JUSTINE CASSELL

ARTIFICIAL INTELLIGENCE FOR A SOCIAL WORLD

The underlying mission is to better understand social interaction and to build machines that work more collaboratively and effectively with humans.

Will artificial intelligence (AI) mean the end of social interaction? I want to say right away—so that you can stop reading if this disappoints you—that I'm not going to answer that question.

I'm not going to answer it because I'm not a futurist, I'm a scientist. I do not believe in making predictions about the future; I believe in understanding what the possibilities are, and then trying to make the best one become the future. I am going to try to convince you that AI can help us better understand social interaction, that it can create a positive future in the form of systems that evoke natural social interaction from people in an increasingly technological world, and that it can preserve one of the aspects of humanity that we care most about—our sociality.

However, before I look at the future, or the present, for that matter, I want to talk a little bit about the past. Our feelings about new technologies always seem to start with hope, and expectation for all the wonderful things that a technology is going to do for us. When televisions first hit the market, the ads claimed that they would raise grades at school and improve family life. However, that joyous expectation eventually turns to fear and worry. In the case of the television, critics claimed that it was going to destroy children's grades, their desire to play outside, even the very fabric of family life. So it was with the printing press, the radio, video games—and so it is with AI.

Just a few years ago, many of us were thrilled at all the benefits that were going to accrue because of AI, and now many of us are terrified about all the disasters that are going to befall us. I began thinking about this curve that starts in hope and ends in fear when I was looking into parents' beliefs about how girls would be affected by their experiences online. I began reading about the history of women and technology and found that the same curve had existed with respect to girls and the telegraph, girls and the telephone, and then girls and the internet. In each case the hope was that these technologies would be improved—made kinder, for example—and then the fear was that girls would be lured into a dangerous world where parents could not control their actions. Thus, the fears we have about new technology such as AI may not be new at all. In which case, what we are experiencing may be fears not about the technology itself, but fears about us, our children, our families, and our workplace; fears about what may become of our society's community values. In my work on girls online, I've described that fear as what is known as a moral panic, in this case a perceived threat to societal values resulting from the possibility of young people, particularly young women, making their own decisions outside the sphere of the family.

AI appears to be evoking a similar kind of moral panic, but in this case it's more about a perceived threat to our capacity for empathy. Have we become more willing to inflict great pain on others without misgivings? Have we lost our sense of responsibility for one another? In that context, my research

program, which has strong continuities with my graduate work in linguistics and psychology, is to understand better how we interact with one another, how we work, play, and learn with our peers, then to understand whether AI is ineluctably making those interactions worse. If it is not, then we can work toward an AI that will maintain the good—perhaps even make it better. In my early years as a researcher, I attempted to better understand how we use language and the body to enhance interactions with others. Today I have added technology to the mix, and have added the challenge of building AI entities that jointly use language and the body to work, play, and learn with others. The knowledge that results from these experiments allows us to build technology that better supports our interpersonal and community goals. And those technologies can be a way to better understand how we work and play and learn.

To describe this in another way, my research began as the study of the relationship between verbal and nonverbal behavior, and what it tells us about cognition and language. Increasingly I became interested in the dynamic and interactive aspects of the relationship between verbal and nonverbal behavior, what is sometimes called joint action, where our communication with one another is analyzed as more than the sum of its parts, and there are phenomena that can be described only as a function of two or more people in conversation. Initially, for me, technology was a way to simulate this concept of units of analysis larger than the individual. I became interested in understanding the concept as a social scientist, and also in using that understanding to implement technological interventions that also engage in joint action with people, and in that way can support us in working, playing, and learning in the world.

This dual perspective on joint action has led me to engage in some broader debates. One is a debate in the social sciences, where the belief that the dyad or group can serve as a basic unit of analysis is not universally held.

Another, more specific to the research I carry out, is to present an alternative to the idea that the goal of AI must be autonomy. This widely held (but, I believe, erroneous) belief can be traced back to a small seminar that constitutes the most broadly cited origin story of AI. The Dartmouth Summer Research Project on Artificial Intelligence gathered 11 men (yes, men) in 1956 for a six-week effort that had as its purpose to "proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it."

However, the quest for AI actually began several years earlier and with a much broader mission. The Macy Conferences on Cybernetics were held between 1946 and 1953. Their goal was (merely) to develop a "general science of the workings of the human mind" and to understand how machines could serve as models of human cognition and, ultimately, how humans and machines might work together. This goal was similar in many ways to that of the Dartmouth workshops. However, the men and women (yes, and women) who attended came from fields as diverse as psychiatry, anthropology, and mathematics, and an important theme running through the conferences was the idea that a speaker and listener established "reflexive feedback loops" that made it possible to see them as one working unit. If the Macy conferences had become the core origin story of AI, we might have avoided a troublesome misunderstanding. If we had recognized at the beginning that humans themselves are not autonomous, but interdependent, the goal of making machines "as autonomous as humans" would not have arisen, and we could have made faster progress toward the more accurate, less troublesome and, ultimately, more useful goal of working toward interdependence. And so I see one goal of my research as providing examples and thus helping to spread the understanding that the mission is not, and need not be, to replace humans with machines, but to build machines that can work in tight interdependence with people.

Positive social interaction turns out to be a good predictor of high levels of task performance in many domains such as peer collaboration, survey interviewing, enrolling patients into clinical trials, and sales.

A third broader debate is to participate in raising up the status of the social sciences in the computational sciences. The mere phrase "soft sciences" is telling as to the status of social science in computer science. For example, computational social science, a growing discipline, that may add important insights into human behavior, unfortunately sometimes is described as a way of improving the rigor of the social sciences. If, however, we are to build machines that work in tight interdependence with people, then engineers need to respect and learn from social scientists who are studying the complexities of human interaction with one another and with technology.

Human behavior

After that long aside, let us turn to what it means to use AI to understand human behavior, and how to go about doing it. One approach is to simulate human behavior in an animated computer character with which a child or adult can interact, and then evaluate the ensuing simulation for how well it resembles reality. One can judge naturalness simply by looking, but more interesting is to look at the behavior of the dyad—that is, the naturalness of the human-machine interaction. Such a simulation of human verbal and nonverbal conversational behavior in an animated computer character is simply a cartoon if it is implemented by animators and found in videogames or on YouTube. However, if the conversational behavior is generated automatically, and adapts to the human interlocutor, then it is called an embodied conversational agent (ECA, pronounced "eh-ka" in Europe), or simply a virtual human. In this instance, as in the vision of the Macy Lectures, the animated character creates a feedback loop in which the conversational behavior of each interlocutor has an impact on the conversational behavior of the other.

An application of the ECA as a way of understanding human behavior is a study I have carried out with graduate students at Northwestern University and Carnegie Mellon (and most recently with the doctoral student Samantha Finkelstein) to better understand the use of marginalized dialects in school. Extensive research and heated debate has focused on what dialects children should be allowed or even encouraged to speak in the classroom. Over a period of at least 50 years, scholars have claimed that children will learn best if they are allowed to brainstorm or think aloud to their peers in the dialect that they speak in their homes. During that same period, other scholars have claimed that only if children speak the mainstream "school dialect" to everybody around them will they learn to think rigorously and to successfully rid themselves of a dialect that many of these scholars think should be eradicated. Note that both positions are independent of the position taken on whether children should learn the mainstream dialect before they go into the work world, or college environment. Both sets of scholars largely agree that outside the classroom and as one grows older, intelligence and abilities are most certainly judged in part based on how one speaks.

One issue that plagues the debate is how difficult it is to run empirical experiments on the subject. Experimenters cannot really control what dialect a child speaks in the classroom with his or her peers in order to assess the impact of that child's dialect on the dialect of other children, and on their learning, as well as on his or her own. It is just as difficult to find a natural experiment where two classrooms differ only in the dialect that the children speak. However, what we can do is create simulations of children that differ only in the dialect they speak, and that are able to interact with children in the school context.

Of course, using ECAs that resemble children—what we call virtual peers—instead of real children depends on the hypothesis that children interact with virtual peers the way they do with real peers. This hypothesis was tested in a number of prior experiments where children's behavior with their peers was compared with their behavior with virtual peers, and few differences were found. The children readily collaborated with the virtual peers on telling stories, conducting science experiments, and other tasks. No children tried to assess how capable the virtual peer was (to "break it"), nor did they ignore it, nor did their gaze and other nonverbal cues to communication differ. The same similarity was found when pairs of children interacted with a single virtual peer. On the other hand, the children did not mistake the virtual peer for a real child. This was demonstrated particularly convincingly in 2001, when we first developed the virtual peer, and a visiting film crew asked us to explicitly ask a group of nine-year-olds whether "Sam" was a "real child," and if not, what it was. I should note that in that experiment, as well as in all our subsequent experiments, we never use a pronoun with study participants to refer to the virtual peer—neither "it" nor "he" nor "she." In those early years, when videogames were less sophisticated, and avatars unknown, the children were clear that this was not a real child, but struggled to describe what it was. One said it was "like a Xerox," another said it was "like a computer game but way more complicated" and a third said it was "kind of like a real person … but not." Finally, one child, leaving the experiment room, asked the experimenter sitting outside the crucial disambiguating question: "Does Sam pee?"

Thus, turning back to the debate on the use of marginalized dialects in American classrooms, Samantha Finkelstein, in her dissertation work, employed in a series of experiments two of our virtual peers that differed only in the dialect they spoke. In one condition of a longitudinal experiment, the virtual peer spoke only mainstream American school English. It therefore used this dialect for brainstorming with children about their collaborative science task, and for practicing a presentation to the teacher about their results (the mainstream-only condition). In the second condition, the virtual child first brainstormed with the child in African American vernacular English (AAVE), and then switched to mainstream American school English for practicing the presentation to the teacher about their

results (the codeswitching condition). In both conditions, and in each science task that the peers collaborated on over a period of five weeks, after the peers had finished the brainstorming, the virtual peer invited the child to practice the presentation, and added, "My teacher likes it when I use my school English."

I should note that, as in many cities, a number of marginalized dialects are spoken in Pittsburgh (including Pittsburghese itself), where the experiments were conducted. As far as I know, all incite the same debate about how/when/where/if they should be allowed in the classroom. AAVE was chosen because I had worked extensively in Chicago with the dialect, and then, with the help of postdoc Brittany McLaughlin, was able to spend a year developing a grammar that captured Pittsburgh's specific instantiation of AAVE.

The results of the longitudinal study demonstrated that when working with the codeswitching virtual peer, the children spoke more dialect themselves, and they made greater use of the kind of science discourse their teachers were looking for (hypothesis formation, adducing evidence, forming conclusions) than those children who worked with the mainstream-only virtual peer. The increase in science content was stronger for children who scored lower on tests of reading skill. Further analysis of the data revealed that the predictive dependent variable was not dialect use in and of itself, but the fact that children felt greater rapport with the dialect-speaking virtual child, which led to improved performance in science.

Building rapport

Rapport is a hard-to-define concept that therefore can be difficult to measure. In this instance, we used a questionnaire developed by Sandra Calvert to assess what she refers to as prosocial interaction between children and cartoon characters. We also employed a technique developed by Nalini Ambady and her colleague Robert Rosenthal to measure underlying psychology states that do not lend themselves particularly well to self-reflection or existent measurement instruments. In what they called "thin slice annotation," everyday people (not researchers or university students or psychologists) are asked to judge the rapport, which is defined for them as a sense of being in sync or in harmony with another person, between two people based on a 30-second "thin slice" of video. Ambady, Rosenthal, and others have shown that judgments based on thin slices are as accurate as judgments based on much longer video clips, that there is high inter-rater reliability, and that the judgments align with self-report. We focused on assessing rapport because it is one of those phenomena mentioned above whose unit of analysis is larger than a single person. Rapport is also a quantifiable metric of positive social interaction. Social interaction—"chit-chat"—is often disregarded in the analysis of dialogues, or the building of conversational systems. However, positive social interaction turns out to be a good predictor of high levels of task performance in many domains such as peer collaboration (as above), survey interviewing, enrolling patients into clinical trials, and sales. In fact, managers often explicitly teach sales staff and other service personnel such as wait staff to build rapport with their clients because, they say, it increases sales or tips.

We have now seen one example where AI does not spell the end of social interaction. On the contrary, in this AI system, positive social interaction, in the form of a sense of rapport with a virtual peer, plays a key role in improving the AI's performance in teaching science. It also played an essential role in our ability to understand something, to my knowledge, not previously discussed concerning children's dialect use in the classroom; specifically the role of shared dialect use (homophily by dialect) in collaborative learning.

Note that rapport is not the same thing as affect. One can like another person but feel out of sync or a lack of harmony. Likewise, one can feel a sense of instant rapport through, for example, a shared smile at a musician's antics on the metro, without liking or even knowing the other person. In order to untangle the state of liking from that of rapport, Amy Ogan, Samantha Finkelstein, and I reanalyzed a dataset collected by Erin Walker for her dissertation on peer tutoring of algebra. In this dataset, children were paired up with somebody of the same age and then took turns tutoring one another. Some of the students showed up with their friends and were paired with them. Those students who came alone were paired with strangers. In analyzing what made peer tutoring successful, Erin had discarded a lot of seemingly irrelevant social interaction talk between the children, and she had not distinguished between friend pairs and pairs of strangers. We wondered whether we could better predict the children's performance on algebra if we included the chit-chat that had been thrown out, and if we looked at the difference between friends and strangers. Indeed, it turned out that by annotating in the dialogues and including in the analysis the teasing and insulting, the private jokes, the making fun of the experimenters, we could better predict learning gains in algebra.

The results, however, showed that what constituted rapport-building behaviors was different for friends vs. strangers. Teasing and insulting correlated with higher learning gains, but only among friends. When strangers teased or insulted one another, their learning gains were lower. How do we understand this? How could teasing and insults—behaviors that seem so counter-productive—improve learning? One interpretation is that learning is a face-threatening situation, where one makes oneself

vulnerable by implicitly admitting that one does not know something, an admission that is particularly threatening to adolescents. If that is true, then rapport building may act as the social lubricant to allow learning to take place between adolescent peers. However, only if a pre-existent bond exists, can teasing and insults—putatively negative social interaction—serve to build greater rapport. In fact, teasing and insulting may serve to highlight the bond, and constitute an index of the specialness of the relationship between two friends. For example, one teen tutor pushed his tutee to do the math problems in the order they were listed on the page. When his tutee refused, he hissed "Well, that's going to do you a lot of good in life!"

Among strangers, on the other hand, tutors were likely to hedge or mitigate the impact of criticism of the tutee's work as a way of creating a bond that could sustain the threat of vulnerability. For example, one teen advising on how to solve an equation said, "well, you kinda might want to add 5 on both sides." The presence of these hedges, among strangers, was what correlated with higher learning gains, as well as more algebra problems attempted.

In subsequent research, we continued this line of investigation and analyzed in even finer granularity the verbal and nonverbal behaviors that signal rapport and evoke rapport, and those behaviors that increase, maintain, and decrease rapport. Our data consisted, once again, of videos of teenagers tutoring one another in linear algebra. The results have identified a number of strong signals of rapport: increased smiles, more mutual gaze, more self-disclosure, more of what is known as "entrainment," such as adopting the other person's speech rate. We use machine learning to conduct some of these analyses. In particular we concentrate on those machine learning techniques that allow us to look at the dynamics of communication—what behaviors at time1 predict the occurrence of other behaviors at time2. This is sequence mining, here using temporal association rules, that can predict an outcome event, such as a learning gain, from the combination of input events, such as teasing followed by mutual eye gaze, or smiles preceded by hedging behavior. What that particular rule says is that when violating a social norm, such as teasing, is followed by reciprocal violation of a social norm, such as teasing back, and those behaviors are followed by mutual smiles, then rapport is likely to be high—but only among friends. The flip side is that if smiles don't follow mutual violation of social norms, then rapport is likely to be low.

Machine-conducted data mining of this sort allows us to understand social behavior in a way that would be hard to do with human observation. The temporal association technique, TITARL, can crunch through transcripts of 300 hours of video, pull out of them thousands of rules, group them, and then use those groupings of rules as input into a model of rapport behavior. Combining this detailed analysis of human behavior with a review of the literature from education, sociology, psychology, ethnomethodology, and many other fields led then doctoral student Ran Zhao and me to develop a more precise model of conversational moves and their impact on rapport.

One of the reasons we conduct the kind of painstaking hand annotation (and, increasingly, annotation by micro-workers, such as Amazon Mechanical Turk) and analysis I describe above is the ease with which it may be translated into a predictive model of rapport-building, and thence into algorithms that drive an AI system. My students and I look at hundreds and hundreds of hours of people interacting in situations that are as ecologically valid as we can make them while still allowing us to collect high-quality video and audio data. Then, rather than taking the shortcut of using data from just a couple of people to feed directly into a computer system, we develop a model that explains the behavior of interest and that is predictive of outcome variables such as learning gains. We know when we finish this model that given X behavior, Y behavior is more likely to occur, and it is that model that gives rise to the algorithms that we implement in systems. My students and I derived these models from observations of a broad range of situations over the years. One child telling stories with another, for example, compared with three children telling stories with one another, children playing with other children, or children playing with their parents. I've looked at graduate students discussing their research, and derived a model, with Candy Sidner and Chuck Rich, of how the movements of the torso predict a shift in topic, and how these shifts can signal topic shifts in ECAs.

We carried out a long study on working couples preparing to buy or rent a home, and the location of social chit-chat in the conversation with a realtor, and another study with Yukiko Nakano on the role of eye gaze in dyads of people giving directions to one another. In each case, the results from these analyses give rise to a model that gives rise to a set of algorithms to drive an AI system. And then, in each case, in order to assess the validity of the

Teasing and insulting may serve to highlight the bond, and constitute an index of the specialness of the relationship between two friends.

model, we ask people to interact with the system and look at whether their behavior resembles their behavior with other people, and whether it has a positive impact on the task on which the person and AI system are collaborating. Unlike describing the behavior of people, here we are building a virtual person from the ground up. Though it is an endless project—ensuring a job for years to come—it is one that allows us to identify some essential parts of human behavior that we have not taken into account in our previous models. These aspects of human behavior many times also have not been documented in the social science literature, so we need to return to investigate more closely the human-human data before going back to modeling and implementing systems.

So far here, we've looked at a couple of AI systems that instantiate a set of social behaviors adaptive to the social behaviors of the human interlocutor in such a way that rapport ensues. The fact that rapport ensues, as judged by objective observers, demonstrates people's willingness to be social with virtual humans. This fact paves the way for AI systems that use social interaction to improve performance on collaborative tasks. In some sense, the social talk that precedes a realtor's request for the financial status of a young couple, or the teasing of her tutor that is interwoven with a teen's attempt to solve an algebra equation, or the careful politeness and hedged suggestions of strangers trying to work together for the first time, all resemble what has been called "water cooler talk." We know that physical spaces that promote water cooler talk also promote a more effective work environment. Here too social interaction between human and AI may be greasing the wheels of effective collaboration.

Under the hood

Until this point, I have been intentionally vague about the technical aspects of AI systems such as these. This is not a discussion about simulated users, or Seq2Seq approaches to dialogue systems, and I do not intend to go into detail about the guts of these systems. However, I do want to touch on some of the aspects of socially aware ECAs to point out the ways in which they have also moved forward the state of the art in AI.

At the most basic level, the ECA is a dialogue system with a body. Dialogue systems are the kind of AI that you speak with when you call United Airlines or AT&T, or converse with when you chat with the assistant on Amazon's webpage. That means the system has the ability to recognize speech and to translate that speech into text (within the rather tight limits of the domain it was built for). It means that the system's natural language understanding (NLU) module looks for particular kinds of utterances (the United Airlines dialogue system can manage requests for flight times, but not requests for baseball scores) and extracts meanings or intentions from them (what the person wants from the system). The airlines agent might say, "Did I understand that you want a flight from Cleveland to Miami?" After the NLU, the dialogue manager maintains the dialogue history and knows what needs to be done next to fulfill the user's request. It also maintains cohesion in terms of how to refer to the items from earlier in the conversation in the most natural way. The system expects the caller to specify a day and time after asking for flights to a particular city, and thus saves the name of the city in memory while waiting for the rest of the information. The dialogue manager sends those requests to a task manager that can look up information from an appropriate database, route calls to different people to deal with, or decide to ask for further information from the caller. Whatever task response is chosen comes back as a set of concepts or commitments to satisfy and is then sent to the natural language generation module to be turned into conversational language. Finally, the language that is generated is sent either to a text-to-speech engine to be spoken, or written out in text for chat. Our virtual people do much the same thing as dialogue systems, except they have bodies. ECAs, however, are not only instantiated in animated characters but use their animated bodies in ways that are attuned to the human interlocutor's nonverbal and verbal behavior, and that add something other than a wow factor to the interaction. For example, ECAs use their animated hands to convey details complementary to those conveyed in accompanying speech, their eyes to indicate that it is the person's turn to speak, and their eyebrows to indicate what is most important in what they are saying.

As for socially-aware ECAs, right around the time that the task manager is working on how to complete the ask that has been set, the social reasoner is planning the kinds of verbal and nonverbal conversational moves that need to be engaged in to grease the wheels of successful task completion. Perhaps teasing followed by a smile, negative self-disclosure followed by a shy look away, adhering to politeness norms or ignoring them. Because they have bodies as well as voices, the ECAs generate not just language but also appropriate nonverbal behavior. Because they are socially aware, their manner of delivering information or requesting further details changes depending on who they are speaking to (a repeat caller or a total stranger), and where they are in the conversation (has the system just introduced itself, or has it been speaking with the human interlocutor for quite some time). Most important, the ECA changes its way of conversing based on the human companion's way of conversing at that point in the conversation. So, much like the teens we observed, our virtual human (named Jayden, to allow the gender- and ethnicity-ambiguous-looking character to be also

ambiguous in name) engages in negative self-disclosure, hedging, and humor to build rapport. Here is an excerpt from an actual interaction:

>> Jayden: No, that's not quite right. Don't worry, I think these are hard too.
>> Student: Okay. Oh, 50.
>> Jayden: Yeah, I think that's right. What do you think you should do next?
>> Student: I will divide by 10, and that will make the answer 5.
>> Jayden: Correct. I'm such a good teacher!

Jayden's first utterance contains a hedge that is followed by negative self-disclosure. Both of those strategies are subtle, but the impact on the student is clear. She relaxes her creased forehead, her voice becomes louder and firmer, and she answers less tentatively. A study conducted by Michael Madaio, Amy Ogan, and me indicates that in situations where the tutor and tutee are not close, this kind of hedging results in the tutee trying to solve more problems and actually learning more. This is particularly the case for tutors who are confident in their abilities, which indicates that the hedging is due to its impact on the tutee, and not due to lack of self-confidence in the tutor. The excerpt ends with teasing, as Jayden indicates that the credit for doing well belongs to the tutor rather than tutee.

Although this way of delivering the tutoring help is only slightly different from any other way, the changes to the system required a fair amount of innovation. For Jayden to be able to adapt to the students in this way, we had to develop a number of new modules for the socially-aware dialogue system. Our study of human behavior enabled us to develop a conversational strategy classifier that can determine with high precision whether a person's utterance is self-disclosure, praise, a question to elicit self-disclosure, a violation of social norms, a following of social norms, and so forth. In addition, a rapport estimator conducts an analysis of the language and nonverbal behavior of the virtual person and real person every 30 seconds to assess the level of rapport in the dyad. After the rapport level is calculated, a social reasoner decides on the conversational strategies to deploy in in order to strengthen or maintain the level of rapport.

Roughly 60 students have been tutored by the linear algebra virtual peer in one of three conditions: task only; a linear increase in rapport over time; and our socially-aware virtual peer, which adapts the language and nonverbal behavior as a function of the interlocutor's behavior and the rapport level of the dyad. Preliminary results suggest that students who worked with the adaptive virtual peer may learned more, particularly in the conceptual domain (understanding of principles, rather than knowledge of simple procedures).

Other applications

I'm going to end this examination of the impact of AI on social interaction with an application of socially-aware technology that opens up social interaction in a way quite different from what has been described thus far. It is an application of AI that I did not initially anticipate but that ended up being one of the most revealing in terms of the positive impact on our understanding of human sociality and on social interaction.

For more than two decades, attendees of my lectures have asked if this technology was available for their children with Asperger's or high-functioning autism, as they thought it would be useful for them. For years my response was that I did not have the expertise to conduct such a project. Then I met graduate student Andrea Tartaro, who was interested and did have expertise. Together we began by conducting observational research in schools and informal environments with 9- to 14-year-olds diagnosed as having Asperger's or high-functioning autism. We paid particular attention to interactions with peers, both neurotypical and autistic. Largely absent from these interactions were contingency—saying something relevant to the prior utterance by the other person—and adapting one's language in any way to the other person. We then invited each young person to play with a virtual peer. The virtual peer started telling a story and occasionally invited participation by saying "and then what happened?" or simply left silences for the child to jump in.

To our surprise, the children's contingency and the sense of a fluid conversation and story were markedly better with the virtual peer than with a real peer. Skills that their parents and teachers described as nonexistent were employed in straightforward ways. For example, the virtual peer started a story about a grandmother baking cookies and then said "umm umm" and then its voice fell off. The child jumped in saying "and she got the flour, and she got the sugar." Of course, none of these children learned those skills over the period of the 20-minute interaction, so what we've found is that performance is different than competence for these children. That is, the social skills or lack thereof that we hear so much about in this population may be due in part to the context rather than the child's lack of ability. That was encouraging, but we do not want children with Asperger's and high-functioning autism to be spending the rest of their lives in interaction with virtual children. That is not the goal. And so our question became, in the same way building

these systems allows us to learn about social interaction, could building and operating these systems allow children with autism and Asperger's to learn about social interaction?

In order to test this theory, we developed a system called the authorable virtual peer (AVP), where we simplified a control panel to operate a virtual peer, gave it to the children, and asked them to control the virtual peer for another child to interact with in another room, visible to the first child on a small TV. We first asked children to direct the virtual peer by selecting from a drag-and-drop menu of behaviors on a control panel. After choosing a set of behaviors, they watched the live interaction between virtual peer and the child in the other room. The children were then asked if they were satisfied with the result, and if not, did they wish to revise and select a different set of behaviors. Subsequently, we asked them if they'd like to record new responses or nonverbal behaviors for the virtual peer, and they were able to choose those as well from their control panel. We found that use of the panel resulted in monitoring and revising. One child made the virtual peer say initially "Do you want to hear a scary story? If yes, that's great. If not, that's too bad, because I'm about to tell one." That demonstrates little attention to the audience, but is quite typical of many individuals with autism. After the interaction, however, the child said, "I should have used more question buttons. Can I try it again?"

> It is in our power to bring about a future with AI where social interaction is preserved and even enhanced.

Teachers who work with these children have told us that there seems to be kind of a heat, or a feeling of being burned, for quite a number of children in this population when they have to interact with other children. Their typical response to close interaction is often to freeze and to end the conversation. However, when acting through a virtual peer, they appeared to feel perhaps less personally threatened and were able to formulate appropriate responses, to attend to gaps in the story and fill them, and even to initiate new parts of the narrative.

Now, the Holy Grail, of course, and the only question that really matters is whether this experience enables the children to have more satisfying interactions with other real children and thus to benefit from the close collaboration that is the basis of so much learning in American schools. A subsequent experiment, therefore, compared the transfer effect of two approaches with teaching social awareness. One approach was "social stories," which was the state of the art when we carried out our study in late 2006-2007 and the technique used by the school we worked with. In this approach, children are read stories that demonstrate rule-governed approaches to social interaction, and then the rules are explained to them. For example, when someone asks you a question, you should ask a question back; when someone asks "How are you?" the answer is "I'm fine, how are you?" We counterbalanced the learning of the set of social skills chosen by the school, such that half of the children learned social skill A with social stories, and the other half of the children learned social skill A by controlling and authoring the virtual peer. The first half of the children then learned social skill B with the AVP, and the second half of the children learned it from social stories, and so forth. We found that those learning through the virtual peer had a higher chance of transferring that learning to subsequent role playing with other children. In fact, there was a higher rate of appropriate responses predicted if the child first interacted with the AVP for the learning session. Especially affected were the social skills of reciprocity and contingency, including "give feedback" and "respond appropriately."

At the top of this discussion I made it clear that I have no predictions about what the future will hold, but I do believe that it is our responsibility to work toward a future we believe in. It is in our power to bring about a future with AI where social interaction is preserved and where social interaction is even enhanced. I believe that can happen through using social AI to understand social interaction, by implementing social AI that concentrates on collaboration rather than replacement, that encourages productive social behavior, and that can teach social skills to those who need and wish to learn them. We're creating an AI that is different from Alexa or Siri or Google Now. We're hearing complaints from parents who say their tech-savvy children are less and less polite, that they don't say thank you. We're developing an AI that may not evoke thank you; it may evoke "That's a stupid thing to say." That, however, is natural social behavior, and that's the goal. Because children then learn from that and grow out of it to become effective social individuals in a social world.

Justine Cassell *is associate dean of technology strategy and impact, and a professor in the Language Technologies Institute, at Carnegie Mellon University. This article is adapted from the Henry and Bryna David Lecture, which she delivered at the National Academy of Sciences in October 2018.*