Chapter 3

Elements of a Philosophy of Mind

Ben Goertzel & Ted Goertzel

1. WHY PHILOSOPHIZE?

Philosophy can seem stodgy, boring, and old-fashioned, especially when compared to ultra-high-velocity, mega-sex-appeal disciplines like Internet AI engineering. But yet, the timelessness of philosophy is probably its greatest strength. The fundamental nature of the universe, which is what philosophy is about, doesn't change over time. In a sense it even pre-exists the notion of time. As I see it, the only way to really understand the latest greatest thing – and the new new new thing, just around the corner – is to tie it in with the truths of the ages.

Of course, philosophy cannot solve science or engineering problems, but it can point science and engineering in interesting and useful directions, saving time that would otherwise be spent exploring conceptually barren "dead ends" and creating whole new areas of investigation. To make a moderately loose analogy, one might say that the relation between philosophy and science/engineering is like the relationship between design and coding in software engineering. Design doesn't tell you how to code, but it gives you the structure according to which the coding will proceed. While coding, you will often discover new things that force you to go back and re-design, but this doesn't mean that the design process was useless. Similarly, philosophy doesn't tell you how to do science or engineering, it just gives structures according to which practical work can proceed; and when the empirical or theoretical results feed back and modify the philosophy, this is merely part of the process.

With this view in the background, this chapter very roughly outlines some of the philosophical ideas that underlie the treatment of mind to be given throughout the book. It is quite sketchy, since philosophy of mind is a big topic and there is only one chapter allocated to it here. But hopefully it will be enough to orient the reader's mind in the right direction. If you approach the notion of Internet intelligence with the presupposition that the mind is a system of logical propositions, or that the mind is a behaviorist stimulus-response system, or that the mind consists of a collection of modules programmed from birth whose parameters are adjusted by experience, then you will find the idea of an intelligent Internet hard to digest. On the other hand, if you approach it from the perspective outlined in this chapter – the mind as a self-organizing, self-producing system, a population of components that create each other by recognizing patterns in each other – then you will find the notion of Internet intelligence, and related specifics such as the Webmind system design, much more approachable.

Much of the chapter consists of a review of ideas from three important and relevant philosophers of mind: Charles S. Peirce, Buddha, and Friedrich Nietzsche. I'll also review aspects of the thinking of a contemporary postmodernist philosopher, Kent Palmer. None of these philosophers are standardly cited in computer science circles, and this is not coincidental. Webmind is not a standard AI system; the Internet is not the kind of system computer scientists would ever have designed intentionally ... it emerged from a vast collection of disparate efforts, turning into what it is today almost in spite of the limited philosophies of most of those who helped create it.

In Chapter 5, I'll present some more formal, mathematical explorations that, not surprisingly, accord very well with these philosophical investigations. The vision of mind as a network of interconnected patterns turns out to be a very deep one, with profound philosophical and psychological implications that are not immediately obvious from the mathematical expressions of the vision. As we proceed to more detailed mind design issue we will need both the formal and the intuitive views of the vision, to help us keep everything straight.

2. PEIRCE'S CATEGORIES OF BEING

Philosophers enjoy dividing the world into categories, and they have done so in very many ways. Of all the category systems that philosophers have come up with, the one that I find the most useful is also one of the simplest ones. The 19'th century American philosopher Charles S. Peirce

believed that on the most fundamental level, the universe was organized numerically, and he divided the universe into categories according to the first three integers, called First, Second and Third. He believed that the small integers - particularly one, two and three - were not just arbitrary human creations, but fundamental organizing principles of the universe. I will follow Peirce and then extend him a bit, introducing a new category called Fourth.

The famous psychologist Carl Jung (1955) used the term *archetype* to refer to patterns which are pervasive and recurrent, which seem to express the fundamental nature of phenomena. Although Peirce did not use that word, he observed the archetypal nature of the small integers:

> Three conceptions are perpetually turning up at every point in every theory of logic, and in the most rounded systems they occur in connection with one another. They are conceptions so very broad and consequently indefinite that they are hard to seize and may be easily overlooked. I call them the conceptions of First, Second, Third.

I believe that Peirce was right on about the archetypality of 1, 2 and 3, and that the other small integers like 0, 4 and 5 also have profound archetypal meaning. These very simple and abstract archetypal considerations turn out to be very useful in cutting through some of the thorny conceptual issues of AI – we will return to them later in the book, for example, when we need to deal with the issue of computer consciousness.

What are the archetypal meanings of the first few integers? As Peirce said in his essay on the "Architecture of Theories,"

> First is the conception of being or existing independent of anything else. Second is the conception of being relative to, the conception of reaction with something else. Third is the conception of mediation, whereby a first and second are brought into relation.... (1935, p.25)

And Fourth, which Peirce did not emphasize, is wholeness or synergy, whereby several relationships are interwoven into a network, in which each relationship relates the others. These general concepts are very simple but they possess significant power to cut through complex issues in cognitive science, as we will see as we proceed.

Let's proceed through the basic numerical archetypes more carefully, beginning not with 1 but with 0:

Naught: The Formless Void.

In Peircean metaphysics, zero corresponds to the original state of the universe, or any other system for that matter. In his cosmology, Peirce

posited that the universe originated in "the utter vagueness of completely undetermined and dimensionless potentiality." He thought that "the initial condition, before the universe existed, was not a state of pure abstract being. On the contrary it was a state of just nothing at all, not even a state of emptiness, for even emptiness is something". This was an odd metaphysical speculation when Peirce wrote it in 1898, rooted in his peculiar numerical metaphysics. I find it truly remarkable that modern scientific astronomy, based on the most rigorous observations and sophisticated mathematical computations, has reached identical conclusions.

The concept of nothingness in Peirce's thought is not quite the same as the integer 0 which we may obtain in arithmetic computations such as 5 - 3 - 2 = 0. In order to do arithmetic of this sort, one has to have numerical logic, which is Secondness in his theory. Naught, the archetypal zero, is a state which exists before arithmetic, before order of any kind has developed. This is also what Carl Jung (1955) had in mind when he thought of the Self as zero. He did not mean the psychosocial self which we develop as we grow up. He meant the unformed, undifferentiated state before the personality begins to develop. The word "naught" captures this archetypal symbolism better than "zero," although arithmetically they are the same thing. Jean-Paul Sartre and his followers captured the difference with a distinction between "nothing" and "nothingness," but that must work better in French than it does in English. The same archetype is expressed in the Buddhist concept of the "formless void." Buddhists believe that considerable religious insight can come from meditating on the concept of nothingness. The same insight is central to other forms of meditation, including Quaker silent worship. Meditation seeks to help us to put aside rigid patterns of thought and open our minds to new insights; it is the pursuit of the Naught archetype.

First: Raw Being.

According to Peirce, "*First is the conception of being or existing independent of anything else.*" For Peirce, First was the first step in the development of any phenomenon out of chaos. In psychology, Firstness is "*feelings, comprising all that is immediately present, such as pain, blue, cheerfulness, the feeling that arises when we contemplate a consistent theory, etc. A feeling is a state of mind having its own living quality, independent of any other state of mind ... an element of consciousness which might conceivably override everything.*" In physics, Firstness is chance behavior, randomness, what we today call the quantum indeterminacy of matter. In philosophy, Firstness is usually called idealism. In complex systems theory, First is simply the state of a dynamical system as observed

either at a particular point in time or over a window of time. It's just there. It's neither stable nor changing.

Second: The Reacting Object.

"Second is the conception of being relative to, the conception of reaction with, something else". In psychology, we have "sensations of reaction, as when a person blindfold suddenly runs against a post, when we make a muscular effort, or when any feeling gives way to a new feeling." In physics, the laws which describe the relationships between different phenomena are Second. In philosophy, Secondness is usually called materialism. Mathematically, Secondness is best represented by a vector, a line with an arrow at the end, representing both magnitude and direction. In complex systems theory, Second is the movement from one state to another. For instance if we have a series of numbers, the numbers are First, they are moments of being. But the movement from one number to another is Second. It's a reaction, bringing us from one thing to another.

Third: The Evolving Interpretation.

Peirce said that *"Third is the conception of mediation, whereby a first and second are brought into relation."* Third is habit. It is abstract thought. In Peirce's view, Thirdness is the inevitable product of the human mind: *"when we think, we are conscious that a connection between feelings is determined by a general rule, we are aware of being governed by a habit...the one primary and fundamental law of mental action consists in a tendency to generalization. Feeling tends to spread; connections between feelings awaken feelings; neighboring feelings become assimilated; ideas are apt to reproduce themselves."*

In the physical sciences, Thirdness exists in general principles or patterns, such as the Theory of Evolution, which are laws of habitual tendency. In philosophy, Thirdness is often expressed in Hegelian historicism and Trinitarian theology. Mathematically, Thirdness is best represented by a triangle. In complex systems theory, Thirdness is a pattern or equation which explains the ways the numbers are related. For instance suppose a series of numbers is generated by the equation $x(n+1) = 3.7\ x(n)\ (1-x(n))$. This is a relation binding together the numbers in the series. Inferring the equation from the series of numbers is a hard mathematical problem.

Thirdness, in general, is patterns that emerge from the series of numbers, whether in the form of a precise mathematical equation or just an

http://www.springer.com/978-0-306-46735-6

Creating Internet Intelligence
Wild Computing, Distributed Digital Consciousness, and the Emerging Global Brain
Goertzel, B.
2002, XVI, 330 p., Hardcover
ISBN: 978-0-306-46735-6