9 Consciousness as a Limitation

Introduction to the Third Section

In the first section of this book, Chapters 1-5, we defined the mind-body problem. The second section, Chapters 6-8, showed how this paradox arises from the operation of the subreality machine in the brain. In our third and last section, Chapters 9-11, we explore a particular aspect of this mental architecture, *consciousness as a limitation*.

By definition, computational machines process information. Further, this information being processed may include details about the internal activity of the computer itself. In other words, computers can be **self-aware**. The question is, how does this type of *computational self-awareness* relate to the human experience of *consciousness*? Is self-awareness sufficient for consciousness, or is something else required? And if something else is required, what is the nature of this "extra thing?"

We begin this chapter with a brief review of the concepts already covered. This leads us to the main topic of this section, the idea that consciousness arises from limitations of our mental capabilities. Our next stop is an examination of the "traditional" view of the mind, and how it is based on a fundamentally incorrect assumption. We end the chapter with a milestone in our quest, a formal definition of consciousness.

Where We Are

We started our journey with an examination of the main tool of science, *the method of reduction*. From this we learned that everything in our reality is composed of only two types of entities, *Information* and *Elements-of-reality*. This is the basis

of modern science, as well as our everyday commonsense. It has allowed us to understand everything from the structure of the universe to the process of life.

The method of reduction has served us well, but when we use it to examine consciousness we come to a disturbing contradiction. This arises because we can observe the mind from two different perspectives, the *third-person* and the *first-person*. The third-person viewpoint sees the mind as pure Information, nothing but the operation of the human brain. In comparison, from the first-person the mind is seen to be one or more Elements-of-reality, such things as qualia, free-will, semantic thought, and the present tense. This paradox is the mind-body problem in its most concise form. It is the heart of what we are seeking to understand, stripped of all that is superfluous and inessential.

This is a milestone in understanding consciousness for two reasons. First, it allows us to condense a wide range of subjective and poorly defined arguments into a single concise definition. Our investigation can then be directed at the root of the phenomenon, rather than its secondary effects. Second, it defines what would count as a solution to the mind-body problem. Since the problem is a paradox between two points of view, the solution must explain how and why this paradox arises. Further, this explanation must be compelling from both perspectives; it must be formulated in rigorous scientific terms, while simultaneously satisfying our introspective judgements. This is the task at hand.

Our next step was to develop a concept called the *Information-Limited Subreality*. This is something that could logically exist in our universe. We understand how it could arise, what its characteristics would be, and how it relates to the known laws of nature. It is based on the idea that reality is defined by observations, such as what we see, hear and feel, as well as what our scientific instruments tell us. For instance, our scientific and everyday observations indicate that we exist in a physical universe consisting of three dimensions of distance and

one dimension of time. This is what we observe; therefore, this is our reality. Lacking evidence to the contrary, we are justified in believing that these observations do indeed arise from an external physical universe, just as they appear to. That is, we conclude that the reality we perceive is genuine.

However, it is clearly within the laws of nature to alter observations by manipulating or distorting information. The Information-Limited Subreality takes this possibility to an extreme by creating a totally artificial reality for an observer. By definition, an observer trapped inside an Information-Limited Subreality has no knowledge of the external physical universe. Rather, this inner observer's reality is consistent with *another* physical universe, one that could exist, but does not. While the inner observer will acknowledge the possibility that he is trapped inside an Information-Limited Subreality, he will dismiss this as an unacceptable belief. Both the inner and the outer observers are justified and compelled to believe that their reality is genuine. Of course, the outer observer knows that the physical universe perceived by the inner observer does not really exist.

Into this setting we bring *The Inner Light*, the story of a scientist who becomes trapped inside an Information-Limited Subreality. As all good scientists do, he uses the method of reduction to classify the entities in his reality as either Information or Elements-of-reality. The problem is, everything that this inner observer classifies as *Elements-of-reality* will be seen as pure *Information* by the outer observer. In spite of this, each of these observers is complying with the most stringent rules of the scientific method, philosophical logic, and plain commonsense. They have reached the correct conclusion for their respective realities. Further, this does not require the observers to be conscious; it is a property of *what* is observed, not *who* is doing the observing. We call this disagreement between the inner and outer observers the *Principle of Relative Reduction*. But what is most important, the Principle of Relative

Reduction is something we fully understand; it may be strange, surprising, and even a little disturbing, but it is not mysterious.

Now we make the critical assertion: *the Principle of Relative Reduction is the solution to the mind-body problem.* This means that the first-person and third-person perspectives view the mind differently because there is an Information-Limited Subreality separating them. The first-person view is inherently from the inside of this Information-Limited Subreality, while the third-person view is from the outside. Introspection is the inner observer, while the world of science is the outer observer.

On the face of it, this explanation has the general form to explain what is needed to be explained. That is, it uses well understood scientific principles to show how introspection can see the mind as one or more Elements-of-reality, while science sees the mind as pure Information. In short, we have shown two things, (1) that the mind-body problem is a certain type of paradox, and (2) that the Information-Limited Subreality has the ability to cause this type of paradox.

However, this explanation requires us to accept a most extraordinary claim: *human consciousness exists within an Information-Limited Subreality.* This is an unsettling notion, completely at odds with our everyday perception of how our minds operate. We instinctively believe that the mind is an observer of the physical world; we seem to be directly aware of objects and events external to ourselves. But the Inner Light theory tells us that this is not true; everything that we consciously perceive is generated by a "subreality machine" within the brain. When we are awake, this inner reality is constructed to coarsely represent the physical world. When we dream, the subreality machine is running amok, creating an inner reality that is disconnected from the outside universe.

This is where we are. Our next task is to take a broader view of these ideas, searching for the general relationship between information processing and this strange thing we call consciousness.

From the Building to the Bricks

The Inner Light Theory asserts that human consciousness is based around an Information-Limited Subreality. This mental architecture accounts for our perception of a detailed and elaborate inner world, our ability to dream, results from change blindness experiments, and the very way that we experience reality. Most important, the Information-Limited Subreality has the ability to make us see pure Information as Elements-of-reality, the key aspect of the mind-body problem.

But now we want to expand our investigation to be as general as possible. We will do this by using a result from the last chapter. As illustrated by our perception of the color yellow, the basic operations used in information processing also have the ability to change Information into Elements-of-reality. This is an inevitable result of presenting a thing, but at the same time hiding how the thing can be reduced to more basic components. To use the metaphor from the last chapter, the Information-Limited Subreality is the *building*, while basic information processing operations are the *bricks*. Taking this further, the ability to change Information into Elements-of-reality resides within the bricks, not the architecture of the building.

To be more specific, there are some aspects of human consciousness that clearly arise from the structure of the Information-Limited Subreality. This includes our perception of a complex inner world, one that is distinct and different from the external universe. However, there are other aspects of our mind that can be adequately explained by much lower level operations. For instance, a full-fledged subreality is not needed to explain why we see yellow as a psychological primary color.

In developing a general theory of consciousness we want our understanding and conclusions to be as broad as possible. In particular, we do not want to define consciousness solely in terms of the mental architecture present in humans. That is, we want to accept the possibility that nonhuman creatures might be conscious, even though their "bricks" may be arranged in a different way.

Accordingly, in the remainder of this book we will carry on the discussion at the level of the "bricks," providing as little restriction as possible on how they are assembled. In short, we are moving toward a definition of consciousness that rests upon low-level information processing, and not the creation of a detailed inner reality. The rationale for this is simple; we want to consider an entity "conscious" if it views itself to be an irreducible thing, regardless of the other properties that it may or may not have.

A good starting point along this path is to revisit the structure of the human brain. It is easy to lose sight of just how complex an organ the brain really is. For instance, one might take the mental architecture we have presented and try to identify corresponding structures within our heads. Naively, we might expect to find a section of the brain that is the conscious observer, surrounded by brain tissue that creates the subreality. But unfortunately this isn't the case; science has found no singular areas of the brain that implement these functions.

It could also be argued that this relatively simple mental architecture is inadequate to explain key aspects of our introspective experience. If the human mind is an observer trapped within a subreality, this would explain how we see Information in the outside world as Elements-of-reality. For instance, this could account for qualia being irreducible. However, this doesn't necessarily explain how the observer could see *itself* as irreducible, such as experiencing semantic thought or mental unity. As an analogy, imagine being trapped with a translucent plastic bubble. Everything in the outside world will look distorted and unclear; however, everything on the inside of the bubble will still look as it truly is.

Figure 9-1 depicts a more realistic picture of the brain's exceedingly complex operation. It is clear from scientific studies that the "observer" is broadly distributed over the brain. For instance, vision is processed and understood in one area, moral judgement in another, initiation of body movement in

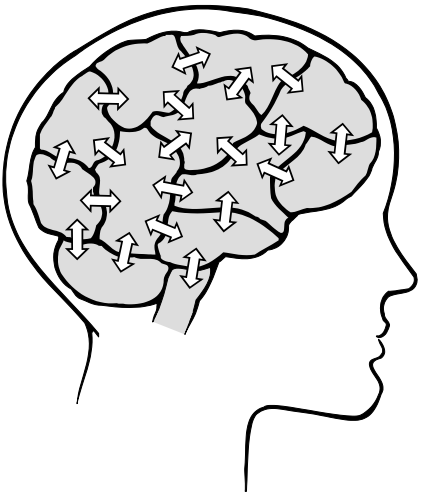

FIGURE 9-1
Distributed consciousness. The "observer" is broadly distributed within the brain, with processed data passing along internal pathways. The Information-Limited Subreality does not surround this observer, but is inherently intertwined with the neural circuits that create the observer.

another, and so on. These various areas are linked together by interconnecting pathways, passing summaries and high-level concepts among the fragmented and discontinuous regions. We have a poor understanding of how these individual regions interact; however, it is clear that there is no central place where it "all comes together." Many regions of the brain are involved in this thing we call "consciousness."

The point is, the Information-Limited Subreality within the brain is not a single bubble around an observer. At the least, it is a large number of smaller bubbles dividing the observer into many isolated regions. More likely, the information processing that creates the subreality is inherently intertwined with the neural circuitry that creates the observer. It may not even be possible in principle to say where one ends and the other begins.

In short, the human mind sees itself as irreducible because of limitations distributed *within* itself. These aspects of the brain are an inherent part of what we are, not some external structure holding us prisoner.

This brings us to our next topic, a discussion of how the traditional view of consciousness is mistaken. We will start with two stories, the *special child* and the *fully-aware being*.

What's so Special About a Special Child?

Suppose that sometime in the future you have a most unusual house guest, an alien exchange student from another planet. Since the goal is to familiarize your guest with humans and their culture, you arrange for the alien to meet a variety of people from different walks of life. One of the activities you arrange is a visit to a care center for mentally retarded children. Of course, political correctness suggests that we refer to these disadvantaged youths as "special" rather than "mentally retarded." Accordingly, you tell the alien that he will have the opportunity to spend a few hours with several special children, without elaborating on what this means.

After the visit you ask the alien what he thinks. He tells you he enjoyed the experience, and was very impressed by just how different and unique these children are. In an attempt to understand their nature better, he asks you to describe the "special" attribute that these children have. He has observed that these children are different in some way that he can't quite describe. He wants your help in identifying and defining exactly what must be added to a normal adolescent to create a special child. His question is very basic and to the point: *Just what is this "special thing" that these children have, that most children do not?*

When we hear this question we realize that the alien has made a fundamental mistake. The alien can clearly see that special children are different from normal children. However, he has incorrectly assumed that this difference results from some "thing" that special children possess, but normal children

do not. But this is not true; a special child is created by taking away abilities from a normal child, not by adding something. The behaviors and unique traits that the alien seeks to explain are a *deficit*, not an *addition*.

Why would the alien make this mistake? Perhaps the primary reason is his lack of experience with *normal* children. He is trying to understand how a special child is different from a normal child, without having a good understanding of what a normal child is like. Given this, it is understandable that he might make a mistake in interpreting the relationship.

In addition, we may have biased the alien by our comments. Our society refers to these children as "special" because the word is soft and without stigma, especially compared to the harshness of "retarded." Unfortunately, this word is somewhat inconsistent with its meaning in other contexts. When we say "special children," we mean that they have special needs. However, the term "special children" could be incorrectly interpreted to mean "exceptional" or "extraordinary," something above and beyond the normal child. Since this mistake has been made by many humans, it is not surprising that it would be made by an alien unfamiliar with our culture.

Lastly, it is common for humans, and presumably aliens, to think about things as a composite of parts. Further, these parts may include voids or missing regions that are treated as components in themselves. For instance, we speak of the "hole" in a doughnut, and an "unfilled" position in a company's personnel roster. Even though these are not actual things, we think of them as such to simplify the description of the overall object or concept. This might predispose the alien to think about the difference between a normal and a special child as a "positive" entity, rather than a void or deficit.

Regardless of these reasons, the fact remains that the alien is wrong. He has made incorrect assumptions, and they have led him to an incorrect conclusion. We will return to this story shortly, but first we need to define an important new concept, the *fully-aware being*.

The Fully-Aware Being

Again we will imagine a scenario occurring in the future. In this case we envision a group of scientists constructing an artificial person, an android that mimics human thought and behavior. They give their creation a body that appears very human-like from the outside, even though it is made from mechanical and electrical components, not biological tissue.

The android's "brain" is an advanced computer, carrying out algorithms, programs, neural networks, and other sophisticated information processing techniques. The android can perceive the world around him by means of his camera-eyes and microphone-ears. Further, he can understand what this sensory data means, being able to recognize objects in the environment and reconcile them with previously learned concepts. He can understand and generate speech, with the ability to carry on intelligent conversations. In short, the scientists design their creation to interact in the world the same way as you and I.

But most important, the android is designed such that he can monitor *everything* about his internal information processing. He knows the exact status of each and every digital bit and analog signal. He can observe the raw information gathered by his electronic senses, monitor its consolidation with previous memories, and examine how it affects his current mental status. There is nothing about his internal computational activities that he does not know. If you offer the android a cup of tea, he will send it away with a wave of his hand, and then apologetically tell you that he does not drink. But then he can discuss with you in the finest detail the billions of computer operations that were needed to carry out these actions. This is what we will refer to as a **fully-aware being**, a computational machine having a complete and detailed knowledge of its internal states.

Of course, such a creation is far beyond our current technology; however, it appears that this is a clear and direct extension of our present capabilities. Those that work in computer science expect that this will come about as computers become more sophisticated, and few knowledgeable people

would disagree. In addition, it is within the realm of possibility that a biological creature could be a fully-aware being. For instance, in the future we may encounter extraterrestrial aliens with the ability to monitor their inner mental operations to the last detail. Even stranger, one day we may be able to modify the human brain to be fully-aware. This premise is the topic of the next chapter.

For now, our concern is with the fully-aware android, something that science will be capable of developing at some time in the future. The question we want to pose and examine is this: *Is this android conscious?*

How the Traditional View is Mistaken

The "traditional view" of consciousness tells us *no*, there is nothing contained within this android that could result in it being conscious. According to this view, consciousness is something above and beyond computations and information processing; it is something "extra" that must be added. To complete their creation, the scientists must open the android's head and pour in a quart of "consciousness stuff," so to speak. Without this extra ingredient the android is nothing but a collection of mindless gears and cogs.

The rationale behind this view is very straightforward. The world of science sees the brain as a machine. In contrast, introspection sees a mind that cannot be reduced to machine operations. In fact, the mind has aspects that cannot be reduced to *anything*; such things as qualia, mental unity, and semantic thought are irreducible. Therefore, according to the traditional view, consciousness must be something in addition to the machine-like operation of the brain.

Of course, this is where the bottom falls out. The problems associated with this traditional view are severe and deep. For instance, if consciousness is something beyond information processing, why is there not the slightest scientific evidence for this "extra thing?" Worse yet, how can something that is not detectable by science interact so easily with the human body?

And just as troubling, why should we have this "consciousness stuff" at all? If information processing is sufficient to control our behaviors for mating, escaping enemies, and finding food, why would evolution give us consciousness in the first place? The traditional view is filled with these types of seemingly unsolvable problems. The more you try to grasp the thing, the more it slips through your fingers.

And here is the reason why. The traditional view of consciousness is based on a flawed assumption, the same error made by the alien visiting the special children. Consciousness is not some entity beyond full-awareness. Rather, it is a limitation, a deficit in one's ability to perceive and understand oneself. Introspection sees the mind as being irreducible because of these limitations, not because an extra entity is present. Consciousness is not created by adding something to full awareness; it is created by taking something away.

As an example of this, our fully aware android perceives the world through his camera-eyes and microphone-ears. Just as in humans, this raw sensory information must be processed before it is meaningful. For instance, the visual field must be broken into regions of similar color and texture, these regions grouped together into objects, and the objects recognized. Lastly, the relevance of the objects must be evaluated. Is this a face? Whose face is it? Is this an enemy or a friend? Hearing and the other senses have a similar hierarchy of information processing.

The important point is that our fully-aware android can perceive and understand each and every step in this process. He can perceive it all, from the raw data, through the intermediate stages, to the final result. If we show him a picture of George Washington, he will not only recognize it, but can tell us in the finest detail *how* he recognizes it. By definition, this is what it means for our android to be *fully-aware*.

But now we want to give our android a human-like mental experience. We do this by blocking his ability to perceive the lower stages of this information processing. We allow him to experience the result of the process, but not the process itself.

To test our modifications we show him the picture of George Washington and ask him what he sees. As before he tells us that the face is of the first president of the United States. But when we ask him *how* he knows this, we receive a blank expression. He does not know *how* he knows, only that he *does* know. The experience of seeing and recognizing the face has come to him without explanation, support, or evidence; it just appears in his mental processes. The experience that "*this is George Washington*" is now an irreducible part of his world. While our fully aware android saw the event as nothing but Information, our "conscious" android experiences it as an Element-of reality. This is the Principle of Relative Reduction in its most basic form, a blockage of Information flow resulting in pure Information becoming an Element-of-reality.

The Inner Light Theory tells us that *human consciousness* is something less than *full-awareness*, not something more. If we were fully-aware beings, we would know each and every operation being carried out by our brains, from the firing of individual nerve cells in our sensory organs, to the large-scale patterns of neural activity that represent our higher thoughts. There would be no mystery to our minds whatsoever; introspection would provide a complete and detailed understanding of exactly what we are.

But of course, this isn't our nature. Our physiology does not allow us to be fully-aware; the information in our brains is segmented into local groups without global accessibility. The low-level workings of the brain cannot be examined by the high-level workings. We do not know *how* we recognize a face, experience pain, or develop a thought, only that we *can* do these things. Our internal mental world appears to us as results without process, conclusions without justification, and things that exist in themselves without a supporting structure. Therefore, all of these things appear to the first-person perspective as irreducible. However, this is not because they are entities above and beyond the brain's activities, but because of the brain's limited ability to perceive its own operation.

Seeing the Forest Between the Trees

Why have we been mistaken about this for so long? Why is it not obvious that consciousness is a limitation and not "something extra?" Perhaps for the same reasons that the alien misunderstood the special children.

First, in order to see consciousness as a limitation, we must compare the human mind with a *fully-aware being*. Trying to compare it with a lesser computational machine, such as a business computer, is meaningless. Unfortunately, no human has ever had direct contact with a fully-aware being; we know them only through our imagination and thoughts. If fully-aware beings lived among us, perhaps it would be obvious that our minds are limited compared to their computational powers, not the other way around. In other words, understanding the nature of the mind requires a reference point, and this reference point is something we have little experience with.

Second, human nature itself predisposes us to think of the mind as something beyond the neural machinery of the brain. As one example, consider how we cope with death. Humans are social creatures, forming their lives around closely woven circles of family and friends. These relationships and bonds are often viewed as the most important things in our lives. But death rips this apart, attacking the survivors on a fundamental level. However, this extreme loss and pain can be minimized by the simplest of acts, merely believing that the mind of the departed still survives in some manner. Nature literally tortures some people into believing that consciousness is something beyond the physical body.

Third, as previously discussed, it is human nature to think of voids, missing regions, deficits, and limitations as positive entities. For instance, a doughnut is thought of as a piece of sweetened bread, *plus a hole*. And there is nothing wrong with this; it simplifies our understanding of the world. The problem is, this predisposition to "positive entities" can bias our analysis of the world. An unexplained phenomenon is inherently viewed

as a "thing," rather than a "void." We must overcome this inherent prejudice to see limitations as they truly are.

But regardless of the reason, the traditional view of consciousness is mistaken. The first-person perspective sees the mind as irreducible because of its limited observational power, not because additional entities are present. This paves the way for stating a formal definition of consciousness, our seventh major teaching:

Major Teaching #7:
The Definition of Consciousness

Consciousness is the irreducible entity a computational machine perceives itself to be, as the result of (1) an ability to observe its own high-level workings, and (2) an inability to observe its own low-level workings.

The Tale of Big Head Bill

This concise definition accounts for consciousness from the third-person view. That is, it provides purely physical reasons why humans claim to have inner experiences involving Elements-of-reality. But now our task is to examine this explanation from the first-person perspective. This places us face-to-face with the most difficult aspect of the mind, explaining the personal and private view we have of ourselves. In the end, each of us will look at the arguments presented and ask the questions: *Does this explain what I feel, what I perceive, what I experience? Does this unify my objective knowledge of science with my subjective knowledge from introspection?* And the most basic question: *Is this really what I am?*

This leaves us with a difficult task, trying to touch one's innermost thoughts and feelings using a grossly inadequate tool, language. How can we explain the feeling of pain, or what it is like to see blue, or what it means to freely make a decision?

The arguments of science, rational as they may be, seem ineffective at doing this. They simply do not connect with our inner world in a way that makes us proclaim, *Yes, this describes what I am.* But if this can't be done through the power of rational arguments, how can it be done at all?

Fortunately, this is not as hopeless as it may sound; artists and poets make their living by invoking and controlling our introspective experiences. And this is the same course we must take to understand the mind from the first-person view. We must use words to invoke and control our introspective imagery, allowing us to experience the concepts, rather than just knowing them by formal logic and rational thought. Such is the strategy of the next chapter, *The Tale of Big Head Bill*.

This is the story of a man being transformed from a normal human into a fully-aware being. In essence, this is a journey across the gap separating the first and third-person views of the mind. Our title character starts with the same introspective experiences as you and I, such indescribable things as free-will, mental unity, semantic thought and so on. But then an alien drug changes his brain structure, allowing him to perceive the mental processes that are blocked in normal humans. Step by step he comes to know the true nature of his introspective world, a hierarchy where thoughts, feelings, and judgements are built upon basic computational processes. As he reaches full-awareness, he perceives and understands his mind in the same way as one observing him from the outside. He has crossed the gap, unifying the first and third person views of the mind. Now let's hear about the journey in his own words.