Turing’s Menagerie: Talking Lions, Virtual Bats, Electric Sheep and Analogical Peacocks

Common ground and common interest are necessary components of engagement

Gary McKeown
School of Psychology
Queen’s University Belfast
Belfast, UK
g.mckeown@qub.ac.uk

Abstract—This theoretical paper attempts to define some of the key components and challenges required to create embodied conversational agents that can be genuinely interesting conversational partners. Wittgenstein’s argument concerning talking lions emphasizes the importance of having a shared common ground as a basis for conversational interactions. Virtual bats suggest that for some people at least it is important that there be a feeling of authenticity concerning a subjectively experiencing entity that can convey what it is like to be that entity. Electric sheep reminds us of the importance of empathy in human conversational interaction and that we should provide a full communicative repertoire of both verbal and non-verbal components if we are to create genuinely engaging interactions. Also we may be making the task more difficult than easy if we leave out non-verbal aspects of communication. Finally, analogical peacocks highlights the importance of between minds alignment and establishes a longer term goal of being interesting, creative, and humorous if an embodied conversational is to be truly an engaging conversational partner. Some potential directions and solutions to addressing these issues are suggested.

Keywords—engagement; common ground; subjective experience; empathy; mind-reading display.

I. INTRODUCTION

Turing’s [1] idea of the imitation game—now more commonly referred to as the Turing test—sought to provide a test of what would become known as artificial intelligence—specifically artificial intelligence that mimics human behavior. The test argued that, within the remit of the game, if someone could not tell the difference between a human and a machine interlocutor then we have a reasonable start for assuming that the machine possesses an intelligence of some sort. Another way to think about this would be that the machine could create a conversational interaction with a human that was engaging in the same way that a human–human conversation could be engaging. It also sought to level the playing field between human and machine by ensuring a reasonably fair comparison given the constraints of technology in the time of Turing and what he could imagine was coming. The leveling was to be achieved by ensuring that any interaction was carried out by using a linguistic question and answer session, one that meant that the interlocutors could not see each other—in modern versions of the test this typically uses an instant messaging like scenario. Turing reasoned that this stopped the game from becoming a beauty competition in which a human would win or a competition that favoured the machine in some way, which he likened to a race between a man and an aeroplane. In constraining the test to a non-visual question and answer session he gave a primacy to the linguistic components of human communication and probably unwittingly removed many important non-verbal channels of communication from the classic Turing test; the thought of having computational machines capable of producing human facial expressions amongst other affective cues was at that time a distant dream that perhaps only Turing and a few others could envisage.

The current state of affective computing is such that machines are probably now at a disadvantage by not providing the relevant affective cues; these are the cues that tend to regulate communicative behavior and encourage polite behavior in human-human interactions. It is easier to engage in aggressive questioning and interrogation if you are not aware that you may be hurting your interlocutor’s feelings—in both human-human and human-machine interactions [2], [3]. This paper will argue that for effective engagement between humans and machines to become a reality there must be: an effective common ground and shared common interest between interlocutors; that a prerequisite for engaging conversation, at least for some, is that there must be a convincing conversational partner that can convey what it is like to be that conversational partner; both verbal and non-verbal aspects of human communication must be convincing; and the common ground and common interest must be aligned in ways that create mutual understanding with knowledge of the conversational expectations of the other interlocutor. If conversation is to excel then the ability to violate these conversational expectations in subtle and interesting ways is also required.

II. ARTIFICIAL LISTENERS

The SEMAINE project was a large European Framework 7 Project that sought to build autonomous artificial listeners—these are embodied conversational agents with an emphasis on listening rather than engaging in full conversation. The project achieved some degree of success in these goals [4], [5]. Towards the end of the project there was an evaluation
phase that involved testing the quality of interactions with the Sensitive Artificial Listeners that had been created as part of the project. The SEMAINE system built four embodied conversational agents based around four characters with different personalities [6]. Participants interacted with each of these personalities in turn and had to evaluate the nature of the interaction. About 80 evaluation sessions were conducted in total and a final evaluation experiment involving 30 participants interacting with two versions of the system and the four characters was published [4]. The two versions of the system involved differing levels of emotional and affective signals and the final experiment showed that for some of the characters an increase in the affective signals led to improvements in levels of behavioural and felt engagement (an example interaction can be observed here [7]). This was an extensive exercise in the evaluation of embodied conversational agents and many lessons were learned but as is often the way these lessons were not always reported or published. One example of an unreported but obvious characteristic of the interaction was that there were strong individual differences in the way people interacted with the system. The author of this paper conducted these evaluations and watched each of the sessions as they occurred; individual differences were obvious and some people found it easy to interact with a virtual character while others remained completely silent. However, despite having given the participants an extensive set of psychometric tests to assess which personality factors may be affecting their interactions and levels of engagement no systematic link with any standard or relevant personality variables were found. As there was no systematic link a simple vehicle for reporting these differences was not available.

The nature of the SEMAINE system employed a conversational trick. The instructions told the participants to lead the conversations and the agents were only listeners who would prompt with questions and provide appropriate feedback but they would not drive the conversations. This enabled some participants who were perhaps more skilled conversationalists and unperturbed by talking to a computer to engage in conversation for up to three minutes before an inevitable poor selection of utterance by the SEMAINE system led to conversational breakdown. As a result of this there was an element of luck in the system, if the utterance sequence was such that the participant could repair the conversation and was inclined to do so then conversation could continue. If the utterance was too bizarre early in the conversation then rapport never really got established and the interaction broke down in the early stages. Despite a relatively large number of participants these more or less random issues probably add sufficient noise to the data to mask any systematic personality variable effects that may have been influencing the interactions.

The sections that follow will introduce some philosophical arguments that provide a basis for theoretical considerations when assessing aspects of engagement between humans and machines. They borrow insights from a number of philosophical arguments relevant to human-human and human-machine interactions that happen to have used animals as examples, often as metaphorical substitutes for machines, and hence I have used the label Turing’s menagerie.

III. TALKING LIONS AND VIRTUAL BATS

A. Talking Lions

In Wittgenstein’s famous “later work” Philosophical Investigations [8] he makes the enigmatic comment “If a lion could talk, we could not understand him.” The usual interpretation of this remark is that there is no common ground between the mental states of a lion and the mental states of a human, that is, there is no shared frame of reference between a lion and a human upon which to build a communicative understanding. As humans we have no idea what it is like to be a lion and following Wittgenstein we assume that lions have no understanding of what it is like to be a human and no motivation towards understanding what it would be like to be a human. The world’s of lions and humans are so different that even if the procedural barriers to a communicative interaction were removed there would be little if anything to actually converse about. Earlier in the work Wittgenstein suggests that “the common behaviour of mankind is the system of reference by means of which we interpret an unknown language,” meaning that we require some common understanding to act as a starting place for our interpretations of another’s communicative meaning or intent. It is the commonality of our shared experience with one another that forms the basis of our communicative interactions and therefore to build an engaging communicative system we must be aware of the need for common ground upon which engagement can grow. This is often implicitly assumed but it is worth explicitly acknowledging the importance of this as, although they are distant from us in terms of mental state, lions have somewhat more common ground with humans than most computational machines have with humans.

There are a number of implications of this thought for the affective computing community. A slightly trivial comment is that we may share some aspects of affective communication with a lion while still lacking the ability to verbally interact; a lion’s roar is still very likely to communicate an intention that will motivate most humans with a will to survive into a fast course of action. This reminds us that Wittgenstein was largely embedded in a language based understanding of communication rather than communication in a broader sense; a broader sense that would incorporate signaling that Darwin [9] would have placed within the human communicative realm. However, the main point Wittgenstein is making is still a very important one for affective computing, because if there is little common ground between a lion and a human there is none between the inanimate circuitry of a computer and a human–neither verbal or non-verbal. This presents a major problem for developing engagement between computers and humans all of the processes for signaling that develop engagement–verbal, non-verbal, and procedural–must be provided by the system developer.
B. Virtual Bats

Within the SEMAINE evaluation phase, one of the important aspects that seemed to be driving engagement was the degree to which the participants could suspend their disbelief and place some value in their conversational partner—the embodied conversational agent. This is both interesting from the aspect of those who could and those who could not suspend disbelief. For those that could, they seemed to enter into a game-playing mode in which they would keep a conversation alive assuming that the reacting partner—the agent—had value as a conversational entity. For those that could not suspend disbelief they saw little value in engaging in a conversation with what amounted in their view to a plastic box filled with circuitry, there was no actual thing present that was worthy of being a conversational partner.

This raises the issue that when people interact with one another in human-human conversation the point of having a conversation is assumed and taken for granted. The person with whom they are interacting might be a nice person with whom they may build a friendship, they may be useful as a contact for the future or they may be a potential romantic partner; it may also simply be the right thing to do to interact with another human being in a polite and socially acceptable manner, as that is how you would like most people who do not know you to interact with you in future circumstances. In human-human interactions there is almost always a point to behaving in the appropriate way for the social context—to seem polite and avoid giving offense. The fundamental motivation driving this set of assumptions is that during a conversation you are talking to a conscious human being with thoughts and feelings, a conversational partner, something that is so obvious and taken for granted that we barely notice it. None of these issues can be taken for granted in the developing engagement with a conversational agent.

In another famous philosophical work of the twentieth century Thomas Nagel—this time addressing the issue of what it is like to be conscious—asked “What is it like to be a bat?” [10]. The “bat” argument is concerned with the special nature of being conscious, and Nagel argued that there is something special about consciousness that means we cannot really deal with it in a reductionist way. It is not like other problems in science where reductionism can make substantial inroads to the problem because in the case of consciousness we have a core issue of subjective experience. He used the example of being a bat and started with the assumption that we all believe that as a bat a mammal it has some sort of experience as it goes about its existence. Bats are useful in this thought experiment as they are close to humans in a phylogenetic space but very different in sensory and perceptual apparatus so presumably their experience is very different too. The argument suggests that we must extrapolate from our own experiences to imagine what it is like to be a bat, however, we are fundamentally limited and cannot hope to actually use our imaginations to know what it is to subjectively be a bat and experience existence in the way a bat experiences its existence. The argument is anti-reductionist, claiming that we cannot learn from tearing down to component parts and building up again as the crucial aspect of subject experience is something that is an intrinsic part of the whole functioning system and not something that we will gain special insight into by knowing each of the component parts in more details. There is something in the nature of subjective experience, in being an entity, that is special to the particular nature of the whole system. This is related to Searle’s Chinese Room thought experiment [11], which argues against the idea that we can simply build a machine that functions in the same way as a living entity and expect it to be conscious and have a subjective experience.

I will not go further into the details of Nagel’s argument. However, I wish to suggest the hypothesis that one of the important dimensions experienced in generating engagement with embodied conversational agents was the degree to which participants had an intrinsic leaning towards a Nagelian worldview. In the sense in which for some people there is a belief that there can never be a virtual agent that has something of a subjective experience that is the same as the subjective experience of what it is like to be a human. Participants with a strong Nagelian worldview will always feel as if they are engaging with a box and circuitry and while it might be an interesting oddity to have computer graphics that are in the form of a human there is little point in engaging in a meaningful interaction with these artificial humans as there is no essence there that can experience the conversational interaction. For this subset of people disbelief cannot be suspended. This is to be contrasted with the other end of the dimension which we may label the Turing end for whom imitation is good enough—if it looks like a bat and acts like a bat then it probably is a bat. For this subset of the population disbelief is easier to suspend and given the right set of stimuli engaging interaction with an artificial agent is just a matter of being presented with the correct set of non-verbal stimuli and a sufficiently convincing, an appropriate set of sequential utterances and subsequently the belief that the agent possesses a subjective experience and some sort of theory of mind becomes relatively easy. In these circumstances within the SEMAINE evaluation there were some participants who went so far as to create fantastical backstories for their agents, the agents would be imagined to go back into the box for the weekend and party with the other characters waiting for their next interactions the following week. This Nagelian to Turing dimension seemed to be an important discriminator along which the participants were placed although none of the psychometric tests that we gave to the participants were able to assess it.

If we extend Wittgenstein’s reasoning to our bats it is extremely unlikely that if we were to talk to a bat we would have anything in common with them to talk about. However, if it were possible to find something to talk about our human curiosity is more likely to steer us towards seeking conversation with an actual bat, seeking information about what it is like to be a bat, rather than towards a virtual bat, seeking information about what it is like to be some circuitry and programming that is pretending to be a bat. Here one would imagine there is something important in actually being a bat that is more alluring as a conversational topic, and consequentially as a conversational partner, than in the pretending to be a bat without any history or experience or
illuminating comment about past bat-style adventures other than those programmed by other humans who also have no experience about what it is like to be a bat. I highlight this somewhat ludicrous scenario to illustrate the need for authenticity in the backstory of a conversational partner to make the conversational partner truly interesting to talk to. If the conversational partner has no interesting tales of seeing flying fish skim across the water on far off oceans—and I suspect very few embodied conversational agents do—then interestingness must come from sources other than the agent itself, they must become interesting for reasons other than their personal autobiography. It would be illuminating to attempt to converse with a bat but somewhat less fulfilling to converse with a virtual bat.

IV. EMPATHY AND ELECTRIC SHEEP

A particularly important aspect of engagement with regard to affective computing is evident in the next animal in Turing's menagerie, electric sheep. Philip K. Dick explores the issues of fake and real sentient entities in his science fiction novel “Do Androids Dream of Electric Sheep,” [12] the book is set in a post nuclear war world where most humans have left the planet and receive a humanoid android when they leave Earth to become colonists. Certain humans—those with radiation induced genetic abnormalities—are not allowed to leave Earth. As technology develops the androids become more human like but are not allowed back to Earth as they are required to be slave-like workers for the colonists. When they do escape and go to earth it is the job of the hero of the book “Rick Deckard” to track them down and present them with an empathy test, as the ability to empathize is the key distinguishing feature between androids and humans. One of the key tensions in the book arises due to the android technology catching up with both human and psychometric abilities to detect the differences in the ability to empathize. The best of the androids become difficult to distinguish from the lower ends of the human scale presenting a moral dilemma, as it is acceptable to kill androids but not humans. The electric sheep are part of the story as it has become a moral duty and a point of social status to keep animals as pets to keep their genetic stock alive, but as real animals are expensive many people are forced to use cheaper artificial animals in their stead to retain an element of social status. The book explores what it means to be fake and real in terms of building artificial animate entities. The exploration raises a number of important issues for the current argument.

For the affective computing community one of the important issues is that it is the ability to empathize and show emotions correctly that distinguishes humans from androids. This contrasts with the view more typical in the preceding decades in which Turing was formulating his test that suggested that the most important and distinguishing aspect of humanity was its rationality and intelligence. Turing was creating the ideas for his imitation game in an era that had been dominated by logical positivism, the Vienna circle and Frege, Russell and the young Wittgenstein’s attempts to make explicit the foundations of mathematics and logic. As a result it was probably easy for Turing in his imitation game to exclude the visual and paralinguistic components of human communication—partly for pragmatic reasons and partly as they could be viewed as part of the beauty contest that a computer could not win. For Dick however, empathy was the distinguishing feature of humanity: “empathy, evidently, existed only within the human community, whereas intelligence to some degree could be found throughout every phylum and order including the arachnida.” The equivalent of the Turing test in Dick’s book comes in a variety of versions of the “Voigt-Kampff scale” where various emotionally compromising situations are recounted to the human or android and the correct—socio-normative—physiological and emotional responses to the socially awkward situation are closely observed for cues to genuine feelings within the participant. Clearly Turing places a primacy on the linguistic aspects of human communication and Dick places a primacy on empathy and the non-verbal components of human communication, those that are the primary concern of affective computing and social signal processing research domains. One of the ultimate questions of the book is whether we can feel “empathy toward an artificial construct,” toward “something that only pretends to be alive?” Could we make a connection with and engage as easily with an electric sheep as a real sheep.

Turing [1] did address these issues quoting Jefferson (1949 cited in [1]) “Not until a machine can write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machine equals brain.” Turing also raises a similar point to Nagel and pre-empts Nagel’s argument “the only way by which one could be sure that a machine thinks is to be the machine and to feel oneself thinking.” A point he rejects as too close to solipsistic thinking.

Two important points arise from the realization of the importance of the affective components of human communication in human machine interaction. Far from leveling the playing field the removal of affective communication makes the task of the machine more difficult. One important aspect is that it removes crucial information concerning the “feelings” of a machine; observation of signals associated with feelings have the important consequence of enforcing the socio-normative behaviours of politeness, and minimizing aggressive questioning that may “upset” the machine [2], [3]. A second point is that much of what convinces us of the interlocutor’s genuine nature comes from these affective cues; they provide signs that there is a “genuine feeler” that is being conversed with and is arguably more important than convincing the interlocutor that there is a “genuine knower” that is a common goal of more traditional Turing tests. To be a convincing conversational human one must both show signs of knowing and delivering from a typical body of human knowledge and being able to feel what it is like to be a human.

A final point more clearly focuses on the question of engagement and that is the realization that for some people to become motivated to engage in a conversation with an embodied conversational agent they must be convinced that there is some genuine felt existence occurring otherwise any
conversation that is not purely based on retrieving information will be without purpose.

V. ANALOGICAL PEACOCKS

This question of the purpose of human communication brings us to an important fourth animal—the analogical peacock. McKeown [13] has argued that we need to reassess our assumptions concerning the purpose of human communication. Traditionally there has been an assumption that human communication has been largely instrumental in nature, it exists mostly to pass information to one another for the purposes of problem solving and cooperation—in this view we talk to each other to provide information and be informative about our mental and felt states. This instrumental provision of information argument has been questioned by some who favour more socio-communicative explanations, the act of communicating itself may serve to foster bonds and the actual information itself is less important than the act of engaging in communication [14], [15]. McKeown [13] argues for a different but related explanation, that we principally communicate to display our mind-reading skills. Mind-reading skills show us to be socially and politically astute animals likely to achieve social status and to rise through the social hierarchy and as a consequence get access to better resources and mating opportunities. Therefore it becomes useful to demonstrate our mind-reading skills to make ourselves more attractive as potential mates and allies. We display our mind reading skills by being creative, humorous and verbally fluent as well as empathetic and cooperative—showing that we know people’s expectations and desires. The peacock part of the hypothesis refers to this display element and the use of creativity and humour adds the analogies to create the analogical peacock hypothesis. In addition to display communications there is a second important aspect to human communication in this hypothesis. To become good mind-readers we must spend a large proportion of our existence learning about the minds of other people. We need to align our minds with those of others in our social groups so that we are aware of the contents of their minds when the opportunities for display arise. This makes us intellectually curious about the knowledge in other people’s minds and the things that might interest them; it also makes us emotionally curious about the feelings of others and the things that might move them.

A further prediction of the analogical peacock hypothesis is that we can display our mind-reading ability not only by being creative, humorous, and empathetic; we can also display our mind-reading ability by showing our awareness of the most interesting and relevant information within the minds of our social group cluster at a given moment in time. This awareness of what is currently fashionable, trending and relevant is known as “taste as display” and is a method of elevating social status. If someone is deemed to have good awareness of the current trends they are likely to be a more engaging and sought after conversational partner.

These factors of intellectual and emotional curiosity are the key components of engagement in communication. According to McKeown [13] our need to align our minds to the minds of others in our social group will motivate our conversational engagement in situations where we are likely to learn about what interests others and what “moves” other emotionally. With regard to embodied conversational agents this presents a problem, if we are not convinced that there is an existing being with genuine interests or genuine ability to feel and be moved, and with sufficient common interests, then the agent cannot satisfy either our intellectual curiosity or emotional curiosity. Additionally, if there is no likelihood of becoming friends or a potential mate with a conversational agent, then the social goals that motivate engaging and protracted conversational interactions are unlikely to be met.

VI. SOLUTIONS AND DIRECTIONS

In the short term there are a number of possible solutions to these issues. The harder problems are those of convincing a human interlocutor that a machine is a genuinely knowing and feeling machine. However the combination of these two communicative aspects to provide a fuller repertoire of the human communicative experience is likely to create more realistic scenarios and result in conversational interactions that will observe the conventions of communications between two humans. These are less likely to involve aggressive conversational interactions and to create interactions that are more oriented towards politeness, prosociality, and the repair of conversations that have become awkward or difficult.

Providing scenarios that have some built in utility can also circumvent the issue of a direct generation of intellectual and emotional curiosity from the agent. In the case of intellectual curiosity search engines can be repurposed to provide a bank of relevant knowledge. The normal way of thinking about a search engine is as a provider of information, you provide it with a clue about the information you want and it seeks more similar information. A reframing of the function of a search engine is as an artificial mind-reader; a user provides a topic area and the search engine attempts to read the user’s mind—given the search query as a starting point—and provide them with the most relevant information. This, of course, would require some degree of interpretation to make any relevant information into a conversational utterance but there are many current endeavours to provide these more engaging interfaces to search engine output, such as personal assistants like Apple’s Siri, Google Now and Microsoft’s Cortana.

A further avenue for providing interesting information that might satisfy intellectual curiosity and increase engagement is to get users to pre-specify channels of interests, this is the strategy adopted by online magazines such as Flipboard. Following [13] this approach is most likely to be useful when combined with a convincing argument that the information provided is not only of interest to the interlocutor but also to the social group that they are likely to be curious about, and in which they seek to gain social status. Presentations that take these factors into account are more likely to increase engagement.
A third important source of intellectual curiosity that already combines these factors is social media. Users of Facebook, Twitter and similar platforms have “friended” and “followed” a certain subset of like-minded individuals as they are the most likely to present them with information that will be interesting and in turn that their social group is likely to find interesting. People seek interesting information from similar-minded people and then display their selective tastes by sharing a subset of these interesting pieces as displays that show that they in turn know what their followers will be interested in. Receiving likes and retweets serves as a confirmation of their judgments. Using a list of followers who are generating interesting information that is relevant to a given social grouping presents a shortcut to gathering interesting material. If an embodied conversational agent is to use these social media platforms as a basis for content generation then they may find a ready-made source of engaging material.

Social media also offers the best opportunity to satisfy emotional curiosity. Many shared posts and tweets include videos and images that contain no instrumental-style information but contain items that have strongly affective content, things that are likely to move us in an artistic or narrative way, make us laugh, or contain an emotional reaction in others to some event. These kinds of posts allow us to gather and selectively display our tastes in affective, emotional, and artistic content to our social group. In a similar way to the use of engaging information outlined in the previous paragraph the use of emotional content derived from social media may present opportunities for an embodied conversational agent to present itself as having greater empathy and understanding without having to actually be empathetic and understand.

For an embodied conversational agent these sources provide possible routes to the appearing to be a provider of engaging content that is useful from a social perspective. Using these “tricks” an embodied conversational agent can present the kind of information that one is likely to gain through a useful and engaging social interaction with a friend. These tricks can function in much the same way as the SEMAINE listener utterances to provide engaging, relevant and socially useful material and topics for conversation. They do not make it more likely that an embodied conversational agent will be viewed as an actual existing entity, but combined with the correct affective and verbal interactional skills they may maximize the likelihood of prolonged engagement.

VII. CONCLUSIONS

The four animals of Turing’s menagerie offer insights into four important issues that must be addressed in order to create engaging interactions with embodied conversational agents. Wittgenstein’s talking lions highlight the necessity for common ground between interlocutors in order to create engaging conversations. Virtual bats emphasize the need for at least some interlocutors to possess the belief that there is an experiencing being that is being talked to, that the conversational partner is an existing entity that can convey something of what it is like to be an entity as part of a conversation. The electric sheep remind us of the importance of getting both the affective and linguistic components correct in order to generate a full interaction in the style of human conversation. Finally, analogical peacocks highlight the need to satisfy both intellectual and emotional curiosity to create engaging conversations. They also create an aspiration of adding creative and humorous components to a conversation to make a more engaging conversational mind-reader as a future goal. If these components can be adequately addressed then our embodied conversational agents may appear to be autonomously interacting and worthwhile conversational partners capable of engaging in prolonged conversational interactions.

REFERENCES