

The Embodied Debate Corpus

Proposed corpus recordings of multi-party multi-embodiment multi-measure debates

Merijn Bruijnes
University of Twente
The Netherlands
m.bruijnes@utwente.nl

Sara Falcone
Fondazione Bruno Kessler
Italy
sara.falcone93@gmail.com

KEYWORDS

Corpus Recordings, Conversational Agents, Embodiment Fidelity, Echoborg, Cognitive Pragmatics

ACM Reference Format:

Merijn Bruijnes and Sara Falcone. 2018. The Embodied Debate Corpus: Proposed corpus recordings of multi-party multi-embodiment multi-measure debates. In *Proceedings of 18th ACM International Conference on Intelligent Virtual Agents' workshop on Methodology and/of Evaluation of IVAs (IVA2018 Workshop on Methodology)*. ACM, New York, NY, USA, 3 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

1 INTRODUCTION

In this paper we outline our plans to record a novel corpus that can be used to investigate how the pragmatics of the interaction, particularly the embodiment of the interlocutor, influence the information perceived by a human and her/his experience during the interaction. The purpose of this paper is twofold: 1) discuss our plans with the community and get feedback on this novel methodology, and 2) it serves as a preregistration of the work.

We will create a public corpus that consists of interactions among three interlocutors debating an ethical dilemma. The participants perform specific roles: moderator, proponent, and opponent. The topic of the debate is the *trolley dilemma* as are many of arguments for and against, and there is no correct answer [17]. To investigate the effects of embodiment and agency, we conceive four variants of interactions in the corpus:

- Human-Human-Human (HHH) Corpus: three participants are instructed before interacting on their role in the debate, their interactions are recorded (audio-video), and afterwards they fill out questionnaires about the interaction and the perceptions they have of their interlocutors;
- Agent-Agent-Agent (AAA) Corpus: virtual agents debating the same topic. The behavior of the agents is based on the linguistic information and pragmatics in the HHH corpus;
- Human-Agent-Agent (HAA) Corpus: one participant (human) interacting with two virtual agents. The agents' behavior is based on the HHH corpus. The human has the role of moderator;

- Human-Echoborg-Agent (HEA) Corpus: one participant, one echoborg, and one virtual agent interact. An echoborg is a confederate who acts as the embodiment of a virtual agent: the virtual agent decides what to say, conveys this covertly to the confederate, and he/she utters these words verbatim. The participant is unaware of the nature of the echoborg [7]. The agent's and echoborg's behaviors are based on the HHH corpus;

Each interaction will be recorded (audio-video) and transcribed. After each interaction, each (human) participant fills out questionnaires about the interaction and the perceptions that they have of their interlocutors. This data will be congregated into a corpus that will be made publicly available.

Next we give some background about our rationale and after that, in section 2, we describe the procedure that we will use to collect the data. In section 2.3 we present the questionnaires that we will administer to the participants in order to compare their perception of the interaction. And in section 3 we discuss the methodology.

1.1 Background

The interpretation of a particular utterance can be almost infinitely variable. It depends on factors such as the identity of the speaker, the physical context of use, and the preceding discourse. The contextual flexibility of language (its pragmatics) and the systematization of its structural features are still open issues, particularly their formalization [9]. Cognitive pragmatics is the study of the mental structures and processes involved in the use of language in communicative contexts. Paradigms of cognitive psychology have been applied to study the abilities of humans to go beyond the literal (inference) and derive meaning in relation to context. Contextual manipulations identify an overt issue of research in cognitive pragmatics [3]. The challenge for cognitive pragmatics, as it is for all psycholinguistics, is to construct experiments that help to identify the obligatory processes occurring beneath human awareness [8]. Our work can aid these fields studying and manipulating a particular aspect of the context, namely the embodiment (or appearance) of an interlocutor.

The physical embodiment has a measurable effect on performance and impression of social interactions [21]. There are studies investigating how a different embodiment of an interlocutor can affect the resulting interaction with a human user (e.g. for a humanoid virtual agent [22], robot [13], or avatar [2]). For example, when a human user interacts with a non-human, regardless the embodiment of the agent, s/he will tend to simplify her/his language and to have low expectation about the interaction [1]. One thing is clear, all of these experiments have to deal with an apparent artificial embodiment and this is likely to affect the user's perception of the agent. This is in addition to the artificial nature of the 'mind' of

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

IVA2018 Workshop on Methodology, November 5, 2018, Sydney, Australia

© 2018 Copyright held by the owner/author(s).

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM.

<https://doi.org/10.1145/nnnnnnn.nnnnnnn>

a virtual agent, which is also likely to influence the perception of a user. Yet, without disentangling the artificial embodiment from the artificial mind the contribution of each of these components to the user's perception remains unclear. We propose that the perception of the 'natural' human embodiment of an echoborg does not affect the perception of the artificial mind. Thus, looking at echoborgs will show more clearly to what extent an artificial mind is perceived by a user as 'human-like'. Our corpus will provide researchers a way to analyze and disentangle the effects of artificialness of embodiment and mind.

2 THE METHODOLOGY

In this section we describe our methodology: the planned procedure for the collection of the corpus, the specifics of the procedure for each of the four parts of the corpus, and the measures.

2.1 Trolley Dilemma

The trolley dilemma is a thought experiment from ethics and will be debated by the participants. The task is to find the most ethical solution to the following problem: "*You see a runaway trolley moving toward five tied-up (or otherwise incapacitated) people lying on the tracks. You are standing next to a lever that controls a switch. If you pull the lever, the trolley will be redirected onto a side track and the five people on the main track will be saved. However, there is a single person lying on the side track.*"¹ Two options exist:

- (1) Do nothing and allow the trolley to kill the five people on the main track.
- (2) Pull the lever, diverting the trolley onto the side track where it will kill one person.

The debating participants are instructed to defend one of the options, where the proponent defends acting (switching) and the opponent defends doing nothing.

2.2 The Procedure

The total time of a session is expected to be around 20 minutes (10 minutes for the debate and 10 minutes to fill the consent form and the questionnaire). The proponent and the opponent will sit opposite each other, with the moderator in the middle. For each participant there is a dedicated microphone and a dedicated camera directed at only that participant. This way we can anonymise/delete the data of participants who do not want to share (parts of) their data without having to remove the data of all participants in that session. For each recording where participants are involved the following is procedure is planned:

- **Recruitment:** Participants are recruited from students and staff from the researchers' university (convenience sample) on a voluntary basis.
- **Welcome and briefing:** After welcoming the participant they are briefed: they are explained what will be recorded, why it is recorded, and they are informed about their rights. After informed consent is obtained, we explain that they will take part in a debate about the *Trolley Dilemma* and what their role is during the debate.

- **Recording:** The recording equipment is turned on and synced. The participants have their discussion (max 10 minutes).
- **Questionnaires:** The participants fill out questionnaires. They do this separated from each other.
- **Data collection and reset.**

2.2.1 HHH Corpus. The participants are instructed to freely discuss the trolley dilemma, but to try and keep the conversation clear by speaking one-by-one. The roles are assigned randomly.

2.2.2 AAA Corpus. The arguments that the participants in the HHH corpus use, are extracted and defined in an agent argumentation engine [15]. The system provides a selection of arguments that are possible (appropriate) for each of the agents. One argument is selected and translated into behavior by an intent planner implemented in Flipper 2.0 [20]. The behavior of the agents is planned by the ASAP realizer [19] and displayed by Unity3D embodiments [11]. The agents' interactions will be recorded with screen capture software. From the possible arguments one is selected randomly which means that multiple 'runs' create different interactions. Crucial in this setup is that any of the agents can be replaced by a human, and the other agents still will be able to debate as if there were three agents. This feature will be used in the other corpus variants, so that a participant can interact with the agents.

2.2.3 HAA Corpus. A participant, who has the role of moderator, interacts with two virtual agents. The two agents take the role of proponent and opponent. The participant will fill out the same questionnaires as the participants in the HHH recordings.

2.2.4 HEA Corpus. A participant, who has the role of moderator, interacts with two virtual agents. One virtual agent has a 'normal' virtual agent embodiment, whereas the other has a human confederate as embodiment. This resulting combination of human and machine is called an echoborg [7]. The virtual agent part of the echoborg decides what it wants to say and do, when to do this, and communicates this covertly to the human embodiment over a hidden ear-piece. The human embodiment uses a technique called speech shadowing to repeat the words of the virtual agent verbatim. The participant is not briefed about the nature of the echoborg and is presumed to assume this to be a 'normal' human counterpart in the debate.

After the interaction, the participant *and* the human playing the echoborg embodiment fill out the same questionnaires as the participants in the HHH corpus. The answers by the echoborg offer a unique insight into the 'experiences' that an agent has while interacting with a human.

2.3 Measures

We selected several existing questionnaires that can provide us with information about how the participants experienced the interaction and the perceptions they have of their interlocutors. The first questionnaire is the *Thomas-Kilman conflict mode instrument* that assesses the participant's predisposition to behavior in a conflict situation on two-dimensions (assertiveness and cooperation) [16]. Next, we present a questionnaire which includes the items from the Godspeed Questionnaire series [4], the Game Experience Questionnaire [10], the Intrinsic Motivation Inventory [12], the Self-Assessment

¹From Wikipedia (October, 2018): https://en.wikipedia.org/wiki/Trolley_problem

Manikin [5] and it analyzes factors such as the Co-presence (Self-reported co-presence, Perceived other's co-presence) [14], Rapport [18], Tele-presence [14], Social presence [14] and Trust [6]. These items measure how the participant perceived her/himself during the interaction and how s/he perceived the others.

3 DISCUSSION

We will collect four variants of a corpus that will hopefully help us in understanding the impact of different embodiments on the human perception of the information that is exchanged in a debate and the experience of the interaction. We hope that our efforts will lead to a corpus that is useful for the community. For this reason, we invite the community to give critical and constructive feedback on our design, and come up with additional questions we should (not) ask the participants.

REFERENCES

- [1] René Amalberti, Noëlle Carbonell, and Pierre Falzon. 1993. User representations of computer systems in human-computer speech interaction. *International Journal of Man-Machine Studies* 38, 4 (1993), 547–566.
- [2] Jeremy N Bailenson, Nick Yee, Dan Merget, and Ralph Schroeder. 2006. The effect of behavioral realism and form realism of real-time avatar faces on verbal disclosure, nonverbal disclosure, emotion recognition, and copresence in dyadic interaction. *Presence: Teleoperators and Virtual Environments* 15, 4 (2006), 359–372.
- [3] Lawrence W Barsalou. 1982. Context-independent and context-dependent information in concepts. *Memory & Cognition* 10, 1 (1982), 82–93.
- [4] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. 2009. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics* 1, 1 (2009), 71–81.
- [5] Margaret M Bradley and Peter J Lang. 1994. Measuring emotion: the self-assessment manikin and the semantic differential. *Journal of behavior therapy and experimental psychiatry* 25, 1 (1994), 49–59.
- [6] Ki-Taek Chun and John B Campbell. 1974. Dimensionality of the Rotter interpersonal trust scale. *Psychological Reports* 35, 3 (1974), 1059–1070.
- [7] Kevin Corti and Alex Gillespie. 2015. Offscreen and in the chair next to you: conversational agents speaking through actual human bodies. In *Intelligent Virtual Agents*. Springer, 405–417.
- [8] G Albyn Davis. 2007. Cognitive pragmatics of language disorders in adults. In *Seminars in Speech and Language*, Vol. 28. Copyright© 2007 by Thieme Medical Publishers, Inc., 333 Seventh Avenue, New York, NY 10001, USA., 111–121.
- [9] Noah D Goodman and Michael C Frank. 2016. Pragmatic language interpretation as probabilistic inference. *Trends in Cognitive Sciences* 20, 11 (2016), 818–829.
- [10] WA IJsselstein, YAW De Kort, and Karolien Poels. 2013. The game experience questionnaire. (2013).
- [11] Jan Kolkmeier, Merijn Bruijnes, Dennis Reidsma, and Dirk Heylen. 2017. An asap realizer-unity3d bridge for virtual and mixed reality applications. In *International Conference on Intelligent Virtual Agents*. Springer, 227–230.
- [12] Edward McAuley, Terry Duncan, and Vance V Tammen. 1989. Psychometric properties of the Intrinsic Motivation Inventory in a competitive sport setting: A confirmatory factor analysis. *Research quarterly for exercise and sport* 60, 1 (1989), 48–58.
- [13] Jonathan Mumm and Bilge Mutlu. 2011. Human-robot proxemics: physical and psychological distancing in human-robot interaction. In *Proceedings of the 6th international conference on Human-robot interaction*. ACM, 331–338.
- [14] Kristine L Nowak and Frank Biocca. 2003. The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments* 12, 5 (2003), 481–494.
- [15] Mark Snaith, Harm Op Den Akker, Tessa Beinema, Merijn Bruijnes, Alvaro Fides-Valero, Gerwin Huizing, Reshmashree Kantharaju, Randy Klaassen, Kostas Konsolakis, Dennis Reidsma, and Others. 2018. A demonstration of multi-party dialogue using virtual coaches: the first Council of Coaches demonstrator. In *Computational Models of Argument (COMMA) (Frontiers in Artificial Intelligence and Applications)*, Vol. 305. 473–474.
- [16] Kenneth W Thomas. 2008. Thomas-Kilmann Conflict Mode. *TKI Profile and Interpretive Report* (2008), 1–11.
- [17] Piercarlo Valdesolo and David DeSteno. 2006. Manipulations of emotional context shape moral judgment. *PSYCHOLOGICAL SCIENCE-CAMBRIDGE* 17, 6 (2006), 476.
- [18] Michel Valstar, Jonathan Gratch, Björn Schuller, Fabien Ringeval, Denis Lalanne, Mercedes Torres Torres, Stefan Scherer, Giota Stratou, Roddy Cowie, and Maja Pantic. 2016. Avec 2016: Depression, mood, and emotion recognition workshop and challenge. In *Proceedings of the 6th International Workshop on Audio/Visual Emotion Challenge*. ACM, 3–10.
- [19] Herwin Van Welbergen, Ramin Yaghoubzadeh, and Stefan Kopp. 2014. AsapRealizer 2.0: The next steps in fluent behavior realization for ECAs. In *International Conference on Intelligent Virtual Agents*. Springer, 449–462.
- [20] Jelte vanWaterschoot, Merijn Bruijnes, Jan Flokstra, Dennis Reidsma, Daniel Davison, Mariet Theune, and Dirk Heylen. 2018. Flipper 2.0 - A Pragmatic Dialogue Engine for Embodied Conversational Agents. In *International Conference on Intelligent Virtual Agents*. ACM.
- [21] Joshua Wainer, David J Feil-Seifer, Dylan A Shell, and Maja J Mataric. 2007. Embodiment and human-robot interaction: A task-based perspective. In *Robot and Human interactive Communication, 2007. RO-MAN 2007. The 16th IEEE International Symposium on*. IEEE, 872–877.
- [22] Ran Zhao, Alexandros Papangelis, and Justine Cassell. 2014. Towards a dyadic computational model of rapport management for human-virtual agent interaction. In *International Conference on Intelligent Virtual Agents*. Springer, 514–527.