The ties that bind

Justine Cassell

THE TIES THAT BIND

Social Interaction in Conversational Agents

Justine CASSELL*[1]

How to cite: Cassell, J. (2020). The ties that bind: Social interaction in conversational agents. Réseaux, no 220-221(2), 21-45. https://doi.org/10.3917/res.220.0021

* Inria Paris, PRAIRIE and Carnegie Mellon University

1. Acknowledgements: The research reported here relies on the work of many talented graduate and undergraduate students and other research collaborators at MIT, Northwestern University, and Carnegie Mellon University. Particular thanks to Kimiko Ryokai and Cati Vaucelle, Francisco Iacobelli and Margaret Echelbarger, Samantha Finkelstein and Amy Ogan. Thanks also to the generous funders who have supported this work, including the National Science Foundation, Heinz Family Foundation, and the IT R&D program of MSIP/IITP in Korea. This work was funded in part by the French government under management of Agence Nationale de la Recherche as part of the "Investissements d'avenir" program, reference ANR-19-P3IA-0001 (PRAIRIE 3IA Institute). Finally, I am grateful to to Bernard Geoghan and Julia Velkovska for precious feedback on the manuscript.

The questions that pre-occupy the public today about Artificial Intelligence concern primarily its dangers: will robots kill people, unintentionally and, most worryingly, intentionally? Will AI systems become smarter than we are, and use that increased intelligence to rule us? Will robots be able to influence us in the way that the smartest marketers do, but better, because they will have all the information about us they need to trick us into thinking that they like us, want us to succeed, are looking out for our best interests? And when they become capable of these expressions of emotion and rapport, will we come to prefer them to human partners?

We know that some of these dangers are real. AI systems in simulation environments have chosen to destroy boats in a fleet because they realized the fleet would move faster without damaged boats (Lenat, 1983; Gladwell, 2009). People might have, instead, saved the sailors on those boats, or fixed the boats. It is important to address these kinds of errors in judgment. Russell refers to making AI human-compatible (Russell, 2019). "Do no evil" was an early formulation of a similar principle (Asimov, 1950). We must ask, however, do no evil to whom, and as defined by whom? How does the self-driving car choose between the old person on one side of the crosswalk and the child on the other? These are thorny and necessary ethical questions. Particularly challenging are cases that involve the social role of AI systems in our world. For example, we know that nurses and doctors obtain better adherence to protocols if they build a bond with their patients (Pinto *et al.*, 2012*)*. Should medical expert systems used in developing nations without easy access to doctors, such as those used to school patients in management of diabetes (Mbogho & Makhubele, 2014), be imbued with the ability to build those same kind of relationships with patients in order to be maximally effective?

Ethics is one approach to struggling with these dilemmas. History provides another. Our feelings about new technologies always seem to start with hope and expectation for the benefits that will accrue to us as individuals and as a society. The introduction of the telegraph, and then radio and then television were each accompanied by widespread expectation about their roles in bringing the family together, and in allowing all children everywhere to

learn, regardless of their family circumstances or the quality of their local schools. However, in each case that joyous expectation eventually turned to fear and worry (Standage, 1998; Cassell & Cramer, 2007; Spigel, 1992; Wartella & Reeves, 1985). In the case of the telegraph, it was first heralded as a way to improve economic outcomes, make the world seem smaller, bring countries closer to one another, and preserve family bonds across time and space (Carey, 1983; Flichy, 1995). A shortage of operators led to the hiring of women, who were thought to bring a certain kindness and grace to the transmission of personal messages. However increasingly parents – and society at large – came to fear the possibility offered by the telegraph for young women to meet extra-familial men, and a number of news stories and novels treated the plotline of young women running away with men they met "on line" (Marvin, 1988).

In the case of the radio, parents worried about the dark stories broadcast on radio, and their impact on children's' thoughts and actions. As one parent wrote, "my child has murder on the mind. It's because of those horrible radio programs. I know it is" (Walter Baruch, 1998). In the case of television, while early enthusiasm focused on family togetherness and, educational content, critics later claimed that it was going to destroy children's grades, their desire to play outside, even the very fabric of family life (Davis, 1976). The US government even weighed in and distributed a pamphlet, written and illustrated by then famous comic strip author Walt Kelly.

> "there are few things to practice not doing. Don't be afraid of it. These things are probably here to stay. Don't be afraid of your child. He's not here to stay. He's a precious visitor. Do not wind your child up and set him to play with it unguided. Do not wind it up and set it to watch your child. A machine is a bad sole companion. It needs help. You can help it. Love your child." (Kelly, 1961).

In fact, the famous TV show Sesame Street was born of frustration with the "wasteland" that television had grown to be seen as in the early 1960s, and a desire to take advantage of the "addictive qualities" of television in order to teach children (Ganz Cooney, 1974). Before the radio and TV the same arc of hope followed by fear existed for many other technologies, all the way back to the printing press (Brann, 1981). Today the fears have pinned themselves on AI.

I have argued elsewhere (Cassell & Cramer, 2007) that these fears are not about technology *per se* but instead are *moral panics* about us and what we

have become in the face of new technologies. Stanley Cohen first described moral panics as

> ... a threat to societal values and interests; its nature is presented in a stylized and stereotypical fashion by the mass media; the moral barricades are manned by editors, bishops, politicians and other right-thinking people; socially accredited experts pronounce their diagnoses and solutions; ways of coping are evolved or (more often) resorted to; the condition then disappears, submerges or deteriorates (Cohen, 1972).

Thus, to apply this concept to the current case, what appears to be fear about the technology, may instead be fears about us, fears about our own capacity to destroy dearly-held societal values, in the face of highly attractive AI systems.

There is another historical perspective that may be useful here, and that is the history of AI itself. The usual origin story concerns the Dartmouth Workshop, that gathered a number of men (yes, men, with occasional visits from their wives or girlfriends) in 1956 for an eight-week summer meeting, convened by John McCarthy. Their goal was, as they phrased it to the Rockefeller Foundation when they requested funding:

> 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 (McCarthy *et al*., 1955)

Sometime before then, however, between 1946 and 1953, the Macy Conferences on Cybernetics were held, sponsored by the Josiah Macy, Jr. Foundation. At the heart of their endeavor was the dream of bringing together scholars from a wide range of disciplines to rethink society, nature, and technology in terms of models largely drawn from computing and the neurosciences (yes, even back then, based on the concept of neural nets as described by neurophysiologist Warren McCulloch and mathematician Walter Pitts in 1943). The group sought to develop a working model of thought and ideas, and a "general science of the workings of the human mind" in order to better understand human relations, and to design future collaborations between humans and machines. Indeed, they saw the effort to think humans and machines in complementary terms as one aspect of building a better world after the machine-aided catastrophes of World War II.

I describe the Macy Conferences here because, while their goal was quite similar in places to the Dartmouth workshops, the men and women who attended (yes, "and women") came from fields as diverse as psychiatry, anthropology and mathematics. An important theme running through the conferences was the idea that a speaker and listener, or any two interacting entities, established "reflexive feedback loops" that made it possible to see them as one working unit. Their initiative was called cybernetics, however, after the book by participant Norbert Wiener, and because of political infighting the Dartmouth Workshops sought another name. And for reasons also having to do with academic politics, although Claude Shannon, the author of Information Theory, attended both workshops, the two groups of scholars interacted very little, and the term and concepts of the Dartmouth workshops were increasingly referred to as the origin of AI.

As a thought experiment, however, imagine if AI researchers—taking a cue from their cybernetic peers—recognized from the very beginning that the goal of making machines autonomous is a dead end, because we're not autonomous, we're deeply interdependent? This might have led us to an AI that did not come to replace humans with machines, but always worked towards machines that can work in tight interdependence with people. Given the participation of Margaret Mead, Gregory Bateson, and others with a long history of studying human behavior, that Macy Conferences approach might very well have highlighted from the beginning the socio-cultural nature of interaction, and therefore the need for computers to interact with us the way we interact with one another in order to be successful.

Neither of these initiatives was independent of their era. This soon after World War II mathematicians, physicists and, indeed, every kind of scientist was forced to reflect on the role technology played in the war. Some reflected on how to improve weapons – a pursuit encouraged by funding agencies who were already planning military initiatives for the Cold War (Kline, 2011) – and others reflected on how to obviate the need for weapons. Norbert Weiner, having worked during the war on the mathematical underpinnings of more accurate anti-aircraft guns, turned towards this latter pursuit. For him Cybernetics, which he first defined in 1948 as "the scientific study of control and communication in the animal and the machine" became a study of feedback loops within and among people, and between people and machines, among societies, and perhaps even between humans, machines, and nature, as well (Turner, 2013). In this light, Wiener saw the importance of ensuring that machines adapted to humans, for the good of the world.

Taking back up that latter challenge, in what follows I will focus in on a particular kind of panic around AI, which concerns the perceived threat to our capacity for empathy and close relationships, and the challenge of designing systems that, on the contrary, maintain our humanity. The questions associated with this specific fear include: have we have become more willing to inflict great pain on others without misgivings? Have we lost our sense of responsibility for one another? Have we lost our ability to distinguish what is unique in human relationships, and lost our ability to value it? A number of scholars have claimed that we have (*inter alia*, Weizenbaum, 1976; Turkle, 2011).

History has given us a window into the etiology of fear concerning AI. The study of human behavior can, I believe, give us a window into how to escape that fear. Here, then, I rely on an interdisciplinary social scientific approach to understanding the role that AI can play in people's lives. I start, perhaps paradoxically, by examining contexts in which AI technology is not present, specifically the way in which individuals interact with one another. This analysis is carried out using the tools of conversational analysis, sociolinguistics and ethnomethodology, as well as contemporary data mining tools – to better understand the moment-to-moment details how we work, play, and learn with one another using language and nonverbal behavior. This understanding of human collaborative and cooperative behavior in the absence of technology is intended to give us a view into healthy and desirable contexts and outcomes of social interaction. In what ways do our conversations with one another – including words, prosody, hand gestures, and facial movements – bring us closer or push us away from one another? In turn, what is the impact of our level of rapport with one another on our ability to work together, and our ability to succeed as teams or groups? The successful social interactions that we uncover in this way can then serve as *desired futures*; futures which we wish to maintain or bring about. We build these successful social interaction protocols into functioning AI systems. And then we evaluate the performance of the systems when they collaborate with people, comparing them to the human-human interaction that spurred their design. Are the human-agent interactions also successful at eliciting productive social interaction – productive in the sense that it evokes a feeling of connection in the real person, and also productive in the sense that it improves the machine's ability to collaborate on an important task.

We can then ask whether AI systems are ineluctably diminishing the quality of social contexts. Must AI diminish social relationships and, if so, will

our collaborations suffer? Alternatively, can we bring about desired futures through the careful design of AI that will maintain, and perhaps improve, the positive aspects of our interactions – with others, with technologies, and with ourselves? And can these rapport-building technologies improve the ability of AI systems to engage in productive collaboration and cooperation with their human users? Here technology plays a double role, as it did for the attendees of the Macy Conferences: as an analytic tool, enabling a view into the workings of human interaction; and as an intervention – relying on human styles of interaction built into technology to bootstrap positive change.

LEARNING WITH A VIRTUAL PEER

An illustration of this approach comes from a technology called the *virtual peer,* an animated cartoon figure of a child driven by an underlying AI infrastructure, and capable of collaborating with children on a task. The initial work on this technology was carried out in the early 2000s, before deep learning techniques were common in AI systems. Nevertheless, the approach was the same. My students and I first examined children's spontaneous collaborative storytelling in the absence of technology, as a way of approaching how children's social interaction might affect the critical skill of literacy. Literacy - knowing how to read and write – begins long before children enter school. One of the key skills to reading and writing is the ability to represent thoughts symbolically and share them in language with an audience that may not necessarily share the same temporal and spatial context for the story. This skill has been called" emergent literacy" (Teal & Sulzby, 1986. Emergent literacy behaviors, specifically what has been called decontextualized language, predict early learning of reading and writing skills (Snow, 1983). It turns out that children learn and practice these important skills when they tell stories with the peers and adults around them, and that their storytelling with peers is more likely to contain decontextualized language than is their storytelling with adults (Goncu, 1993). This is thought to be because other children are quick to interrupt the storyteller when they don't understand something, whereas adults are more forgiving of parts of the story that are not really understandable by somebody outside the child's own head (Preece, 1992). We therefore looked for the moment-to-moment verbal and nonverbal storytelling behaviors that characterize emergent-literacy in children's collaborative storytelling around a toy castle, collecting enough data about children's peer storytelling to be able to construct a model of what peer storytelling behaviors are most successful in helping children to acquire literacy.

We then built the most successful of these emergent literacy behaviors into a virtual peer system called "Sam the CastleMate." Sam was projected lifesize on a very large screen located behind a real physical toy castle. A toy figurine was designed that could exist in either the physical world or on the screen, to allow Sam and the child to pass the story back and forth between their worlds. In fact, the system detected a child's presence in front of the castle through a microphone and a floor mat. When the child was playing with the castle and narrating, the system used audio threshold detection to determine when to give feedback (backchannels such as ''uh-huh,'' nods, and explicit prompts such as ''and then what happens?''). Sensors embedded in each room of the castle and on the figurine told the system in which room in the castle the figurine was located in so that Sam could give contextually appropriate feedback. Finally, a switch in the door to the castle tower allowed the system to sense whether the figurine was inside of the magic tower or outside of it. This allowed the toy figure to exist in either the physical or virtual world, but never both at a time. Passages from the child-child data that resulted in emergent literacy behaviors were recorded by a real child, and built into the system to be uttered by Sam in response to children's stories adults (Cassell *et al*., 2000). We chose to make the Sam character anthropomorphic so as to elicit the unconscious conversational behaviors that even very young children use. However, while Sam was capable of natural verbal and nonverbal responses, in no way was Sam lifelike or realistic. The graphics were intentionally cartoon-like in order to constrain the human partner's expectations, and to avoid any ambiguity about whether Sam was "real" or not. Most importantly, Sam was not photo-realistic because the focus of the research was on the child's behavior and not the virtual peer's.

Finally, we evaluated whether the Sam system evoked natural social interaction behaviors in children, and whether it improved their emergent literacy skills. We asked some children to play with Sam by themselves, and other children to play with Sam and one other child, in a triad. We compare these interactions to children playing with another child, or telling stories by themselves. First it was very clear that children's interaction with Sam was natural, and that they enjoyed the interaction. In fact, sometimes the children even coached Sam in how to tell stories, as in the case of one boy who told Sam "Try to make a longer story next time. It's like this. The little boy was outside. . ." The key question, however, is whether there is any benefit to Sam's presence? Can a system like this improve literacy, or is it simply as good as children collaborating with one another. For, if simple human peer social

interaction improves emergent literacy behaviors, we might prefer encouraging children to tell stories with other children, in the absence of technology. The answer to that question is that the Sam system did what real peers do, but better. The dyads of children, playing without Sam, sometimes told complete stories with decontextualized emergent literacy language. They also sometimes told stories that devolved into arguments or breaking parts of the castle. The stories with Sam, on the other hand, were more likely to show the important decontextualized language that predicts later literacy, and these instances of decontextualized language increased with each subsequent story that they told in collaboration with Sam. In fact, when dyads of children played with Sam, their stories with one another showed more of these emergent literacy behaviors – linguistic behaviors, but also the kind of pro-social collaboration that allowed them to get maximal educational benefit from the storytelling, unlike their stories with one another adults (Cassell, 2004). We can conclude, then, that systems like the virtual peer can have real and concrete positive impact both on child-child social interaction, and on children's learning. This demonstrates that AI need not diminish the quality of social interaction, nor worsen collaboration. It can serve as a tool to improve human-human collaboration and learning, as well as the collaboration between human and machine. It also serves as a valuable tool to investigate human behavior.

AI AND POLICY DEBATES AROUND LEARNING

More recently, the virtual peer was deployed in another, more contentious and more high-stakes educational context. Here too the approach was to examine children's interaction with one another, build models of their interaction, focusing particularly on features that predicted positive educational outcomes, incorporate the models into AI systems, and then evaluate the performance of those systems, in encouraging social interaction, and in improving real task outcomes, such as classroom learning. In this context, too, the technology served as tool to understand better human behavior, as well as intervention to potentially improve human performance. In this instance the focus was the politically sensitive topic of dialect and learning. The United States Department of Education has consistently shown a significant gap in literacy skills between African American and European American children (NAEP, 2019). More recently, several studies have pin-pointed one source of that achievement gap in the dialect spoken by children: there is a robust negative correlation between the use of African American Vernacular English and reading and writing literacy, independent of the learner's socio-economic

status (Gatlin & Wanzek, 2015). However, these findings have not resulted in consistent interpretations about the source of the correlation, nor about how dialect should be treated in the classroom. In fact, over a period of at least 50 years, scholars have gone back and forth on the topic. On the one hand, some 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 home (Lee, 1997). Others have claimed, on the contrary, that only if children speak the mainstream "school dialect" to everybody around them will they learn to read and write (Wiley & Lukes, 1996). Note that both positions are independent of whether one believes that children should learn the mainstream dialect before they go into the work world, or the university environment. Both sets of scholars largely agree that outside of the classroom and as one grows older intelligence and abilities are unfortunately most certainly judged in part based on how one speaks. What is being asked here is not whether children should learn the mainstream dialect, but whether they should speak it while also trying to learn subject matter domains.

A pervasive issue with research on this topic is the difficulty of running empirical experiments. Young children cannot be commanded to speak a dialect. In fact, they may not even be aware of what dialect they are speaking at a given moment, and thus we cannot look at the nature of learning when a child is working with the same peer speaking one versus another dialect. Likewise, it is virtually impossible to find a natural experiment where two classrooms differ only in the dialect that the children speak, independent of other factors such as quality of the school system, dialect spoken at home, socio-economic status, ethnicity, and so forth. We can, however, create virtual peers designed to look identical and act identically, but who use dialect differently. Those virtual peers can then serve to investigate the role of peer dialect on learning.

As with all of the research in this paradigm, the first step was to collect data on peer interaction in the absence of a virtual peer or any other digital technology. The data were collected in two mid-size American cities, in schools in where more than 90% of the children were African-American, and the majority were from low-income families (as measured by the percentage of children eligible for a free or reduced-cost lunch, a common measure of socio-economic status in the US). Observational studies of the schools demonstrated that African American Vernacular English was the dialect most often spoken by the students, and that teachers often corrected the students who spoke dialect, and insisted that they speak "correct English" (Rader, Echelbarger& Cassell, 2011; Finkelstein *et al*., 2013). Those observational studies also concluded

that teachers often divided the students into dyads or small groups to work together on various school tasks. Children were asked to work with their classmates on one of several similar science tasks. In one task, the pairs of students were asked to brainstorm about how to build a bridge out of Lego blocks, and then to take turns playing teacher so that the other student had an opportunity to practice explaining to a teacher the decisions made in building the bridge. (Cassell, 2009). In another study pairs of students wereasked to brainstorm together to answer questions about a series of imaginary animals pictured in their natural (imaginary) habitat (Finkelstein *et al*., 2013). These kinds of tasks, common in US science classrooms for children around the age of 8 years, require reasoning from data and being able to point out evidence for one's conclusions (for example, in the imaginary animal task, concluding that if the animal is shown without claws or sharp teeth, it probably won't be able to eat meat). The students were then, here too, asked to take turns playing the teacher, so that each child had an opportunity to practice explaining the answers she/he had come up with about the imaginary animal's characteristics. Analyses of these data focused on two kinds of behavior. One was the way they described the science assignment. The second set of analyses focused on the children's use of dialect. In many of the dyads, the children changed their language when pretending to be a teacher, often reducing the use of AAVE and introducing some school English into the conversation. Some of the students, then, were able to *code-switch*, or shift from one dialect to another, as well as able to use the school-ratified style of science discourse.

The children's language and science discourse features were incorporated into the virtual peer's speech, which was subsequently used in a series of experiments on the correlation between dialect use in peers and the use of what is known as "science discourse" (more accurately called "school-ratified" or teacher-approved science discourse). Science discourse is the use of language that indicates that children are observing, asking questions and reasoning about what they observe, and gathering evidence to support their beliefs. Science discourse is about reasoning and is independent of the dialect that is spoken. In order to reduce as much as possible any source of difference between the two virtual peers, their speech – both AAVE and Mainstream English was recorded by the same bi-dialectal voice actress.

Based on these data, two virtual peers were developed, capable of engaging in exactly the same science tasks as had been completed by the child-child dyads. Their look was identical and ethnicity-ambiguous (as assessed beforehand

with children of the same age group), they were dressed identically, and both were named Alex, but their use of dialect was different, although the content of what they said was the same. In one experimental condition (mono-dialectal), the virtual peer spoke only mainstream American English, both while brainstorming with an 8-9 year old child about the science task that they were working on together, and while practicing a presentation to the teacher about their work. In a second condition (bi-dialectal), the virtual child first brainstormed with the real child in African American Vernacular English and then switched to mainstream English for practicing the presentation to the teacher. In both conditions, right before beginning to practice the presentation to the teacher, the virtual peer said "my teacher likes it when I use my school English, so let's practice our presentation that way." An assessment of the children's use of science discourse before and after the study allowed us to assess change in science discourse by condition. We also assessed the rapport (entente) between the child and virtual peer, as a quantifiable metric of positive social interaction. Following work by social psychologist Nalini Ambady (Ambady & Rosenthal, 1992), rapport was assessed by naïve viewers, who were presented with a video of the child-virtual peer interaction which had been divided into randomized 30 second slices, and told simply "rapport is a feeling of harmony or getting along. Please judge each 30 second slice for its level of rapport (1-7)."

Results demonstrated that, when working with the bi-dialectal Alex agent, children showed a greater increase in the use of science discourse than did the children who worked with the mono-dialectal virtual peer. However, a closer analysis revealed that there was a mediating factor that was responsible for most of the results, and that was the children's significantly stronger sense of connection with the dialect-speaking virtual child. In other words, when children felt rapport with the virtual peer, they were more likely to use science discourse – regardless of the dialect the agent spoke. However, they were much more likely to feel rapport with the bi-dialectal agent – an agent who spoke as they did. In fact, this result was strongest for children who were under-performing in school – children whose grades on reading were lowest, and who were most in need of support in classrooms where there were not enough teachers to provide individual attention.

A subsequent similar study carried out over a 6 week time period likewise found that children working with the bi-dialectal virtual peer produced more science discourse during the practice of the presentation to the teacher.

This result was in part due to the fact that, unlike the cooperative relationship between the child and bi-dialectal agent, many children working with the mono-dialectal virtual peer simply refused to practice a presentation. As one student said to the virtual peer "I hate doing this every week with you, you know that?" These two virtual peer systems demonstrate places where AI does not spell the end of social interaction. On the contrary, in these systems, the virtual peer's ability to engage in positive social interaction, in the form of building rapport, plays a key role in improving the system's task performance (teaching literacy or science) as well as an essential role in the system's ability to help scientists understand children's language use in the literacy and science classrooms.

THE EFFECTS OF ALIGNING ONESELF WITH A TEACHABLE AGENT

Whereas virtual peers like those described above may serve all the roles that real children do in peer collaboration, *teachable agents* are a kind of virtual peer that is designed specifically to allow children to learn through teaching. Teachable agents have been shown to increase students' confidence and self-efficacy and their motivation to learn, as students come to feel responsible for their teachable agent's success, and improve their confidence in their own ability to learn (Rohrbeck *et al*., 2003; Cohen *et al*., 1982; Chase *et al*, 2009). Teachable agents can also help children increase their knowledge in a domain through what is known as the *tutor effect*, whereby peer tutors tend to learn more than their tutees (Biswas *et al*, 2005). It has been posited that the tutor effect is due to students reflecting on how best to teach a given topic, and reworking or elaborating material that the teachable agent doesn't understand. Prior work on teachable agents did not look, however, at the interaction between tutoring behavior, social behavior, and the student's learning gains.

In what follows, teachable agents were used to better understand the mechanism behind the interaction between social interaction and learning gains. Students learning algebra were introduced to a teachable agent named Stacy, implemented on the SimStudent platform (Matsuda *et al*., 2011), and were told that their goal was to help Stacy learn how to solve equations with variables on both sides to help her pass four sections of a quiz. Students worked with Stacy for two 90 minute sessions using the SimStudent platform to demonstrate problems in linear equations, and give feedback to Stacy as she tried to solve problems in order to accomplish this goal. A pre-test and post-test provided information concerning the students' learning of the material

through teaching Stacy. We also asked the students to talk aloud while they were teaching Stacy. "Think aloud" protocols like this one are good ways to access the process underlying cognitive performance. In this instance, we were interested in whether the students would engage Stacy socially whiles demonstrating linear equations, and whether that social interaction would correlate with the student's own learning gains in algebra.

The students had little trouble verbalizing their thoughts as they worked. They described how they were approaching the tutoring, and what worked and what did not in the tutoring process. We also found that some students talked aloud about Stacy's performance, referring to her in the third person, for example "OK, Stacy doesn't understand the distributive party." We called this way of referring to Stacy an "outside alignment." Others talked directly *to* Stacy about her performance, referring to her in the second person. We called this an "inside alignment." Some of these inside alignment comments were positive, for example, "you got it, Stacy. Congratulations!" or "Oh, Stacy, you were so close!" Other comments were quite negative, for example "you got lucky Stacy" or "Arggh, you annoy me so much!" Some students stuck to one alignment throughout their time with Stacy while others switched from inside to outside alignment and back again.

To our surprise, the results demonstrated that, unlike previous studies by other scholars, elaboration of linear algebra material negatively rather than positively correlated with learning gains. That is, when students spoke about what Stacy understood or didn't, and how they were going to tutor her, they were less likely to learn. A closer analysis, however, showed that the negative correlation between elaboration and learning gains was accounted for by the fact that when students began to elaborate on Stacy's understanding of the material, they also switched to an outside alignment perspective, turning away from referring to her as "you" and beginning to refer to her as "she" or "it." In fact, overall, we found that shifting away from direct communication with Stacy and instead talking about her was more negatively correlated with learning than any other behavior we observed. Looking more closely at the inside alignment stance towards Stacy we also found an unexpected result: negative social moves, such as teasing and face-threat, made directly to Stacy, were the most positively predictive of learning gains. These are comments such as "Stacy, what are you doing!" or "that's terribly not right." While unexpected, these results, when taken together, highlight the importance of collaboration, cooperation, and alliance-building in human-system interaction. It is when

the students pulled away from Stacy and began to treat the system as an "it" rather than a "you" that they learned less. It is when they felt enough rapport with her to tease her that they learned more. Stacy was an early version of a teachable agent, unable to respond in kind. How much more rewarding could the interactions be, from a social and a task perspective, if Stacy could respond.

CONCLUSION: DESIGNING SOCIALLY-EVOCATIVE AI

Some have claimed that digital technologies are spelling the end of conversation and hence of empathy (*inter alia* Turkle, 2015). I believe, on the contrary, that it is not the technologies that hold the power to shut down conversation. It is us. And empathy is born not of speaking to others, but of listening to them. The kind of technological determinism that blames machines for a lessened desire to converse can get in the way of bringing about positive change in the nature of technology design. For example, note that in the case of the two rapport-building virtual peers, Sam and Alex, and the teachable agent Stacy, what is meant by social interaction is not in any way chit-chat or mere small talk that happens around a water cooler, or waiting for a bus. Social interaction in the task contexts I've described involves building a bond around collaboration on that task. Unlike current chat-bots such as Xiaoice (Zhou *et al*, 2019), or smart speakers such as Alexa, the construction of an alliance is something that relies on joint task and social goals, and individual goals that change as a function of one's partner's goals, and not merely talk to pass the time. But in order to understand how to implement machines that can engage in tight collaboration of this sort, we need to do some listening ourselves to understand how it is actually done among humans, and then implement those capacities into machines that are cooperative. Social AI of the kind I've described here can play an important role in discovering this kind of human behavior, and an equally important role in sustaining it in human-machine interaction. This was the goal of Gregory Bateson, Norbert Weiner, Margaret Mead, and the other participants in the Macy Conferences – to build systems that could engage in feedback loops – what we would call today "adaptive systems." Systems that are adaptive not only to the changing environment, but to the task and social needs of their human partners.

ABSTRACT

The article argues for a genre of AI capable of building social bonds with humans. The argument's starting point is the two competing origin stories of Artificial Intelligence. In one, the goal of AI was to create machines that could simulate every aspect of human intelligence. In the other, it was to build machines that adapt closely to natural human behaviour. While the first story is better known, it is argued that the second would have been more fruitful, as it places the human at the heart of the endeavour. Based on this historical perspective, the article provides several examples of conversational agents that engage in this kind of adaptive social behaviour. Results of experiments with these social agents find that they do in fact improve relations between people and the systems. Additionally, they improve performance on the task that the human and the conversational agent are conducting together.

Keywords: Artificial intelligence, social artificial intelligence, human-computer interaction, conversational agents, virtual agents, conversation analysis

BIBLIOGRAPHY

AMBADY N., ROSENTHAL R. (1992), "Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis", *Psychological Bulletin*, n° 111(2), p. 256-274.

ASIMOV I. (1950), *"Runaround". I, Robot*, The Isaac Asimov Collection (ed.), New York City: Doubleday, p. 40.

BISWAS G., SCHWARTZ D., LEELAWONG K., VYE N., TAG-V. (2005), "Learning By Teaching: A New Agent Para-digm for Educational Software", *Applied Artificial Intel-ligence*, vol. 19, p. 363-392.

BRANN N. (1981), *The Abbot Trithemius (1462-1516): The Renaissance of Monastic Humanism,* Vol. 24 of Studies in the history of Christian thought, Leiden: Brill.

CAREY J.W. (1983), "Technology and Ideology: The Case of the Telegraph", *Prospects*, vol. 8, p. 303-325.

CASSELL J (2004), "Towards a Model of Technology and Literacy Development: Story Listening Systems", *Journal of Applied Developmental Psychology*, vol. 25 (1), p. 75-105.

CASSELL J. (2009), Culture as Social Practice: Being Enculturated in Human-Computer Interaction, *in* C. Stephanidis (Ed.) *Proceedings of HCII*, (published as Universal Access in HCI, Part III. Berlin Heidelberg: Springer-Verlag), p. 303-313.

CASSELL J., ANANNY M., BASU A., BICKMORE T., CHONG P., MELLIS D., RYOKAI K., VILHJALMSSON H., SMITH J., YAN H. (2000), "Shared Reality: Physical Collaboration with a Virtual Peer", *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, April 4-9, Amsterdam, NL, p. 259-260.

CASSELL J., CRAMER M. (2007), "Hi Tech or High Risk? Moral Panics about Girls Online", *in* T. MacPherson (Ed.), *Digital Youth, Innovation, and the Unexpected: The MacArthur Foundation Series on Digital Media and Learning*, Cambridge, MA: MIT Press, p. 53-75.

CHASE C., CHIN D., OPPEZZO M., SCHWARTZ D. (2009), "Teachable Agents and the Protégé Effect: Increasing the Effort Towards Learning", *Journal of Science Education and Technology*, vol. 18, p. 334-352.

COHEN S. (1972), *Folk Devils and Moral Panics*, London: MacGibbon and Kee.

COHEN P., KULIK J., KULIK C. (1982), "Educational Outcomes of Tutoring: A Meta-analysis of Findings", *Journal of Education-al Research*, vol. 19(2), p. 237-248.

COONEY J. G. (1974), "Foreword", *in* G. S. Lesser (ed.), *Children and Television: Lessons from Sesame Street*, New York: Vintage Books.

DAVIS R.E. (1976), *Response to innovation: A study of popular argument about new mass media*, New York: Arno Press.

FINKELSTEIN S., YARZEBINSKI E., VAUGHN C., OGAN A., CASSELL J. (2013), "The effects of culturally-congruent educational technologies on student achievement", in *Proceedings of Artificial Intelligence in Education (AIED)*, July 09-13, Memphis, TN.

FLICHY P. (1995), *Dynamics of modern communication. The shaping and impact of new communication technologies*, London Sage.

GATLIN B., WANZE J. (2015), "Relations Among Children's Use of Dialect and Literacy Skills: A Meta-Analysis", *Journal of speech, language, and hearing research : JSLHR*, vol. 58(4), p. 1306-1318, on line : https://doi.org/10.1044/2015_JSLHR-L-14-0311, Retrieved 06/03/2020.

GLADWELL M. (2009), "How underdogs can win", The New Yorker Magazine. *Published in the print edition of the May 11, 2009, issue*. https://www.newyorker.com/magazine/2009/05/11/how-david-beats-goliath, retrieved 10/3/2020.

GONCU A. (1993), "Development of intersubjectivity in the dyadic play of preschoolers", *Early Childhood Research Quarterly*, vol. 8, p. 99-116.

KLINE R. (2011), "Cybernetics, Automata Studies, and the Dartmouth Conference on Artificial Intelligence", in *IEEE Annals of the History of Computing*, vol. 33, n° 4, April, p. 5-16.

LEE C.D. (1997), Bridging Home and School Literacies: Models for Culturally Responsive Teaching, A Case for African American English: Research on Teaching the Communicative and Visual Arts. in *A Handbook for Literacy Educators: Research on Teaching the Communicative and Visual Arts*, Macmillan Publishing Company, New York, p. 334-345.

LENAT D. (1983), "EURISKO: A program that learns new heuristics and domain concepts", *Artificial Intelligence*, vol. 21(1-2), p. 61-98.

MARVIN C. (1988), *When Old Technologies Were New*, New York: Oxford University Press.

MATSUDA N., YARZEBINSKI E., KEISER V., RAIZADA R., STYLIANIDES G., COHEN W. W. (2011), Learning by Teaching SimStudent – An Initial Classroom Baseline Study comparing with Cognitive Tutor, in *Proc AIED 2011*, Springer, p. 213-221.

MBOGHO A., DAVE J., MAKHUBELE K. (2014), Diabetes Advisor – A Medical Expert System for Diabetes Management, in Bissyandé T., van Stam G. (eds) *e-Infrastructure and e-Services for Developing Countries*, AFRICOMM 2013, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 135, Springer, Cham.

MCCARTHY J., MINSKY M., ROCHESTER N., SHANNON C.E. (1955), *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence.* (retrieved from Ray Solomonoff archive: http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf, retrieved 09/03/2020.

NATIONAL CENTER FOR EDUCATION STATISTICS (2019), National Assessment of Educational Progress: an overview of NAEP. [Washington, D.C.]: National Center for Education Statistics, Institute of Education Sciences, U.S. Dept. of Education. NCES 2019-153, https://nces.ed.gov/nationsreportcard/subject/about/pdf/NAEP_Overview_Brochure_2018.pdf, retrieved 09/03/2020.

PINTO RZ., FERREIRA ML., OLIVEIRA VC., FRANCO MR., ADAMS R., MAHER CG., FERREIRA PH. (2012), "Patient-centred communication is associated with positive therapeutic alliance: a systematic review", *J Physiotherapy*, n° 58(2), p. 77-87.

PREECE A. (1992), "Collaborators and critics: The nature and effects of peer interaction on children's conversational narratives", *Journal of Narrative and Life History*, n° 2(3), p. 277-292.

RADER E., ECHELBARGER M. CASSELL J. (2011), "Brick by Brick: Iterating Interventions to Bridge the Achievement Gap with Virtual Peers", *in Proceedings of the CHI'11 Conference*, May 9-12, Vancouver, BC.

ROHRBECK C. A., GINSBURG-BLOCK M. D., FANTUZZO J. W., MILLER T. R. (2003), "Peer-assisted learning interventions with elementary school students: A meta-analytic review", *Journal of Educ. Society*, n° 95(2), p. 240-257.

RUSSELL S. J. (2019), *Human Compatible: Artificial Intelligence and the Problem of Control*, Viking.

SNOW C.E. (1983), "Literacy and Language: Relationships during the Preschool Years", *Harvard Educational Review*, n° 53, 2, p. 165-189.

SPIGEL L. (1992), *Make Room for TV: Television and the Family Ideal in Postwar America*, University of Chicago Press.

STANDAGE T. (1998), *The Victorian Internet: the remarkable story of the telegraph and the nineteenth century's online pioneers*, New York: Walker and Co.

TEALE W.H., SULZBY E. (1986), *Emergent Literacy: Writing and readin*g, Norwood, NJ: Ablex.

TURKLE S. (2011), Alone together: why we expect more from technology and less from each other, New York: Basic Books.

TURKLE S. (2015), *Reclaiming Conversation: The Power of Talk in a Digital Age*, Penguin Press.

TURNER F. (2013), *The Democratic Surround: Multimedia and American Liberalism from World War II to the Psychedelic Sixties*, Chicago: University of Chicago Press.

WALT K. (1961), Pogo Primer for parents (TV Division). Children's Bureau Headliner Series, number 2. US Dept. of Health Education and Welfare, Social Security Administration, Children's Bureau.

WALTER BARUCH D. (1998) [1949], The Play's the Thing, *in* H.Jenkins, (Ed.). (1998), *The Children's Culture Reader*, NYU Press, p. 493-495.

WARTELLA E., REEVES B. (1985), Historical trends in research on children and the media:1900-1960, *Journal of Communication*, n° 35, p. 118-33.

WEIZENBAUM J. (1976), *Computer Power and Human Reason: From Judgement to Calculation*, New York: Freeman and Co.

WIENER N. (1948), *Cybernetics, or, Control and communication in the animal and the machine*, Cambridge MA: Technology Press.

WILEY T., LUKES M. (1996), English-Only and Standard English Ideologies in the U.S, *TESOL Quarterly*, n° 30(3), p. 511-535.

ZHOU L., GAO J., LI D., SHUM H.-Y. (2019), "The Design and Implementation of XiaoIce, an Empathetic Social Chatbot", ArXiv:1812.08989 [Cs], September 14, 2019. http://arxiv.org/abs/1812.08989, retrieved 09/03/2020.