pre-publication of a chapter in Emotions, Technology, Design and Learning. S. Y. Tettegah & M. Gartmeier (eds). ISBN: 978-0-12-801856-9. Elsevier, New York, 2016.

Engagement, Emotions and Relationship: On Building Intelligent Agents

Candace L. Sidner
Worcester Polytechnic Institute
candy.sidner@alum.mit.edu
617-332-4192

Keywords: engagement, emotions, dialogue, collaboration, relationships, intelligent agents, robots, social interaction

Summary
This article reports on research in the development of collaborative intelligent agents. The paper argues that intelligent agents, both virtual and robotic, must know about more than just the task they are performing during collaboration with a human. They must also be capable of managing engagement with their human partners, expressing affect as it pertains to the collaboration, and modeling the relationship the agent develops with the user over time. The article draws on examples of human-to-human collaborations, to illustrate affective expressions play a role in the outcome of the collaboration. It also presents numerous examples of research with agents that include engagement interaction during dialog, model relationship while being collaborators with human users.

Introduction

The Collagen Collaboration project research has centered on the nature of collaboration between people, where collaboration is defined as two or more people agreeing on a shared goal and undertaking to perform it. The phenomenon of collaboration is valuable because it is inherently fascinating; however, understanding collaboration is essential for building computers that can collaborate with people. Figure 1 illustrates the nature of the collaboration in the Collagen Collaboration Project. The human and the "smiley" agent can collaborate in the same circumstances that two human partners do, that is, they can communicate with each other, can each interact with the shared objects, and each observe what the other is doing with the shared objects.

Figure 1: Collaboration

The study collaboration in this paper differs from the approach of an anthropologist or psychologist; computational models of collaboration benefit from insights from these fields, but they look towards nature of the computation of collaboration. Computation demands designing the algorithms and architectures needed to create an intelligent computer agent that takes the place of a human and performs reasonably well. The agent is not exactly like a human (that is still too difficult), but participates in collaborations without holding up the human, getting in his or her way, or doing odd things that keep the human partner from operating as he or she normally would with a collaborator. Most critically, the agent must know about the task, what it can contribute and how to proceed when it is missing information that is vital to the collaboration.

This paper will focus on collaborations and what is needed to make an intelligent computer agent a useful participant in collaborations. It discusses how non-verbal behavior plays a role, and delves into emotional expression as a non-verbal behavior in collaborations, and to what purposes it can and must serve to collaborators. As the reader shall see, emotional expression provides much more than a signal of the cognitive state of the one who expresses emotions.

Collaboration concerns doing things in the world--for example, fixing something, cooking a meal together, making decisions in a group meeting about a company's products, instructing a person on how to do task or playing a game such as soccer (Tambe, 1997). In all these collaborations people talk, but they also perform actions, sometimes together, such as lifting a heavy object, but just as often, each participant acts on his or her own. Understanding collaboration requires accounting for both the joint and individual behaviors in the collaboration process. Some collaborations involve nothing more than conversation between the collaborators. For example, counseling sessions, lectures on a topic, and even conversations that are just to pass the time of day, are all collaborations. They each have a shared purpose among the collaborators and a known way of getting that purpose to come to pass. Each

participant has certain tasks to do, and they all know what those tasks are and they are committed to them (Cohen & Levesque, 1990).

Though collaborations are diverse in their goals, they have many common properties (Grosz & Sidner, 1986; Grosz and Kraus, 1996): a shared goal, shared beliefs among the actors about the state of the world and how to do accomplish the collaboration, individual intentions to perform actions needed to accomplish the shared goal, intentions that others do their parts, and shared beliefs about their intentions.

The theories of collaboration cited above are computational in nature. They offer constraints that can be expressed in computational terms. These theories can be and have been used to build intelligent computer agents, both virtual and robotic ones that can participate in collaborations with people (generally with one human at a time). Rich et al (2001) provides detail about the collaboration agent Collagen discussed here. Work by Allen et al (2007) reports on an agent that can be taught by a person to explore information on the Internet. These human-computer collaborations use dialogs between the person and the agent during the collaboration. Examples of collaborations built using Collagen, with the accompanying computer interfaces, are shown in figure 2. They include air travel planning, scheduling TV shows, instruction in the operation of a gas turbine engine, assistance with reading and writing email, demonstrating a new laboratory invention, and guidance on the operation of home appliances. These interfaces depict the computer as an agent, shown either as a smiling face, a penguin robot (talking to a person), or with an icon of a human shape (in the home appliance interface). The depiction of the agent makes clear its role in the interaction with the user.

Figure 2: Example Collaborative Agent Interfaces

A related effort with a full-face virtual, collaborative agent, illustrated in figure 3 was developed to counsel people on the adoption of new health related behaviors such as exercising more or eating more fruits and vegetables (Bickmore et al, 2011).

Figure 3: An agent for counseling on health change behaviors

The role of engagement in collaborations

Certain basic properties of interactions, which will be dubbed "engagement properties" must hold for a collaboration to take place and succeed, and these properties can be built into intelligent computer agents when they interact with people. These properties are necessary, but are not sufficient to ensure success in collaboration. Yet without them, the collaboration will either never take place or will fail. The collaboration will not be able to get started, or if it does, the lack of these properties will quickly de-motivate the collaborators from continuing to work together.

When people collaborate together, they converse, both about the collaboration and about whatever else they deem relevant or appropriate to their partners. However, before they even start talking, they must ***engage*** with one another. Here **engage** means to establish a perceived connection between two interactions. Engagement begins with glances to each other, facing one another, and then addressing some dialogue from one to the other, often starting with a greeting. This process, the i**nitiation** of engagement, is fundamental to their collaborative interaction. All through their subsequent collaboration, the participants must **maintain** this perceived connection to one another, and when they decide to terminate their interaction, they must **terminate** this perceived connection (Sidner et al, 2005). Bohus & Horvitz (2009) have developed an agent that can interact with two or more people.

Much of what is necessary to initiate engagement and to maintain it happens without explicit mention of the perceived connection. Participants accomplish engagement by verbal and non-verbal means. The verbal means is conversation, while the non-verbal includes looking, gazing, body stance, and making facial and hand gestures. Non-verbal cues as well as the very fact of verbal exchanges is evidence that the partners are engaged even though these behaviors are not directly discussed nor are they used solely for engagement. All of the non-verbal behaviors are serving double-duty because each of them also allows a person to gather information or to communicate about the state of the conversational situation and their own mental states. Lack of these behaviors, especially looking, glancing and taking part in the conversation or mis-use of these will allow the conversational partner to conclude that the other does not want to continue to the interaction.

To understand these engagement behaviors better, my colleagues and I studied videotapes of people in collaborations. We catalogued four major activities, which are called "connection events," that support the engagement between the participants (Rich et al, 2010). The four types are: directed gaze, mutual facial gaze, adjacency pairs and backchannels. In directed gaze (Kendon, 1967), one person (the initiator) looks and optionally points at some object or group of objects in the immediate environment, following which the other person (the responder) looks at the same object(s). Mutual facial gaze (Argyle and Cook, 1976) occurs when one person (the initiator) looks directly at the face of the other person (the responder), and the responder returns his or her gaze to the initiator. An adjacency pair is a two-part exchange in which the second utterance is functionally dependent on the first, as exhibited in conventional greetings, invitations, and requests (Schegoloff, 2007. During an communication from an initiator to a responder, a backchannel (Yngve, 1970) is an event in which the responder directs a brief verbal or gestural communication back to the initiator. Typical examples of backchannels are nods and/or saying “uh, huh.”

Using a formal description of these connection events that indicates what each person does over a given time period, intelligent agents can keep track of

connection events in conversation and to assess both its own level of engagement and that of its human partner during collaborations. A robot was programmed to interpret connection events from a human partner and to generate appropriate ones when making a tangram (a puzzle figure) together with a person (Holroyd et al, 2011). The robot had a set of policies for generating connection events, which it uses to decide when to look at the human partner, when to look and point at the puzzle pieces, which referential description to use for the puzzle piece ("the pink triangle" versus "that one") and when to track the human partner's hands because he or she is pointing to a puzzle piece. The robot also had an algorithm for recognizing connection events generated by its human partner. The picture in Figure 4 illustrates the robot pointing and looking at one of the puzzle pieces with a human partner.

Figure 4: Putting a Tangram together

Does the inclusion of engagement connection events make a difference in collaborations? Three of our studies indicate that it does. In the first study, people interacted with the penguin robot in Figure 5, as it collaborated with a person to demonstrate a laboratory invention [Sidner et al, 2005]. The collaboration was required because the penguin did not have “hands” with which to maniplulate the glass on the table surface to show to the user. One version of the robot produced no looking gestures at all during the conversational interaction while the robot with normal looking gestures (called the moving robot) held the same kind of conversation. Thus the non-moving robot never looked at the objects that it was demonstrating to the user, nor did it look at objects when the person made use of them, but the moving robot did all these things as well as made gazes at the person's face and tracked it as the person moved about. Not only did people report that the moving robot had more appropriate gestures, but they held significantly longer conversations with the moving robot and looked back at it significantly more often.

Figure 5: Mel the robotic penguin

In a second study (Holroyd et al, 2011), with the robot that collaborated on tangram puzzles, people interacted with either a robot that produced connection events or one that did not. People offered significantly higher agreement to statements that the robot looked at the person in a natural and appropriate way, and that the person could easily tell what object the robot pointed at and looked at. A third study (O’Brian et al, 2011) discovered that a robot that looked at its human partner when conversing and pointed appropriately during collaboration in comparison with one that did not received much higher approval as a collaborator from its human partners.

Affective expression

Of all of the connection events, the one that is the most relevant for the remainder of this paper is mutual facial gaze. When individuals gaze at another's face, they each gain information not only about his or her attention to their partner, but also about his or her affect, simply because faces express emotions. Simply said, the face is an important source of this information. How well individuals can read facial expressions of emotion may vary. Furthermore, individuals choose to display more or less of their emotion state in their faces (as well as their bodies), depending on their situation. Individuals can also read the faces of others when there is no mutual facial gaze, but in mutual facial gaze, each agent is aware of the face providing that information.

Affect in the face, as well as the body, has been understood by many scientists (for example, Ekman (1993) and Scherer (1986b)) as reflecting the internal emotional state of a person. This paper explores the need for creating intelligent agents that can express affect and read it in its human counterparts. This section explores the role that affect seems to play in human collaboration with an eye towards creating agents that can use this information and produce it as well. The term "affect" or "affective expression" will be used to discuss emotional displays because "emotion" connotes high intensity displays, whereas the full range of facial displays includes many that do not appear as intense. Indeed facial displays are often subtle, even when the underlying emotional state of the expressor is intense.

It has long been understood that affect plays an important role in social interaction. Matsumoto et al (2010) argue that evolutionist approaches to emotions (beginning with Darwin) hold that facial emotions have an important social function, especially in solving social problems. Facial affect expression provides information about the expressor's affect as well as about his or her relationship with person receiving the affective display. It is particularly important to gain insight into the ways in which facial expression as well as other affective devices of the body contribute to the collaboration process between two or more people.

Faces alone do not give all the critical information about affect. The tenor of the voice, including pitch, duration, and voice quality, also provide information about affect (Scherer, 1986a). Faces and voices (and bodies) work together to provide the information that humans use to discern a person's affective state. Furthermore, people are better able to determine affective state when the face and voice are congruent (Jaywant & Pell, 2012). Creating computer agents that can do likewise is an active area of research. Producing spoken utterance with affect has been an active area of research since Cahn's work (1990).

A great deal of scientific literature has been directed at understanding what affective expressions people can recognize (Ekman, 1993), what muscles of the face are used in which emotions (see (Matsumoto et al, 2010) for an extensive list), and the dynamics of facial expression (Sherer and Ellgring, 2007; Basilli, 1979). These efforts have inspired computational scientists to develop intelligent agents that can

use their faces to convey affect (using when possible the same kind of simulated "muscles" that human faces do), and reproducing the dynamics.

Creating intelligent agents that can discern a human's affective state and produce affect in the face, voice and body is still a work in progress. Researchers have been creating virtual agents that can reliably produce affective expressions through the face, the body and the voice (Pelachaud, 2009; Wagner et al, 2011 ; Gratch et al, 2009). Expressive robot faces, which may or may not bear a close resemblance to a human face are also being explored. Figure 6 presents the faces for the commercial iCat robot (Breeman et al, 2005), and for the robot head Einstein created by Hanson Robotics (http://www.youtube.com/watch?v=vx35zMyFJ94).

Figure 6: The iCat and Einstein robots

Some researchers have attempted to program intelligent agents to have affective internal state, notably the work of Gratch and Marsella (Gratch et al, 2009) using appraisal theory (Sherer et al, 2001). The ability of an intelligent agent to discern affect is more difficult, due to the challenges of visual processing to recognize affect, although research is also taking place on this front (see (el-Kaliouby & Robinson, 2005; Malatesta et al, 2009; Tscherepanow et al, 20009; Schroder at al, 2011)).

Nonetheless, if such capabilities are available, will they be needed? Given an interest in collaboration, one must ask: what role does affect play for collaborators on shared goals and tasks? One way to answer this question is to ask and answer a related one: Do human collaborators express affect in their faces, with their bodies and their voices? If they do, when and why?

Affective Expression and Collaboration

Our pre-scientific, everyday experiences make clear to us that we have all been in collaborations where affective displays occur. We are building something together, and it falls apart. This unexpected event causes us and our partners to at least express surprise. We can imagine just how our face feels: our eyes widen, our brows and eyebrows go up, and perhaps our jaw drops to open our mouths. We may voice a sound of surprise or words that express surprise. When some task we wish to finish fails, we can express our disappointment with some unhappy expression in our faces as well as in our voices. We may frown or look exasperated. We may express our displeasure with sounds or an utterance. If we are called on to do something distasteful, we express our distaste, in mild or extreme form in our faces, voices and bodies. There is not sufficient space in this article to give a catalog or calculus of just what affective displays accompany which circumstances, though many behavioral scientists and physiologists have pursued the problem of how a face shows emotion, and whether those expressions are universal across culture (such as (Ekman & Friesen, 1986) or (Izard, 1971)). Far more remains to be explored on how the mind generates and reflects outward affect, as well as the

effects of individual variation. Whatever further study may reveal, it is clear that when collaborations have difficulties, we express an affective response to it.

What happens when collaborations are not failing, that is, the parts of the collaboration are proceeding smoothly? The collaboration has no distasteful chores or serious disappointments. That is, the collaboration goes on in just the way the collaborators expect. Do people need to take into account affective displays in the normal course of things in collaboration?

This question can be addressed using the common scientific approach for developing new theories and models: studying the data, in this case, human data. Three different datasets of collaborations between people were carefully observed to determine which, if any, affective displays take place and the circumstances in which they do. These datasets were collected for other research projects, but were revisited for this article to determine if and how affect arises during collaborations.

The first of these videos was collected in a quasi-experimental situation in which two people were putting together a porch swing. The participants had their own reasons for doing this task, and their interaction was videotaped, with their permission, in one of their homes. The session lasts about forty-five minutes. The second is a series of observations of pairs of people making canapés in a laboratory setting. One person was asked to teach a second person how to make canapés, and then the "student" taught a third person. Each teaching episode takes about eight minutes, and the pairs know each other and are students at my university. There are three such triples of observations. The last video dataset was also collected in a laboratory setting by videotaping three different groups, each consisting of three people who are in a sales encounter. One person is buying a cellphone, the second is the buyer's friend who is along to give advice, and the third person is the sales person who provides information and advice about which phone to buy. Each session lasts between five and ten minutes. Data from the canapés dataset was used to develop the model of connection events discussed earlier in this paper.

Affective expressions, as one might guess, occur in all these encounters. Every single person pair or triple has occasions on which they show affect. Some of their affective displays result from surprise, disappointment, or embarrassment as discussed above. For example, in the porch swing task, the following exchange occurs as the partners, C and D, finish reading the instructions about how to put the swing together. The last part of the instructions provides the concluding tasks for the swing. When D reads the last instructions, he is expecting them to indicate that the nuts are to be tightened at that time. What he finds is just the opposite.

D reads: Do not tighten nuts at this time.
D says: No more directions.
D laughs and then C laughs.
D says: I think it means **tighten** any loose nuts at this time. D laughs.
C says: Entirely possible.

D and C's initial laughter is an expression of surprise at the final part of the instructions. D's subsequent laugh seems to be amusement. C's comment "Entirely possible" does not respond to amusement affectively, but to the content of D's utterance and conveys agreement on how the last part should be done. Thus affect and task are in this example intertwined.

In one cellphone sales encounter presented below, the salesperson displays embarrassment when the friend (F) asks the salesperson (S) about data plans for cell phones. Because the salesperson has not carefully studied his notes about the cell phones, he does not know how to answer this question, and he stalls by saying "that's a good question." His expressions just after seem to convey embarrassment, and his partners smile, but work hard not to look at him. They are as embarrassed as he is. At the end of this excerpt, he discovers the information that F was looking for.

F says: Um what are the data plans offered for um
S says: Um.
S is smiling.
S says: That's a good question
S then looks at the cell phone for information and finds none.
F and B (the buyer) begin smiling but look at their notes and then the cell phone counter.
B puts her hand over her smiling mouth.
S makes a wry face.
S says: Darn good question.
S smiles then looks at his information sheet.
F says: Okay.
S says: Oh, right then I just. Um the Geophone data plan first year is....

In the reminder of the instances of affect in the three datasets, participants' expressions involve smiling, and their voices often indicate laughter and amusement. In the sales interactions, all three groups, which consist of members who did not know each other before the encounter, begin their interaction with greetings and smiles. Their affective expressions are the familiar ones that accompany greetings and express welcome in the socially conventional way.

In many occurrences of laughter and smiling, amusement provides a social function to create some sense of solidarity (Spencer-Oatey, 1996; Wheeless, 1978), that is, the sense that the participants are in this situation together. For example, in the canapés dialog below, the participants, who are engineering students, make jokes about the term canapés, about what canapés are, and about the use of lace doilies ("fancy orientation") for the plate for serving canapés. In each case, one participant reaches out to the other with a joke about canapés. Their jokes also indirectly serve the collaboration because they provide relevant information for the collaboration. For example, when one participant P1 responds to the jokes about canapés, he

explains what they are and how one dresses up the serving platter (which is referred to by the odd term "fancification"). However, the jokes do more, by making it clear that the participants both know they are involved in an odd activity, for in general, engineering students do not make canapés.

Canapé Dialog
P1: Um, Today we're making canapés. P1 smiles
P2: (smiling) What is a canapé?
P1: Um a canapé is (P1's face changes to a serious look) it's a sort, what you see ah sort of ah upscale function that goes plates and they are sort of like hors d'oeuvres type thing.
P2: (serious face) right.
P1: Um, and so, so they start off with ah, crackers, followed by cream cheese or hummus, um and then they put on a smaller topping, such as pimentos, olives or raisins or something like that. And then they are sort of put onto a plate, in a sort of fancy orientation and they are served for dinner.
P2: What's the other stuff?
P1: Those come in in the fancification of the plate.

Affective expressions accompany what are mildly embarrassing events. In one of the canapés encounters, one participant smiles at his partner when the partner drops an olive while trying to put it on top of the canapé. Then the partner smiles back. On another occasion, the partner cannot open a jar; the instructor participant then grins at her while noticing the error. These encounters are not just to show that one participant looks ridiculous, and the other is amused at the first's expense. The fact that they look at each other and smile conveys that the observer is expressing empathy, and a sense of solidarity at the situation. Empathy also conveys that the observer sees the situation as understandable. That determination is important in maintaining and building trust between the collaborators.

A related purpose of smiling and amusement occurs in one of the sales encounters. After deciding which phone to purchase, the buyer (B) asks the salesperson (S);

B says: well, can I have my phone now? B laughs.
S says: Sure, I'll ring you up. S laughs.
B and S laugh and walk away.

Their shared joke centers on the fact that they aren't in a real store and that no real purchasing is going to happen, but that instead they are all play acting. In effect, they deny that the next task is going to actually occur, and make clear what they are doing together. Their whole exchange again expresses their solidarity in dealing with the situation at hand.

Another example of affective responses occurs in both the canapés tasks and the sales task. When the participants finish, they smile at one another. These behaviors are social, a means of saying "look we're done" and also providing the social need to

acknowledge success of the group. Affective expression works both for task purposes and social ones.

In all of these encounters a large portion of the affective displays represent what can be called "a serious working face." The participants have tasks to do, determining how to build a porch swing from written instructions and then actually doing so, learning how to make canapés and doing so, or gathering information about various cell phones and deciding which best fits the buyer's needs. During these efforts, when not making jokes, looking embarrassed, or the other affective expressions discussed above, the participants do not wear blank expressions on their faces. Their facial expressions reflect their attendance to the tasks at hand and signal that they are working. It seems likely that closer inspection of these periods of expression will reveal not just one kind of expression, but several expressions that change as they move from learning to making something or making a decision. These faces are not the ones normally associated with high intensity affective moments. They do, however, convey to other participants the mental state of the expressive participant, and they convey his or her involvement in the collaboration.

The necessity for the expression of affect and social behavior in collaboration

As the previous discussion of simple and commonplace collaborations illustrates, affective expressions reflect the function of solidarity and trust among collaborators. Affective expressions in response to problems in performing a task are useful in conveying one collaborator's assessment of circumstances of the collaboration. One can ask whether virtual agents and robots should produced and recognize affective expression during collaborations. Further, one can ponder whether social behaviors are necessary to collaboration or just a side benefit to interaction.

Consider the following observations based on human behavior. First, recognizing and expressing affect in response to problems in collaborations contributes to the smoothness of the collaboration and getting the work done efficiently. The use of facial expressions and voice tone (as well as body expressions, which are not discussed here) provide additional ways to communicate about tasks and how they are going. People do not ignore this information or fail to recognize it in their encounters. It tells them how their partners view the current situation, and gives them further insights for their own evaluations. A person who fails to provide such information requires his or her collaborator to either ask directly or to fail to recognize something about the circumstances that the partner has already ascertained. A person who fails to use affective expressions to convey his or her assessment of the situation must instead express that assessment by the semantic content of an utterance. However, early assessments of problems often do not come with a clear sense of what is wrong, just that something is, or that things are not going as expected. A collaborator who does not use the affective mechanisms available must invent language that will do the same. Using language that way is harder, and takes more time than the affective response.

Second, social interaction in collaboration may be necessary for two reasons. One rests on the undeniable fact that people are social beings who have always done their work together in social networks and social groups. The other reason concerns the ongoing requirement for assessing trust and solidarity between collaborators. Concerning humans as social beings, our Western emphasis on thought as our raison d'etre would suggest that humans are foremostly problem solvers, tool makers and manipulators of the all that surrounds them. In fact virtually all human activity takes place in the context of families, tribes, clubs, companies, neighborhoods, and nations, simply said, the in context of social organizations, large and small. That context is easy to overlook because it is so commonplace.

With all those social institutions, come conventions, rules and knowledge for how to pursue work, how to get it done together, and how to come to value the group that has formed for the work. It also provides the knowledge of how to work together. One aspect of that knowledge is the assessment of the other person as a collaborator. Are they trustworthy? Can they be relied on to do their part? Are they "on the same page with me" even if they are not known to me? As the collaboration happens, collaborators assess their partners, both in what work they do and how they accomplish it, but also in the sense of whether they are going to continue to be valuable as partners. Said another way, trust and solidarity are not fixed for all time, but evolve over time. As they evolve, they make it possible for the collaborators to see each other as part of a group, and to make use of whatever social knowledge that group has to support their effectiveness.

Should interactive intelligent agents supply affective responses, and if they do, for what purposes?

Work by Bickmore (Bickmore et al, 2005) has demonstrated that computer agents who express affective faces and who talk with human users about social matters during their collaboration and discussions of exercise tasks create a sense of alliance with their human users. Using standard psychological measures of alliance, Bickmore showed that people who interact with his "social" agents report significantly higher sense of alliance than those who work with agents that show no affect and offer no social talk, and they report even less alliance with computer interfaces that have no agent at all.

That people feel a sense of alliance to computer agents that act socially, strengthens the case for social agents. However, one can ask about the cost to completing the collaboration successfully if agents do not express affect and do not produce and assess social relationship features such as trust and solidarity. It is essentially the same problem as having a human who cannot produce and assess affect and what it means socially. That is, if agents ignore these responses, they are missing out on factors that can influence what the person decides to do next. They miss signals about how the human is challenged or disturbed by collaboration failures, and they

miss unspoken responses to their own difficulties. If agents cannot produce such responses themselves when such events occur, they slow down the collaboration in the same ways that humans would do if they failed to produce them. People will not get the signals of problems or the indications that the agent is going to be a good collaborator. In sum, failing to recognize affective responses and failing to express affect about problems in the collaboration as well as failing to conveying social information makes the agent less able to participate in the full range of factors that humans rely on in making decisions. Furthermore, as intelligent agents clearly are not people, their human partners are likely to see their flawed abilities as evidence of their being stupid and unreliable.

Relationships with Intelligent Agents

So far, the article has been hinting at an underlying but significant aspect of collaboration, namely the role of relationship among collaborators. The role of affective expression in collaboration serves not only the collaboration itself, but the formation and evolution of the collaborators as a group, that is, as having some relationship between the group members. Producing social behaviors, from the most basic ones of being engaged in the interaction, to conveying affective expression, to using social dialog during work together are all capabilities that support the developing relationships between people. The relationships in turn support collaboration by offering each collaborator a sense that the other is going to be team player as the collaboration unfolds.

At present, few intelligent agents have any depth of knowledge about relationships with people. Should they have much at all? Why would it matter to a person if an intelligent agent knew about relationships? If intelligent agents collaborate with people, should they not then know enough about groups to understand what membership in a group entails? However, if a collaboration is short lived, maybe group relationship knowledge is not so vital.

My research colleagues and I are currently investigating intelligent agents that use a basic model relationships and explicitly reason about their evolving relationship with a person. These agents will not have just short-term interactions with users. They will instead have long-term collaborations with people, over a period of months, and be present in people's homes all the time. The agent is "always-on," which means that whenever the agent senses the person's presence, it makes itself available to the person (Sidner et al, 2013). In figure 7, the situation of the older adult and the agent is illustrated, as well as the actual virtual agent after she has greeted her human partner in the morning.

Figure 7: Always-on agents

The target population for interaction with the always-on agent is healthy, isolated older adults who live alone. The agent's tasks for working with these adults are tasks that will provide a bit of day-to-day companionship, as well as support the adult's personal connections to the other people, both in cyberspace and the physical neighborhood of the adult. During a typical day, the agent might talk about the day's weather, or the appointments in the adult's daily calendar. It can play a hand of cards in a social way, with talk about how the game is going as well as other topics, such as how a favorite sports team is doing in the current week. For connecting the adult to others, the agent acts as an exercise coach (using work by Bickmore & Picard (2005) to keep encouraging the adult to exercise. As a side benefit, such walks can get the adult out and about in the local neighborhood. In cyberspace, the agent can provide the adult with the means to connect to family members and old friends via social networking tools such as Google Hangout, a video calling application.

Figure 8 provides a screen capture of the current the virtual agent in a conversation with a person as they play cards, and the figure shows the equivalent robotic agent when it has just provided the user with a list of possible activities to do. The virtual agent has just spoken the words displayed in the thought balloon. The person's choices of what to say in response are shown below the agent in clickable boxes. The state of the card game is shown on the left of the figure.

Figure 8: Always-on virtual and robotic agents

The always-on agent reasons about its relationship with its human partner. It judges when the relationship moves from being one of strangers, to being acquaintances, to being in a relationship that one might call companions (Coon et al, 2013). As strangers, the agent will undertake only the most basic social encounters such as discussing weather or playing games. As the relationship changes, the agent will offer to participate in scheduling events in the user's calendar, telling stories and connecting the adult to friends and family in cyberspace. The agent waits until it determines it has become "companionship familiar" with the person before undertaking to discuss more personal topics, such as exercise or asking the user to tell a personal story.

An always-on agent can be a virtual agent that appears on a computer screen, as shown in Figure 8, but for a small number of homes, the agent will be the small robot with a head and face, shown in Figure 8. The agents will produce some affective expressions. It remains an unresolved challenge for this project to develop algorithms that will determine the types of affect to use and the timing of any affective expressions that always-on agents will display in their daily tasks with their human partner.

Closing thoughts

Intelligent agents will soon be part of our every day lives. While this claims sounds like a bit of imaginary science fiction, this paper has explored some of the research that will make it possible. Agents will understand how to engage with us, how to collaborate with us, how to express and interpret affective expressions, and will think about and form relationships with us.

Acknowledgements

My thanks to my collaborators on much of the work reported in this paper: Charles Rich, Tim Bickmore, Brett Ponsler, Will Coon, Aaron Holroyd, Bahador Nooraei, Dan Schulman, Mohammad Shayganfar, Morteza Behrooz, Lazlo Ring, and Zessie Zhang.

This work is supported in part by the National Science Foundation under awards IIS-0811942 and IIS-1012083. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

Argyle, M., & Cook, M. (1976). *Gaze and Mutual Gaze*. New York: Cambridge University Press.

Allen, J., Chambers, N., Ferguson, G., Galescu, L., Jung, H., Swift, M. & Taysom, W. (2007). PLOW: A Collaborative Task Learning Agent. *Proceedings of the twenty-second AAAI Conference on Artificial Intelligence*. Menlo Park, CA: AAAI Press, pp1514-1519.

Bassili, J.N. (1979). Emotion recognition: The role of facial movement and the relative importance of upper and lower areas of the face. *Personality and Social Psychology*. 37(11): 2049–2058,

Bickmore, T., Caruso,L., K. Clough-Gorr,, K. & Heeren, T. (2005). 'It's just like you talk to a friend' Relational Agents for Older Adults. *Interacting with Computers* 17(6) : 711-735.

Bickmore, T. & Picard, R. (2005). Establishing and Maintaining Long-Term Human-Computer Relationships. *ACM Transactions on Computer Human Interaction (ToCHI)*. 12(2):293-327.

Bickmore, T., Schulman, D. & Sidner, C. (2011). Modeling the Intentional Structure of Health Behavior Change Dialogue. *Journal of Biomedical Informatics*. 44:183–197.

Breeman, A.V., Yan, X., Meerbeek, B. (2005). iCat: an animated user-interface robot with personality. in *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*. New York: ACM Press, pp. 143–144.

Bohus, D. & Horvitz, E. Learning to Predict Engagement with a Spoken Dialog System in Open-World Settings. *Proceedings of SIGDIAL 2009: the 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue*, Stroudsburg, PA: ACL Press, pp. 244–252.

Cahn, J.E. (1990). The Generation of Affect in Synthesized Speech. *Journal of the American Voice I/O Society*, 8: (July) 1-19.

Coon, W., Rich, R. & Sidner, C. (2013). Activity Planning for Long-Term Relationships (extended abstract), *Proceedings of the 2013 International Conference on Intelligent Virtual Agents*. New York: Springer, pp 425-427.

Cohen, P. R. & Levesque, H. (1990). On acting together. In *Proceedings of the Eighth National Conference on Artificial Intelligence*, pg. 94-99, Menlo Park, CA: AAAI Press.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist*. 48(4):384-392.

Ekman, P. & Friesen, W. (1986) A new pan-cultural facial expression of emotion. *Motivation and Emotion*. 10(2):159-168.

el-Kaliouby, R. & Robinson, P. (2005) Real-time Inference of Complex Mental States from Facial Expressions and Head Gestures, in B. Kisačanin, V. Pavlović and T. Huang (eds.) *Real-Time Vision for HCI*. New York: Springer, pg. 181-200.

Gratch, J., Marsella, S., Wang, N. & Stankovic. B. (2009). Assessing the validity of appraisal-based models of emotion. The Third International Conference on Affective Computing and Intelligent Interaction. Amsterdam: IEEE Press, pp 1-8.

Grosz, B.J. & Kraus, S. (1996) Collaborative plans for complex group action. *Artificial Intelligence*. 86(2):269–357.

Grosz, B.J. & Sidner, C. L. (1986) Attention, Intention and the Structure of Discourse. *Computational Linguistics*. 12(3):175-204.

Holroyd, A., Ponsler, P., Rich, C. & Sidner, C. (2011). Generating Connection Events for Human-Robot Collaboration. *Proceedings of Ro-Man 2011, the 20th IEEE International Symposium on Robot and Human Interactive Communication*. New York: IEEE Press, pp 241-246.

Izard, C. E. (1971) *The face of emotion*. New York: Appleton-Century-Crofts.

Jaywant, A. & Pell, M. D. (2012). Categorical processing of negative emotions from speech prosody. *Speech Communication*, 54(1): 1-10.

Kendon, A. (1967). Some functions of gaze direction in two person interaction. *Acta Psychologica*. 26:22–63.

Malatesta, L., Murray, J., Raouzaiou, A., Hiolle, A., Cañamero, L. & Karpouzis, K. (2009). Emotion Modelling and Facial Affect Recognition in Human-Computer and Human-Robot Interaction. In J. Ponce, A. Karahoca (eds.), *State of the Art in Face Recognition*. Rijeka, Croatia: InTech Publishers.

Matsumoto, D., Keltner, D., Shiota, M., O'Sullivan, M. & Frank, M. (2010). Facial expressions of emotion, in *Handbook of Emotions, Third Edition*. M. Lewis, J. M. Haviland-Jones, L.F. Barrett (eds), Guilford Press.

O'Brian, K., Sutherland, J., Rich, C. & Sidner, C. (2011). Collaboration with an autonomous Humanoid Robot: A Little Gesture Goes a Long Way. *Proceedings of the Sixth ACM/IEEE International Conference on Human Robot Interaction*, pp 215-216.

Pelachaud, C. (2009). Modelling Multimodal Expression of Emotion in a Virtual Agent, *Philosophical Transactions of Royal Society B Biological Science*, B364:3539-3548.

Rich, C. ,Holroyd, A., Ponsler, P. & Sidner, C. (2010). Recognizing Engagement in Human-Robot Interaction. Proceedings of the Fifth ACM/IEEE International Conference on Human Robot Interaction. pg. 375-382.

Rich, C., Sidner, C. & Lesh, N.B. (2001). COLLAGEN: Applying Collaborative Discourse Theory to Human-Computer Interaction. *Artificial Intelligence Magazine*, 22(4):15-25.

Schegeloff, E.A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis.* Cambridge, UK: Cambridge University Press.

Scherer, K. R. (1986a). Vocal affect expression: Review and a model for future research. *Psychological Bulletin*, 99:143-165.

Scherer, K. R. (1986b). On the nature and function of emotion: A component process approach. In K. R. Scherer and P. E. Ekman, (eds.), *Approaches to emotion*. Hillsdale, NJ: Erlbaum, pp 293-317.

Scherer, K.R. (2001). Appraisal Considered as a Process of Multilevel Sequential Checking. In Scherer, K .R. , Shorr, A., & Johnstone, T. (Eds.). *Appraisal processes in emotion: theory, methods, research*. Canary, NC: Oxford University Press, pp 92-120.

Scherer, K. R.&, Ellgring, H. (2007). Multimodal Expression of Emotion. *Emotion*, 7(1), 158–171.

Schroder, M., Bevacqua, E., Cowie, R., Eyben, F., Gunes, H., Heylen, D., Maat, M., McKeown, G., Pammi, S., Pantic, M., Pelachaud, ., C., Schuller, B., Sevin, E., Valstar, M.F. & Woellmer, M. (2011). Building Autonomous Sensitive Artificial Listeners. *IEEE Transactions on Affective Computing*. 3(2):165-183.

Sidner, C., Lee, C., Kidd, C., Lesh, N., Rich, C. (2005). Explorations in Engagement for Humans and Robots. *Artificial Intelligence*. 166(1-2): 140-164.

Sidner, C., Bickmore, T., Rich, C., Barry, B., Ring, L. Behrooz, M. & Shayganfar, M., (2013). Demonstration of an Always On Companion for Isolated Older Adults. *The Fourteenth Annual Meeting of the Special Interest Group in Discourse and Dialog*. Stroudsburg, PA: ACL Press, pp. 148-150.

Spencer-Oatey, H. (1996). Reconsidering power and distance. *Journal of Pragmatics*. 26(1):1-24.

Tambe. M (1997). Towards Flexible Teamwork, *Journal of Artificial Intelligence Research*, 7: 83-124.

Tscherepanow, M., Hillebrand, M., Hegel, F., Wrede, B., Kummert, F. (2009). Direct Imitation of Human Facial Expressions by a User-Interface Robot. *9th IEEE-RAS International Conference on Humanoid Robots*. New York: IEEE Press, pg 154 - 160.

Wagner, J., Lingenfelser, F., Bee, N. & André, E. (2011). Social Signal Interpretation (SSI) - A Framework for Real-time Sensing of Affective and Social Signals. *Kunstliche Intelligence*. 25(3):251-256.

Wheeless, L. (1978). A follow-up study of the relationships among trust, disclosure, and inter-personal solidarity. *Human Communication Research*. 4:143–157.

Yngve, V. (1970). On getting a word in edgewise. *Papers from the Sixth Regional Meeting [of the] Chicago Linguistic Society*, Chicago: Chicago Linguistic Society, pg 567-578.