Considering Social and Emotional Artificial Intelligence

Marc Schröder[1] and **Gary McKeown**[2]

Abstract.

This paper presents the concepts of social and emotional intelligence as elements of human intelligence that are complementary to the intelligence assessed by the Turing Test. We argue that these elements are gaining importance as human users are increasingly conceptualising machines as social entities. We describe an implementation of Sensitive Artificial Listeners which provides a hands-on example of technology with some emotional and social skills, and discuss first elements of test methodologies for such technology.

1 INTRODUCTION

In his seminal paper, Alan Turing [31] proposed to operationalise the question "Can machines think?" in terms of an "imitation game", a written dialogue between an interrogator and an entity which could be either a human or a machine imitating a human's behaviour. The idea in this "Turing Test" is that the machine can be said to be "intelligent" if the interrogator cannot reliably distinguish the machine from the human based on the written interchange. Turing uses poetry, maths, and chess as examples of intelligent behaviours that could be assessed via the written dialogue, and claims that "the question and answer method seems to be suitable for introducing almost any one of the fields of human endeavour that we wish to include" (p. 435). He proposed to focus on a written interchange so as not to "penalise the machine for its inability to shine in beauty competitions, nor to penalise a man for losing in a race against an aeroplane" (p. 435). In other words, the aim was to allow the human-machine interaction to focus on the relevant aspects of intelligence by leveling the playing field in other respects that are not so relevant for intelligence.

The present position paper argues in favour of a different perspective on intelligence in general and of machine intelligence in particular. Section 2 briefly describes a relevant aspect of human intelligence that is not covered by Turing's method, namely social and emotional intelligence. Section 3 discusses the metaphors used by humans to conceptualise machines, and tries to corroborate the claim that the conceptualisation of machines as "social entities" is becoming increasingly important in the lives of people, thus motivating the need for social and emotional intelligence in machines. Section 4 describes a current endeavour to build a "Sensitive Artificial Listener", a machine that has some emotional intelligence and conversational skills but little else. We conclude with a discussion of possible elements of a test of emotional machine intelligence.

2 SOCIAL AND EMOTIONAL INTELLIGENCE

The concept of intelligence has evolved since 1950. Wechsler [33] states that "intelligence is the aggregate or global capacity of the individual to act purposefully, to think rationally, and to deal effectively with his environment" [33, p. 7]. Out of this generic notion of a *general intelligence*, Turing appears to have focused on the capability to think rationally, the intellectual aspect of intelligence. In the human sciences, concepts have since been developed which are explicitly complementary to this purely rational notion of intelligence; among them are the notions of social intelligence and emotional intelligence. Indeed dual-processing accounts from diverse literatures in cognitive and social psychology posit that rational thought is an architecturally and evolutionarily distinct mechanism sited amongst a much broader system or collection of systems that deal with most of the implicit, nonverbal and emotional aspects of human life and interaction [12]. Although Turing sought to level the playing field by ignoring these aspects there are likely to be important interactions between rational thought and this broad base of complementary mechanisms that could if removed tilt the playing field against a machine. A rationality test without social and emotional context would result in reduced tolerance and opportunity for repair. These are normally offered to humans due to cordiality, etiquette and a desire not to upset other individuals which normally incurs some form of social sanction.

Social intelligence is considered to encompass "our abilities to interpret others' behaviour in terms of mental states (thoughts, intentions, desires and beliefs), to interact both in complex social groups and in close relationships, to empathize with others' states of mind, and to predict how others will feel, think and behave" [1, p. 1891]. Humans can lack social intelligence while having a very high general intelligence, as can be the case for people with autism [1].

The "social brain hypothesis" suggests that the rational intelligence Turing sought evolved as an extension and an ability to manipulate and manoeuvre through the intricacies of social dominance relations and social hierarchies [10]. Almost synonymous with social cognitive abilities is the recognition of a Theory of Mind [23], the ability to read another's mental state and understand another individual as an intentional agent like oneself.

One relevant aspect of social intelligence is the ability to have a conversation with other people. This requires the respect of social conventions – hello and goodbye rituals, appropriate turn-taking, speaking at the right time, with appropriate loudness, choice of words, and gaze behaviour, depending on the situational context, the relation with the interlocutors and many other factors [4].

Emotional intelligence was proposed as "the subset of social intelligence that involves the ability to monitor one's own and others' feelings and emotions, to discriminate among them and to use this information to guide one's thinking and actions" [25, p. 189]. It includes the capability to become aware of, identify and label one's own or another person's affective state; the capability to reason in terms of the appraisals that lead to affective responses, and to predict possible future actions from the affective state; and finally, a set of capabilities related to the regulation of the affective states, be it the

[1] Deutsches Forschungszentrum für Künstliche Intelligenz (DFKI) GmbH, Germany, email: marc.schroeder@dfki.de

[2] Queen's University Belfast, UK, email: g.mckeown@qub.ac.uk

capability to hold back a socially inappropriate own emotion, or to act in a certain way so as to influence the emotions of another person. There is evidence that the capability of feeling one's own emotion is an essential element of a range of seemingly unrelated capabilities. For example, lesions in emotion-related brain regions leave people unable to feel emotions, engage in simple decision-making or make socially appropriate choices [6]. Capgras syndrome disconnects emotional from rational areas leading people to conclude impostors have replaced friends or family [14].

What role should the concepts of social and emotional intelligence play in machine intelligence? The answer to this question naturally depends on the concept of machine intelligence one chooses to adopt, which is related to the perspective on machines in general. From a philosophical point of view, one could ask basic questions such as whether a machine can potentially be conscious. We will approach the topic from a more pragmatic perspective here, and view machines from a utilitarian point of view. From this viewpoint, we can say that a machine is "intelligent" if it is useful, if it is good at its job. Given the fact that humans have created machines to do work for them, one can say that a machine's job in general is to, in one way or another, make the life of humans easier.

The following section discusses metaphors that people seem to use in defining their relation with machines, and how this relates to the utilitarian view on machine intelligence.

3 MACHINES AS SOCIAL ENTITIES

Humans can conceptualise machines in terms of a range of metaphors, including the *tool* and the *social entity* (cf. [24]). Simple machines such as staplers, loudspeakers or vacuum cleaners are likely to be conceptualised as tools, not much different from a hammer. The machine, electronic or not, is perceived as an extended hand or arm [16], as a thing with predictable properties that one can fully control if one has learned how to use it. In the hands of an expert, more complex machines such as mechanical clocks or desktop computers may also be approached from the "tools" metaphor: in operating the device, the human feels in control, and if something doesn't work as expected, the understanding is that there must be something wrong in the mechanism (be it the clock's cog wheels or the computer's programming) which could potentially be fixed.

As the mechanisms of machines become more complex and the causal chains of functioning become opaque humans need to replace simple cause and effect models with more complex mental models of functioning. A natural second metaphor for machines is that of a social entity. Even though people often explicitly deny anthropomorphism regarding machines and computers, they still interact with them as social entities in states of "mindlessness" [19] – states similar to the implicit/automatic systems in dual-processing theories [12]. In fact, Reeves and Nass have shown [24] that people tend to interact with computers, new media and the likes as they do with people: they are polite, behave differently with computers that speak with a male vs. a female voice, they use proximity-regulating behaviour with faces on the screen, and much more. It seems that people apply rules similar to those governing social behaviour when interacting with new technology beyond a certain threshold of complexity.

What determines, for a given machine, whether a given person will view it as a tool or a social entity, whether he or she will "use" or "interact with" the machine? We can only begin to discuss this point. Biocca et al. [3] provide elements of a definition of "social presence", of the sense of "being with" another social entity. According to them, a sense of social presence requires, firstly, a co-presence, where one perceives the other and is perceived by the other; some "psychological involvement" in the sense that the user forms a mental model of the machine as having "some minimal intelligence in its reactions to the environment and the user" [3, p. 7]; and an element of behavioural engagement, including interaction and synchronisation between user and machine. From that point of view, then, the more reactive, interactive, and "intelligent" a machine becomes, the more likely users would perceive it as a social entity.

It seems that users generally prefer to feel "in control" when interacting with machines [18, 7]. For simple and predictable machine behaviour, a sense of control may best be covered by the tool metaphor. Where the machine's complexity exceeds a certain threshold, it is common for users to feel incompetent and to use associated coping strategies [18]. One way to avoid this reaction is to *design* a user interface according to a social entity metaphor that mitigates the information overload [17]; this then raises the expectation that behaviour should be predictable based on social conventions [7]. Whether or not the same mechanism is unconsciously applied by people when dealing with complex machines in general, even machines not designed to behave as a social entity, remains to be investigated.

Society at large is clearly moving towards ever-more complex technology, with ever-fewer people in the position to really understand and control that technology. Technology is becoming more autonomous, from self-updating software programs on the desktop to autonomous vacuum cleaner robots. Often they take over roles previously filled by human service staff, as is the case for train ticket vending machines or airline self check-in terminals. These machines show some awareness of the user, they interact and exhibit some limited intelligence, so they seem to fulfill the minimal criteria for a feeling of social presence. If unpredictable behaviour is added to that, some users may indeed feel that they are interacting with a social entity.

At the same time, the functions filled by some of these machines have a high relevance for their users. If I need my train ticket or boarding card on time, the utilitarian "intelligence" of the machine has a high impact on my well-being. If it doesn't do its "job" right, the machine becomes an obstacle between me and my goal, and anger or a feeling of helplessness will result [26].

This discussion is intended to show that there are reasons to believe people will increasingly use the "social entity" metaphor when interacting with machines. Whether or not they are designed explicitly for that, a machine's actions may be interpreted increasingly as social actions. Human users might, for example, attribute personality traits based on appearance, reaction speed, ease of use, clarity etc. They may apply their own social intelligence in an attempt to infer the machine's state of mind – is it thinking, grumbling, unfriendly, or why is it not responding? The machine may be perceived as arrogant if the user's situation of distress is peacefully ignored by a pre-recorded friendly voice. The machine may be perceived as hectic or distressed if it moves too fast, or as sluggish or bored if it moves too slow, etc. Basically, the expectation is that human users will automatically compare the machine's behaviour with a somehow adapted version of their mental model of other social entities that they know, such as other humans, pets, cartoon characters or similar.

A valuable challenge for computer science, from this point of view, would therefore be to endow machines with the kinds of intelligence they need to be perceived as useful, helpful and intelligent in this utilitarian way, by providing them with the capabilities to perceive, predict and generate socially and emotionally relevant signals.

The following section describes an example of such an endeavour: to provide a machine with the "soft skills" needed to sustain an emotionally-coloured conversation with a human user.

4 SENSITIVE ARTIFICIAL LISTENERS

The European project SEMAINE (www.semaine-project.eu) aims to build a "machine" that possesses some social and emotional intelligence: a Sensitive Artificial Listener (SAL) [9]. A SAL is a multimodal real-time dialogue system with a focus on the "soft skills" required to make the interaction feel natural for the user. Through these skills, the system aims to sustain a conversation with the user for a considerable number of minutes even though its verbal understanding is extremely limited.

A SAL perceives the user's voice and face, analyses these mostly in terms of the user's non-verbal behaviour, and builds an internal model of the current state of the user and the dialogue. In parallel, it updates its own (SAL agent) state, plans multimodal behaviour including multimodal listener feedback and verbal utterances, and generates this behaviour through a 3d head and a synthetic voice.

In the SAL scenario [9], there are four characters, each with a distinct emotionally-defined personality: Poppy is cheerful; Prudence is pragmatic; Spike is aggressive; and Obadiah is gloomy. Each SAL agent's utterances are chosen from a script designed to "drag" the user's emotion towards that of the SAL: Poppy tries to cheer up the user, Spike tries to make them upset, etc. The utterances are chosen as a function of, in particular, the user's current emotion. While the user is speaking, the SAL shows multimodal listener behaviour in real-time which is intended to be consistent with their personality. For example, if the user is talking while a positive-active user emotion is detected, Poppy may give listener feedback signalling agreement and interest, for example by nodding, smiling and saying "yeah!" [21]. Obadiah may react to the same user state either not at all or with a frown and a head shake to signal disagreement.

The SEMAINE-2.0 system is a first complete and fully autonomous implementation of the SAL concept. It is built as a distributed system on top of a middleware and component integration framework based on standard representation formats, the SEMAINE API [27]. OpenSMILE [13] is used for analysing the user's emotion from the voice and for keyword spotting. Facial analysis components [22] determine the user's affective state from the facial points and determine whether the user is nodding or shaking the head. A set of interpreter components [30] consolidate these analyses in context and determine the system's "current best guess" regarding the state of the user and the dialogue. Action proposer components for speaker and listener behaviour [2] continuously update a list of possible actions, and send candidate actions when the dialogue state seems appropriate (for example, listener actions are usually appropriate only while the user has the turn). In particular, verbal utterances are sent when the agent decides to take the turn [29]. An action selection component [8] filters the possibly conflicting requested actions from the different action proposers, to make sure that only a single action is carried out at any given time. Finally, the selected actions are generated using the text-to-speech system MARY [28] and the 3D conversational agent Greta [20], using custom expressive faces and voices.

In its current state, the system is fully functional, but several components are based on preliminary training data or ad hoc rules grounded more in expert intuition than in solid evidence. To remedy that situation, a database of SAL-type human-to-human dialogues [5] has been recorded in high quality [15], simultaneously via several high-resolution video cameras and microphones. Much of the data has been annotated for emotion, epistemic state, and interaction processes [32] and is now being used to improve system components.

The system and the database are publicly available from the project website www.semaine-project.eu. The middleware and most components have been released as open source so that they can be reused as building blocks of other emotion-oriented systems.

A crucial question is to what extent a system like the SEMAINE SAL system is "intelligent", in the sense of the word as defined in Section 2, and how this can be verified. Methods of evaluation of such systems cannot rely on classic tests of intelligence but must be capable of addressing the variety of domains of intelligence so a multidimensional approach should be adopted. Where possible an evaluation should be appropriate to the domain that it seeks to evaluate, there is little value in asking for a highly linguistic feedback of aspects of an interaction that may be implicit and difficult to verbalise. A number of evaluations have been developed within the SEMAINE project to assess interactions with SAL characters [32].

The criteria against which the SAL system's intelligence is to be determined are focused on its emotional competence and its skills to sustain a conversation. Is the SAL character acting naturally (like a human might act) in the conversation? Are its non-verbal and verbal actions plausible in the context where they are generated? To what extent is a good-willed user at a loss on how to continue the conversation? Does the SAL character appear as a consistent personality? Does it seem aware of the user's presence, of the user's actions?

The SEMAINE project assesses the quality of the interaction using both evaluations of recorded interactions and concurrent evaluations embedded in the interaction. Evaluations of recorded interactions do not interfere with the interaction and can create a strong multidimensional picture of where rapport develops or breaks down. Techniques include expert annotation of recordings such as continuous trace style ratings of relevant dimensions or FACS annotation [11], large group evaluations of interactions or character personality and analysis of the emotional verbal content. Embedded techniques aim to assess the global feeling or certain domain specific aspects of an interaction. A "yuck" button approach asks a user to press a button when conversation feels inappropriate or when a system utterance seems particularly incongruous, this requires minimal verbal processing and can be influenced by implicit aspects that might be hard to verbalise. Self-report sub-dialogues built into the scenario, allow interaction assessment between SAL characters without breaking the social interaction to complete a questionnaire.

Completely inappropriate for judging the competences of the SEMAINE system, on the other hand, would be the verbal criteria of the Turing Test. Since the SAL system possesses no world knowledge and only very limited understanding of the user utterances' verbal content, the system would certainly fare very badly under that criterion. Conversely, however, any system based on textual interaction could not even begin to address the criteria by which the SEMAINE system is evaluated.

5 CONCLUSIONS FOR AN EMOTIONAL MACHINE INTELLIGENCE TEST

This paper has pointed out aspects of human intelligence that are complementary to the intelligence concept embodied in the Turing Test, and has argued in favour of their relevance for current and future technology. To the extent that machines are conceptualised by humans as social entities, their effectivity depends on the capability to deal appropriately with social and emotional factors. Machines should be able to interpret a user's behaviour in terms of the user's mental states, predict how the user will feel, think and behave in a given situation, and respect social conventions regarding interaction. Specifically, they should be considerate of the user's emotional state, predict how the emotion may change given relevant events in the

interaction, and verify those predictions based on actual reactions. Machines should be capable of regulating behaviour so as to have a positive effect on the user's emotion.

The competences required for these kinds of capabilities are numerous, and many of them are not sufficiently understood for immediate realisation in technology. The SEMAINE system has made some small steps in that direction, by providing Sensitive Artificial Listeners with some awareness of the user's emotion, and some simple strategies for influencing it while aiming for competence in the social skill of sustaining a conversation.

The consequences of these observations for an emotional machine intelligence test seem to be twofold. First, it will have to deal with real-time, multimodal human-machine interaction, such as a conversation or a joint task. Second, the binary answer envisaged by Turing ("machine or human?") misses many opportunities. Instead, multi-faceted evaluation methods will be required, and the assessment should be a matter of degree. For example, a judge could fill in a form after an interaction, containing questions such as:

- "How appropriate was your interlocutor's eye gaze / turn taking behaviour / ..., on a scale from 1 to 10?"
- "How 'normal' did your interlocutor behave in the conversation?"
- "How much did your interlocutor seem to care about your feelings?", etc.

A practical test of emotional machine intelligence will need to determine suitable questions to assess, and find new ways of avoiding to penalise either humans or machines in real-time interaction.

ACKNOWLEDGEMENTS

The research leading to these results has received funding from the European Community's 7th Framework Program (FP7/2007-2013) under grant agreements 211486 (SEMAINE) and 231287 (SSPNet).

REFERENCES

[1] S. Baron-Cohen, H. A. Ring, S. Wheelwright, E. T. Bullmore, M. J. Brammer, A. Simmons, and S. C. R. Williams, 'Social intelligence in the normal and autistic brain: an fMRI study', *European Journal of Neuroscience*, **11**(6), 1891–1898, (1999).

[2] E. Bevacqua, K. Prepin, E. de Sevin, R. Niewiadomski, and C. Pelachaud, 'Reactive behaviors in SAIBA architecture', in *AAMAS 2009 Workshop Towards a Standard Markup Language for Embodied Dialogue Acts*, pp. 9–12, Budapest, Hungary, (2009).

[3] F. Biocca, J. Burgoon, C. Harms, and M. Stoner, 'Criteria and scope conditions for a theory and measure of social presence', in *Presence 2001*, Philadelphia, (2001).

[4] P. M Brunet, G. McKeown, R. Cowie, H. Donnan, and E. Douglas-Cowie, 'Social signal processing: What are the relevant variables? and in what ways do they relate?', in *Proc. IEEE Intl. Workshop on Social Signal Processing*, Amsterdam, The Netherlands, (2009).

[5] R. Cowie, G. McKeown, and C. Gibney, 'The challenges of dealing with distributed signs of emotion: Theory and empirical evidence', in *Proc. Affective Computing and Intelligent Interaction*, pp. 1–6, Amsterdam, The Netherlands, (2009). doi:10.1109/ACII.2009.5349542.

[6] A. Damasio, *Descartes' Error: Emotion, Reason, and the Human Brain*, Grosset/Putnam, New York, 1994.

[7] K. Dautenhahn, 'Robots as social actors: Aurora and the case of autism', in *Proc. CT99, The Third International Cognitive Technology Conference*, pp. 359–374, San Francisco, (1999).

[8] E. de Sevin and C. Pelachaud, 'Real-Time backchannel selection for ECAs according to users level of interest', in *Intelligent Virtual Agents*, pp. 494–495, (2009). doi:10.1007/978-3-642-04380-2_59.

[9] E. Douglas-Cowie, R. Cowie, C. Cox, N. Amir, and D. Heylen, 'The sensitive artificial listener: an induction technique for generating emotionally coloured conversation', in *LREC2008 Workshop on Corpora for Research on Emotion and Affect*, Marrakech, Morocco, (2008).

[10] R. Dunbar, 'The social brain: Mind, language, and society in evolutionary perspective', *Annual Review of Anthropology*, **32**, 163–181, (2003). doi:10.1146/annurev.anthro.32.061002.093158.

[11] P. Ekman and W. Friesen, *The Facial Action Coding System*, Consulting Psychologists Press, San Francisco, 1978.

[12] J. Evans, 'Dual-processing accounts of reasoning, judgment, and social cognition', *Annual Review of Psychology*, **59**, 255–278, (2008). doi:10.1146/annurev.psych.59.103006.093629.

[13] F. Eyben, M. Wöllmer, and B. Schuller, 'openEAR – Introducing the Munich open-source emotion and affect recognition toolkit', in *Proc. Affective Computing and Intelligent Interaction*, Amsterdam, The Netherlands, (2009). doi:10.1109/ACII.2009.5349350.

[14] W. Hirstein and V. Ramachandran, 'Capgras syndrome: a novel probe for understanding the neural representation of the identity and familiarity of persons.', *Proceedings of the Royal Society B: Biological Sciences*, **264**(1380), 437–444, (1997). PMID: 9107057.

[15] J. Lichtenauer, M. Valstar, J. Shen, and M. Pantic, 'Cost-Effective solution to synchronized Audio-Visual capture using multiple sensors', in *Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, Genova, Italy, (2009). doi:10.1109/AVSS.2009.92.

[16] J. Lockman, 'A Perception-Action perspective on tool use development', *Child Development*, **71**(1), 137–144, (2000). doi:10.1111/1467-8624.00127.

[17] P. Maes, 'Agents that reduce work and information overload', *Commun. ACM*, **37**(7), 30–40, (1994). doi:10.1145/176789.176792.

[18] D. Mick and S. Fournier, 'Paradoxes of technology: Consumer cognizance, emotions, and coping strategies', *Journal of Consumer Research*, **25**(2), 123–143, (1998). doi:10.1086/209531.

[19] Clifford Nass and Youngme Moon, 'Machines and mindlessness: Social responses to computers', *Journal of Social Issues*, **56**(1), 81–103, (2000). doi:10.1111/0022-4537.00153.

[20] R. Niewiadomski, E. Bevacqua, M. Mancini, and C. Pelachaud, 'Greta: an interactive expressive ECA system', in *Proc. 8th AAMAS*, pp. 1399–1400, (2009).

[21] S. Pammi and M. Schröder, 'Annotating meaning of listener vocalizations for speech synthesis', in *Proc. Affective Computing & Intelligent Interaction*, Amsterdam, (2009). doi:10.1109/ACII.2009.5349568.

[22] S. Petridis, H. Gunes, S. Kaltwang, and M. Pantic, 'Static vs. dynamic modeling of human nonverbal behavior from multiple cues and modalities', in *Proceedings of the 2009 international conference on Multimodal interfaces*, pp. 23–30, Cambridge, MA, (2009). doi:10.1145/1647314.1647321.

[23] D. Premack and G. Woodruff, 'Does the chimpanzee have a theory of mind?', *Behavioral and Brain sciences*, **1**(4), 515–526, (1978).

[24] B. Reeves and C. Nass, *The media equation: how people treat computers, television, and new media like real people and places*, Cambridge University Press New York, NY, USA, 1996.

[25] P. Salovey and J. Mayer, 'Emotional intelligence', *Imagination, cognition and personality*, **9**(3), 185–212, (1990). doi:10.2190/DUGG-P24E-52WK-6CDG.

[26] K. R. Scherer, 'Appraisal theory', in *Handbook of Cognition & Emotion*, eds., T. Dalgleish and M. Power, 637–663, John Wiley, (1999).

[27] M. Schröder, 'The SEMAINE API: towards a standards-based framework for building emotion-oriented systems', *Advances in Human-Computer Interaction*, **2010**(319406), (2010). doi:10.1155/2010/319406.

[28] M. Schröder, S. Pammi, and O. Türk, 'Multilingual MARY TTS participation in the blizzard challenge 2009', in *Blizzard Challenge 2009*, Edinburgh, UK, (2009).

[29] M. ter Maat and D. Heylen, 'Turn management or impression management?', in *Intelligent Virtual Agents*, pp. 467–473, (2009). doi:10.1007/978-3-642-04380-2_51.

[30] M. ter Maat and D. Heylen, 'Using context to disambiguate communicative signals', in *Multimodal Signals: Cognitive and Algorithmic Issues*, pp. 67–74. Springer, (2009). doi:10.1007/978-3-642-00525-1_6.

[31] A. M Turing, 'Computing machinery and intelligence', *Mind*, **59**(236), 433–460, (1950).

[32] M. Valstar and G. McKeown. SEMAINE deliverable d7b: Public access to labelled SEMAINE data, 2009. http://tinyurl.com/SemaineD7b.

[33] D. Wechsler, *The measurement and appraisal of adult intelligence*, Williams & Wilkins, Baltimore, MD, USA, 1958.