WONDER AND UNDERSTANDING

by Margaret A. Boden

Abstract. Wonder is a root of the religious experience, and the desire to understand drives science. If wonder and understanding are fundamentally opposed, religion and science will be also. But only if wonder is limited to the contemplation of magic or mysteries is religion in principle opposed to science. The aim of science is to explain how something is possible. Understanding how something is possible need not destroy our wonder at it. Recent scientific theories of the human mind—albeit based in computer technology—increase our wonder at its richness and power.

In its broader context, this issue of *Zygon* on artificial intelligence and human consciousness is about religion and science. That is to say, it is about wonder and understanding—and whether they can coexist.

Wonder is a prime component of the religious sentiment. Anyone devoid of wonder is incapable of a religious response to the world and to our experience in it. Such a person might assent to certain pronouncements or dogmas of a "religious" nature, but lacking this emotional dimension their assent cannot be religious in spirit. Thus Thomas Aquinas insisted that his "Five Ways to Prove God's Existence" cannot force a Christian commitment, or indeed any form of emotional assent: they offer some intellectual understanding of a nonpersonal First Cause, and no more.

Intellectual understanding, of course, is the prime goal of science. There is some disagreement over what type of understanding science should seek. The natural sciences typically enable us to predict, and often to control, future events; they describe reality in quantitative mathematical terms. Whether prediction, control, and mathematical elegance (quantitative or not) are also possible in the human sciences is a controversial question—to which several papers in this *Zygon* issue are relevant. But the basic aim of science in general is to make the world intelligible by showing how various phenomena are possible.

Margaret Boden is professor of philosophy and psychology, University of Sussex, Falmer, Brighton, Sussex BN1 9QN, England. She presented this paper at the annual conference ("From Artificial Intelligence to Human Consciousness") of the Science and Religion Forum, Canterbury, England, in April 1984.

[*Zygon*, vol. 20, no. 4 (December 1985).]
 ISSN 0044-5614

To ask how something is possible, one must have been surprised by it in the first place. So it cannot be denied that science grows *from* wonder—or at least from curiosity, which is very closely related. But must it grow *out of* wonder? That is, are wonder and understanding not merely different, but fundamentally opposed?

Wonder is an emotional response to someone or something seen as valuable and also as beyond our reach. (Both aspects distinguish wonder from mere curiosity.) The way in which the object of wonder is beyond our reach varies. We wonder at beauty and saintliness and at praiseworthy human achievements of many kinds. A person who rises above suffering, not allowing it to dull human sensibilities (perhaps even using it to sharpen them) is cause for our wonder, as is the person who achieves distinction in the arts or sciences—or even on the battlefield.

"Distinction" here does not mean mere "difference," or Genghis Khan would be the most wonderful of men and Satan the most wondrous angel. Positive evaluation of what is wondered at is crucial; without it, wonder shades into fear. But it is often accompanied by fear, especially if the object of wonder is perceived as being more powerful than oneself or is associated with some greater power. So it is not absurd to say, "The fear of the Lord is the beginning of wisdom." Wonder is closely related also to awe and worship, for awe is a high degree of wonder, in which fear and respect are prominent. And worship is a deliberate expression of awe, which in theistic religions is directed to personal objects deserving gratitude from the worshippers.

We commonly wonder at things we do not understand, things which are "beyond our reach" insofar as they are unintelligible. So it may appear that science and religion must be radically opposed. For the aim of science is precisely to understand. It is not merely that science offers alternative "facts" or "explanations," which conflict with religious text, tradition, or dogma. Indeed, one might argue that this is impossible, on the grounds that religious statements—*qua* religious—do not attempt worldly explanations and should not be interpreted as though they did. Rather, it seems that even if the specific findings of science do not conflict with religion, the scientific attitude does: understanding must drive out wonder.

Sometimes, to be sure, it does. When wonder is based on pure ignorance or on error and illusion, it must fade in the light of understanding. Science is therefore opposed to superstition. But understanding may lead in turn to a new form of wonder, which cannot be so easily destroyed.

One of the early cyberneticians tells the story of how as a very young child he was for a while utterly fascinated by circles. He would collect

circular things—coins, bottle-tops, tins—which he kept in his toy-cupboard, and which he used to draw circles of many different sizes. One day his parents told him that there is an instrument which can draw any circle whatever (below a certain size). He wondered greatly at this idea and could hardly wait to receive this marvel as a gift. He thought of it as some sort of magically changing item which could transform itself into equivalents of all the different objects he had collected in his cupboard.

Then, he was given a compass. He was horribly disappointed, for there was nothing magical at all about the compass. It was boringly simple, and its "power" was transparent even to an infant. He still remembers this day of disillusion as a traumatic event in his childhood.

But he now has the maturity to see that the compass was indeed wonderful—and the mathematical principle it embodied is even more so. Its simplicity is "boring" only to those who feel that baroque confusion is a necessary mark of the wonderful. Even theologians have never believed this.

A certain sort of simplicity, capable of generating many superficially varying cases, is universally accepted as a criterion of "good" scientific theory. It is harder to define this simplicity than to cite examples of it or recognize its intuitive appeal; but for our purposes this does not matter. Galileo Galilei declared mathematics to be "the language of God," and theoretical simplicity has won many battles in the scientist's fight to understand the physical universe. (Some of these battles were engaged in by people carrying a theological or quasi-religious banner: think of Isaac Newton's views on God or Johannes Kepler's faith in the five perfect solids.)

Science has transmuted some of the world's mysteries into puzzles, many of which have been solved, to boot. Thomas Kuhn's description of normal science as puzzle-solving is a forceful reminder of the absence of wonder in much scientific practice. This is not surprising, for wonder that is based on ignorance or illusion must retreat in the face of understanding.

To retreat, of course, is not to disappear. With the advance of physics and chemistry, the focus of "ignorant" and "illusory" wonder has gradually shifted toward the biological and psychological end of the scientific spectrum. The morphology and behavior, and above all, the conscious experience, of living things have all been represented as being inexplicable by scientific means, and therefore as being proper objects of wonder. Moreover, these aspects of life are greatly valued by us. It is no accident, then, that these matters are often highlighted in discussions on science and religion.

Animal behavior, for instance, is something which causes most people to wonder. For centuries, it was referred to as "instinct": some-

thing marvelously akin to intelligence but nevertheless distinct, since only human beings were allowed to be really intelligent. This vague usage covered both the insect's ability to find other insects and the chimpanzee's communication with its fellows. It did not distinguish carefully between the many different forms of contextually appropriate animal behavior. Still less did it offer an intelligible explanation of how such behavior is *possible*.

Ethology and psychology are now helping us both to appreciate the range of animal behavior and to understand some of the *prima facie* surprising facts about it. In some cases, this added understanding may tempt us to react much as the infant cybernetician did to the compass. That is, the newly discovered simplicity may drive out our wonder, leaving us with only the "brute facts" of science.

This response is especially common when the animal's adaptive behavior is found to be rigidly determined by specific environmental stimuli; for then there is no room for "intelligence," nor even for "instinct," insofar as this is conceived as something very close to intelligence.

Consider the hoverfly, for example. A hoverfly is able to fly to meet another hoverfly in midair—which is just as well, since they need to be at the same place if they are to mate. One might initially assume that it does this in a way analogous to that in which a person (intelligently) alters direction on recognizing a friend on the other side of a city-square, adjusting the path as necessary if the friend suddenly swerves. Sentimentalists would thereby be led to expound on the wonders of nature, as illustrated by the marvelous powers of the humble hoverfly. Even more sober souls (given the assumption in question) might feel some sympathy for such a view. It turns out, however, that this assumption about how the hoverfly manages its social life is false.

On closer examination, there is nothing like the flexible selection and variation of means which characterizes truly intelligent behavior. For the fly's flight-path appears to be determined in accordance with a very simple and inflexible rule. This rule, which could conceivably be built into the insect's brain, transforms a specific visual signal into a specific muscular response. The fly's change of direction is found to depend on the particular approach-angle subtended by the target at the time. The creature, in effect, always assumes that the size and velocity of the seen target are those corresponding to hoverflies ("in effect," because there is no reason for thinking that it really *assumes* anything at all). When initiating its new flight-path, the fly's angle of turn is selected on this rigid, and fallible, basis. Moreover, the fly's path cannot be adjusted in midflight, there being no way in which it can be influenced by feedback from the (unpredictable) movement of the target animal.

Certainly, this evidence must dampen the enthusiasm of the person who had originally wondered at the similarity between the hoverfly's behavior and the ability of human beings to intercept their friends. The hoverfly's powers have been demystified with a vengeance, and it no longer seems worthy of much respect.

As with the compass, one may see a deeper beauty if one looks at the general, evolutionary principles which enabled this simple behavioral mechanism to exist. But the fly itself cannot properly be described in anthropomorphic terms. Even if we wonder at evolution, we can no longer wonder at the subtle mind of the hoverfly.

Many people believe that this "dehumanization" is a foretaste of what science must say about human behavior too. Indeed, behaviorism has already likened us to rats, which it described much as though they were furry hoverflies. Notoriously, this psychological theory offered an image of human beings which made wonder difficult, if not impossible. Behaviorism ignored consciousness, represented mental processes as strings of associated responses, and denounced concepts such as "freedom" and "dignity" as socially harmful mystifications.

This common belief—that science must deny the mind or regard it as a superfluous epiphenomenon—is mistaken. Current scientific ideas have helped to *increase* our respect for the human mind. In particular, ideas drawn from artificial intelligence have helped many psychologists to a greater appreciation of the wonder that is the human mind (see Boden 1977).

How can this be? How can a computer-based enterprise such as artificial intelligence have anything at all to say about psychological phenomena? And how can *any* scientific understanding of the mind increase our appreciation of it as a *wonder*?

The basic reason why artificial intelligence can have something of interest to say about psychological matters is that its central concept is *representation*, or *meaning*. Correlatively, because the natural sciences such as physics and chemistry—and even biology—have no place for this concept, they are in principle unable to do justice to the mind. For the mind is essentially symbolic, or representational.

Humanist philosophers have always insisted, rightly, that we conceive of the mind as an origin of subjectivity. Psychological descriptions and explanations concern subjective rather than objective truths. This is why it is not logically absurd for socialists to study the psychology of fascism, or atheists the psychology of religion. (By contrast, it would be absurd for biologists to hunt for the Loch Ness monster if they were convinced no such animal exists.) Both Sigmund Freud and Carl Jung, for example, studied the source and content of religious beliefs, and both agreed that they have important psychological effects. But neither

held that these effects ensue because of the objective truth of the beliefs concerned: Freud because he held theism to be objectively false, Jung because he held that religion is not a matter of *objective* truth at all.

These distinctions simply could not be made without the concept of representation, symbolism, or meaning. But representations, and their transformations, are precisely what artificial intelligence is about. Computational psychology, influenced by this new technology, conceives of the mind first and foremost as a representational system. It asks what sorts of representations human (and animal) minds can hold and how they are built, compared, and transformed in the service of the organism's interests. Concepts like purpose, choice, reasoning, knowledge, and even creativity have come back into "scientific" psychology accordingly.

To be sure, computer scientists borrowed the ideas of representation and symbol from our everyday psychological language in the first place. But this does not affect my argument: much as physics borrowed "force" and thereby helped us to understand force better, so computer science has borrowed "representation" and is helping us to understand representation better. This is not to deny that there are unsolved philosophical problems regarding representation. Still less could one deny that current computer science leaves many practical and theoretical questions unanswered—and even unasked. But the crucial point is that artificial intelligence promises to advance scientific understanding of the mind, because it studies representation.

However, this promise is perhaps also a threat. Surely, this increase in scientific understanding must destroy our wonder for the mind, much as ethology undermines one's respect for hoverflies?

Not so. Ethology informs us that the mind of the hoverfly is much *less* marvelous than we had imagined. Our previous respect for the fly's intellectual prowess is thereby shown to have been mere ignorant sentimentality. But computational studies increase our respect for human (and some animal) minds, by showing them to be enormously *more* complex and subtle than we had previously recognized.

Admittedly, poets and novelists have long had an intuitive sense of some of the subtleties concerned: consider Marcel Proust's depiction of memory, for example. Theoretical psychologists such as Freud had relied on similar insights in discussing symbolism within the dreams and dramas of everyday life. But such notions have remained "literary" and intuitive, rather than being scientifically rigorous. Moreover, Freud was not representative of theoretical psychology: the behaviorists, as already noted, offered a very different picture of psychological reality. Even Freud grossly underestimated the degree of complexity of the mental processes he described.

Artificial intelligence offers us, for the first time, a way of investigating these matters in a precise and detailed way. Psychologists now have a much greater appreciation of the complexities involved in natural language or vision, for example. Because we use language effortlessly, it is natural to assume that the psychological processes responsible—whatever they may turn out to be—are fairly simple in kind. This assumption has been shown to be mistaken by attempts to write computer programs capable of using English or French sensibly. Likewise, vision turns out to be a much more complicated achievement than anyone had previously realized. In short, underlying these everyday powers are myriad psychological processes, which generate and interrelate our representations of the world, including our representations of other people's minds and meanings.

These "myriad psychological processes" are of course largely unconscious, but the mental phenomena which interest us most of all include many which find expression as conscious experiences. Self-consciousness is normally taken to be essential to the concept of a human individual, or person. Moreover, consciousness plays a significant role in the religious response to life. I am not thinking here of "religious experience" in particular (visions, conversion experiences, or "oceanic feeling"), but of our wonder at the phenomenon of consciousness in general. Pierre Teilhard de Chardin is not alone in finding it awe-inspiring or in according it a deep religious significance. Indeed, consciousness and the soul are closely linked concepts.

A theoretical psychology that denied consciousness, or ignored it, would to this extent be at odds with the religious attitude. It might seem to follow that a scientific psychology based in computational studies could not easily coexist with our religious life.

This conclusion would be mistaken. Cognitive psychologists are the first to admit that most of the computational processes they posit are not introspectible. They can afford to do so, for they are not wedded to the philosophical view that only conscious phenomena (and mental processes which could become so) are truly psychological. But of course they allow that a theoretical psychology should help us understand the nature of consciousness and also why some processes are conscious while others are not.

There is even more disagreement about how computational ideas might apply to consciousness than there is about the computational processes underlying language. A number of intriguing suggestions have been made; two examples should suffice to show that consciousness does not have to be ignored—and might even be illuminated—by this new scientific approach.

Many psychologists have pointed out that consciousness seems to arise when we are performing a familiar task but suddenly meet with

some difficulty. If our shoelaces get in a tangle, for example, we stop tying them "unthinkingly" and start to concentrate on the details of what we are doing. It has been suggested that this phenomenon may have something in common with the switch from a "compiled" to an "interpreted" program, at points of difficulty during execution.

A compiled program is one whose instructions (originally written in a high-level programming language) have all been translated into the machine code of the computer that is going to be used. When the computer runs the program, only these unvarying, machine-code instructions are available to it. An interpreted program is one in which each high-level instruction is translated into machine code at the point when it is reached during run-time. This enables "one and the same" high-level instruction to be translated differently at different times, according to the specific circumstances of execution. Because time is needed to translate each instruction, a program takes longer to run when interpreted than it does when compiled. So efficiency is best served by compiled programs. Flexibility, however, is best served by interpreted programs.

So as to get the best of both worlds, some programs are run in their compiled form until a difficulty is encountered, at which point the interpreted version of the program takes over. The interpreted version allows for execution details to be varied so as to overcome the difficulty—if necessary, by on-the-spot planning which works out a new series of operations to fit the particular difficulty involved. For this to happen, the execution details must themselves be made the focus of information processing. This would have been impossible in the compiled form, since a compiled program runs "automatically" without variation according to unforeseen circumstances.

Analogously, the physical details of one's shoelaces are not normally noticed (not accessible to consciousness) when one is tying them in a bow. But if one's fingers are prevented from carrying out their normal, automatic tying movements, one will then pay attention to the laces. One may notice, for example, that there is a knot in one of them. If so, one will have to either undo the knot before carrying on "as usual," or deliberately (*sic*) tie the laces in such a way that the knot ends up at a position where it does not impede the tying of a secure bow.

The second example of how computational ideas might help us to understand consciousness relates to psychiatric cases of so-called dual personality, or dissociation of consciousness.

The fictional Dr. Jekyll and Mr. Hyde are less strange than some of the actual clinical cases, for neither of the "personalities" depicted by R. L. Stevenson knew of the existence of the other. The well-known case concerning Eve White and Eve Black is more complicated and that

of Sally Beauchamp more complex still. Eve Black apparently had access to the consciousness (the thoughts and emotional responses) of Eve White, but not vice versa. Moreover, Eve Black sometimes played tricks on Eve White: for instance, she once bought a highly unsuitable dress and left it in the wardrobe so that Eve White would have a difficult time explaining it away to her husband.

The familiar way of conceiving the mind—as a strictly unitary theater of consciousness—gives us no way of understanding how such a phenomenon is even possible. But if we think of the mind on the analogy of a complex computer program, understanding how such a thing can happen seems relatively easy. The memory data and operations of one subroutine can be accessible to another, which may use this information in selecting its own operations, some of which may even input new data to the first. However, the first subroutine need have no comparable access to the second, and may not even "know" of its existence.

To the extent that one finds this example persuasive, one's wonder at such psychiatric cases will be diminished, for it was initially based largely on ignorance and incomprehension.

However, our wonder at consciousness in general may not be destroyed. Many would say that the greatest wonder is not that our consciousness is thus-and-so (however surprising the thus-and-so), but that it exists at all. It is this wonder which underlies the religious response to the general phenomenon of consciousness which I mentioned earlier. It cannot be diminished by any scientific psychology, whether computational or not, which leaves the basic philosophical conundrum of consciousness unsolved.

Many would claim that science in principle cannot solve or even shed light on this still-unsolved philosophical question. But suppose that some future philosophy of mind is developed, with or without the help of science, which makes intelligible the existence of consciousness in a material universe. Would consciousness therefore cease to be a proper cause for wonder?

Surely not, for to say so would be to commit "the fallacy of the compass." The temporary disaffection of the infant cybernetician was irrational. Grown to maturity, he is still capable of *valuing* circles for their superficial simplicity and aesthetic form. In addition, he now understands the simple mathematical principle which relates one circle to another and which enables them to be generated at will by anyone with a compass to hand. This appreciation of the underlying principle has made it possible for him to wonder at circles in a deeper sense than before.

Likewise, it is *prima facie* inconceivable that mere matter could generate consciousness. But a scientifically informed philosophy of mind

that explains how this is possible would give us reason to wonder at matter *more* not *less*. Matter, organized according to certain principles, would then have been shown to be the "compass" for generating consciousness. Our valuation of consciousness as such need be in no way diminished. But we would value matter more, for we would have to allow that matter is not quite so "mere" as we had thought.

Charles Darwin made a similar point in response to those who said that his theory of evolution was antireligious. Rather than denying our wonder at God's creation, evolutionary theory can increase it: "[I think it] an idea from cramped imagination, that God created the Rhinoceros of Java and Sumatra, that since the time of the Silurian he has made a long succession of vile molluscous animals. How beneath the dignity of him, who is supposed to have said 'Let there be light,' and there was light" (Gruber 1981). On the contrary, Darwin saw the more magnificent view as that all these creatures, with their more aesthetically appealing cousins, have been produced by the body's laws of harmony.

Undeniably, however, Darwin's theory of natural selection contributed to the weakening of certain forms of religious belief in the late nineteenth century, and the battle between science and religion sparked off by his theory was only one skirmish in a long-standing war. History shows us many examples where religious faith has been destroyed by advances in scientific knowledge. Sometimes, the religious faith in question was an entire pattern of responses to the world and a total life-style. In other cases, the religious faith destroyed was belief in specific items of dogma or trust in (certain sections of) holy writ.

If my argument has been correct, only the second class—in which specific illusions gave way to new knowledge—was inevitable. Faced with William Wilberforce and Thomas Henry Huxley, one can cheer only for Huxley. It is of course understandable that Wilberforce's cause as a whole should have suffered, given the rout of such an unworthy champion. But the giving up of an overall religious attitude was not demanded by logic, nor even by psychology.

In summary, the wonder that is based on ignorance and/or illusion is indeed incompatible with understanding. Insofar as religion is motivated by, or justified in terms of, this sort of wonder, it is threatened by science. But wonder that is based on a positive valuation of the world, and on its awe-inspiring richness and harmony, cannot fall prey to science. This sort of wonder may even be increased by scientific understanding.

REFERENCES

Boden, M. 1977. *Artificial Intelligence and Natural Man*. Cambridge, Mass.: Harvard Univ. Press.

Gruber, Howard E. 1981. *Darwin on Man*. Chicago: Univ. of Chicago Press.