Galanter, and Pribram 1960 for precursors), by the hierarchical arrangement of behaviors (Albus 1992), or by some more intricate principle of composition. Techniques have also been proposed (Kaelbling 1988) that use off-line symbolic reasoning to derive reactive behavior modules with guaranteed real-time on-line performance.

A third architectural paradigm, explored by researchers in distributed artificial intelligence, is motivated by the following observation. A local subsystem integrating sensory data or generating potential actions may have incomplete, uncertain, or erroneous information about what is happening in the environment or what should be done. But if there are many such local nodes, the information may in fact be present, in the aggregate, to assess a situation correctly or select an appropriate global action policy. The distributed approach attempts to exploit this observation by decomposing an intelligent agent into a network of cooperating, communicating subagents, each with the ability to process inputs, produce appropriate outputs, and store intermediate states. The intelligence of the system as a whole rises from the interactions of all the system's subagents. This approach gains plausibility from the success of groups of natural intelligent agents, for example, communities of humans, who decompose problems and then reassemble the solutions, and from the parallel, distributed nature of neural computation in biological organisms. Although it may be stretching the agent metaphor to view an individual neuron as an intelligent agent, the idea that a collection of units might solve one subproblem while other collections solve others has been an attractive and persistent theme in agent design.

Intelligent-agent research is a dynamic activity and is much influenced by new trends in cognitive science and computing; developments can be anticipated across a broad front. Theoretical work continues on the formal semantics of MENTAL REPRESENTATION, models of behavior composition, and distributed problem solving. Practical advances can be expected in programming tools for building agents, as well as in applications (spurred largely by developments in computer and communications technology) involving intelligent agents in robotics and software.

See also BEHAVIOR-BASED ROBOTICS; COGNITIVE ARCHITECTURE; FUNCTIONAL DECOMPOSITION; MODULARITY OF MIND; MULTIAGENT SYSTEMS

—Stanley J. Rosenschein

References

Albus, J. S. (1992). RCS: A reference model architecture for intelligent control. *IEEE Comput.* 25(5): 56–59.

Brooks, R. A. (1986). A robust layered control system for a mobile robot. *IEEE Trans. Rob. Autom.* 2: 14–23.

Genesereth, M. R. (1983). An overview of metalevel architecture. *Proceedings AAAI* 83: 119–123.

Georgeff, M., and A. Lansky. (1987). Reactive reasoning and planning. *Proceedings AAAI* 87.

Kaelbling, L. (1988). Goals as parallel program specification. *Proceedings AAAI* 88.

Miller, G., E. Galanter, and K. H. Pribram. (1960). *Plans and the Structure of Behavior.* New York: Henry Holt and Company.

Newell, A. (1990). *Unified Theories of Cognition.* Cambridge, MA: Harvard University Press.

Russell, S., and E. Wefald. (1991). *Do the Right Thing.* Cambridge, MA: MIT Press.

Simon, H. A. (1969). *The Sciences of the Artificial.* Cambridge, MA: MIT Press.

Intentional Stance

The *intentional stance* is the strategy of interpreting the behavior of an entity (person, animal, artifact, or the like) by treating it as if it were a rational agent that governed its "choice" of "action" by a "consideration" of its "beliefs" and "desires." The distinctive features of the intentional stance can best be seen by contrasting it with two more basic stances or strategies of prediction, the physical stance and the design stance. The physical stance is simply the standard laborious method of the physical sciences, in which we use whatever we know about the laws of physics and the physical constitution of the things in question to devise our prediction. When I predict that a stone released from my hand will fall to the ground, I am using the physical stance. For things that are neither alive nor artifacts, the physical stance is the only available strategy. Every physical thing, whether designed or alive or not, is subject to the laws of physics and hence behaves in ways that can be explained and predicted from the physical stance. If the thing I release from my hand is an alarm clock or a goldfish, I make the same prediction about its downward trajectory, on the same basis.

Alarm clocks, being designed objects (unlike the rock), are also amenable to a fancier style of prediction—prediction from the design stance. Suppose I categorize a novel object as an alarm clock: I can quickly reason that if I depress a few buttons just so, then some hours later the alarm clock will make a loud noise. I do not need to work out the specific physical laws that explain this marvelous regularity; I simply assume that it has a particular design—the design we call an alarm clock—and that it will function properly, as designed. Design-stance predictions are riskier than physical-stance predictions, because of the extra assumptions I have to take on board: that an entity is designed as I suppose it to be, and that it will operate according to that design—that is, it will not malfunction. Designed things are occasionally misdesigned, and sometimes they break. But this moderate price I pay in riskiness is more than compensated for by the tremendous ease of prediction.

An even riskier and swifter stance is the intentional stance, a subspecies of the design stance, in which the designed thing is an agent of sorts. An alarm clock is so simple that this fanciful anthropomorphism is, strictly speaking, unnecessary for our understanding of why it does what it does, but adoption of the intentional stance is more useful—indeed, well-nigh obligatory—when the artifact in question is much more complicated than an alarm clock. Consider chess-playing computers, which all succumb neatly to the same simple strategy of interpretation: just think of them as rational agents that want to win, and that know the rules and principles of chess and the positions of the pieces on the board. Instantly your problem of predict-

ing and interpreting their behavior is made vastly easier than it would be if you tried to use the physical or the design stance. At any moment in the chess game, simply look at the chessboard and draw up a list of all the legal moves available to the computer when it is its turn to play (there will usually be several dozen candidates). Now rank the legal moves from best (wisest, most rational) to worst (stupidest, most self-defeating), and make your prediction: the computer will make the best move. You may well not be sure what the best move is (the computer may "appreciate" the situation better than you do!), but you can almost always eliminate all but four or five candidate moves, which still gives you tremendous predictive leverage.

The intentional stance works (when it does) whether or not the attributed goals are genuine or natural or "really appreciated" by the so-called agent, and this tolerance is crucial to understanding how genuine goal-seeking could be established in the first place. Does the macromolecule really want to replicate itself? The intentional stance explains what is going on, regardless of how we answer that question. Consider a simple organism—say a planarian or an amoeba—moving nonrandomly across the bottom of a laboratory dish, always heading to the nutrient-rich end of the dish, or away from the toxic end. This organism is seeking the good, or shunning the bad—its own good and bad, not those of some human artifact-user. Seeking one's own good is a fundamental feature of any rational agent, but are these simple organisms seeking or just "seeking"? We do not need to answer that question. The organism is a predictable intentional system in either case.

By exploiting this deep similarity between the simplest—one might as well say mindless—intentional systems and the most complex (ourselves), the intentional stance also provides a relatively neutral perspective from which to investigate the differences between our minds and simpler minds. For instance, it has permitted the design of a host of experiments shedding light on whether other species, or young children, are capable of adopting the intentional stance—and hence are higher-order intentional systems. Although imaginative hypotheses about "theory of mind modules" (Leslie 1991) and other internal mechanisms (e.g., Baron-Cohen 1995) to account for these competences have been advanced, the evidence for the higher-order competences themselves must be adduced and analyzed independently of these proposals, and this has been done by cognitive ethologists (Dennett 1983; Byrne and Whiten 1991), and developmental psychologists, among others, using the intentional stance to generate the attributions that in turn generate testable predictions of behavior.

Although the earliest definition of the intentional stance (Dennett 1971) suggested to many that it was merely an instrumentalist strategy, not a theory of real or genuine belief, this common misapprehension has been extensively discussed and rebutted in subsequent accounts (Dennett 1987, 1991, 1996).

See also COGNITIVE DEVELOPMENT; COGNITIVE ETHOLOGY; FOLK PSYCHOLOGY; INTENTIONALITY; PROPOSITIONAL ATTITUDES; RATIONAL AGENCY; REALISM AND ANTI-REALISM

—Daniel Dennett

References

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind.* Cambridge, MA: MIT Press.

Byrne, R., and A. Whiten. (1991). *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes and Humans.* New York: Oxford University Press.

Dennett, D. (1971). Intentional systems. *Journal of Philosophy* 68: 87–106.

Dennett, D. (1983). Intentional systems in cognitive ethology: the "panglossian paradigm" defended. *Behavioral and Brain Sciences* 6: 343–390.

Dennett, D. (1987). *The Intentional Stance.* Cambridge, MA: Bradford Books/MIT Press.

Dennett, D. (1991). Real patterns. *Journal of Philosophy* 87: 27–51.

Dennett, D. (1996). *Kinds of Minds.* New York: Basic Books.

Leslie, A. (1991). The theory of mind impairment in autism: evidence for a modular mechanism of development? In A. Whiten, Ed., *Natural Theories of Mind.* Oxford: Blackwell.

Intentionality

The term *intentional* is used by philosophers, not as applying primarily to actions, but to mean "directed upon an object." More colloquially, for a thing to be intentional is for it to be *about something*. Paradigmatically, mental states and events are intentional in this technical sense (which originated with the scholastics and was reintroduced in modern times by FRANZ BRENTANO. For instance, beliefs and desires and regrets are about things, or have "intentional objects": I have beliefs about Boris Yeltsin, I want a beer and world peace, and I regret agreeing to write so many encyclopedia articles.

A mental state can have as intentional object an individual (John loves *Marsha*), a state of affairs (Marsha thinks that *it's going to be a long day*) or both at once (John wishes *Marsha were happier*). Perception is intentional: I see John, and that John is writing Marsha's name in his copy of *Verbal Behavior*. The computational states and representations posited by cognitive psychology and other cognitive sciences are intentional also, inasmuch as in the course of computation something gets computed and something gets represented. (An exception here may be states of NEURAL NETWORKS, which have computational values but arguably not representata.)

What is at once most distinctive and most philosophically troublesome about intentionality is its indifference to reality. An intentional object need not actually exist or obtain: the Greeks worshiped Zeus; a friend of mine believes that corks grow on trees; and even if I get the beer, my desire for world peace is probably going to go unfulfilled.

Brentano argued both (A) that this reality-neutral feature of intentionality makes it the distinguishing mark of the mental, in that all and only mental things are intentional in that sense, and (B) that purely physical or material objects cannot have intentional properties—for how could any purely physical entity or state have the property of being "directed upon" or *about* a nonexistent state of affairs? (A) and (B) together imply the Cartesian dualist thesis that no