ARTIFICIAL INTELLIGENCE, RELIGION, AND COMMUNITY CONCERN

by Matt J. Rossano

Abstract. Future developments in artificial intelligence (AI) will likely allow for a greater degree of human-machine convergence, with machines becoming more humanlike and intelligent machinery becoming more integrated into human brain function. This will pose many ethical challenges, and the necessity for a moral framework for evaluating these challenges will grow. This paper argues that community concern constitutes a central factor in both the evolution of religion and the human brain, and as such it should be used as the organizing principle for moral evaluations of AI technologies.

Keywords: artificial intelligence; community concern; evolution; morality.

God's people are commanded to keep the Sabbath holy. In the Jewish tradition "keeping" the Sabbath meant suspending work and gathering in the synagogue to contemplate and discuss God's laws (see Sanders 1992; 1993). While pious Jews believe Sabbath observance to be divinely mandated, that should not obscure its practical and adaptive value. Keeping the Sabbath has strengthened community by the setting aside of individual labors for group discussion, prayer, worship, and meals. Judaism is hardly singular in the fact that its traditions and rituals serve a community-building function. Religious festivals accounted for no fewer than 120 days of the year in ancient Greece (Burkert 1985). A church typically served as the central meeting place of a medieval village. Muslims gather on Fridays for communal prayer, as do Christians on Sundays. Throughout history religion has served as one of humanity's chief cohesive forces. The very term *religion* may be derived from the verb *religare*, which means "to bind

Matt J. Rossano is an Associate Professor in the Department of Psychology at Southeastern Louisiana University, SLU 10831, Hammond, LA 70402. His e-mail address is mrossano@selu.edu.

[*Zygon*, vol. 36, no. 1 (March 2001).]
 ISSN 0591-2385

together." In this paper I argue that it is precisely the community-building and -strengthening function of religion that will provide its most rational basis for addressing the moral issues associated with the emerging technologies of artificial intelligence, especially those that have a direct impact on brain function.

RELIGION AND COMMUNITY: EVOLUTION

In contemplating his theory of natural selection, Charles Darwin found little on which to ground religious faith. The struggle and brutality of evolutionary history seemed, for him, to leave scant room for a benevolent God (Darwin 1958). Most of Darwin's protegés concurred with this assessment. T. H. Huxley made a clear separation between human morality and the amorality of nature (Birx 1991; Huxley [1894] 1989). Richard Dawkins continues this line of thinking today. "Be warned that if you wish, as I do, to build a society in which individuals cooperate generously and unselfishly towards a common good, you can expect little help from biological nature. Let us try to teach generosity and altruism, because we are born selfish" (Dawkins 1976, 3).

But this brutish view of evolution owes some of its currency to Darwin's preoccupation with Malthusian economic theory as opposed to an entirely passionless assessment of nature. In 1902 Peter Kropotkin published *Mutual Aid*, which examined the cooperative tendencies of many species. This tradition continues today in the work of Frans deWaal (1996), Lee Dugatkin (1997; 1999), and others (see Ridley 1996 for a review). This research highlights the fact that some species have found that the struggle for survival is best dealt with by forming cooperative communities. That natural selection sometimes favors the best cooperators in the competition for survival and reproduction provides the foundation for the evolution of morality and religion.

Cooperative and even self-sacrificial behavior is an integral aspect of the natural world. James Rachels (1990) identifies the most basic form of altruism as an expression of kin selection, in which close family members work together in service of the perpetuation of their genes. A mother protecting her young from a predator is guarding her own genetic material from destruction. Worker honeybees slavishly give themselves to the hive in order to propagate the queen's offspring with whom they are more closely related than the offspring of other workers (Ratnieks 1988; Oldroyd, Smolenski, Cornuet, and Corzier 1994). Among birds, some jays and wrens will remain at the nest and help mom and dad raise the younger siblings. These instances show how nature can squeeze a limited form of unselfishness out of organisms whose ultimate concern is the continuation of their own genetic line.

As groups grow larger, however, encompassing not just immediate kin but nonrelatives as well, natural selection must find ways to forge cooperation among organisms that share relatively little familial connection. A dark cave teeming with vampire bats seems an unlikely place to uncover the basis for this cooperation, but that is exactly where Gerald Wilkinson found it. His painstaking investigations of the behavior of vampire bats revealed that after a nightly outing in search of sustenance, bats in the cave would regurgitate blood to one another. This sharing, however, was not indiscriminate. A bat was most likely to share with one who would reciprocate at a later time. A successful blood run, it turns out, is not a sure thing, and if one bat has come up dry, it's nice to have a friend who has scored with the mother lode. Access to another's wealth, though, is predicated on one's own history of similar generosity (Wilkinson 1984; DeNault and McFarlane, 1995). This notion of mutual back scratching is called "reciprocal altruism" (Trivers 1971), and researchers have found evidence for it in numerous group-living species. Troops of monkeys and apes have highly sophisticated reciprocal systems for sharing food, sex, and power within the group (deWaal 1996; Harcourt 1992). Reciprocity of this sort often serves as the basis for many human relationships as well.

Reciprocity does, however, have its limitations. Sometimes reciprocal relationships among individuals can serve to undermine group stability and cohesion, as when "coalitions" of chimpanzees challenge each other for dominance within the group. Furthermore, establishing reciprocal relationships can sometimes be costly and time consuming. Sharing with one who refuses to return the favor can foster hard feelings and even violence, which weaken the larger group. Yet these potential pitfalls only highlight the creative power of natural selection. DeWaal (1996) notes how some members of a chimpanzee troop will step in and act as mediators in order to de-escalate a mounting crisis. More important is the fact that social creatures very quickly become selective in determining those with whom they establish reciprocal relationships. The community is strengthened if all or many members know not to depend on certain other members; thus they avoid establishing relationships that will likely produce intragroup conflict. Richard Alexander (1987) uses the term "indirect reciprocity" to describe how reputations based on observation are established within a social community. A smart chimp, like a smart human being, will carefully observe those within the group who fulfill obligations versus those who take advantage of the kindness of others. Those with good reputations tend to be the ones with whom others seek to establish relationships. An individual with a solid reputation can reap the benefits of numerous reciprocal relationships, thus increasing his or her own status within the group, at the same time contributing to general group cohesion and stability.

The evolution of cooperation seems to have produced a desirable state of affairs. For selfish reasons social organisms seek to establish and maintain good personal reputations. That reputation leads to the establishment of reciprocal relationships with others in the group (as others desire relationships with those of good repute), thus benefiting both the individuals involved and the group as a whole. In human societies, however, it is not this simple. Matching those of good social stature with those seeking relationships can be resource intensive. Even in the most close-knit communities it is not possible to monitor directly how so-and-so fared in his or her numerous relationships. Nor is it possible to keep tallies on how many times various individuals fulfilled or reneged on promises. This problem was especially onerous in our evolutionary past when the failure to reciprocate could have life-threatening consequences. Suppose you had shared meat with a neighbor only to have him greet you with clubs and stones later when you were starving and his family had food. It would be helpful to have some sign or indicator announcing the extent to which an individual's reputation was important to him or her. As Robert Frank (1988) has termed it, we need something to address the commitment problem—a costly-to-fake sign assuring others that an individual will guard his or her reputation even under difficult circumstances.

William Irons (1996) theorizes that this commitment problem forms the evolutionary basis of religion. A devotee to a religious tradition is one who accepts certain precepts as being in some fashion divinely mandated. Thus, to call myself a Christian is to imply that I am committed to certain ideals and behaviors, such as loving my neighbor, avoiding self-aggrandizement, and being honest and concerned about the welfare of others. Assuming that others simply accept my commitment to these principles, my reputation is probably a good one and my status within the group relatively high. Labels, however, are cheap and easy to fake. Prudent group members tend to look for further signs or indicators of one's commitments. The Mandan of North America used excruciatingly painful initiation rites, such as hanging someone from punctures in the chest and shoulders, to ensure that the would-be warrior was thoroughly committed to his beliefs and tribe. Such extreme signs of commitment are rare, but they underscore the role that religious tradition and ritual have served as a public forum for costly-to-fake signs of commitment. To announce one's seriousness about certain Christian principles, being a regular and active member of a church, sacrificing time and money to appropriate causes, or adopting a celibate lifestyle (in the case of priests or nuns, for example) would all tend to serve as reasonably convincing signs of commitment. The obligations, rituals, and sacrifices demanded by religious traditions announce to other community members the seriousness with which certain individuals take their principles and commitments.

From this perspective religion can be understood as a natural outgrowth of community concern. As a social structure it helps to organize community life and promote a common set of ideals and right behaviors among community members. This in turn helps reduce intragroup conflict and increase group stability and cohesion. Religion is a naturally selected means of building and maintaining strong human communities.

Religion and Community: Scripture

It is not very difficult to find evidence of community concern in the Judeo-Christian religious tradition. Chapter 19 of Leviticus expands upon the Ten Commandments, setting rules for a wide range of community life and social interactions. Verse 9 (NEB) commands that one is not to harvest to the very edge of the land or to strip completely the vine of grapes or to glean those that fall. These are to be left for the poor and the outsider. Prohibitions against slander, holding grudges, exhibiting favoritism to either rich or poor, causing distress to the disabled, and seeking revenge are also expressed. The rules state that one is to adhere to strict justice in dealing with countrymen and to reprove them frankly but never to harbor anger or hatred. Pious Jews are reminded not to oppress the alien within their land, for they themselves were once aliens in Egypt. Ecclesiastes lauds cooperative effort over individual effort: "I saw emptiness under the sun: a lonely man without a friend . . . two are better than one; they receive a good reward for their toil, because if one falls the other can help his companion up again (Ecclesiastes 4:8–10 NEB). A significant portion of the book of Ecclesiasticus (Sirach) is devoted to wisdom teachings on social conduct and upright community behavior, much of which is based on the notion of reciprocity and reputation, for example: "A devout man lends to his neighbor; by supporting him he keeps the commandments. Lend to your neighbor in his time of need; repay your neighbor punctually. Be as good as your word and keep faith with him, and your needs will always be met" (Ecclesiasticus 29:1–3 NEB).

Jesus is depicted as continuing this tradition in Matthew 18:15–17, where he tells his disciples how to handle a dispute with a brother who has sinned. First the matter is to be dealt with personally—"take the matter up with him strictly between yourselves." If this approach is fruitless, Jesus instructs bringing before the errant brother one or two others who will support the case. If the brother is still unmoved, the matter should be taken before the congregation, and if even this produces no results, the brother is to be treated as a pagan or tax collector (he is to be shunned and future reciprocal relationships with him avoided). The message appears to be one of limiting the damage to as small a circle as possible so as to bring as little harm as possible to the community.

Community concern was also a prominent theme among the early Christians, who often displayed a high degree of egalitarianism and communalism. In Acts 4:32, we are told that "the body of believers was united heart and soul," with all possessions held in common. A little later, in Acts 5:13, we are told that the believers met by common consent in Solomon's Portico and that people in general spoke highly of them. Chapter 6 of Acts tells of a dispute that arose between the Hellenists and the Hebrews of the early Christian movement. The dispute was settled by calling together the "whole body of disciples" (verse 2) and appointing different individuals to separate tasks. Verse five states that "this proposal proved acceptable to the whole body."

In his letter to the Romans, Paul encourages believers: "Let love for our brotherhood breed mutual affection. Give pride of place to one another in esteem" (Romans 12:10 NEB). "Contribute to the needs of God's people, practice hospitality" (v. 13). "Care as much about each other as about yourselves" (v. 16). "If possible, so far as it lies with you, live at peace with all men" (v. 18). "Let us then pursue the things that make for peace and build up the common life" (14:19). To the church in Philippi, Paul writes: "fill up my cup of happiness by thinking and feeling alike, with the same love for one another, the same turn of mind, and a common care for unity. There must be no room for rivalry and personal vanity among you, but you must humbly reckon others better than yourselves. Look to each other's interests and not merely your own" (Philippians 2:2–4 NEB).

It appears, then, that a substantial proportion of the Judeo-Christian tradition concerns itself with social relations and the maintenance of group stability and cohesion. The virtues of justice, honesty, and humility, so often praised in sacred scripture, serve to foster strong communities as well as earn potential heavenly rewards.

Community and the Evolution of the Brain: What's a Big Brain for, Anyway?

Somewhere between 5 million (Sarich and Wilson 1967) and 7.7 million years ago (Sibley and Alquist 1984) the hominid line that produced modern human beings split from that of the chimpanzees. Over the course of hominid evolution there has been a general tendency toward increasing brain size, an increase that is not fully attributable to increases in body size. As Richard Passingham (1982) has noted, the brain of the modern human being is about three times as large as would be expected for a primate of our build.

Australopithecus emerged approximately 3 or 4 million years ago, with the fossil skeleton named "Lucy" being one of the most celebrated (although not the oldest) of this group (Johnson and Edey 1981). Detailed analysis of Lucy's bone structure indicates that the Australopithecines were small of stature, walked upright, and had somewhat humanlike faces (Lovejoy 1988). Their brains, however, were not much larger than ape brains, averaging about 450 cubic centimeters (Blumenberg 1983; Corballis 1991). About 2 million years ago *Homo habilis* emerged. With an average size of nearly 660cc (ranging from 500 to 800cc), the brain of *H. habilis* was dramatically larger than that of *Australopithecus*. The brain structure of *H. habilis* was also far more humanlike, with squared frontal lobes and an enlarged parietal region that contrasted with the more apelike organization of *Australopithecus* (Falk 1982). About 1.8 million years ago *Homo erectus* emerged. Fossil evidence reveals a cranial capacity of *H. erectus* somewhere between 800 and 1000cc (or 750 and 1250, depending on which fossils are placed in the category), thus indicating another substantial increase in brain size compared to the earlier *H. habilis* (Brown, Harris, Leakey, and Walker 1985; see Lieberman 1998 for a summary). The size differential between males and females was also smaller in *H. erectus* than in *H. habilis*, suggestive of a more cooperative social order (Corballis 1991). Merlin Donald (1991; 1993) points to the emergence of *H. erectus* as a landmark event in human cognitive evolution, arguing that with *H. erectus* came a form of cultural exchange called *pure mimesis*. This refers to the ability to remember and reproduce refined voluntary motor movements that can be used as a vehicle for the transmission of behavioral skills, to engage in common social activities such as dance, and to facilitate communication of intentions and emotions.

Hominid brains underwent another dramatic increase in size with the emergence of the archaic *Homo sapiens* about .5 million years ago. Although their skulls were shaped differently from those of modern human beings, these human ancestors had a cranial capacity within modern limits, ranging from 1200 to 1400cc (Pfeiffer 1973). The Neanderthals, who emerged about 200,000 years ago, also possessed brains approximately the same in overall size as modern human brains, if not slightly larger. Donald (1991; 1993) considers this a second landmark moment in human cognitive evolution in that with the archaic *H. sapiens* (and to a somewhat lesser extent, Neanderthals) came the mythic culture. The brains of these hominids allowed for an elaborate symbolic system of communication that produced narrative tales and oral traditions that were used as a means of social organization. These traditions provided purpose and meaning to existence and became the groundwork for religion.

The transition from archaic *Homo sapiens* to modern human beings (*Homo sapiens sapiens*) took place about 100,000 years ago and appears to have involved changes in brain structure more than overall size, although the modern human brain is slightly larger (ranging from 1200cc to 1700cc). These changes, along with other anatomical modifications, allowed for more efficient hunting, foraging, and communication. In fact, it has been argued that even modest advantages in hunting strategies and communicative abilities afforded by the modern human brain would have been enough to displace the Neanderthals completely over time (Zubrow 1990; Lieberman and Shay 1994; Lieberman 1998).

Simple increases in brain size are not the entire story, however. The evolution of the hominid brain significantly involves proportional enlargement of the neocortex (the very top layer of brain), and it is this change that is most implicated in the evolution of intelligence (Dunbar 1992;

Jerison 1973; Passingham 1975; Sawaguchi 1989). As noted earlier, the transition from *Australopithecus* to *H. habilis* involved important structural changes that produced enlargement of the frontal and parietal lobes of cortex. The frontal lobes of modern human beings are much larger proportionally than those of other primates and are functionally distinct as well (Jerison 1973; Sutton and Jurgens 1988). Terrence Deacon (1988) has pointed out that the human prefrontal cortex is twice the size one would expect if it were just an enlarged version of an ape brain. These structural modifications may have been of critical value in endowing *H. sapiens sapiens* with decisive cognitive advantages over the Neanderthals.

Recently Robin Dunbar (1992; 1996) has done extensive work in isolating the factors necessitating hominid brain growth. He argues that the increasing complexity of hominid community life required greater mental capacity. After analyzing a range of behavioral ecological variables, he found that only group size was significantly related to neocortical size in seventy primate species. With frugivorous monkeys, Sawaguchi (1992) also found that the relative size of the neocortex (RSN) was significantly related to troop size, whereas variations in habitat (such as arboreal versus terrestrial) were not. Among nonprimates, dolphins and vampire bats possess relatively large neocortices and also are known to have sophisticated social relationships. Vampire bats, as discussed earlier, live socially and maintain extensive reciprocal relationships. Dolphins live in a "fission-fusion" type of social structure with complex alliances being formed as a means of competing for mates (Connor, Smolker, and Richards 1992; Dunbar 1996).

Cognitive neuropsychologist Michael Corballis concurs with these assessments of the link between sociality and brain size:

> Although the manufacture of stone tools may have given the initial kick to an increase in [hominid] brain size . . . subsequent development must have had other causes. One possibility is that the hominids were forced increasingly into a mode of existence that required cooperation rather than competition. . . . The cooperative foraging for sources of meat would have favored more sophisticated forms of communication and perhaps the development of specialized skills within the social group. It may have been these factors that were largely responsible for the increase in brain size. (Corballis 1991, 65–66)

Studies of highly successful (adaptive) people today confirm the critical importance of cooperative, communicative ability. In a recent book, Daniel Goleman (1998) has reviewed data from nearly two hundred corporations and other organizations looking for the key factors to individual success. The most important factor was something he called "emotional intelligence," or the ability to handle personal relationships effectively. Goleman estimates that emotional intelligence accounts for 85 to 90 percent of the success of the very top business and professional leaders. Thus, to answer the question posed at the beginning of the section, a big brain serves the purpose of constructing and participating in complex human communities.

Religion, Brain, and Community

The purpose of the preceding sections has been to establish both an empirical and a scriptural basis for the ways that religion and the brain are intimately interconnected with community. Religion evolved in service to community cohesion and stability, and this is well reflected in the expressions and teachings of sacred scripture. As the brain evolved, it provided the capacity for individuals to engage in complex social interactions and relationships. If community is of central concern to both religion and the brain, then when searching for a firm moral basis on which to judge the emerging technologies of artificial intelligence, religion should concentrate on the consequences to community. Put bluntly, a purpose (maybe *the* purpose) of the human brain is to allow for the establishment and maintenance of long-term relationships with other human beings. The success of these relationships is critical to individuals in their singular interests for happiness, success, and security; and it is critical to human societies at a broader level so that they can be stable, supportive, healthy environments in which individuals can thrive. Religion has traditionally been one of the frameworks upon which a healthy adaptive society has been constructed. A concern for maintaining healthy communities should motivate religion to a thoughtful, bold, and widely defensible critique of the advances and potentials of AI.

The Future of Artificial Intelligence

In his recent book, *The Age of Spiritual Machines* (1999), Raymond Kurzweil speculates about the nature of artificial intelligence in the next century. In his view, the next decades will be marked by increasing human-machine convergence and increasing ambiguity over what constitutes a human being in contrast to a machine. A number of factors will drive this process. First, computers will become increasingly more powerful, surpassing the processing capacity of the human brain and eventually surpassing the processing power of all human brains combined. Second, computer design will increasingly be based upon the parallelism of human brain function, thus mimicking human cognitive flexibility and learning. These factors are relatively uncontroversial among observers of AI and will in all likelihood serve to create machines that are more humanlike in their functioning. Current debates in AI have tended to focus on whether these more powerful and parallel machines will possess a humanlike consciousness (Churchland and Churchland 1990; Dreyfus 1979 and 1992; Searle 1992). For Kurzweil the issue is irrelevant, because, whether they are conscious or not, they will be sophisticated enough to convince us that they are:

> The machines will convince us that they are conscious, that they have their own agenda worthy of our respect. We will come to believe that they are conscious much as we believe that of each other. More so than with our animal friends, we

will empathize with their professed feelings and struggles because their minds will be based on the design of human thinking. They will embody human qualities and will claim to be human. And we'll believe them. (1999, 63)

However, it is not the issue of machines' becoming more human that is the most provocative aspect of Kurzweil's vision. The convergence of human beings with machines is also driven by factors that will integrate intelligent technology more intimately with the human brain. These include (1) advances in neural implant technologies and (2) miniaturization based on nanotechnology.

Neural implant technology has been a practical reality for about two decades, with sensory prosthetics leading the way. More than ten thousand patients have received cochlear implants for treating deafness, with varying degrees of success (Brown, Dowell, and Clark 1987; Owens 1989). Research on retinal implants and direct visual cortex stimulation is holding promise for visually impaired and blind patients; however, the challenges here are more daunting (Dobelle 1977). Experimental work suggests that implant technology may also provide a means of treating the symptoms of Parkinson's disease (Benabid et al. 1996), and recently the FDA approved a device called the neurocybernetic prosthesis system, which provides intermittent stimulation of the vagus nerve to help relieve epileptic seizures (Finn 1997). Kurzweil assumes that these technologies will eventually evolve beyond exclusive use for overcoming disabilities and will be adapted for use in enhancing normal perceptual and cognitive functions. "Once a technology is developed to overcome a disability, there is no way to restrict its use from enhancing normal abilities, nor would such restriction necessarily be desirable" (Kurzweil 1999, 150). A possible example of such adaptation is the research underway to develop an artificial hippocampus (a brain region important in learning and memory), which might aid in human memory function. One scientist predicts practical results within twenty years (Finn 1997).

According to Kurzweil, not only will the neural implant technology of the future allow for enhanced mental abilities, but it also will provide a means of directly interacting with our computers. With wireless technology, computer images and sounds can be made to stimulate the brain directly. As the "realism" of virtual reality progresses, it will be possible to experience computer-generated worlds indistinguishable from those that actually impinge upon our sense organs.

This will be the essence of the Web in the second half of the twenty-first century. . . . A typical "web site" will be a perceived virtual environment, with no external hardware required. You "go there" by mentally selecting the site and then entering the world. Debate Benjamin Franklin on the war powers of the presidency at the historical society site. Ski the Alps at the Swiss Chamber of Commerce (while feeling the cold spray of snow on your face). Hug your favorite movie starlet at the Columbia Pictures site. Get a little more intimate at the Penthouse or Playgirl site. Of course, there may be a small charge. (Kurzweil 1999, 144)

Miniaturization through nanotechnology is a second important factor in the convergence of human being and machine, according to Kurzweil. In this context, nanotechnology refers to the construction of machines on a molecular level. The leading pioneer of nanotechnology is Eric Drexler, who, along with a group of other prominent scientists including AI giant Marvin Minsky, has formed the Foresight Institute, an organization dedicated to the advancement of nanotechnology. In his book *Nanosystems: Molecular Machinery, Manufacturing, and Computation* (1992), Drexler provides a detailed description of the feasibility of microengineering as well as some specific designs. One possibility emanating from this research is the creation of computers small enough to serve as neural implants (Drexler 1986). Although most scientists consider Drexler's notions of nanotechnology more fiction than science (Stix 1996), a growing number of companies and research labs are actively engaged in nanotechnology research, most notably the Center for Nanoscale Science and Technology, at Rice University, headed by Nobel Prize winner Richard Smalley. If Kurzweil is correct, this research eventually will lead to miniaturized implantable chips that will directly connect the brain to supercomputers that will enhance human mental functions. Kurzweil takes the technology and its implication to the ultimate end. Those with the nanotechnology-neural implant-enhanced mental functions will command the highest salaries in the marketplace. The free market will thus drive a process of the human brain's becoming decreasingly neurally based as the economy favors those who are more intelligent (as it already does). Furthermore, an enhanced brain will have an ever more intense experience of virtual worlds, which most people will find desirable:

> By the fourth decade [of the twenty-first century] we will move to an era of virtual experiences through internal neural implants. . . . With this technology, you will be able to have almost any kind of experience with just about anyone, real or imagined, at any time. It's just like today's online chat rooms, except that you don't need any equipment that's not already in your head, and you can do a lot more than just chat. You won't be restricted by the limitations of your natural body as you and your partners can take on any virtual physical form. Many new types of experiences will be possible: A man can feel what it is like to be a woman, and vice-versa. (Kurzweil 1999, 148)

Should Kurzweil Be Taken Seriously?

There are reasons to take Kurzweil seriously. He is a Massachusetts Institute of Technology graduate who has founded a number of successful high-tech companies and created numerous AI-related inventions. He has received a plethora of honorary doctorates and awards, including the Dickson Prize, Carnegie-Mellon University's top science award, and the MIT Inventor of the Year award. In addition to his extensive knowledge and experience in the AI field, he also has a proven record of prediction. His earlier book *The Age of Intelligent Machines* (1990) made a number of predictions concerning AI in the 1990s, many of which have proven correct. Also, Kurzweil's predictions are not entirely unique. Hans Moravec (1988), Daniel Crevier (1993), and Eric Drexler (1986), among others, have speculated along similar lines. His ideas may be highly speculative, but he is an authoritative source from both an academic and a technological standpoint, and worthy of a thoughtful ear.

Despite these credentials, there are reasons to be cautious about Kurzweil's future vision. First, prediction is always difficult; unforeseen and (at present) unknowable factors can confound the best of experts. Second, solid arguments can be made against some of his assumptions. For example, Kurzweil's future relies heavily on advances in nanotechnology that most scientists would (at present) consider unrealistic.

Kurzweil is also convinced that human perceptual and cognitive abilities can be enhanced by developing neural technologies. It is unclear, however, what form this "enhancement" would take. There is, for instance, no evidence that sensory systems become more acute when they are more relied upon. Decades of research on the sensory abilities of the blind have debunked the notion that loss of vision renders the remaining intact systems supersensitive (Hollins 1989). From a theoretical perspective this makes sense, given that, as Hollins points out, human sensory systems have evolved to be as acute as they can usefully be. The auditory system of the average person is nearly sensitive enough to detect the Brownian motion of air molecules in a completely silent room. Detecting this stimulus would produce a constant background of "white noise" that would muddle other, more critical auditory signals. Randolf Neese and George Williams (1994) describe the ways that many human systems (such as the immune system) have evolved to a "balancing point"—potent enough to perform their required task but not so potent that they cause the organism more harm than good. Even human emotions, they maintain, normally stride along this delicate, evolutionary line. Emotions serve the function of helping human beings regulate and modulate their behaviors within a social context. Too little emotional response can leave one unable to empathize with or relate to others effectively, thus leading to dangerous isolation. Too much emotional response may lead to anxiety or depressive disorders. It seems likely that human perceptual and cognitive abilities are also balanced on tight evolutionary ropes. Attempts at enhancement seem just as likely to produce debilitation as advantage.

One can also question the extent to which the "direct linkage" that Kurzweil envisions between machines and human beings (a linkage that is important for the enhancement of human mental functions) will ever come to pass. Work by Philip Kennedy and Roy Bakay with stroke and ALS patients (Kennedy and Bakay 1998; also see a brief report in the February 1999 issue of *Discover*) seems optimistic in this regard. Implanted electrodes onto which neurons can grow have enabled a patient to use his thoughts to control computer-cursor movements. Other researchers, though, see limits to this approach. Peter Fromhertz of the Max Planck Institute notes that the informational semantics used by the brain and the computer are vastly different, and getting each to understand the other is an "enormous" problem. He claims that any ethical concerns over direct brain-to-computer links are premature (Fromhertz 1997). Human being-to-computer communication is likely to become less mediated as keyboards and monitors are replaced by less obtrusive and more natural means of information exchange, but the complete elimination of mediation seems at present doubtful.

General Trends

In addressing the future challenges of AI, it seems most reasonable to concentrate on a few general and probable trends and to propose general questions and guidelines to use as a basis for moral judgments.

1. *Advances in power and design will make our interaction with computers more seamless and humanlike.* As Kurzweil points out, the current state of the art in continuous speech recognition (CSR) by computers would probably surprise most people. The increasing use of neural-net and parallel-processing technology is likely to produce machines that are more flexible and responsive and thus more similar to human responsiveness. It does not seem unreasonable to suppose that in the next decades, when we interact with machines by speaking to them, they will respond in kind. Whether these interactions will pass the Turing Test (and thus be indistinguishable from interactions between human beings) is an open question.
2. *Virtual-reality technology will advance and allow for more realistic and compelling experiences.* Increasingly the kind of sensory experience one can obtain from computer-generated environments will approximate that of the real world. Furthermore, virtual-reality technology will probably be integrated into communication systems such that one can "visit" far-off places and people and have the experience feel similar to (if not the same as) actually being there.
3. *Increasingly miniaturized interfacing devices will allow for greater control over proximal inputs.* Whether it be with the aid of implants, contact lenses, lightweight glasses, or other relatively inconspicuous sensory devices, it seems reasonable to assume that computer-generated information can be made to impinge more directly upon our sensory systems. This can be done when excluding "real" environmental input, such as when one wants a virtual-reality experience, or a computer-generated image can simply be another element of the environmental scene, as is generally the case today when we are working with computers. The simultaneous presence of computer-generated

images and real objects may also allow for intriguing sensory mixtures, such as carrying on a conversation with someone via the computer and projecting the person's image into the local environment so as to create the illusion that he or she is in the same room. One need only imagine how these combinations could be manipulated and modified according to one's whims.

4. *Barriers of time, place, and distance will become increasingly irrelevant.* As communication and virtual-reality technologies advance and merge, it will be easier to communicate with far-off people and "experience" far-off places in a manner approximating real interactions. Additionally, these interactions can be brought under greater control by the user such that they occur in the most comfortable and convenient manner. This is already the philosophy being adopted by the emerging "virtual universities," which employ extensive distance-learning technology to allow students to take classes where and when they choose. The only element missing is the experience of actually being "in" the class itself.

Questions and Guidelines Concerning the Morality of AI and Its Future

If religious thinkers are to contribute constructively to the development of AI technologies and the social attitudes surrounding their use, they must avoid the extremes of either blanket condemnation or uncritical acceptance. It is certainly true that not everything that can be done should be done, but it is also true that not everything that can be done will necessarily be threatening or dehumanizing. What is most important is the construction of a framework that can facilitate moral evaluation. In this regard, understanding the *right questions to ask* about emerging technologies may be the best first step. Given what we know about the evolution of both religion and the brain, I propose that those questions be ones that address how human communities will be affected by AI advances. Questions should take the form of: How will this affect an individual's ability to establish and maintain significant long-term personal relationships? How will this affect an individual's ability to contribute constructively to his or her local community? How will this affect an individual's ability to give of him- or herself to others? How will this affect family and local community stability and cohesion? Even if the most extreme of Kurzweil's predictions comes true, and we are capable of expanding human cognitive ability by interlinking (or even replacing) the brain with enormously powerful computer intelligence, these questions will still be relevant. By posing such questions and seriously contemplating the answers, religious thinkers can find both rational and faith-based arguments for distinguishing between moral and immoral uses of AI technologies and exert some influence on the development and social standing of those technologies.

Potential Benefits. It becomes clear, if we ask the right questions, that some of the developments in AI are currently, or have the potential to be, quite beneficial and highly moral. For example, the neural-implant work cited earlier, involving epileptic and paralyzed patients, easily qualifies as ethical on many grounds. However, for purposes of this discussion, it does so because the neural implants heighten the individual's ability to contribute in a social context. One who is paralyzed and unable to communicate is severely limited in the ability to establish and maintain deep social relationships. Intelligent machines may serve to reintegrate individuals into the community and allow them to contribute to it.

Advances in AI that break down barriers of distance also hold promise for strengthening human communities. In centuries past, extended families were more the norm than the exception. So too were small, relatively stable local communities where individuals were born, lived, and died among familiar people and surroundings. As the world became more industrialized, people moved to where the work was, and family and community stability often suffered. With the world's moving into a technological age, there exists a real possibility of reinvigorating the extended family and the close-knit community. However, this possibility depends critically upon family and community proximity's taking priority over work or professional proximity. If technology truly makes physical travel superfluous in the future, religious leaders might push hard to ensure that the work and professional relationships, and not the personal relationships, use technology for their maintenance. Let business meetings be conducted using technology, not Thanksgiving dinners. Let one move back to one's hometown and call in to work (if feasible) rather than moving away to work and calling home.

Potential Concerns. Community and social life may erode and suffer from some aspects of AI. Consider the consequences of interactions between human beings and computers that are difficult to distinguish from interactions between human beings. Some people may prefer the company of their computers to that of other persons; some argue that this has already taken place. We may expect this sort of social isolation to become even more problematic as machines take on increasingly human characteristics and our interactions with them require less effort. This could lead (and some argue has already led) to increases in intolerance and bigotry. People's personal characteristics and social dynamics vary, and dealing with them can be challenging. Computers can be especially designed to be cooperative and compliant and to posses whatever characteristics their owners desire. The ease of interaction with them may be seductive and may cause the "outside" world to look ever more brittle and unwelcoming. In confronting this potential hazard, religious thinkers may need to remind people that, from a scientific standpoint, it was the challenge of complex social interactions that led to the enlargement of our brains; in other words, we are equipped for the treacherous waters of human exchange and not the "soft sailing" of machine interaction. From a religious perspective, we are called to serve others (as Jesus served, regardless of the cost), not to hide in comfort from them.

Social isolation could also be of the group variety and not just personal. Intelligent communication systems of the future could very well include translation capabilities that would allow input of one language and output of another language. An English speaker could therefore call a French speaker, and the two could converse freely over the telephone. Although this would seem to enhance cultural integration and understanding, it could possibly have the opposite effect, especially if the technology becomes ubiquitous. Some may opt to let the machines do the "heavy lifting" of social communication. Instead of learning other languages (and the historical-cultural appreciation associated with that), there may be an increasing trend toward using smart machines as "crutches," thus perpetuating ignorance and misunderstanding among different groups. "Love thy neighbor" does not mean allowing a machine to communicate with someone and otherwise ignoring that person.

Potentially the most significant negative effect of AI on community comes from the increasing power to control our proximal worlds (the world that impinges on our senses). In recent years cognitive scientists have come to appreciate more keenly the importance of the external environment in our mental functioning. Environmental structure plays such a crucial role in the encoding, processing, and retrieval of information that researchers are now questioning the appropriateness of construing the mind as something that resides entirely within the head. Daniel Dennett (1996) has referred to this as the "offloading" of mind into the world and uses the following scenario as a description.

Suppose one is searching through boxes for a lost item. When confronted with a box to examine, one must determine whether or not it has already been searched through. To do this, one could open the box, and if the items look familiar one could conclude that the box has already been searched. If, however, one has been putting a check mark on the boxes as they are searched, there will be no need to reopen those boxes. Thus, an internal mental process (recognition memory) would be externalized by manipulating the structure of the environment. In many and often unrecognized ways, we human beings have exploited and manipulated the structure of the environment around us so that the world can share some of the burden of our mental processing. Environmental structure in turn impresses itself upon the mind, providing an interpretive framework for encoding and processing information, as the long history of context effects in memory have shown (Godden and Baddeley 1975; Tulving and Thompson 1973). What we learn and remember is powerfully affected by the environmental context in which these processes occur. This fact compels philosopher Andy Clark to ask: "Where, then, is the mind? Is it indeed 'in the head,' or has mind now spread itself, somewhat profligately, out into the world? . . . Every thought is had by a brain. But the *flow* of thoughts and the adaptive success of reason are now seen to depend on repeated and crucial interactions with external resources" (Clark 1997, 68).

The intimacy of mind and world that cognitive science has revealed to us calls attention to how profoundly human communities depend on common spaces, experiences, and activities. The shared world around us not only influences mental processes but in a very tangible way is integral to those processes. A shared space or experience interweaves itself into the minds of those present, providing for them a commonality not only of sense but also of mental structure. So what becomes of this commonality if we can increasingly alter, modify, and even occlude the common distal world about us? If we can effortlessly and idiosyncratically supplement our proximal input, or even withdraw entirely into proximal worlds, what becomes of community?

In the future, churches and religious rituals may be one of the last few shared spaces and experiences left to bind a community together. Religious leaders may want to consider carefully the value of "virtual" attendance at services or ceremonies. The maintenance of strong communities may require the effort of being physically present in a shared place with attention focused on a single purpose in a common act. In the future this may be a hard sell to an ever-more-complacent public, but it may be the only way to ensure the commonality of mind necessary for a vibrant community life and for a community to have a fruitful relationship with God.

References

Alexander, Richard D. 1987. *The Biology of Moral Systems.* New York: Aldine De Gruyter.

Benabid, A. L., P. Pollak, D. Gao, D. Hoffmann, P. Limousin, E. Gay, I. Payen, and A. Benazzouz. 1996. "Chronic Electrical Stimulation of the Ventralis Intermedius Nucleus of the Thalamus as a Treatment of Movement Disorders." *Journal of Neurosurgery* 84:203–14.

Birx, James. 1991. *Interpreting Evolution: Charles Darwin and Teilhard de Chardin*. Buffalo, N.Y.: Prometheus Books.

Blumenberg, B. 1983. "The Evolution of the Advanced Hominid Brain." *Current Anthropology* 24:589–623.

Brown, A. A., R. C. Dowell, and G. M. Clark. 1987. "Clinical Results for Postlingually Deaf Patients Implanted with Multichannel Cochlear Prosthetics." *Annals of Otology, Rhinology, and Laryngology* 96 (Suppl. 128): 127–28.

Brown, F., J. Harris, Richard Leakey, and A. Walker. 1985. "Early *Homo Erectus* Skeleton from West Lake Turkana, Kenya." *Nature* 316:788–92.

Burkert, W. 1985. *Greek Religion*. Cambridge: Harvard Univ. Press.

Churchland, Paul M., and Patricia S. Churchland. 1990. "Could a Machine Think?" *Scientific American* (January), 32–37.

Clark, Andy. 1997. *Being There: Putting Brain, Body, and World Together Again.* Cambridge: MIT Press.

Connor, R. C., R. A. Smolker, and A. F. Richards. 1992. "Dolphin Alliances and Coalitions." In *Coalitions and Alliances in Humans and Other Animals*, ed. A. H. Harcourt and F. B. M. deWaal, 415–44. Oxford: Oxford Univ. Press.

Corballis, Michael C. 1991. *The Lopsided Ape.* New York: Oxford Univ. Press.
Crevier, Daniel. 1993. *AI: The Tumultuous History of the Search for Artificial Intelligence*. New York: Basic Books.
Darwin, Charles. 1958. *The Autobiography of Charles Darwin*. New York: W. W. Norton.
Dawkins, Richard. 1976. *The Selfish Gene*. Oxford: Oxford Univ. Press.
Deacon, Terrence. 1988. "Human Brain Evolution II: Embryology and Brain Allometry." In *Intelligence and Evolutionary Biology*, ed. H. J. Jerison and I. Jerison, 383–416. Berlin: Springer-Verlag.
DeNault, L. K., and D. A. McFarlane. 1995. "Reciprocal Altruism Between Male Vampire Bats, *Desmondus rotundus.*" *Animal Behavior* 49:855–56.
Dennett, Daniel C. 1996. *Kinds of Minds*. New York: Basic Books.
deWaal, Frans B. M. 1996. *Good Natured: The Origins of Right and Wrong in Humans and Other Animals*. Cambridge: Harvard Univ. Press.
Dobelle, W. H. 1977. "Current Status of Research on Providing Sight to the Blind by Electrical Stimulation of the Brain." *Journal of Visual Impairment and Blindness* 71:290–97.
Donald, Merlin. 1991. *The Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Cambridge: Harvard Univ. Press.
———. 1993. "Human Cognitive Evolution: What We Were, What We Are Becoming." *Social Research* 60:143–70.
Dreyfus, Hubert L. 1979. *What Computers Can't Do: The Limits of Artificial Intelligence*. New York: Harper and Row.
———. 1992. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge: MIT Press.
Drexler, Eric K. 1986. *Engines of Creation*. New York: Doubleday.
———. 1992. *Nanosystems: Molecular Machinery, Manufacturing, and Computation*. New York: John Wiley and Sons.
Dunbar, Robin I. M. 1992. "Neocortex Size as a Constraint on Group Size in Primates." *Journal of Human Evolution* 20:469–93.
———. 1996. *Grooming, Gossip, and the Evolution of Language*. Cambridge: Harvard Univ. Press.
Dugatkin, Lee A. 1997. *Cooperation Among Animals: An Evolutionary Perspective*. New York: Oxford Univ. Press.
———. 1999. *Cheating Monkeys and Citizen Bees: The Nature of Cooperation in Animals and Humans*. New York: The Free Press.
Falk, Dean. 1982. "Mapping Fossile Endocasts." In *Primate Brain Evolution: Method and Concepts,* ed. Este Armstrong and Dean Falk, 217–26. New York: Plenum.
Frank, Robert H. 1988. *Passion with Reason: The Strategic Role of the Emotions*. New York: W. W. Norton.
Finn, Robert. 1997. "Neural Prosthetics Come of Age as Research Continues." *The Scientist* 11:13–16.
Fromhertz, Peter. 1997. "Neuron-Silicon Junction or Brain-Computer Junction?" In *Ars Electronica Festival*, ed. G. Stocker and C. Schopf, 158–61. Vienna: Springer.
Godden, D. R., and A. D. Baddeley. 1975. "Context Dependent Memory in Two Natural Environments: On Land and Underwater." *British Journal of Psychology* 66:325–31.
Goleman, Daniel. 1998. *Working with Emotional Intelligence*. New York: Bantam Books.
Harcourt, A. H. 1992. "Coalitions and Alliances: Are Primates More Complex than Non-Primates?" In *Coalitions and Alliances in Humans and Other Animals*, ed. A. H. Harcourt and F. B. M. deWaal, 445–72. Oxford: Oxford Univ. Press.
Hollins, Mark. 1989. *Understanding Blindness: An Integrative Approach*. Hillsdale, N.J.: Lawrence Erlbaum.
Huxley, Thomas H. [1894] 1989. *Evolution and Ethics*. Princeton: Princeton Univ. Press.
Irons, William. 1996. "Morality, Religion, and Human Evolution." In *Religion and Science: History, Method, and Dialogue,* ed. W. M. Richardson and W. J. Wildman, 375–400. New York: Routledge.
Jerison, Harry J. 1973. *Evolution of Brain and Intelligence*. New York: Academic Press.
Johnson, Donald C., and M. Edey. 1981. *Lucy: The Beginnings of Humankind.* London: Granada.
Kennedy, Philip R., and Roy A. Bakay. 1998. "Restoration of Neural Output from a Paralyzed Patient by a Direct Brain Connection." *Neuroreport* 9:1707–11.

Kropotkin, Peter. 1902. *Mutual Aid: A Factor in Evolution*. London: Allen Lane.
Kurzweil, Raymond. 1990. *The Age of Intelligent Machines*. Cambridge: MIT Press.
———. 1999. *The Age of Spiritual Machines*. New York: Viking.
Lieberman, Daniel, and John Shay. 1994. "Behavioral Differences Between Archaic and Modern Humans in the Levantine Mousterian." *American Anthropologist* 96:300–332.
Lieberman, Philip. 1991. *Uniquely Human: The Evolution of Speech, Thought, and Selfless Behavior*. Cambridge: Harvard Univ. Press.
———. 1998. *Eve Spoke*. New York: W. W. Norton.
Lovejoy, Owen. 1988. "The Evolution of Human Walking." *Scientific American* 259:118–25.
Moravec, Hans. 1988. *Mind Children: The Future of Robot and Human Intelligence*. Cambridge: Harvard Univ. Press.
Neese, Randolf, and George Williams. 1996. *Why We Get Sick*. New York: Times Books.
Oldroyd, B. P., A. J. Smolenski, J.-M. Cornuet, and R. H. Corzier. 1994. "Anarchy in the Beehive." *Nature* 371:749.
Owens, E. 1989. "Present Status of Adults with Cochlear Implants." In *Cochlear Implants in Young Deaf Children*, ed. E. Owens and D. K. Kessler, 25–52. Boston: Little, Brown.
Passingham, Richard E. 1975. "The Brain and Intelligence." *Brain, Behavior, and Evolution* 11:1–15.
———. 1982. *The Human Primate*. San Francisco: W. H. Freeman.
Pfeiffer, J. E. 1973. *The Emergence of Man.* London: Book Club Associates.
Rachels, James. 1990. *Created from Animals: The Moral Implications of Darwinism*. New York: Oxford Univ. Press.
Ratnieks, F. L. 1988. "Reproductive Harmony via Mutual Policing by Workers in Eusocial Hymenoptera." *American Naturalist* 132:217–36.
Ridley, Matt. 1996. *The Origins of Virtue*. New York: Viking.
Sanders, E. P. 1992. *Judaism: Practice and Belief 63 BCE–66 CE*. London: Penguin Books.
———. 1993. *The Historical Figure of Jesus*. London: Penguin Books.
Sarich, V. M., and A. C. Wilson. 1967. "Immunological Time Scale for Hominid Evolution." *Science* 158:1200–1203.
Sawaguchi, T. 1989. "Relationships between Cerebral Indices for 'Extra' Cortical Parts and Ecological Categories in Anthropoids." *Brain, Behavior, and Evolution* 43:281–93.
———. 1992. "The Size of the Neocortex in Relation to Ecology and Social Structure in Monkeys and Apes." *Folia Primatologica* 58:131–45.
Searle, John R. 1992. *The Rediscovery of the Mind*. Cambridge: MIT Press.
Sibley, Charles G., and Jon E. Alquist. 1984. "The Phylogeny of Hominid Primates as Indicated by DNA-DNA Hybridization." *Journal of Molecular Evolution* 20:2–15.
Stix, Gary. 1996. "Waiting for Breakthroughs." *Scientific American* (April), 94–99.
Sutton, Dwight, and Uwe Jurgens. 1988. "Neural Control of Vocalization." In *Comparative Primate Biology: Neurosciences*, ed. H. D. Steklis and J. Erwin, vol. 4:625–47. New York: Alan D. Liss.
Wilkinson, Gerald S. 1984. "Reciprocal Sharing in the Vampire Bat." *Nature* 308:181–84.
Trivers, Robert L. 1971. "The Evolution of Reciprocal Altruism." *Quarterly Review of Biology* 46:35–57.
Tulving, E., and D. M. Thompson. 1973. "Encoding Specificity and Retrieval Processes in Episodic Memory." *Psychological Review* 80:352–73.
Zubrow, Ezra. 1990. "The Demographic Modelling of Neanderthal Extinction." In *The Human Revolution: Behavioral and Biological Perspectives on the Origin of Modern Humans*, ed. P. Mellars and C. B. Stringer, vol. 1:212–31. Edinburgh: Edinburgh Univ. Press.