The 2016 AAAI Fall Symposium Series:
Artificial Intelligence for Human-Robot Interaction
Technical Report FS-16-01

Psychological Science in HRI: Striving for a More Integrated Field of Research

Patrícia Alves-Oliveira,[1 2] **Dennis Küster,**[3] **Arvid Kappas,**[3] **Ana Paiva**[1 4]
[1]INESC-ID, Lisboa, Portugal
[2] Instituto Universitário de Lisboa (ISCTE-IUL), CIS-IUL, Lisboa, Portugal
[3]Jacobs University, Bremen, Germany
[4]Instituto Superior Técnico, Universidade de Lisboa, Portugal

Abstract

Human-Robot Interaction (HRI) is a highly multidisciplinary endeavor. However, it often still appears to be an effort driven primarily by technical aims and concerns. We outline some of the major challenges for fruitful interdisciplinary collaboration in HRI, arguing for an improved integration of psychology and applied social sciences and their genuine research agendas. Based on our own disciplinary backgrounds, we discuss these issues from vantage points mostly originating in applied engineering and psychology, but also from relevant related fields such as sociology, communication sciences, philosophy, arts, and design. We take a project-case as an example to discuss grounded and practical challenges in HRI research, and to propose how a combination of artificial intelligence advances and a better conceptual definition of the role of social sciences in HRI research may prove to be beneficial. Our goal is to strengthen the impact and effectiveness of social scientists working in HRI, and thereby better prepare the field for future challenges.

Introduction

Although robots have captured our shared collective imagination in science fiction literature and movies since at least the early 1950s (e.g., (Asimov 1950)), their transition into everyday life is a much more recent phenomenon. Today, we can observe a great industrial effort to develop all kinds of robotic and automated systems: from self-driving cars (e.g., Google Self-Driving Car, https://www.google.com/selfdrivingcar/) to drones (e.g., Amazon Prime Air, https://www.amazon.com/b?node=8037720011) and social companion robots (e.g., Jibo https://www.jibo.com/). Outside of the laboratories, social robotics is bound to fundamentally transform how we live and interact with and through technology, with an increasing demand of multidisciplinary attention from the research community. In many ways, the field of HRI appears ready for psychological scientists to have a stronger place in the field by developing work that relates intrinsically with their own research agendas, but still this does not appear to be a common practice.

By definition, HRI is considered a "*field of study dedicated to understanding, designing, and evaluating robotic systems for use by or with humans*" (Goodrich and Schultz 2007). It started as a multidisciplinary field in the mid 1990s and early 2000s, bringing together major categories of academic disciplines, such as applied sciences (computer science, mechanical engineering, robotics, etc), social and psychological sciences (psychology, linguistics, etc), philosophy, and the arts (e.g., designers) (Goodrich and Schultz 2007; Dautenhahn 2007). This multidisciplinary nature of HRI has always been an essential characteristic and driver of the field. Nevertheless, HRI as a whole often appears to be advancing so rapidly at a technical level, that other disciplines often have only a rather limited or constrained time frame to analyze the deeper mechanisms of the human side of HRI, before the technical part advances again. An overarching challenge for psychology in HRI is therefore the question of how to effectively integrate more advanced and psychology-driven research agendas into this context.

Obstacles toward more psychology-oriented HRI research include an array of practical challenges and required tools as well as some more high level structural and political challenges. For example, HRI funding opportunities at the European level still typically tend to relegate psychology into a secondary, supportive role that can help to make the robot seem better – yet often at the risk of sacrificing or postponing psychological research that might be able to reveal more about the psychological underpinnings of human behavior in HRI. In our view, there is little doubt today that psychology can already make significant contributions to HRI, including in the problem definition and requirements specification, which precedes the technical development, by informing design and engineering even before prototypes are created and tested. Yet, for a broader spectrum of psychology to truly thrive in the field, we argue that more could also be done to facilitate psychological research once a given system is fully operational.

Our goal in the present paper is to analyze and discuss the top 3 challenges towards this general aim, starting to debate different research agendas over practical challenges involving design and data analyses in multidisciplinary research projects. For this, we rely on our own disciplinary backgrounds in psychology and applied engineering, and provide grounded examples from our practice using a project-case as a way to sustain our arguments. For each of these challenges, we will provide some suggestions of how psychological work and psychologists' roles might be advanced and improved in the future, making HRI a more integrated multidisciplinary field.

The EMOTE project

EMOTE (http://www.emote-project.eu/) was a recent 3-year multidisciplinary research project aiming to create and develop an autonomous empathic robotic tutor to assist and support 11-14 year old students in geography-curricular topics (Aylett et al. 2014; Castellano et al. 2013). It required intense collaboration between computer sciences, engineering, psychology, and education to achieve its aims. We shall refer to some of these challenges in order to improve the ongoing dialogue between fields, as well as propose possible solutions to some of the problems.

HRI Research Challenges

Psychology and social sciences are much more diverse and varied in expertise, perspectives, and aims than what is currently visible in the HRI community, making the role of these researchers enter in a "gray area" for their own fields. Once more capable social robots start entering the consumer market, we can anticipate a substantial broadening of the HRI "user base" among psychologists and social scientists. This new generation of researchers will pose inherently psychological and social questions, thus they will likely still require a type of infrastructure that is usable by non-engineers. If this can be provided, a larger and more diverse community will be able to ask more refined questions about social issues in HRI, such as how we perceive, operate, interact with, and morally evaluate robots and automated intelligent devices, as well as their actions in the real world.

In this section, we discuss some of the major challenges related to the psychological design, implementation and evaluation of effective HRI experiments, including fundamental differences between disciplinary perspectives and methodological training.

Challenge 1 - Research agendas

Research agendas in HRI differ substantially across disciplines, even if this is not always made explicit. Psychologists and behavioral scientists frequently play a role in validating HR-interactions, including laboratory testing and the evaluation of autonomous HRI systems with human subjects. As such, the "traditional" role of psychologists in HRI is often that of a task force that contributes knowledge on studies design and methodology refinement. In comparison, researchers from applied engineering are more focused on the success of the implementation of a given algorithm and require studies that target engineering-driven questions that translate the results of their implementation. Psychological and behavioral scientists thus place greater emphasis on the human impact of social robots and much less on the creation of new technologies; whereas in philosophy the goal is more oriented towards societal understanding of robots and their place in society; and the arts usually focuses more on the study of robots' aesthetics and materials. While a rich context for research emerges here, the tricky part is that designing HRI studies that fulfill all the research agendas becomes a challenging task.

Furthermore, the methodological training of social scientists has generally been designed to answer social science questions, not engineering questions – and vice versa. In consequence, tools and methods need to be adapted on both sides – ideally with a balanced shared view and understanding of one another's aims. Indeed, the appropriation of methods used in other domains reinforces how recent HRI is as a field of study and translates the lack of specific HRI tools that can be used robustly to infer a deeper understanding of an HR-interaction. For social, psychological and behavioral scientists, philosophers, or artists, more mature tools and stable systems would be invaluable to conduct more advanced domain specific research outside of the relatively narrow spectrum of roles in applied HRI design and evaluation.

Taken together, there is still uncertainty related not only to the actively pursued disciplinary research agendas, but also about the ultimate aims, disciplinary perspectives, and overall potential of psychological sciences agendas that are not focused on the design and evaluation of systems still under construction. Such psychological agendas in HRI are likely to be comparatively more geared towards the human side of HRI involving, e.g., social context, sustained long term effects, and group-based interactions with intelligent, expressive, communicative, and empathic social robots. As obvious as this insight might appear, these differences in ultimate aims imply important distinctions not only in the goals of the research but also in the types of tools and applications that need to be used by researchers. Ideally, psychologists would often envision to work with rather low-tech front-ends that could further be interoperable with a very large number of robotic products (e.g., a *friendly* version of ROS (http://www.ros.org/)) to support flexible experimentation with different types of robotics platforms and conditions.

Challenge 2 - Behavior design of robots

The design of HR-interactions capable of engaging and interacting with users is a very challenging task that can only be successfully addressed via multidisciplinary efforts. Psychologists in HRI often work in this area, including the authoring of effective and ecologically valid behaviors for the robot. In this sense, we are often responsible for basic components of the behavior design, and more generally, for the ability of robots to apropriately communicate with humans in a given context.

Psychology can, in principle, contribute to HRI design and evaluation, with a substantial expertise and body of research knowledge. For example, pertinent psychological research has investigated human perception of smile dynamics, such as "false" vs. "genuine" smiles (Krumhuber and Kappas 2005; Krumhuber et al. 2009; Krumhuber, Kappas, and Manstead 2013) that could also be implemented in suitable social robots capable of exhibiting humanlike facial expressions. However, there are still technical and conceptual interdisciplinary gaps to be bridged. A technical challenge in this case is that of translating expressive behavior to different embodiments. Although some tools already exist toward this purpose (e.g., (Ribeiro and Paiva 2012)), more user friendly interfaces as well as interdisciplinary awareness of such tools would still be required.

To provide a more specific example, the design of behaviors for social robots is a highly complex task, typically involving large and complex strands of interactions and possible branches that need to be anticipated. The empirical basis for this design process is often still relatively thin, and thus requires a lot of creativity and imagination in order to fill in the gaps. For example, in the EMOTE project, the robot was intended to autonomously interact with pairs of children in schools for a period of four sessions (considered a long-term interaction of 30min per session, one session per week). Towards this goal, we created a database of *utterances*, constituted by approximately 1000 utterances. Below is an example of an utterance created for the autonomous robotic tutor of the EMOTE project:

```
  <GAZE(/currentPlayerRole/)>
<ANIMATE(PointPlayer/currentPlayerSide/)>
/currentPlayerName/, it's your turn to play
<GAZE(clicks)>
```

The design of the behaviors of the robot, in this case, included not only verbal behavior content (*i.e., "It's your turn to play"*), but also the non-verbal behavior for the robot in each utterance (*i.e.,* in the example the robot starts talking while symultaneously looking at the current player and performing the animation of pointing to the player to emphasize whom it is addressing). The complex design of all of these behaviors was inspired in educational literature as well as on studies performed in school with real teachers and students interacting in the same educational task. Further, in order to be believable in a given social context, the behavior design of the robot needs to take into account previously performed utterances (*i.e.,* consider memory of previous behaviors), and modify its behavior accordingly.

While this complexity is a low level and fine-grained challenge, it serves as an example for a case in which more mature design tools could be immensely helpful for typical design processes in HRI. It is furthermore an example of a case where interdisciplinary interaction can be very fruitful and lead to cumulative progress. However, since such currently existing tools are out of the scope of traditional psychological and communication research, they are usually not very user-friendly, and risk to become confusing and error prone.

New and re-usable software could significantly enhance this process in the future. For example, such software could facilitate insertion of non-verbal behavior in the design of utterances in a more user-friendly way, e.g., by including emojis in a selected part of the verbal content, thus not hindering the legibility during the design process. Additionally, having to categorize each utterance in a category and subcategory implies hand-work that is time consuming and complex for long-term interactions. A possible solution would be to program software that autonomously attributes or suggests categories and subcategories according to each utterance. As the field progresses, these types of advances in the non-technical aspects of content creation could furthermore open the doors towards more independently managed subprojects in HRI, such as those typically conducted by young researchers in psychology and related disciplines.

Challenge 3 - Data treatment and analysis

Psychologists are trained to design and test experimental studies, a skill that can be very valuable in evaluating user responses in projects such as EMOTE. Psychological sciences researchers can further bring a variety of established tools and measures, help with questionnaires validation or construction of new scales (e.g., the perception of empathy towards and from robots, or psychophysiological measures in the laboratory). There are many other examples of exciting collaboration with psychologists within HRI research, such as studying the role of interindividual differences (e.g., personality), or novel means of measuring human responses in real time (e.g., classification of expressions, or gaze tracking). However, this new type of working conditions also leads to a set of unique new challenges to psychologists.

Even psychologists and psychophysiologists with expertise in highly pertinent fields, such as fine-grained analysis of facial responses (e.g., facial electromyography, or coding of facial action units) have to adapt and refine their methodological compass when the aim is to provide, analyze, and act upon the data in real-time HRI. Suddenly, the familiar dependent variables are no longer an aggregated output towards specific manipulations in ANOVA-type experimental designs – but instead aim to serve as an input to machine learning. The implications of this seemingly subtle shift are manyfold, starting with requirements for the logging and real-time streaming of complex data, and leading up to understanding and responding to constraints introduced by working with machine learning instead of strict testing of psychological theories. At the same time, we need to define an appropriate anchoring of the social context of the interaction to ensure that the data will be interpretable.

This constitutes a novel apparatus of research for psychologists as they have to deal with different types of datasets to treat and analyze an HRI study, such as the logging and annotation of multiple parallel streams of data, including not only dynamic behavioral data but also subjective user responses and system states. A way to counter this effect could be to include user-friendly tools to verify data synchronization and sampling rates across measures. This would not only ease the process of data collection, but also save enormous amounts of time spend on data treatment before the actual data analyses begin.

Questionnaires are arguably still the most frequently used form of assessment in HRI. However, there is still a lack of properly validaded instruments for analysing interactions between humans and robots, wherein the majority of validated questionnaires targets human-human interactions. Thus, the majority of studies adapts questionnaires that have been validated only in another domain, or uses entirely adhoc questionnaires. Apart from addressing this gap in the research by additional validation efforts, video coding of the interaction is often seen as a promising alternative. However, comprehensive manual video annotations tend to be extremely time consuming, in particular in more exploratory studies where the spectrum of potentially relevant behaviors to be coded is yet to be defined. Simultaneously, automatic video analyses (e.g., via Kinect) are limited in scope, agnostic of the context, and often lack reliability. Nevertheless, more intelligent annotation tools could help to import, manage, and filter information from automatic analyses so that they are more human-readable and more easily combined with human annotation efforts. It would also be useful to develop a standardized set of behaviors to be annotated in order to facilitate comparison between studies that perform video analyses as a means to extract results and test theories, by, e.g., creating frameworks of behaviors that can be reused for comparisons across studies.

Summary

In this paper, we have discussed a number of practical and conceptual challenges faced by psychologists and social scientists in HRI. We have argued for a number of possible solutions that begin with a better mutual awareness of implicit disciplinary research agendas and aims. Specifically, we have presented a set of 3 challenges that address research agendas, behavior and study design, up to data analysis of HRI studies. We have proposed a number of concrete tools and requirements that could facilitate the work and engagement of psychologists in HRI, and we have aimed to initiate an informed discussion in the field, about the type of both present and future roles and contributions, that may be expected from psychological and behavioral sciences.

We conclude by suggesting that future research in HRI might soon follow along the footsteps of developments in Human-Computer Interaction (HCI). As can be seen in HCI, psychological and social issues have the potential to play an increasingly more prominent role. In HCI, many essential research tools can already be found in accessible form for a broader audience - e.g., for topics such as *Internet psychology*, or concerning the presence of well established methods (Kim 2012). As a result, HCI has seen a flourishing of many kinds of social sciences research, reflected in numerous handbooks (e.g., (Bryant and Oliver 2009; Joinson 2007), and the rise of new interdisciplinary approaches to social media such as computational social science (Lazer et al. 2009).

As the field of HRI matures further, we envision not only the emergence of more focused sub-fields in HRI but also a more large scale shift in attention toward social topics that will increasingly allow psychological and behavioral sciences to take a more proactive role in defining future HRI research agendas.

Acknowledgments

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013 and by the EU-FP7 project EMOTE under the grant agreement no. 317923. P. Alves-Oliveira acknowledges a FCT grant ref. SFRH/BD/110223/2015. The authors show their gratitude to all the involved schools for taking part of the studies.

References

Asimov, I. 1950. J, robot. *New York: Fawcett*.

Aylett, R.; Barendregt, W.; Castellano, G.; Kappas, A.; Menezes, N.; and Paiva, A. 2014. An embodied empathic tutor. In *2014 AAAI Fall Symposium Series*.

Bryant, J., and Oliver, M. B. 2009. *Media effects: Advances in theory and research*. Routledge.

Castellano, G.; Paiva, A.; Kappas, A.; Aylett, R.; Hastie, H.; Barendregt, W.; Nabais, F.; and Bull, S. 2013. Towards empathic virtual and robotic tutors. In *International Conference on Artificial Intelligence in Education*, 733–736. Springer.

Dautenhahn, K. 2007. Methodology and themes of human-robot interaction: a growing research field. *International Journal of Advanced Robotic Systems*.

Goodrich, M. A., and Schultz, A. C. 2007. Human-robot interaction: a survey. *Foundations and trends in human-computer interaction* 1(3):203–275.

Joinson, A. 2007. *Oxford handbook of internet psychology*. Oxford University Press.

Kim, J. W. 2012. Human computer interaction. *Ahn graphics*.

Krumhuber, E., and Kappas, A. 2005. Moving smiles: The role of dynamic components for the perception of the genuineness of smiles. *Journal of Nonverbal Behavior* 29(1):3–24.

Krumhuber, E.; Manstead, A. S.; Cosker, D.; Marshall, D.; and Rosin, P. L. 2009. Effects of dynamic attributes of smiles in human and synthetic faces: A simulated job interview setting. *Journal of Nonverbal Behavior* 33(1):1–15.

Krumhuber, E. G.; Kappas, A.; and Manstead, A. S. 2013. Effects of dynamic aspects of facial expressions: a review. *Emotion Review* 5(1):41–46.

Lazer, D.; Pentland, A. S.; Adamic, L.; Aral, S.; Barabasi, A. L.; Brewer, D.; Christakis, N.; Contractor, N.; Fowler, J.; Gutmann, M.; et al. 2009. Life in the network: the coming age of computational social science. *Science (New York, NY)* 323(5915):721.

Ribeiro, T., and Paiva, A. 2012. The illusion of robotic life: principles and practices of animation for robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, 383–390. ACM.