Human-Agent Interaction: Challenges for Bringing Humans and Agents Together

Rui Prada and Ana Paiva

INESC-ID and Instituto Superior Técnico, Universidade de Lisboa
Av. Prof. Doutor Aníbal Cavaco Silva 2744-016 Porto Salvo, Portugal
http://www.gaips.inesc-id.pt

Abstract. There has been an increasing interest in the area of Human-Agent Interaction (HAI), as demonstrated by the recent organisation of workshops and conferences dedicated solely to this topic (and this workship is a very good example of this). However, as illustrated by the topics in some of these events, there are different visions of what the area of Human-Agent Interaction is. In this paper we argue that while human computer interaction (HCI) studies the interaction between humans and computers, the field of Human-Agent Interaction should be focused on the systems, models, applications and methodologies that allow for humans and "agents" to interact, collaborate, negotiate and even become peers. Differently from HCI in HAI we have "autonomy" on the side of the machines, making the interaction much richer, decentralised and emergent. With this idea in mind, in this paper we try to draw some challenges to the field, which we need to address if we aim at having humans and agents together interacting as peers.

1 Introduction

The world is changing and autonomous technological artefacts are now becoming a reality. Self-driving cars, autonomous vacuum cleaners, nurse robots, or virtual sales-people can now act autonomously to support human activities. Yet, as these new technologies enter our social world, they must be able to interact with humans in a natural way.

There has been a recent interest in the area of Human-Agent Interaction (HAI), as demonstrated by the recent organisation of workshops, such as this particular one [1] and conferences such as the HAI and the HRI, [2] dedicated solely to this topic. Although, at first glance it may appear that this topic is fairly recent in the Intelligent Agents and Multiagent's community, this is not the case, and many avenues have been investigated by researchers over the past twenty years. In fact the Autonomous Agents Conference (one of the predecessors of the AAMAS conference) that started in 1997, had several papers in Human-Agent Interaction, devoting a few sessions to that topic (see for example papers

[1] See for example http://haidm.wordpress.com/

[2] See for example http://hai-conference.net/ihai2013/ or the well know HRI seris in http://humanrobotinteraction.org/

on personality [7], dialogue [14] and language [7], interaction for learning [13], collaboration [12], autonomy [3], believability [6] and methodologies [11] among others). Now, more than fifteen years after that innovative and groundbreaking conference, it seems as though this sub-field has emerged with a new vigour. However, this recent vigour found in the community suggests that the field is not only growing significantly but pehaps re-visiting many of the old questions, and, as such, one needs to reflect upon the visions and challenges that we are facing, which in fact can be inspired from some of such earlier work.

In general, while human computer interaction (HCI) studies the interaction between humans and computers, the field of Human-Agent Interaction should be focused on the systems, models, applications and methodologies that allow for humans and "agents" to interact, collaborate, negotiate and even become peers. Differently from HCI in HAI we have "autonomy", on the side of the machines, making the interaction richer, decentralised and emergent. Nevertheless, HAI has its down side, because as humans interact with agents, expectations are high, making it harder for an agent to keep with the high expectations raised. So, interactions between agents and humans my lead to miscommunication, deadlocks, problems with turn taking and so forth. So, the HAI must consider not only the design of systems where humans interact with agents (be them embodied or not), but also the study of the types of interactions and interaction styles that emerge when humans and agents get together. On the other hand one of the goals of HAI lies on improving the interactions between users and agent-based systems by making those systems more usable and receptive to the user's demands, ways of reasoning, and in general to human specific characteristics. For this reason, this field is highly multidisciplinar relying on expertise and work not only from the autonomous and multiagent systems community but also on work coming from cognitive science, HCI, psychology, behaviour economics, affective computing, and sociology, among others.

The main goal of the area of Human-Agent Interaction is to design and develop mechanisms in agents to promote rich and natural interactions between humans and agents. In this position paper we will discuss two main aspects in the design of HAI: (1) the interaction design, and what aspects of that interaction are shaped by the presence of agents and the role that such agents place in that interaction; and (2) the design of models for the agents to behave in an autonomous and social way. These two aspects combined, are essential for the creation of rich Human-Agent interactions. Then, from these two main departure topics we will present some challenges that we believe our community faces.

2 On Interaction...

Interaction is an ubiquitous concept studied across many different fields (e.g. psychology, computer science, physics). It takes different forms according to the context used, but it has two properties that prevail: interaction involves reciprocal action and mutual influence.

First, interaction engages different parts (actors) into taking actions in a common channel that connects them. Second, the actions taken have some effect on the actors. Furthermore, interaction is, in fact, a process that involves coordinated sequences of actions and effects along time.

An interaction becomes as complex as the number of actors involved, the type and number of actions used, the number of active channels and the time it takes.

To achieve a successful interaction it is important that the actions taken, by all parts, are related to the history of the interaction (e.g. the previous actions). Since interactions can be quite complex it is natural that this property does not always hold. According to Rafaeli [10], the quality of an interaction, which can be framed as its degree of interactivity, has three different levels: (1) it can be *non-interactive* when there is no relation between the actions, (2) it can be *reactive* when the actions are only related to the immediately previous ones, and (3) it can be *interactive* when actions are related to several previous ones and to their relations.

Our goal as designers of systems with humans and agents interacting is, therefore, to create an highly interactive interaction experience for all the actors (in particular for the humans). Hence, we want the actions taken by the actors in a given point of the interaction to be highly related to the previous ones.

To achieve this we have to take into account the fact that people ascribe human qualities, in particular intentionality, to artefacts they interact with [8]. The effect is stronger if the artefacts present autonomous and proactive behaviour, which is, by definition, the case of agents. This means that the actions taken by the agents, must be coherent with the expectations that people have regarding the agents' human-like qualities. Thus, because of this, we should take inspiration from social sciences that study human social interaction when building the agents.

One of the first things to note is that the actors, when engaging in a social interaction, always intent some of the effects of their actions. Hence, an intentional stance is required by part of the agents in order to truly engage in a good interaction. This is reinforced by the fact that people expect it anyway. Another important aspect is the fact that a social interaction is highly subject to the subjectivity of perceptions of the actors involved. The intention of an actor may not be perceived as such. For example, one can make an energetic move towards someone with a friendly intention, which can be perceived as an aggressive move. The interpretation is context dependent and is made according to an expected baseline.

Social actors need to coordinate at different levels in order to engage in a successful social interaction [1]:

- **The content of the interaction**: actors needs to agree on the meaning and the nature of the interaction, which includes: common agreement on the symbols used (e.g. language and topic addressed), the type of activity they are engaging (e.g. completing a task) and rules of interaction (e.g. what actions are acceptable).

- **Dimensions of relationship**: actors need to agree on the mutual expectations of their relationship in the interaction. For example, they need agreement on the *roles* that each actor takes, they need to adapt the type of actions to the levels of *intimacy* they share and they need to agreed on their relative levels of *dominance* (e.g. who takes more initiative and take decisions).
- **Timing**: actors need to coordinate the timing of their actions (e.g. by taking turns), which is is crucial if using speech. Timing is important to avoid periods of inactivity (e.g. long periods of silence) and to avoid interference (e.g. actors interrupting each other). Timing is also part of a general coordination mechanism common in solving collaborative tasks (e.g for joint action).
- **Sequence of behaviour**: the previous items define underlying principles that determine acceptable and expected responses from the actors to certain actions (e.g. norms or rituals). For example, a greeting should be reciprocated, a question should be followed by an answer and a laugh should follow a joke.
- **Non-verbal responsiveness**: the coordination of the interaction at the different levels is maintained by behavioural cues, which are typically non-verbal. These cues involve maintaining distance, posture, gaze, gestures, eye contact and facial expression.
- **Emotional tone**: actors should converge to compatible emotional states. For example, if the interaction is cooperative it is not acceptable for someone to be quite happy while others are unhappy. The emotional tone should also fit to the situation. This leads to empathic and emotion contagion processes among the actors.

In a typical interaction, failures in the coordination of the items described above occur frequently. However, people are quite good at recovering from these situations. An interaction should, therefore, be considered as a continuous process of control and feedback, where actors perform actions, with given intentions, that they monitoring to check if the intended results were achieved.

The main challenge for an agent is not to totally avoid failures in an interaction with humans but to take appropriate measures when the, unavoidable, failures occur.

3 On modeling each other...

Agents need to adapt themselves to the humans interacting with them in order to successfully coordinate at the different levels described in the previous section. The whole interaction process is a dynamic one that relies on the capabilities of the actors to adjust themselves to each other. Yet, it is not only the users/humans that learn and adapt themselves to the agents. Agents also need to learn and adapt to the user, creating a kind of symbiotic relation.

However, to endow our agents with the capability of modeling humans, we need to, first of all, consider *what to model*. Secondly, we need to consider *how to*

perceive the other actors (e.g. the human), *how to interpret* their behaviour, and then *update the created models* in a dynamic manner, responding and adjusting the interaction accordingly.

From a methodological point of view, deciding what to model may be *driven by data*, for example, by observing humans interacting with humans in the same situations as agents will, and then extract the particular features that are essential for achieving the desired interactions. Another approach is *theory driven*, where psychological and sociological theories are used as the ground truth for modeling. Then, agents are endowed with the capabilities to replicate or interpret behaviour in light of the processes described in those theories. For example, if the theory adopted is the prospect theory [5], we may build a model to endow our agents with predictive capabilities about how humans choose between alternatives that involve risk.

One of the challenges of this area is the creation of models that are both data driven and theory driven, taking advantage of the theories that explain human behaviour and processes, and combine them with data that is fine grained enough to allow for parametrization and adjustment of the real-time models created for our agents.

Further, and depending on the application domain or the task, the models created may capture different facets from the human users. In fact, we should rely on work in different research areas that have been investigating how to model users. In particular, we must consider:

Goals and Preferences: For agents to really be able to interact with users, they must learn about the user and adapt to the user's preferences. The area of "user modeling"[3] focuses exactly on that by developing systems that learn the preferences of users, and then provide the right recommendations. Recommender systems try to model the user by learning patterns of preferences and relations in data gathered and adjusting the recommendations.

Expertise and Skills: In learning applications there is the need to model how well the user is doing in a learning task and how far they have achieved in a certain topic. In computer games, in particular, there is a need to model the opponent in order to predict what he/she will do and thus endow the game's NPC (Non-Player Characters) with strategies that make them have a chance to win the game.

Emotions: Emotions colour our senses and dictate our actions. Without emotions, we, humans would not be able to act in a "rational" way [2]. As such, to model agents that interact with humans, not only agents need to be able to perceive the emotions of the humans, but also, take decisions, and simulate decision making taking into account the impact that such emotions have in many of the other cognitive processes. The area of "Affective Computing" [9] deals with these problems not only from a perceptive of sensing, but also modeling and acting in an emotional manner. In Human-Agent interaction we must rely on the recent findings from this community and embrace their models and systems into our agents.

[3] See the journal http://www.umuai.org/ as a good reference of the area.

Decision Making Many researchers are investigating the creation of models that capture human decision making in order to predict what people will do in concrete tasks [4] (such as social dilemmas). Yet, since people do not usually follow fully rational strategies, to build more realistic models of humans we must consider that psychological factors such as empathy, altruism, fairness and others are needed in order to develop a good model of true human behaviour for our agents. The area of "Behaviour Economics" should serve as inspiration for the modeling of these decision making processes.

4 Challenges

Developing agents that are able to successfully interact with people is very challenging. There are still many different problems to address and these will keep researchers and developers busy for many years to come.

We can find (virtual) agents in interactive environments and games, in particular for serious purposes, that support part of the social dimension of the virtual world. There the agents can play different roles and have different functions, such as, guiding, tutoring, supporting and challenging us. In the real world we see a similar although, a bit slower. loneliness social support For example agents that encourage and promote health behaviour. Agents that play games or act in serious games to promote different kinds of behaviours. Affective artificial tutors that need to talk to students and explain content or robots that are used in hospitals to support patients.

The major challenge is to endow these agents with capabilities (social skills) that allow them to take roles that are usually only taken by humans, thus, make them less as tools (or slaves) that people can use and more as peers that people can relate and trust. To achieve this we need to work in different aspects of the problem. In particular, we foresee challenges at four different levels.

Methodological Challenges. Some of the challenges are related to the methodology used in the development of the agents.

- *Develop models of interaction dynamics.* First of all, we need to develop theories and models that take interaction (with humans) as a core concept. Agents cannot focus only on atomic actions or see interaction just as following a fixed protocol. Interaction should be seen as a flexible and dynamic process, and the social effects of the actions must be taken into account. There should be a careful integration between the practical (instrumental) dimension of the interaction and the socio-emotional one.
- *Take a user-centred approach.* We need to involve the human in the loop. This means the we need to take a more user-centred design approach that considers the needs and characteristics of the humans in the development process.
- *Use data, but also theories* There is some debate on the need to use data to guide the creation of agents (the data driven vs theory driven). The main challenge here is to create agent models that are both data driven and theory

driven, taking avantage of the theories that explain human behaviour and processes, and combine them with data that is fine grained enough to allow for parameterization and adjustment of the real-time models created for our agents.
- *Develop methods for assessment.* We need to develop techniques for evaluating and comparing agent-based systems in terms of the interaction they establish with humans. Agents are currently, mostly, evaluated in terms of performance in concrete tasks that they take. However, although, task performance may be important in some cases, the acceptance of users is the ultimate goal. This implies the use of less objective measures that require a Turing test like approach. We argue, however, that we do not need the agent to be fully (undistinguishable) perceived as human. Most important is the fact that the agent is accepted as a social actor and this can happen even if its condition of artificial computational entity is salient. Furthermore, to support the assessment it is essential to define canonical scenarios and test cases (and even competitions) that people can share and, in addition, to develop common and validated methods of measure (such as believability scales).

Situation Awareness Challenges. A major class of challenges is related to giving the ability to the social actors to understand the contextual situation of the interaction and its actors.

- *Develop computational mechanisms to understand others.* As already discussed (in this paper) and recognized by many researchers, a very important part of social skills involves understanding the other social actors. Agents need to model others in order to be able to establish a baseline to predict the effects of their actions (on others). The model of others should address several levels (e.g. intentions, expertise, preferable style of interaction, mood, etc.), but it does not need to be very specific, and precise, in all situations. It can be, for example, based on social categorization and stereotyping.
- *Care for user understanding.* Humans also need to understand the agents. We need to develop interfaces and communication strategies that enable users to model the agents in the very same way we want agents to model users (e.g. intentions, mood, etc.). Particularly, it is important to consider the embodiment of the agent, because it will allow the implementation of non-verbal cues that support interpretation (e.g. gestures, gaze, facial expression). If speech is used, paralinguistic features, such as, intonation, should be considered.
- *Develop computational mechanisms to understand the context.* It is important to be able to understand the broader view of the situation as well. For example, agents need to understand the activity that is framing the interactions, the tasks, goals and role structure that are expected, the cultural norms that are suitable and the kind (and valence) of social relations that are important. The understanding of the situation is, of course, interrelated with the understanding the social actors.

Interaction Dynamics Challenges. Some other challenges are related to the type and characteristics of the interactions that the agents can engage.

- *Engage in long-term interactions.* Keeping the human-agent interaction successful beyond short interactions is still a big challenge. The quality of interaction can, initially, have a boost due to the novelty effect, sustained by people's pleasure of discovery, but will drop significantly, after, if the agent is not able to keep the interest. The effect is, actually, also common in human-human interactions, so it is part of the nature of interaction. To maintain the interaction, both in terms of time and frequency, the agent should be able to remember the user, adapt the style of interaction, develop a relationship (e.g. trust) and pay attention to her needs (e.g. providing emotional support and help on a task). However, it should also share its own interests and needs, thus, for example, requesting help from the user in order to achieve a more symbiotic interaction and relationship.
- *Consider group dynamics.* Most of the research that is already being conducted around some of the challenges we discussed here is focused on dyadic interaction (e.g. the work on embodied conversational agents). However, we should not forget that this is a very particular situation. Most of the social situations engage more than two social actors. Even if the interaction is, for some moments, focused in one-to-one interaction the observers will also interpret the actions and "suffer" their effects. Affiliation to groups, reputation and status dynamics are an important part of social interactions and should be considered.
- *Including the five senses.* In human social interactions the five senses have a role to play, although, it is arguable that taste and smell are always important. Vision and hearing are particular important and have being quite explored in human-agent interaction. But, touch, which is also quite relevant (in particular in certain cultures) not so much. There are, of course, some (important) technology limitations, for example, it is quite difficult for robots to really feel the user and we do not have yet strong computational models (and interfaces) for smell and taste. Nevertheless, we should keep this in our mind. For example, when working with virtual agents we can consider, carefully, the representation of both the user and the agents in the virtual world in order to be able to simulate touch.

Societal Challenges. Finally, that there are some challenges regarding the societal impact of having agents interacting frequently with people in the world.

- *Account for social responsibility.* To be integrated in our society the agents should be social responsible. This means that we need to work on computational morality, thus, creating mechanisms for the agents to deal with moral issues, and reason about moral values and blameworthiness of actions (and effects).

5 Final Discussion

This is indeed a challenging and rich area of research that many researchers have already undertaken some of the challenges that we discuss here. Our idea was not to present an overview of the state of the art, but rather to provide a position paper that gives voice to some of challenges we face in designing and developing successful human-agent interactions. We hope that this discussion will foster deeper reflection regarding the work already performed for the many years, reflecting upon the obvious links with other research areas, and to motivate our community to embrace further developments that will lead agents and humans to co-exist in a symbiotic manner.

6 Acknowledgments

This work was partially supported by national funds through FCT - Fundação para a Ciência e a Tecnologia, under project PEst-OE/EEI/LA0021/2013. It was also partially supported by the European Community (EC) with the EMOTE project (FP7 ICT-317923). The work of Rui Prada was supported by a Grant from The Netherlands Institute for Advanced Study in the Humanities and Social Sciences (NIAS). The authors are solely responsible for the content of this publication. It does not represent the opinion of the EC, the NIAS or the FCT, which are not responsible for any use that might be made of data appearing therein.

References

1. M. Argyle. *Social Interactions*. Transaction Publishers, USA, 2009.
2. A. Damasio. *Descartes' error: Emotion, reason, and the human brain*. Vintage, 2006.
3. B. Friedman and H. Nissenbaum. Software agents and user autonomy. In *Proceedings of the first international conference on Autonomous agents*, pages 466–469. ACM, 1997.
4. Y. Gal, B. Grosz, S. Kraus, A. Pfeffer, and S. Shieber. Agent decision-making in open mixed networks. *Artificial Intelligence*, 174(18):1460–1480, 2010.
5. D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pages 263–291, 1979.
6. J. C. Lester and B. A. Stone. Increasing believability in animated pedagogical agents. In *Proceedings of the first international conference on Autonomous agents*, pages 16–21. ACM, 1997.
7. A. B. Loyall and J. Bates. Personality-rich believable agents that use language. In *Proceedings of the first international conference on Autonomous agents*, pages 106–113. ACM, 1997.
8. C. Nass, J. Steuer, and E. R. Tauber. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 72–78. ACM, 1994.
9. R. W. Picard. *Affective computing*. MIT press, 1997.

10. S. Rafaeli. Interactivity: From new media to communication. *Sage annual review of communication research: Advancing communication science*, 16:110–134, 1988.
11. N. Reilly and W. Scott. A methodology for building believable social agents. In *Proceedings of the first international conference on Autonomous agents*, pages 114–121. ACM, 1997.
12. C. Rich and C. L. Sidner. Collagen: When agents collaborate with people. In *Proceedings of the first international conference on Autonomous agents*, pages 284–291. ACM, 1997.
13. J. Rickel and W. L. Johnson. Integrating pedagogical capabilities in a virtual environment agent. In *Proceedings of the first international conference on Autonomous agents*, pages 30–38. ACM, 1997.
14. M. A. Walker, J. E. Cahn, and S. J. Whittaker. Improvising linguistic style: Social and affective bases for agent personality. In *Proceedings of the first international conference on Autonomous agents*, pages 96–105. ACM, 1997.