Can human beings *prevent* it from becoming embodied? As long as intelligence does not have the characteristics of life—self-reproduction, self-maintenance, and self-definition within a survival contest—one may hope that it will remain within human control, following human values and directives. But, as soon as it becomes an agent in its own right, we will have lost any guarantee of control.

Even if superintelligence can be circumscribed to remain a human tool, machines will so far outstrip human capabilities that we will only be their undeserving nominal masters. It will make sense to defer to them the actual control of our world. In the past, humanity could justify its domination of other species by virtue of superior intelligence. We also sometimes used the dubious rationale that only human beings are conscious—that animals are "just machines." Now that machines have replaced animals as beasts of burden, how will we view them when they become vastly more capable than us? What will be our place in a world dominated by superior

machines? Our only reason for being will be our *consciousness*—the precious unique thing we have that machines do not.

But how precious is it, and how unique? We used to believe that we are conscious because (unlike animals) human beings have a soul that survives death. Science now holds that there is no such thing and that consciousness is a bodily function that dies with the organism. That means, of course, that "I" die when my body does. The fact that this "I" is extremely attached to its own continued existence reflects no more than a biological function, a survival instinct. Consciousness is a high-level cognitive strategy of some organisms, which ceases to have a reason for being when the organism ceases to function. It is no more (or less) precious than the organism it serves. In a world dominated by superior machines, human beings will have to ask themselves how precious is natural life compared to artificial life? If, further, the machines have evolved into superior *organisms*, they will have superseded us as the dominant form of life. They may ask themselves the same question and reach a different answer. Their only reason for keeping us around might be nostalgia—as pets or creatures in a zoo.

16. The scientist as epistemic subject

We've seen that experience depends on real events in the external world and also on the mind's responses. How it responds will depend on what "matters" to the organism, even the artificial organism. It is no coincidence that the English word *matter* is both a noun and a verb, for the material environment must loom in importance to the creature—as food, as predator, as resource, or simply as objects to navigate around or otherwise use. Nature at large is the context for human life, and natural materials are the basis of human technology and culture.[4]

Moreover, the root of the word *real* means *thing.* We are surrounded by objects, which are both real things and objects of our perception. What makes things real is that they can affect us and we them; they have their own existence quite apart from how we perceive them and try to control them. In other words, natural reality is mind-independent. The real external world for us is our natural environment. Whatever our thoughts and experiences, we live in the physical reality of nature at large, upon which we are crucially dependent as living organisms. Science is our best attempt to understand this mind-independent reality of the external world. It is a paradoxical enterprise, since science—like perception—is obviously an activity of the mind!

The modern science of physics was originally concerned with the behavior of matter, described from an objective point of view, which is essentially a space-time description. The primary physical variables are space coordinates and their changes over time: position, motion, velocity, acceleration. But matter has other properties besides extension in space; so these spatial quantities are compounded with others—such as *mass* and *temperature*—resulting in derivative concepts such as *force, energy,* and *field*. Physicists realized that different forms of energy could

[4] 'Matter' comes from the same root word as 'wood', which was the first natural material exploited by human beings.

be converted from one to another, and that mass and energy are also interchangeable. As these discoveries were made, the original and intuitive notion of *matter* became increasingly abstract and removed from any reference to sensory qualities and any connection with the scientific observer as a living organism.

Yet, the physicist *is* a living organism, with a literal and figurative point of view. The success of science depends on factoring out the subjective aspects of human perception and thought. But, to what extent is that possible, and at what price? The implicit program of science is not only to "understand" nature, but also to use and control it for human benefit. Such a goal reflects the subjective needs and desires of the human organism, which science is supposed to factor out. We seek to understand what the natural world is in its own right in order to better make it serve us. If knowledge is bent to serve subjective need, how objective is it? The plain fact is that we *cannot* know reality as it truly is, independent of sensory perception and its substitutes, but only through concepts and theories that are shaped by human needs.

Kant recognized this fact in the late 18th century. He referred to the notion of reality-as-it-truly-is (the *noumenon*) as "the world in itself" (in contrast to the *phenomenon*, which is how we actually experience it.) This distinction was not applied within science, perhaps because scientists assumed that what their mathematical theories described *is* the world-in-itself. After all, if the subjective aspects of experience have been removed, isn't what remains objective? However, we have seen that what actually remains is description in terms of the visual sense, and what was excluded is the input from other senses. Kant's essential point was missed: that *all* sensory input and all concepts based on it necessarily fail to access the world-in-itself, which is beyond knowledge and perception. Or, to put it another way, "knowledge" is *not* about the world-in-itself but about our *interactions* with it—whether through ordinary sensory experience and motor behavior, through measurement by instrument and active probing, through mathematical theorizing, or through engineering and industrial exploitation.

The scientist is in the same cognitive position as the ordinary perceiver, whose brain is encased in the skull and who can only "fly by instrument." The metaphor is all the more apt for the scientist, who relies heavily on actual instruments for measurement and on experimental apparatus for interaction. Instruments replace the natural senses; machines and technology replace the natural muscles. Science does not provide a glimpse of the noumenon, but only an alternative version of the phenomenon. As in the Equation of Experience, there is no avoiding the contribution of the subject factor; it merely takes a different path in science.

While measuring instruments substitute for the natural senses, and carefully designed experiments replace the interactions of the body, a mathematically defined theoretical model replaces the mind's perceptual model based on sensory input. It is then this mathematical model that is studied in place of the natural thing. This substitution is not trivial in its consequences, because the mathematical model is an idealization that may depart in significant ways from the reality it represents. Its advantage is to allow us to predict the behavior of a well-defined "system" (a conceptual machine), even though that system is a fiction. In many cases, the fiction is close enough to reality that the predictions are accurate enough for specific purposes. The disadvantage is that the only systems studied are those simple enough to be idealized that way.

The phenomena initially deemed "fundamental" were those that could be easily treated with the mathematics of the day. At the outset of science, this included systems such as colliding billiard balls, (two) gravitating bodies, a swinging pendulum, a stretching spring, etc.—all essentially artifacts. Since the development of computers, far more complex systems can be mathematically described. Yet, an underlying assumption remains: that the mathematical idealization describes the real thing. In effect, the real thing is presumed to *be* a well-defined mathematical idealization. On this misunderstanding, some scientists go so far as to claim that physical reality itself consists of mathematics! While this runs counter to our ordinary experience of the material world, it makes sense in a way. Just as the subject in our thought experiment has only a *theoretical model* to rely upon to guide interaction with the inscrutable external world, so it is with the scientist, whose best theories are mathematical. From the point of view of the subject in both cases, for all practical purposes the model *is* the reality!

So, let us review our metaphor of navigating by instrument. We left our cognitive subject in a sealed windowless "vehicle" (imagine a submarine, perhaps), with the task to figure out how to operate this thing and arrive at a "vision" of the outside world that permits the subject's continued existence. Now imagine that this subject is a scientist, who wants (impossibly) to know what is "really" out there. Like the ordinary subject, the best he or she can hope for is a theory that is not contradicted by the reality. For an organism in its evolutionary development, its "theory" is confirmed (or at least not contradicted) so long as its kind continues to exist. Its own individual existence does not necessarily confirm this, because it could be a matter of luck. But what is mere chance on the individual level cannot persist on the species level, which is a statistical effect involving large numbers of individuals.

The situation is only slightly different for the scientist. Rather than random "experiments" conducted through genetic changes, only some of which survive, the scientist can conduct well-considered experiments that do not risk personal annihilation. However, here too there is a collective (statistical) question. Our civilization has faith—perhaps too glib—that science and technology can ensure the long-term survival of the species. We are literally betting our descendants' lives on our current theories and practices. Because of the selective blinders mentioned above, we could be missing something essential that would make the difference between success and extinction.

Like the brain and the encapsulated subject in the thought experiment, the scientist has only readings of sensing instruments from which to interpret the world-in-itself. For the scientist, the "truth" of the natural world boils down to how it responds to probing with scientific instruments. The scientist can know only the results of such probing, in which observer and observed both play a part. Yet, the premise of scientific method is to know the world as it is "in itself," apart from the intervening roles of observation, experiment, and theory—as though the scientist were a mere fly on the wall.

The pretense of standing outside of nature while being part of it works well enough when studying a confined system upon which one can look without influencing it. However, that sort of system is another fiction, to which reality approximates only in limiting cases. A system can only be relatively closed to outside influence. In truth, nothing is perfectly shielded from anything

else. If it were totally closed to the rest of the world, then it would be closed to observation as well.[5] When the system of study is the universe as a whole, there is simply no place for the observer to stand outside it. At the scale of cosmology, the logic of isolated systems can only lead to contradiction.

17. Conclusion

Without self-consciousness there is simply the world. *With* self-consciousness, there is *experience*—of things in the world and also of things in the mind. That duality leads to an understanding that all experience (which includes thought, feeling, ideas, concepts, imagination, etc.) is a product of inside and outside together: the "equation" of experience. Yet, it is not only consciousness that is a joint function of subject and object. All that we *do* also involves both an input from the world and an input from the self. *That* understanding of behavior is crucial for ethics—not only the norms and rules for a given society, but now a species-level ethics capable of managing the future of the planet.

For, every creature lives by responding to its environment in such a way that permits its existence. Its perception and behavior represent a highly "interested" interaction of self and world. We are hardly different. What we take to be real or true or important is shaped not only by objective reality but also by our organic needs and desires. To act responsibly, the perennial challenge for us as social creatures is to distinguish what comes from without and what from within.

Survival—as individuals and as a species—depends, of course, on meeting our needs. But this depends in turn on how we relate to and interact with the natural environment upon which we are crucially dependent. It also hinges on how we relate to our conspecifics, the others who constitute our human environment. It seems we live in two worlds: one reasoned and idealistic, the other determined by our brutish biological inheritance. This is the basic conflict in our nature that must be resolved if our kind is to persist.

Other creatures on this planet are kept in mutual balance by their relatively limited capacity to impact the biosphere as a whole through their natural and unconscious behaviors. *Through* their consciousness, humans have exceeded those limits and become a force disrupting that natural balance. Whatever chance there is for us to restore and maintain that balance now depends on our consciousness, which means our ability to see objectively and impartially what needs to be done. This applies as much on a personal basis as on the species level, which must be integrated in an updated ethics. We must seek the objectively best for all, such as we can understand it. And to see objectively is to be clearly aware of our subjectivity.

---

[5] In some ways a black hole resembles such a totally closed system. However, it "leaks" Hawking radiation, matter and energy can fall into it (if not out), and its gravitational influence permeates the event horizon.