Software Social Organisms: Implications for Measuring AI Progress

Kenneth D. Forbus

■ *In this article I argue that achieving human-level AI is equivalent to learning how to create sufficiently smart software social organisms. This implies that no single test will be sufficient to measure progress. Instead, evaluations should be organized around showing increasing abilities to participate in our culture, as apprentices. This provides multiple dimensions within which progress can be measured, including how well different interaction modalities can be used, what range of domains can be tackled, what human-normed levels of knowledge they are able to acquire, as well as others. I begin by motivating the idea of software social organisms, drawing on ideas from other areas of cognitive science, and provide an analysis of the substrate capabilities that are needed in social organisms in terms closer to what is needed for computational modeling. Finally, the implications for evaluation are discussed.*

Today's AI systems can be remarkably effective. They can solve planning and scheduling problems that are beyond what unaided people can accomplish, sift through mountains of data (both structured and unstructured) to help us find answers, and robustly translate speech and handwriting into text. But these systems are carefully crafted for specific purposes, created and maintained by highly trained personnel who are experts in artificial intelligence and machine learning. There has been much less progress on building general-purpose AI systems, which could be trained and tasked to handle multiple jobs. Indeed, in my experience, today's general-purpose AI systems tend to skate a very narrow line between catatonia and attention deficit disorder.

People and other mammals, by contrast, are not like that. Consider dogs. A dog can be taught to do tasks like shaking hands, herding sheep, guarding a perimeter, and helping a blind person maneuver through the world. Instructing dogs can be done by people who don't have privileged access to the internals of their minds. Dogs don't blue screen. What if AI systems were as robust, trainable, and taskable as dogs? That would be a revolution in artificial intelligence.

In my group's research on the companion cognitive architecture (Forbus et al. 2009), we are working toward such a revolution. Our approach is to try to build software social organisms. By that we mean four things:

First, companions should be able to work with people using natural interaction modalities. Our focus so far has been on natural language (for example, learning by reading [Forbus et al. 2007; Barbella and Forbus 2011]) and sketch understanding (Forbus et al. 2011).

Second, companions should be able to learn and adapt over extended periods of time. This includes formulating their own learning goals and pursuing them, in order to improve themselves.

Third, companions should be able to maintain themselves. This does not mean a 24-hour, 7-day-a-week operation — even people need to sleep, to consolidate learning. But

 ISSN 0738-4602

they should not need AI experts peering into their internal operations just to keep them going.

Fourth, people should be able to relate to companions as collaborators, rather than tools. This requires companions to learn about the people that they are working with, and build relationships with them that are effective over the long term.

Just to be clear, our group is a long way from achieving these goals. And this way of looking at the problems is far from standard in AI today. Consider for example IBM's Watson. While extremely adept at factoid question and answering, Watson would not be considered an organism by these criteria. It showed a groundbreaking ability to do broad natural language processing, albeit staying at a fairly shallow, syntactic level much of the time. But it did not formulate its own learning goals nor maintain itself. It required a team of AI experts inspecting its internals constantly through development, adding and removing by hand component algorithms and input texts (Baker 2011). Another example are cognitive architectures that started as models of skill learning, like ACT-R (Anderson and Lebiere 1998) or SOAR (Laird 2012). Such architectures have done an impressive job at modeling a variety of psychological phenomena, and have also been used successfully in multiple performance-oriented systems. However, using them typically involves generating by hand a model of a specific cognitive phenomenon, such as learning to solve algebraic equations. The model is typically expressed in the rule language of the architecture, although for some experiments simplified English is used to provide declarative knowledge that the system itself proceduralizes. The model is run multiple times to satisfy the conditions of the experiment, and then is turned off. More ambitious uses (for example, as pilots in simulated training exercises [Laird et al. 1998], or as coaches/docents [Swartout et al. 2013]) work in narrow domains, for short periods of time, and with most of the models being generated by hand. Creating systems that live and learn over extended periods of time on their own is beyond the state of the art today.

Recently, more people are starting to work on aspects of this. Research on interactive task learning (Hinrichs and Forbus 2014, Kirk and Laird 2014) is directly concerned with the first two criteria above, and to some degree the third. Interactive task learning is a sweet spot in this research path. But I think the importance of the shift from treating software as tools versus collaborators should not be underestimated, both for scientific and for practical reasons. The scientific reasons are explained below. As a practical matter, the problems humanity faces are growing more complex, while human cognitive capacities remain constant. Working together fluently in teams with systems that are enough like us to be trusted, and have complementary strengths and weaknesses, could help us solve problems that are beyond our reach today.

The companion cognitive architecture incorporates two other scientific hypotheses. The first is that analogical reasoning and learning, over structured, relational representations, is ubiquitous in human cognition (Gentner 2003). There is evidence that the comparison process defined by Gentner's structure-mapping theory (Gentner 1983) operates across a span of phenomena that includes high-level vision and auditory processing, inductive reasoning, problem solving, and conceptual change. The second hypothesis is that qualitative representations are central in human cognition. They provide a level of description that is appropriate for commonsense reasoning, grounding for professional knowledge of continuous systems (for example, scientists, engineers, analysts), and a bridge between perception and cognition (Forbus 2011). These two hypotheses are synergistic, for example, qualitative representations provide excellent grist for analogical learning and reasoning.

These two specific hypotheses might be correct or might be wrong. But independent of them, I think the concept of software social organisms is crucial, a way of reframing what we mean by human-level AI, and does so in a way that suggests better measurements than we have been using. So let us unpack this idea further.

Why Software Social Organisms?

I claim that human-level AI is equivalent to sufficiently smart software social organisms. I start by motivating the construction of organisms, then argue that they need to be social organisms. A specification for the substrate capabilities that are needed to be a social organism is proposed, based on evidence from the cognitive science literature.

Why Build Organisms?

There are two main reasons for thinking about building AI systems in terms of constructing software organisms. The first is autonomy. We have our own goals to pursue, in addition to those provided by others. We take those external goals as suggestions, rather than as commands that we run as programs in our heads. This is a crucial difference between people and today's AI systems. Most AI systems today can't be said to have an inner life, a mix of internally and externally generated plans and goals, whose pursuit depends on its estimation of what it should be doing. The ability to punt on an activity that is fruitless, and to come up with better things to do, is surely part of the robustness that mammals exhibit. There has been some promising work on metacognition that is starting to address these issues (Cox and Raja 2011), but the gap between human abilities and AI systems remains wide.[1]

Another aspect of autonomy is the separation of internal versus external representations. We do not have direct access to the internal representations of children or our collaborators. (Cognitive science would be radically simpler if we did.) Instead, we communicate through a range of modalities, including natural language, sketching, gesture, and physical demonstrations. These work because the recipient is assumed to have enough smarts to figure them out. The imperfections of such communications are well known, that is, the joint construction of context in natural language dialogue involves a high fraction of exchanges that are diagnosing and repairing miscommunications. To be sure, there are strong relationships between internal and external representations: Vygotsky (1962), for example, argues that much of thought is inner speech, which is learned from external speech. But managing that relationship for itself is one of the jobs of an intelligent organism.

The second reason for building organisms is adaptation. Organisms adapt. We learn incrementally and incidentally in everyday life constantly. We learn about the world, including learning on the job. We learn things about the people around us, both people we work and play with and people who are part of our culture that we have never interacted with and likely never will (for example, political figures, celebrities). We learn about ourselves as well: what we like and dislike, how to optimize our daily routines, what we are good at, bad at, and where we'd like to improve. We build up this knowledge over days, weeks, months, and years. We are remarkably good at this, adapting stably — very few people go off the rails into insanity. I know of no system that learns in a broad range of domains over even days without human supervision by people who understand its internals. That is radically different from people, who get by with feedback from the world and from other people who have no privileged access to their internals.

Having autonomy and adaptability covers the second and third desiderata, and can be thought of as an elaboration of what is involved in achieving them. Communication through natural modalities is implied by both, thereby covering the first at least partly. But to complete the argument for the first, and to handle the fourth (collaborators), we need to consider why we want social organisms.

Why Social Organisms?

People are social animals. It has been proposed (for example, Tomasello [2001]) that, in evolutionary terms, being social provides a strong selection bias toward intelligence. Social animals have to track the relationships between themselves and others of their species. Being social requires figuring out who are your friends and allies, versus your competitors and enemies. Relationships need to be tracked over time, which involves observing how others are interacting to build and maintain models of their relationships. Sociality gives rise to friendship and helping, as well as to deceit and competition. These cognitive challenges seem to be strong drivers toward intelligence, as most social creatures tend to be more intelligent than those that are not, with dolphins, crows, and dogs being well-known examples.

A second reason for focusing on social organisms is that much of what people learn is from interactions with other people and their culture (Vygotsky 1962). To be sure, we learn much about the basic properties of materials and objects through physical manipulation and other experiences in the world. But we can all think about things that we have never experienced. None reading this lived through the American Revolutionary War, for example, nor did they watch the Galápagos Islands form with their own eyes. Yet we all can have reasonably good models of these things. Moreover, even our knowledge of the physical world has substantial contributions from our culture: how we carve the mechanisms underlying events into processes is enshrined in natural language, as well as aspects of how we carve visual scenes up into linguistic descriptions (for example, Coventry and Garrod [2004]).

A number of AI researchers have proposed that stories are central to human intelligence (Schank 1996, Winston 2012). The attraction and power of stories is that they can leverage the same cognitive capacities that we use to understand others, and provide models that can be used to handle novel situations. Moral instruction, for example, often relies on stories. Other AI researchers have directly tackled how to build systems that can cooperate and collaborate with people (Allen et al. 2007; Grosz, Hunsberger, and Kraus 1999). These lines of research provide important ingredients for building social organisms, but much work remains to be done.

Hence my claim that human-level AI systems will simply be sufficiently smart software social organisms. By sufficiently smart, I mean capable of learning to perform a broad range of tasks that people perform, with similar amounts of input data and instruction, arriving at the same or better levels of performance. Does it have to be social? If not, it could not discuss its plans, goals, or intentions, and could not learn from people using natural interaction modalities. Does it have to be an organism? If not, it will not be capable of maintaining itself, which is something that people plainly do.

Substrate Capabilities for Social Organisms

This equivalence makes understanding what is needed to create social organisms more urgent. To that end, here is a list of substrate capacities that I believe will be needed to create human-level social organisms. These are all graded dimensions, which means that incremental progress measures can be formulated and used as dimensions for evaluation.

(1) Autonomy. They will have their own needs, drives,

and capabilities for acting and learning. What should those needs and drives be? That will vary, based on the niche that an organism is operating in. But if we are wise, we will include in their makeup the desire to be good moral actors, as determined by the culture they are part of, and that they will view having good relationships with humans as being important to their own happiness.

(2) Operates in environments that support shared focus. That is, each participant has some information about what others can sense, and participants can make their focus of attention known to each other easily. People have many ways of drawing attention to people, places, or things, such as talking, pointing, gesturing, erecting signs, and winking. But even with disembodied software, there are opportunities for shared focus, for example, selection mechanisms commonly used in GUIs, as well as speech and text. Progress in creating virtual humans (for example, Bohus and Horvitz [2011] and Swartout et al. [2013]) is increasing the interactive bandwidth, as is progress in human-robotics interaction (for example, Scheutz et al. [2013]).

(3) Natural language understanding and generation capabilities sufficient to express goals, plans, beliefs, desires, and hypotheticals. Without this capability, building a shared understanding of a situation and formulating joint plans becomes much more difficult.

(4) Ability to build models of the intentions of others. This implies learning the types of goals they can have, and how available actions feed into those goals. It also requires models of needs and drives as the wellsprings of particular goals. This is the basis for modeling social relationships.

(5) Strong interest in interacting with other social organisms (for example, people), especially including helping and teaching. Teaching well requires building up models of what others know and tracking their progress. There is ample evidence that other animals learn by observation and imitation. The closest thing to teaching in other animals found so far is that, in some species, parents bring increasingly more challenging prey to their young as they grow. By contrast, human children will happily help adults, given the opportunity (for example, Liszkowski, Carpenter, and Tomasello [2008]).

This list provides a road map for developing social organisms of varying degrees of complexity. Simpler environmental niches require less in terms of reference to shared focus, and diminished scope for beliefs, plans, and goals, thereby providing more tractable test beds for research. I view Allen's Trips system (Ferguson and Allen 1998), along with virtual humans research (Bohus and Horvitz 2011, Swartout et al. 2013), as examples of such test beds. As AI capabilities increase, so can the niches, until ultimately the worlds they operate in are coextensive with our own.

Implications for Measuring Progress

This model for human-level AI has several implications for measuring progress. First, it should be clear that no single test will work. No single test can measure adaptability and breadth. Single tests can be gamed, by systems that share few of the human characteristics above. Believability, which is what the Turing test is about, is particularly problematic since people tend to treat things as social beings (Reeves and Nass 2003).

What should we do instead? I believe that the best approach is to evaluate AI systems by their ability to participate in our culture. This means having AI systems that are doing some form of work, with roles and responsibilities, interacting with people appropriately. While doing this, it needs to adapt and learn, about its work, about others, and about itself. And it needs to do so without AI experts constantly fiddling with its internals.

I believe the idea of apprenticeship is an extremely productive approach for framing such systems. Apprenticeship provides a natural trajectory for bringing people into a role. They start as a student, with lots of book learning and interaction. There are explicit lessons and tests to gauge learning. But there is also performance, at first with simple subtasks. As an apprentice learns, their range of responsibilities is expanded to include joint work, where roles are negotiated. Finally, the apprentice graduates to autonomous operation within a community, performing well on its own, but also interacting with others at the same level. Apprentices do not have to be perfect: They can ask for help, and help others in turn. And in time, they start training their own apprentices.

Apprenticeship can be used in a wide variety of settings. For example, we are using this approach in working with companions in a strategy game, where the game world provides a rich simulation and source of problems and decisions to make (Hinrichs and Forbus 2015). Robotics-oriented researchers might use assembly tasks or flying survey or rescue drones in environments of ever-increasing complexity.

An example of a challenge area for evaluating AIs is science learning and teaching. The scientific method and its products are one of the highest achievements of human culture. Ultimately, one job of AIs should be helping people learn science, in any domain and at any level. The Science Test working group[2] has proposed the following trajectory, as a way of incrementally measuring progress. First, evaluate the ability of AI systems to answer questions about science, using standardized human-normed tests, such as the New York Regent's Science Tests, which are available for multiple years and multiple levels. Second, evaluate the ability of AI systems to learn new scientific concepts, by reading, watching videos, and interacting with people. Third, evaluate the ability of AI systems to communicate what they know about science across multiple domains and at multiple levels. We conjecture that this provides a scalable trajectory for evaluating AI systems, with the potential for incremental and increasing benefits for society as progress is made.

This challenge illustrates how useful the apprenticeship approach can be for evaluation. The first phases are aimed at evaluating systems as students, ensuring that they know enough to contribute. The middle phase focuses on being able to contribute, albeit in a limited way. The final phase is focused on AIs becoming practitioners. Notice that in each phase there are multiple dimensions of scalability: number of domains, level of knowledge (for example, grade level), and modalities needed to communicate. (We return to the question of scalable evaluation dimensions more generally below.) Progress across these dimensions need not be uniform: some groups might focus entirely on maximizing domain coverage, while others might choose to stick with a single domain but start to focus early on tutoring within that domain. This provides a rich tapestry of graded challenges. Moreover, incremental progress will lead to systems that could improve education.

Scalable Evaluation Dimensions

A productive framework should provide a natural set of dimensions along which progress can be made and measured. Here are some suggestions implied by the software social organism approach.

Natural Interaction Modalities

Text, speech, sketching, vision, and mobility are all capabilities that can be evaluated. Text can be easier than speech, and sketching can be viewed as a simplified form of vision.

Initial Knowledge Endowment

How much of what a system knows is learned by the system itself, versus what it has to begin with? What the absolute minimum initial endowment might be is certainly a fascinating scientific question, but it is probably best answered by starting out with substantially more knowledge and learning how human-level capabilities can be reached. Understanding those pathways should better enable us to understand what minimal subsets can work. It is very seductive to start from scratch, and perhaps easier, if it could be made to work. But the past 50 years of research suggests that this is much harder than it seems: Look at the various "robot baby" projects that have tried that. Arguably, given that IBM's Watson used more than 900 million syntactic frames as part of its knowledge base, the 5 million facts encoded in ResearchCyc might well be considered a small starting endowment.

Level of Domain Knowledge and Skill

Prior work on learning apprentices (for example, Mitchell et al. [1994]) focused on systems that helped people perform better in particular domains. They started with much of the domain knowledge that they would need, and learned more about how to operate in that domain. In qualitative reasoning, many systems have been built that incorporate expert-level models for particular domains (Forbus 2011). Breadth is now the challenge. Consider what fourth graders know about science (Clark 2015), and the kinds of social interactions they can have with people. AI systems are still far from that level of accomplishment, nor can they grow into expertise by building on their everyday knowledge, as people seem to do (Forbus and Gentner 1997).

Range of Tasks the System Is Responsible For

Most AI systems have focused on single tasks. Being able to accomplish multiple tasks with the same system has been one of the goals of research on cognitive architecture, and with interactive task learning, the focus is shifting to being able to instruct systems in new tasks, an important step toward building systems capable enough to be apprentices.

Learning Abilities

Software social organisms need to learn about their jobs, the organisms (people and machines) that they work with, and about themselves. While some problems may well require massive amounts of data and deep learning (for example, speech recognition [Graves, Mohamed, and Hinton 2013]), people are capable of learning many things with far fewer examples. Office assistants who required, for example, 10,000 examples of how to fill out a form before being able to do it themselves would not last long in any reasonable organization. There are many circumstances where children learn rapidly (for example, fast mapping in human word learning [Carey 2010]), and understanding when this can be done, and how to do it, is an important question.

Summary

I have argued that the goal of human-level AI can be equivalently expressed as creating sufficiently smart software social organisms. This equivalence is useful because the latter formulation makes strong suggestions about how such systems should be evaluated. No single test is enough, something which has become very apparent from the limitations of Turing's test, which brought about the workshop that motivated the talk that this article was based on. More positively, it provides a framework for organizing a battery of tests, namely the apprenticeship trajectory. An apprentice is initially a student, learning from instructors through carefully designed exercises. Apprentices start working as assistants to a mentor, with increasing responsibility as they learn. Eventually they start working autonomously, communicating with others at their same level, and even taking on their own apprentices. If we can learn how to build AI systems with these capabilities, it would be revolutionary. I hope the substrate capabilities for social organisms proposed here will encourage others to undertake this kind of research.

The fantasy of the Turing test, and many of its proposed replacements, is that a single simple test can be found for measuring progress toward human-level AI. Part of the attraction of this view is that the alternative is both difficult and expensive. Many tests, involving multiple capabilities and interactions over time with people, all require substantial investments in research, engineering, and evaluation. But given that we are tackling one of the deepest questions ever asked by humanity, that is, what is mind, this should not be too surprising. And I believe it will be an extraordinarily productive investment.

Acknowledgements

I thank Dedre Gentner, Tom Hinrichs, Mike Tomasello, Herb Clark, and the Science Test Working Group for many helpful discussions and suggestions. This research is sponsored by the Socio-Cognitive Architectures and the Machine Learning, Reasoning, and Intelligence Programs of the Office of Naval Research and by the Computational and Machine Intelligence Program of the Air Force Office of Scientific Research.

Notes

1. Part of the gap, I believe, is the dearth of

broad and rich representations in most AI systems, exacerbated by our failure as a field to embrace existing off-the-shelf resources such as ResearchCyc.

2. The Science Test Working Group includes Peter Clark, Barbara Grosz, Dragos Margineantu, Christian Lebiere, Chen Liang, Jim Spohrer, Melanie Swan, and myself. It is one of several groups working on tests that, collectively, should provide better ways of measuring progress in AI.

References

Allen, J.; Chambers, N.; Ferguson, G.; Galescu, L.; Jung, H.; Swift, M.; and Tayson, W. 2007. PLOW: A Collaborative Task Learning Agent. In *Proceedings of the Twenty-Second AAAI Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press.

Anderson, J. R., and Lebiere, C. 1998. *The Atomic Components of Thought*. Mahwah, NJ: Erlbaum.

Baker, S. 2011. *Final Jeopardy! Man Versus Machine and the Quest to Know Everything*. New York: Houghton Mifflin Harcourt.

Barbella, D., and Forbus, K. 2011. Analogical Dialogue Acts: Supporting Learning by Reading Analogies in Instructional Texts. In *Proceedings of the Twenty-Fifth AAAI Conference on Artificial Intelligence,* 1429–1435. Palo Alto, CA: AAAI Press.

Bohus, D., and Horvitz, E. 2011 Multiparty Turn Taking in Situated Dialog: Study, Lessons, and Directions. In *Proceedings of the SIGDIAL 2011 Conference, The 12th Annual Meeting of the Special Interest Group on Discourse and Dialogue*. Stroudsburg, PA: Association for Computational Linguistics.

Carey, S. 2010. Beyond Fast Mapping. *Language, Learning, and Development.* 6(3): 184–205. dx.doi.org/10.1080/15475441.2010.484379

Clark, P. 2015. Elementary School Science and Math Tests as a Driver for AI: Take the Aristo Challenge! In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence,* 4019–4021. Palo Alto, CA: AAAI Press.

Coventry, K., and Garrod, S. 2004. *Saying, Seeing, and Acting: The Psychological Semantics of Spatial Prepositions*. London: Psychology Press.

Cox, M., and Raja, A. 2011. *Metareasoning: Thinking about Thinking*. Cambridge, MA: The MIT Press. dx.doi.org/10.7551/mitpress/9780262014809.001.0001

Ferguson, G., and Allen, J. 1998 TRIPS: An Intelligent Integrated Problem-Solving Assistant. In *Proceedings of the Fifteenth National Conference on Artificial Intelligence and Tenth Innovative Applications of Artificial Intelligence Conference*. 567–573. Menlo Park, CA: AAAI Press..

Forbus, K. 2011. Qualitative Modeling. *Wiley Interdisciplinary Reviews: Cognitive Science* 24 (July/August): 374–391. dx.doi.org/10.1002/wcs.115

Forbus, K., and Gentner, D. 1997. Qualitative Mental Models: Simulations or Memories? Paper presented at the Eleventh International Workshop on Qualitative Reasoning, Cortona, Italy, June 3–6.

Forbus, K.; Riesbeck, C.; Birnbaum, L.; Livingston, K.; Sharma, A.; and Ureel, L. 2007. Integrating Natural Language, Knowledge Representation and Reasoning, and Analogical Processing to Learn by Reading. In *Proceedings of the Twenty-Second Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press.

Forbus, K. D.; Klenk, M.; and Hinrichs, T. 2009. Companion Cognitive Systems: Design Goals and Lessons Learned So Far. *IEEE Intelligent Systems* 24(4): 36–46. dx.doi.org/10.1109/MIS.2009.71

Forbus, K. D.; Usher, J.; Lovett, A.; Lockwood, K.; and Wetzel, J. 2011. Cogsketch: Sketch Understanding for Cognitive Science Research and for Education. *Topics in Cognitive Science* 3(4): 648–666. dx.doi.org/10.1111/j.1756-8765.2011.01149.x

Gentner, D. 1983. Structure-Mapping: A Theoretical Framework for Analogy. *Cognitive Science* 72(2): 155–170.

Gentner, D. 2003. Why We're So Smart. In *Language in Mind,* ed. D. Gentner and S. Goldin-Meadow. Cambridge, MA: The MIT Press. dx.doi.org/10.1207/s15516709cog0702_3

Graves, A.; Mohamed, A.; and Hinton, G. 2013. Speech Recognition with Deep Recurrent Neural Networks. In *Proceedings of the 2013 IEEE International Conference on Acoustic Speech and Signal Processing.* Piscataway, NJ: Institute for Electrical and Electronics Engineers. dx.doi.org/10.1109/ICASSP.2013.6638947

Grosz, B.; Hunsberger, L.; and Kraus, S. 1999. Planning and Acting Together. *AI Magazine* 20(4): 23–34

Hinrichs, T., and Forbus, K. 2014. X Goes First: Teaching Simple Games through Multimodal Interaction. *Advances in Cognitive Systems 3:* 31–46.

Hinrichs, T., and Forbus, K. 2015. Qualitative Models for Strategic Planning. Proceedings of the Third Annual Conference on Advances in Cognitive Systems, Atlanta, May.

Kirk, J., and Laird, J. 2014 Interactive Task Learning for Simple Games. *Advances in Cognitive Systems 3:* 13–30.

Laird, J. 2012. *The SOAR Cognitive Architecture*. Cambridge, MA: The MIT Press

Laird, J.; Coulter, K.; Jones, R.; Kenney, P.; Koss, F.; Nielsen, P. 1998. Integrating Intelligent Computer Generated Forces in Distributed Simulations: TacAir-SOAR in STOW-97. Paper presented at the Simulation Interoperability Workshop, 9–13 March, Orlando, FL.

Lizkowski, U.; Carpenter, M.; and Tomasello, M. 2008. Twelve-Month-Olds Communicate Helpfully and Appropriately for Knowledgeable and Ignorant Partners. *Cognition* 108(3): 732–739. dx.doi.org/10.1016/j.cognition.2008.06.013

Mitchell, T.; Caruana, R.; Freitag, D.; McDermott, J.; and Zabowski, D. 1994. Experience with a Personal Learning Assistant. *Communications of the ACM* 37(7): 80–91. dx.doi.org/10.1145/176789.176798

Reeves, B., and Nass, C. 2003. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Palo Alto, CA: Center for the Study of Language and Information, Stanford University.

Schank, R. 1996. *Tell Me a Story: Narrative and Intelligence*. Evanston, IL: Northwestern University Press.

Scheutz, M.; Briggs, G.; Cantrell, R.; Krause, E.; Williams, T.; and Veale, R. 2013. Novel Mechanisms for Natural Human-Robot Interactions in the DIARC Architecture. In *Proceedings of the Twenty-Seventh AAAI Conference on Artificial Intelligence*. Palo Alto, CA: AAAI Press.

Swartout, W.; Artstein, R.; Forbell, E.; Foutz, S.; Lane, H.; Lange, B.; Morie, J.; Noren, D.; Rizzo, A.; Traum, D. 2013. Virtual Humans for Learning. *AI Magazine* 34(4): 13–30.

Tomasello, M. 2001. *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.

Vygotsky, L. 1962. *Thought and Language*. Cambridge, MA: The MIT Press. dx.doi.org/10.1037/11193-000

Winston, P. 2012. The Right Way. *Advances in Cognitive Systems 1:* 23–36.

Kenneth D. Forbus is the Walter P. Murphy Professor of Computer Science and Professor of Education at Northwestern University. His research interests include qualitative reasoning, analogy, spatial reasoning and learning, sketch understanding, natural language understanding, cognitive architecture, reasoning system design, intelligent educational software, and the use of AI in interactive entertainment. He is a fellow of AAAI, ACM, and the Cognitive Science Society.