Response

EMBODIED AI, CREATION, AND COG

by Anne Foerst

Abstract. This is a reply to comments on my paper *Cog, a Humanoid Robot, and the Questions of the Image of God*; one was written by Mary Gerhart and Allan Melvin Russell, and another one by Helmut Reich. I will start with the suggested analogy of the relationship between God and us and the one between us and the humanoid robot Cog and will show why this analogy is not helpful for the dialogue between theology and artificial intelligence (AI). Such a dialogue can succeed only if both our fascination for humanoids and our fear of them are equally accepted. Any avoidance of these emotions, as well as any rejection of the possibility that Cog might one day be humanlike, destroy the dialogue. The interpretation of both scientific theories and religious metaphors as stories replaces seemingly "rational" arguments with the confession of the respective commitments to a body of stories and opens up a space for exchange and friendship between AI-researchers and theologians—an option that usually remains closed.

Keywords: artificial intelligence; Cog; dialogue; Mary Gerhart; image of God; K. Helmut Reich; robot; Allan Melvin Russell; scientific and religious stories.

The comments of Mary Gerhart and Allan Russell and of K. Helmut Reich highlight different features of the Cog research.

Gerhart and Russell (1998) point out the intriguing parallel between (1) the relationship between the creator God and us human beings as his creations and (2) the relationship between some artificial-intelligence (AI) researchers and their creation Cog. The argument that AI can never be humanlike because it is always created by human beings is one of the

Anne Foerst is a Postdoctoral Fellow at the Artificial Intelligence Laboratory, Massachusetts Institute of Technology, 545 Technology Square, NE 43–812, Cambridge, MA 02139. She is also a Research Associate at the Center for the Study of Values in Public Life, Harvard Divinity School. Her e-mail address is annef@ai.mit.edu. This research was funded by the German Research Society (DFG).

[*Zygon*, vol. 33, no. 3 (September 1998).]

traditional arguments against AI research and its possible success. Because Cog is the first serious attempt to create a humanoid with not only the wits but also the features of a human being, this argument gains a new quality, which Gerhart and Russell correctly point out.

There is an old Talmudic question of whether the concept of the image of God means that the human bodily form is analogous to the form of God. This issue is mirrored in the question of whether Cog must have a humanlike body to be like us (a question passionately debated within the field of AI right now).

Also, the theory that Cog has to interact with human beings in order to become human can actually deepen the meaning of *image of God*, as Gerhart and Russell suggest: if we interact with God we have the chance to become like God. This proposition, however, implies not only an opportunity but a threat: a threat posed to God by the Godlike human being and, likewise, a threat to human beings that might be posed by a realized AI creature. In this regard, Gerhart and Russell refer to the humanoid HAL in Stanley Kubrick's *2001: A Space Odyssey.* This film takes up the theme of Mary Shelley's *Frankenstein* (see Foerst 1996, 681–93)—whose motif has appeared in science fiction so often that Isaac Asimov has dubbed the widespread animosity against humanoids the "Frankenstein Complex " (Asimov 1950).

There is finally the fact that the builders of Cog construct the robot in their own image so as to learn more about human beings—especially the development of human cognitive abilities right after birth—and thus to learn more about themselves. As I have written elsewhere, the dialectic of subject and object within AI leads to certain contradictions and limits the descriptive powers of AI systems with regard to human beings (see Foerst 1997, 6).

Nonetheless, the idea that God might have created human beings to learn more about Godself is delightful and immediately brings to mind Georg Wilhelm Friedrich Hegel's interpretation of Genesis. Hegel (and his followers, including Paul Tillich) interprets the Fall as a development necessary for the actualization of true human nature: the capability of thought and rational reflection. With the Fall, Hegel suggests, human beings became the antithesis of God; the synthesis, he believes, is promised for the end of time.

Following through on this analogy, we might conclude that Cog is as necessary for us as we are necessary for God. Such an interpretation might also explain the age-old and cross-cultural fascination of human beings with their artificial counterparts. One might easily derive other exciting speculations from the parallel between the relationships between God and us and the connections between AI researchers and Cog.

Because this parallel is so fruitful both for the concept of the image of God and for the understanding of the Cog project, it is indeed appropriate to ask why I did not draw it. The answer is that my intention was to enter into dialogue with researchers from AI; such dialogue cannot work if the dialogue partners base their arguments on assumptions not shared by both. Within a dialogue between AI researchers and theologians, it is useless to argue that the relationship between Cog and human beings is like the relationship between human beings and God. The argument assumes that there is a God, and this assumption is *not* shared by the builders of Cog. The paper attempts to avoid repeating the strategy of most AI opponents, who insist on either a God or a qualitative and unbridgeable difference between human beings and an AI system.

To work within the field of religion and science, one must situate oneself between academics who pursue religious inquiries and scientists who do not use theological arguments for their research. This means one has to make a basic decision: whether (1) to argue from one's own worldview, (2) to argue from a metalevel, or (3) to become part of the other community, learning to speak its language, adopt its beliefs, and operate from there. My sympathy for a more constructivist approach is no coincidence, for I understand my work within the Cog project as in the tradition of Steve Woolgar and Bruno Latour (1986). I attempt to enter dialogue with AI researchers by buying into their basic assumption that Cog can in principle become humanlike. All commentators think that I am overoptimistic regarding the success of Cog, and I am gladly using this opportunity to clarify my point of view.

I personally do not think that Cog will be successful. I think that our technology is not advanced enough to build something so complex as a human being. Rodney Brooks has built insectlike robots that are among the best ever constructed, but these creatures fail when compared with real insects—and how much more complex are human beings! The technology used for Cog is advanced and fascinating, the results are very promising, and I personally am excited about the project; nonetheless, if I compare the robot's abilities with those of a newborn infant, I am convinced that Cog in the end will fail. At the same time, I am convinced that the Cog project shows much more promise than any other AI project ever, and has the best chance to succeed eventually. Therefore, I am extremely hesitant to use this conviction as an argument against Cog! With increasing progress in science and technology, every argument based on a qualitative difference between human beings and AI systems has to retreat.[1] Entering into dialogue with AI researchers for me means to take them and their fundamental beliefs seriously and to operate from there. It is true, as Gerhart and Russell point out, that my paper is indeed intended to deal with relationships, but I am not primarily concerned with how relationships between Cog and its builders parallel those between human beings and God. In fact, my chief concern is the relationship between theologians and AI researchers.

Many theologians as well as other academics experience ambiguous feelings when confronted with artificial-intelligence research, especially with projects such as Cog. These mixed reactions are a result of the fact that the Cog project's criteria for humanlike intelligence differ from those of classical AI. As Reich correctly states, the ultimate goal of Cog as well as of any other AI project is to pass the Turing test.[2] Instead of building symbol- and rule-based systems with knowledge bases, however, the Cog project attempts to reach this goal by having the robot follow the steps of human development insofar as that is currently possible. Cog does not have explicit knowledge but experiences the world around it, experiences its own system, and learns from interaction—interaction among its own parts and between itself as an entity and its surroundings, which also include human beings.

The studies done by Sherry Turkle and my own experiences suggest that a robot that interacts and behaves as if it had "a mind of its own" seems to be more threatening to our intuitive sense of specialness than does a disembodied, intelligent database.[3] To enter into dialogue with the Cog researchers, then, means that one must not only stay away from any Dreyfusian argumentation but must also overcome such fears as these.

Only if I take the Cog project seriously, in this way, can I ask whether it is possible to use my own theological concepts in this dialogue. The basic decision to buy temporarily into the foundational assumptions of the Cog project prohibits me from taking the image of God as existing fact. But I think what Gerhart and Russell fail to notice is that I also do not accept as fact the Cog project's underlying anthropological assumptions that human beings are basically functional and mechanistic systems. I want to do justice to the Cog project by taking its ideas seriously without immediately attacking them. I do not have to convince some of my readers of the validity of the "image of God" concept, but I would like to invite these same readers as well as others to see the reasonability of the Cog project on its own terms.

The concept of the image of God and the AI understanding of human beings are consistent and mutually exclusive. They cannot both be considered factual. I can accept either one or the other as fact, but not both at once. That would create a situation in which no dialogue is possible, since the beliefs are fundamentally different.

It is this reality that led me to develop what I call the "symbolic approach." Like all the commentators, I am convinced that my description of the "Cartesian approach" is simplistic and that my suggestions for the symbolic approach are superficial. The paper aims to describe and advance dialogue with the people from the Cog project, not to lay down the foundation for a new epistemology (which cannot be done in a journal paper anyway). Nevertheless, I am grateful for all the suggestions that my commentators made and again would like to use this opportunity to clarify some of my points of view.

The symbolic approach is intended as a middle way between what I have characterized as Cartesian on one hand and Constructivism on the other. I do not want to give up the notion of a reality that exists independent from us. I am as convinced, however, that there is no way to describe this reality in an objective and universally acceptable way. We use symbols, analogies, and metaphors to describe the reality around us. We use images, myths, and stories to make sense out of what happens and to give our experiences meaning.

The description of reality one ultimately adopts is the one that correlates best with one's life experiences and one's character, that offers the best answer to existential questions, whose symbols seem most convincing to us, and whose stories move us the most. Because the description of reality we accept is so closely linked to who we are, we do not just rationally decide that this is the description we will use, but we make a commitment. Philosophy of science has repeatedly pointed out that every single theory about reality creates epistemological circles; we enter each circle with a commitment. In this respect, the theological circle (see Tillich 1951) is not greatly different from the epistemological circle created by the theory of embodied AI.

But even more significantly, a circle I do enter constitutes a community. Whatever circle I enter, I do not enter alone; every theory creates a community of people who all share the same commitment. The description of reality is therefore not just a personal construction but a transpersonal event. Reich (1998, 256) inquires after my understanding of transcendence. In fact, such personal and existential commitment to a body of symbols and the transpersonal community that this commitment creates, open up levels of reality that are otherwise closed to us (Foerst 1998, 97), and thus point toward the transcendent nature of symbols.

The Cog project rests upon a personal commitment to one description of humankind. This account is as convincing—and yet as limited—as the stories told in Genesis 1 and 2. The researchers involved in the Cog project do not just formulate their theories on cognition so as to enlarge their engineering choices; they also hope to get more insight into who they themselves are and how they function. For them, the stories they can formulate by using Cog as a thinking tool are fully convincing.

The Cog group's descriptions of humankind not only answer questions about who we are but create a community; this transpersonal aspect can be shared only if one first enters the circle created by commitment to the theories around embodied AI. Only if one takes the Cog story seriously and does not immediately confront it can one discover the descriptive power of the project: Cog *can* tell us something about who we are. And only by becoming part of the group (by joining the research group or simply by sympathizing with its assumptions and theories for a while) can one confront the Cog story with other stories (including the one about the image of God) and gain a new and enriched insight into who we are.

If we introduce the concept of the image of God in this very way, we can finally have a real conversation about religious experience. Paul Tillich, in his concept of the theological circle, argues strictly in the framework of theology of revelation: God gives us the answer, and governed by this answer we may ask about the meaning of our existence. In my paper (Foerst 1998, 106) I quote Oswald Bayer, who explains this circle further: God exists for us only in this answer because God does not want to exist for us otherwise. God's commitment to us—so the biblical promise maintains—is universal and valid; our commitment establishes the relationship.

AI researchers are commited to another circle. Someone who happens to believe in a God can be confident that the AI researchers are nonetheless under God's promise. Accepting them as fellow images of God cannot mean forcing them into our circle, however. Rather, we must respect the answers they have chosen for themselves—and create, *together*, a body of stories, symbols, and metaphors that will help us all to master the challenges of the (technological) developments of the twenty-first century.

NOTES

[1] I often refer in this context to the work of Hubert L. Dreyfus, who wrote in 1979 his famous book entitled *What Computers Can't Do: The Limits of Artificial Intelligence.* In 1992 he had to admit that many of the limits for computer intelligence he had proposed had in fact been overcome, and he entitled his next book *What Computers Still Can't Do: A Critique of Artificial Reason*. One can imagine how this kind of retreat could go on endlessly. I would like to thank Ray Kurzweil for the following comments on the IBM machine Deep Blue, which also attack the Dreyfus strategy. At the time Garry Kasparov beat the machine in 1996, everyone had said a machine could never beat a grandmaster in chess. Chess was seen as the ultimate limit for a machine that would run only basic calculations and lacked human creativity and inspiration. As soon as the machine indeed beat the human world champion, many commentators, who still insisted on a principal difference between us and a machine, then degraded chess. That way, human intelligence could continue to be regarded as intact, something special, with chess no longer seen as part of this specialness. It is needless to ask what will remain of this specialness with the rapid development of increasingly smart machines.

[2] This test was developed by Allan Turing as an intelligence test for computers: a human being is connected with two computers via keyboard. The person can ask questions of both and receives answers on a computer screen. The trick is now that the answers from one system are produced by a human being and the others by an intelligent program. If the human being cannot tell with certainty, after a while, which computer is used by a human being and which one is run by AI, the intelligent system has passed the test and has to be called intelligent. Because the topics for discussion are not specified and the human being can choose to talk about religious experiences as well as love or the weather, no AI system has ever come close to passing this test.

3. For a thorough sociological study of reactions toward these systems, see Turkle 1995. I also have experienced such reactions from various groups of visitors and when giving talks in numerous academic and church settings.

REFERENCES

Asimov, Isaac. 1950. *I, Robot*. Greenwich, Conn.: Fawcett.

Dreyfus, Hubert L. 1979. *What Computers Can't Do: The Limits of Artificial Intelligence*. New York: Harper and Row.

_____. 1997. *What Computers Still Can't Do: A Critique of Artificial Reason*. Cambridge: MIT Press.

Foerst, Anne. 1996. "Artificial Intelligence: Walking the Boundary." *Zygon: Journal of Religion and Science* 31 (December): 681–93.

_____. 1997. "Why Theologians Build Androids." *Insights: The Magazine of the Chicago Center for Religion and Science*, August, 1–7.

_____. 1998. "Cog, a Humanoid Robot, and the Question of the Image of God."*Zygon: Journal of Religion and Science* 33 (March): 91–111.

Gerhart, Mary, and Allan Melvin Russell. 1998. "Cog Is to Us as We Are to God: A response to Anne Foerst." *Zygon: Journal of Religion and Science* 33 (June): 263–69.

Reich, K. Helmut. 1998. "Cog and God: A Response to Anne Foerst." *Zygon: Journal of Religion and Science* 33 (June): 255–62.

Tillich, Paul. 1951. *Systematic Theology*. Vol. 1. Chicago: Univ. of Chicago Press.

Turkle, Sherry. 1995. *Life on the Screen: Identity in the Age of the Internet*. New York: Simon and Schuster.

Woolgar, Steve, and Bruno Latour. 1986. *Laboratory Life: The Construction of Scientific Facts*. Princeton, N.J.: Princeton Univ. Press.