An experimental approach to study the physiology of natural social interactions

Thierry Chaminade

An experimental approach to study the physiology of natural social interactions

Thierry Chaminade

Institut de Neurosciences de la Timone, Centre National de la Recherche Scientifique - Aix-Marseille Université, UMR 7289, 27 Bd Jean Moulin, 13005 Marseille, France.

Abstract

The classical experimental methodology is ill-suited for the investigation of the behavioral and physiological correlates of natural social interactions. A new experimental approach combining a natural conversation between two persons with control conditions is proposed in this paper. Behavior, including gaze direction and speech, and physiology, including electrodermal activity, are recorded during a discussion between two participants through videoconferencing. Control for the social aspect of the interaction is provided by the use of an artificial agent and of videoed conditions. A cover story provides spurious explanations for the purpose of the experiment and for the recordings, as well as a controlled and engaging topic of discussion. Preprocessing entails transforming raw measurements into boxcar and delta functions time series indicating when a certain behaviour or physiological event is present. The preliminary analysis presented here consists in finding statistically significant differences between experimental conditions in the temporal associations between behavioral and physiological time series. Significant results validate the experimental approach and further developments including more elaborate analysis and adaptation of the paradigm to functional MRI are discussed.

Introduction

This manuscript presents a new approach to investigate scientifically the physiological bases of natural social interactions. This approach is pertinent for "second-person social neuroscience" (Schilbach 2010, Schilbach et al., 2013), which puts forward the importance of studying real-time social cognition in truly interactive scenarios. It is now agreed that such interactive approaches are necessary to understand social cognition and its disorders (Rolison et al., 2015). Efforts in the field of social neuroscience are therefore currently put in developing more ecological paradigms. Here, an experimental paradigm that allows recording of behavioral and physiological measures while two individuals discuss together is presented as proof-of-concept, with preliminary results demonstrating the potential use of the paradigm as well as validating the approach.

Believing that I am interacting live with a human partner compared to an autonomous robot is sufficient to activate brain areas involved in the attribution of mental states (Krach et al., 2008; Chaminade et al., 2012). Considering that other humans' behaviour is controlled by hidden mental states - intentions, desires, beliefs etc - is called "adopting the intentional stance" after philosopher Daniel Dennett (Dennett, 1996). Adopting the intentional stance is a defining aspect of social interactions, and is absent when interacting with an artificial agent, as described in an extensive review of the use of artificial agents in the study of the physiological bases of social cognition (Wykowska, Chaminade, Cheng, 2016). In everyday life, we adopt the intentional stance in response to the bottom-up information we naturally gather in real social interactions. This idea is the core of the Turing Test (Turing, 1950): it is through the content of the written interactions that one participant decides whether he's interacting with a human or an artificial intelligence. This importance of bottom-up information for adopting the intentional stance is at the core of the current experimental approach: it compares behavior and physiology when participants interact with a human agent, a real social interaction for which they adopt the intentional stance, and when they have a similar interaction with an agent for which they don't adopt the intentional stance, a robot or, in the present case, an embodied conversational agent.

According to the classical experimental method, the deterministic relation between a cause and a consequence is investigated by comparing the effect of two conditions controlled over all experimental variables except the one being tested as the possible cause. Natural social interactions can't easily be approached with such method. The first reason is theoretical: an uncountable number of events influence our decisions in real life, from our individual temperament to the minute-to-minute external events and physiological homeostasis. The second reason is practical: if participants are aware of the objectives of the experiment, their behaviour becomes unnatural. This phenomenon is known as the reactivity effect in psychology. Knowledge of being observed alters the performance of the participant, most often in order to fulfil the expectations of the experimenter (French & Sutton, 2010). A second feature of the experimental approach presented here addresses this issue by keeping the social interaction as natural as possible. Data acquisition is performed during a natural interaction while the two participants believe they are doing another task, not directly related to the study of social interactions.

A cover story providing spurious explanations of various elements of the experimental procedure has therefore been developed for the present approach. The use of a cover story to hide the true purpose of an experiment from the participants is commonly used in social psychology. A seminal experiment by Chartrand and Bargh (1999) investigated implicit mimicry (the "Chameleon effect") while participants believed they were participating to the development of a new psychological scale for which they were to describe in pairs the content of pictures. One of the pair of participants was a confederate of the experimenter whose role was to perform certain actions, rub his face or shake his foot. The behaviour of the discussant was rated to identify increase of the same target action. But as one discussant was a confederate,

one can imagine a “clever Hans effect” (Pfungst, O., 1911). Clever Hans, at the beginning of the 20th century, was a horse believed to perform arithmetic tasks. But further investigation showed that he was actually reacting to subtle postures and expressions from the humans asking questions that indicated when he reached the correct answer, while the humans themselves were not aware of communicating these cues. To avoid this, the two volunteers discussing together in the present experimental approach are both naive to the objective of the experiment.

The experimental approach proposed here is grounded in the comparison between a condition in which two humans discuss together and a condition in which one-participant discusses with an artificial agent, in order to identify behavioral and physiological features that are specific to interacting with a human, hence to adopting an intentional stance, to those that are preserved when the interacting agent is artificial. Meanwhile, the direct comparison of the interaction with a human and an artificial agent is also useful to investigate the social competence of the artificial agent (Chaminade and Cheng, 2009). Artificial agents are increasingly present in our society. Embodied Conversational Agents (ECAs) are used as web agents for e-commerce or tutors in e-learning applications. It is believed that ECAs’ behaviours should be endowed with communicative and emotional expressiveness to sustain long-term interactions with humans (Pelachaud, 2009). Humanoid robots have been proposed to intervene in cognitive therapies for children with autism spectrum disorder (see Diehl et al., 2012 for a review) and are more generally believed to become increasingly present in contact with humans. Yet, few objective measures exist to evaluate their social competence. As a matter of fact, "Can artificial agents be social?" is a conundrum, as the adjective “social” refers to behaviours taking place between humans. The social acceptance of robots is usually addressed with questionnaires, for example the Negative Attitude towards Robot Scale (NARS, Nomura et al., 2006). Such an approach can be useful to compare various artificial agents in terms of the subjective response they elicit, but are not sufficient to be interpreted in terms of their social acceptance, which requires a comparison with humans. While it is not its primary objective, the approach proposed here, that compares human behavioral and physiological responses during natural interactions with a human or an artificial agent, allows us to investigate how different dimensions of social competence are impacted by the adoption of the intentional stance (Wykowska, Chaminade, Cheng, 2016).

Summary of the Experimental Approach

The experimental approach comparing behavioural and physiological responses when a participant has a natural social interaction with a human or an artificial agents is presented in details in the methods section. In a nutshell, pairs of naive participants are tested together in an experimental setup using videoconferencing to support the discussion. Importantly, videoconferencing is known to preserve a strong sense of presence (Hauber et al., 2005). The first objective of the experiment being the exploration of natural social interactions, it is mandatory that participants are not aware of this objective. A believable cover story provides credible, but spurious, explanations to most aspects of the experimental setup.

The second objective is–to compare natural social interaction conditions to control conditions. An Embodied Conversational Agent presented as autonomous is used as control: it reproduces human behaviour superficially, but because of its artificial nature, humans interacting with this agent don’t adopt an intentional stance. Hence, by definition, interactions with an artificial agent aren’t social, providing a valid control condition. Actually, the experimenter has a long experience of using artificial agents as control conditions to study social cognitive neuroscience (humanoid robots and computer animations; see e.g. Chaminade et al., 2007, 2009, 2012, 2015). The other factor is the bidirectional nature of the interaction. Videos from live interactions are played back and participants are asked to try to interact with the video. Having live interaction and discussions with a video allows the comparison of behaviour between two conditions having the exact same sensory (visual and auditory) input to the participant but very different

experience (interactive or not). In summary, the first factor controls for the nature of the agent, social or not, while the second controls for the bidirectionality of the interaction, interactive or not.

The last objective is to investigate not only behavioural - speech produced, face and head movements, eye gaze - but also physiological correlates of natural social interactions. The underlying assumption is that physiological events, and in particular skin conductance (Dawson et al., 2007), reflect autonomous system responses that can't be voluntarily controlled. In the current approach, these physiological events are causally related not to deceit, but to behavioral features of the interaction. For example skin conductance can be a marker of the emotion felt (e.g. Khalfa et al., 2002), of orientation of attention (Frith & Allen, 1983), or of cognitive load and stress (Kilpatrick, 1972). In agreement, the last objective is to characterize how physiological events are temporally correlated with behavioural events. A preliminary attempt to test these correlations using Bayes theorem is presented here.

The next sections describe the proof-of-concept of this experimental approach that is extended to other behaviours and physiological responses in the discussion.

Methods

Participants

Because only a female voice was available for the artificial agent, only women were included to avoid mixing the gender of the two agents discussing. A total of six pairs of women volunteered to participate in this experiment, but one was excluded a posteriori given the poor quality of recorded data (data missing for more than one condition in each of the recorded measures). All participants gave informed consent in agreement to the declaration of Helsinki. The final sample comprises 10 female students recruited by word-of-mouth (mean age 22.7 years, standard deviation 6.4 years).

Experimental Paradigm

Cover Story

The cover story was fundamental in this experiment. It provided a common goal for the two interacting agents as well as a topic for the discussion. It has also been developed to provide spurious explanations for the main elements of the experimental paradigm. The fact that videoconferencing was used, a requirement in order to record the face from the front and to present the artificial conversational agent, was presented as a necessity to control precisely the time the two participants discuss together. This time pressure - one minute per condition - was important to avoid wavering.

For the sake of keeping the instructions natural, the experimenter presented, apparently informally, the goal and setting of the experiment to the pair of naive participants, who didn't know each other, upon arrival. It took 15 to 30 minutes to provide all required information, depending on the questions participants asked during the presentation. Italics represent phrases that were systematically provided orally to all participants.

The experiment was presented as a *neuromarketing experiment*. Participants were told that *we (the experimenters) are hired by an advertising company in order to validate a central assumption of a forthcoming campaign*. *You will see 3 images from an advertisement campaign without any written information and will have to find out the message of the campaign by discussing together* (Figure 1). *The images are naturally ambiguous* and *the company wants to validate their assumption that discussion between people is necessary to understand correctly the message*. *In order to validate this assumption, they hired us to run scientifically controlled experiments with an experimental psychology approach*. This

presentation helped to justify why they had to discuss, why we used an experimental psychology methodology, and introduce the topic of the discussion.

-- Figure 1 around here --

In order to *control the amount of information the two participants are able to exchange* we choose to have *the discussions through Skype so that their duration is exactly 1 minute for each of the three images*. With only three minutes of discussion altogether, *it is important to start the discussion without hesitation*. For this reason we required that *the participant in the testing room initiates the discussion every time;* this implied that she didn't know whether the condition was live or videoed at the onset of each trial but discovered it during the discussion.

Then they were introduced to the artificial embodied conversational agent GRETA (experimental factor 1: Nature of the Agent). GRETA was presented as *an autonomous agent having knowledge about the advertisement campaign. You are encouraged to discuss with GRETA to gather helpful information to understand the message of the advertisement campaign*. Finally we explained that *for experimental purposes, half of the conditions present a video of a previous interaction* (experimental factor 2: Bidirectionality of the Interaction).

Next we presented the recordings that would be made. We pretended recordings would take place when the participant looks at the images, while in reality we recorded during periods of discussion. *We will be using eyetracking to record what parts of the images you are looking at. We will also record your physiological response (heart rate and electrodermal activity).* Importantly participants were not told at this point that we would also record the audio and the video of their discussion, and that the real purpose of the experiment was to characterize how their behaviour would change during the discussion phase as a function of the experimental condition. While some participants were puzzled by the experimental procedures, none reported doubts about the cover story or the actual purpose of the experiment.

Experimental Conditions

Experimental conditions were defined by a 2 by 2 factorial plan. The first factor was the nature of the agent the participant discussed with, referred to as the discussant, a fellow Human or the Artificial embodied conversational agent GRETA, presented as fully autonomous. The second factor was the bidirectionality of the interaction, either Live or Videoed, the latter being a replay of the video of the previous live discussion with the same agent on the same image. The four conditions were therefore Human/Live, Artificial/Live, Human/Videoed, Artificial/Videoed.

As there are three images per series and four conditions, there were 12 trials per experiment. It was decided to alternate between human and artificial agents to avoid surprise about the nature of the agent at the onset on each trial. Importantly, the behaviour of both discussants was recorded in live conditions and played back in videoed condition, implying that for each given image and agent, the live trial preceded the videoed trial. Finally, to make sure that recognizing videoed conditions wasn't straightforward, the live and videoed condition for one image couldn't be consecutive for a given agent. There was a necessary imbalance in the temporal distribution of conditions, so that one order of the 12 trials was created to optimize the organisation despite these constraints and used for all participants.

Experimental Setup

Embodied Conversational Agent

The embodied conversational agent (ECA) GRETA used for this project was developed at the LTCI (Laboratoire Traitement et Communication de l'Information, mixed Télécom ParisTech & CNRS UMR

5141, Paris). GRETA is an experimental platform specifically dedicated to investigate verbal and nonverbal aspects of human-machine interactions (Pelachaud, 2009) and is particularly relevant for the current project as it is able to reproduce human emotional states and generic behavioral feedbacks (Ochs et al., 2012). A voice synthesizer from company CereProc was used to generate speech.

In order to fulfil its function, a simple Wizard of Oz (WoZ) procedure was programmed. In the field of human-computer interactions, a Wizard of Oz procedure corresponds to a human controlling an artificial agent directly while pretending that the artificial agent is autonomous. That allows the artificial agent to have an adapted behaviour without the requirement to program an autonomous behaviour. It has been used repeatedly in the study of human-robot interactions (Riek, 2012).

To achieve the WoZ procedure, around 80 simple behaviours were first constructed in the form of control files encoding upper body and head movements (e.g. nodding or shaking the head), facial expression of emotion or feeling (e.g. smiling or frowning) and verbal behaviour. There were two categories of verbal behaviours: half were non-specific feedbacks that could be used for all images (e.g. “Yes”, “No”, “Maybe”, “I think you’re right” etc…) and the other half were feedbacks specific of each campaign (e.g. for series 1: “They look like superheroes”, or for series 2: “It looks like they had a fight”) or specific of each image (e.g. for the first image of series 1: “It looks like the apple has Spiderman eyes”). Note that the limited number of possible feedbacks is a consequence of the cover story, as the discussion focuses on a controlled topic in order to use a circumscribed and well-controlled vocabulary. These control files were called online by the experimenter: sitting in the same room, he could hear what the participant was saying and therefore respond accordingly by typing the number attributed to each of the conversation feedbacks on a silent keyboard.

Physical setup

There was a room for the participant and another room for the human discussant, connected by ethernet (the setup is schematized in Figure 2). In the Participant room, the recorded participant sat comfortably on a chair in front of a computer screen topped by the webcam used for the videoconference discussion (using Skype). The two cameras of the eyetracker Facelab5 (SeeingMachines technology) used for gaze tracking were located under the screen and connected to a computer dedicated to gaze tracking. This system doesn’t require physical constraint so that participants remained free of their head movements. The left hand of the participant was fit with a photoplethysmograph sensor on the thumb to record blood pulse and the two electrodes of the electrodermal activity sensor (both on Biograph from ThoughtTechnology Ltd.) on the index and middle fingers according to electrodermal activity measurement guidelines (Roth et al., 2012), connected to the Biograph box (long dashed arrow with filled circle). A photodetector was fixed on the bottom left of the screen with opaque adhesive tape and connected to Biograph box (short dashed arrow with filled circle). The box itself was connected to a computer running Biograph and dedicated to the recording of physiology (Dashed arrow), and also outputted a synchronisation signal transformed electronically into a button click on the computer dedicated to gaze tracking (long dash and dot arrow). The Control computer was connected to the screen and the webcam (bidirectional dotted arrow) and the participant’s headphones. Headphones were used so that the speech from both participants were acquired separately. In addition, GRETA control and the WoZ program were installed on this Control computer. Two experimenters were present, one running the Control computer including the control of GRETA through the WoZ procedure in the “Artificial/Live” condition, the other one controlling the recordings of Facelab and Biograph data. The installation of the discussant room consisted in a slave computer controlled by the Control computer. This computer ran Skype and was connected to the discussant screen, webcam and headphones (bidirectional dotted arrow). A third experimenter stayed with the human discussant to inform her of upcomings Human/Live trials.

-- insert Figure 2 around here--

Experimental recording

Upon arrival, the pair of volunteers was briefed together about the goal of the experiment and the procedures (including the cover story). Then one of them went into the discussant's room with a number of questionnaires to fill, while the one in the participant's room was installed and fit with the captors. The eyetracker was also initialized for this participant with standard procedures. One participant was attributed to campaign 1 (superheroes) and the other to campaign 2 (rotten fruits).

On the participant's point of view, each trial consisted in viewing an image for 10 seconds, followed by 3 to 5 seconds of black screen, and then one minute during which they talked with the discussant (depending on the experimental condition). After each trial, the participant was asked whether the condition was live or videoed.

On the experimenter's' point of view, each trial started by launching Biograph and Facelab recording, then running one of four scripts. Each script first started recording the participant's screen and microphone, and showed the image for 10 seconds full screen on the participant's screen. In the condition Human/Live, the same procedure (start screen recording and show image for 10 seconds full screen) was also launched on the discussant's screen. The videoconference call launched on the participant's side was automatically answered on the discussant's side, and the script put both videoconference windows full screen. After one minute, the script stopped the videoconference and the screen recordings. Then the experimenter edited the audio and video of the discussant, in order to have a one minute audio/video file that was later used in the Human/Videoed condition.

In the Artificial/Live condition, after the image presentation and the black screen, the script launched GRETA and put it full screen on the participant's screen. During the interaction, the experimenter used the WoZ procedure to either respond to the participant's question, or provide her with new ideas when she was not talking. After one minute, the script stopped GRETA and screen recording, and the experimenter edited the screen recording on the participant's side, therefore corresponding to the audio and video of GRETA, in order to have a one minute audio/video file starting with GRETA window going full screen, that would be used in the Artificial/Videoed condition.

In both Artificial/Videoed and Human/Videoed, after the image presentation and the black screen, the script launched the video recorded during the previous live condition with GRETA and the discussant, respectively, on the same image. Altogether the audio and the video of both agents was recorded in every condition. At the end of the first experiment, we asked the participant what she concluded the message of the advertisement campaign was. Then the two volunteers changed room and role, and the second participant was tested.

When the two participants of a given pair were recorded, they were questioned to verify they still believed the cover-story, and then debriefed about the actual purpose of the experiment. They were informed of the audio and video recording and asked whether we could use them in our research. All still believed they participated to a neuromarketing experiment.

Data preprocessing

The objective was to characterize behavioural events that were temporally associated with physiological events. Preprocessing included the precise synchronisation of the behavioural and physiological time series acquired independently and the extraction of events from the time series. Binary time series describing events will be noted with brackets ([]) and can take two forms: boxcar functions for events lasting in time and delta functions for instantaneous events.

Electrodermal Activity

The example of electrodermal activity is used to illustrate analysis of physiological data. Using a Matlab toolbox for the analysis of electrodermal activity data (Ledalab; Benedek & Kaernbach, 2010), the raw data was decomposed into phasic and tonic components. Tonic components were deconvoluted in order to identify the timing of the event responsible for each tonic-response. The timing of the electrodermal activity events was used to construct a 30 Hz time series, called [isElectrodermalEvent],–indicating with delta functions when events giving rise to electrodermal responses happened.

Eyetracking

Synchronisation of eyetracking time series was performed by transforming the photodetector signal into a FaceLab label. Screen x and y voxel coordinates of the direction of the gaze and of the face on the screen was extracted. Eye closure and saccades were also extracted for filtering out unusable data. Preprocessing included windowing exactly 1 minute of data after the FaceLab label indicated the discussion was started, and excluding the data that was not usable given saccades and eyes closures. Finally, time series were downsampled from 60 to 30 Hz to match the frequency of the other recordings.

Conversation Behaviour

A screen recording software was used to record the video and audio of the conversation. In all conditions but Human/Live, synchronisation between the two rooms was automatic given that the participant's room computer recorded the video and audio of the two agents. In the Human/Live condition, synchronisation between the video and audio feeds recorded by each of the computer was ensured by using the small "self" video in videoconference window, that was hidden from the participant's view by the opaque tape holding the photosensor. The onset corresponded to the luminance reaching the level that activated the photosensor (the synchronisation device also used by FaceLab and biograph computer), then the audio and the video files lasting exactly one minute were produced.

The audio files were processed automatically using SPPAS (Bigi et al., 2014), resulting in two boxcar time series, one indicating when the participant is speaking ([isParticipantSpeak]) and another one indicating when the discussant is speaking ([isDiscussantSpeak]).

The video data was analyzed to extract facial features for each frame. A face recognition algorithm (Facial Feature Detection & Tracking; Xiong & De La Torre, 2013) was run frame by frame to identify the face present in the image. Screen x and y voxel coordinates of 49 keypoints on the face were recorded ("face tracking") as well as the rotation of the face mask in relation to the screen normal vector.

Face tracking results were combined with gaze tracking data to provide boxcar 30 Hz time series indicating gaze information for each frame. First, using face tracking coordinates, the position of the face, the eyes and the mouth of the discussant on the screen were calculated for each frame and used to define regions of interest. Regions of interest were ellipses centered on the center of mass of the points representing the nose, mouth or the totality of the points, with the major and minor radius equal to 1.5 times the distance from the center to the most extreme point (in the horizontal and vertical axis of the face template). Circles were used for the eyes, with similar geometrical features. Then, using gaze tracking coordinates from the participant, 30 Hz boxcar time series were created indicating where the participant was looking at (is she looking at the screen [isData], the face [isFace], the eyes [isEyes] or the mouth [isMouth]?).

-- insert Figure 3 around here --

Statistical Analysis

Several 30Hz binary time series were produced during preprocessing, corresponding to speech ([isParticipantSpeak], [isDiscussantSpeak]), to the direction of the participant's gaze ([isFace], [isMouth], [isEye]), and to physiological events ([isElectrodermalEvent]). The goal is to identify temporal relationships between physiology and behaviour (Bach & Friston, 2013). A probabilistic approach was privileged under the assumption that it is adapted to the ecological type of relationships expected here, which are multidimensional (speech, face and eye movements, physiology) and, because of the experimental method, noisy. The exploration of these relationships between these time series was performed with a direct application of Bayes theorem. It is particularly well suited to approximate the posterior probability of certain behaviours giving rise to certain physiological responses while keeping track of events that are not controlled in terms of their probability and temporal distribution. Given □ the physiology and □ the behaviour, the posterior probability of □ happening in the context of □ was calculated:

$$(1) \qquad \square(\square/\square) = \frac{\square(\square/\square).\square(\square)}{\square(\square)}$$

For each trial, P(□) is the number of ones divided by the total number of time intervals. P(□) is the division of the number of physiological events by the number of time intervals-P(□/□) is the division of the number of physiological events for which □=1 within the time interval divided by the total number of physiological events. These probabilities can be calculated for each data point, corresponding, at the frequency of 30 Hz used in the time series, to an interval of 33 ms. But given the noise in electrodermal deconvolution and the timing of behavioural events, co-occurrence between them are unlikely to take place in such a small time interval. In the absence of a priori knowledge about the relevant time interval, time intervals between 100 and 500 ms were tested empirically. Five-sampling frequencies, 2, 3, 5, 6 and 10 Hz were used, corresponding to time intervals of 500, 333, 200, 167 and 100 ms respectively, during which co-occurrence of behavioral and physiological events were investigated.-These 5 frequencies were divisors of 30 Hz so that no extrapolation was required for downsampling time-series.

Results

Effect of Temporal Resolution of the Analysis

First we calculated the posterior probability of having an electrodermal event as a function of three linguistic and three oculomotor behaviours at different frequencies, for each subject and trial. Missing data points (maximum one per measure and participant) were replaced by empty cells. These probabilities were analyzed with repeated-measures analysis of variance using mixed models in SPSS in order to identify effects of the nature of the agent (Human/Artificial) and the bidirectionality of the interaction (Live/Videoed) as well as the interaction between these two factors on the posterior probabilities.

Results on the significance of the factors on posterior probabilities at different sampling rates are presented graphically on figure 4. Despite the small sample, significant effects (at $p < 0.05$) were found (see next sections). It should be noted that the analysis is affected by the sampling rate used for the analysis. Most of the largely non significant effects ($p > 0.25$) are quite similar at all sampling rates used. In the case of effects reaching significance, significance isn't found at all sampling rates used, implying that the sampling rate used for the analysis is crucial for an effect to reach significance. More interestingly, in all cases when significance is reached, the result obtained at 5 Hz, corresponding to a time window of 200 ms, always had the lowest p values. In a number of cases significance only reaches $p < 0.05$ threshold for 5 Hz, the other sampling rates being only marginally significant. Further analysis of significant effects was therefore limited to the 5 Hz resampling of the time series.

-- insert Figure 4 around here --

Relation with Verbal Behaviour

Figure 5 presents the posterior probability of a physiological occurring when a particular behaviour is taking place ($\square(\square/\square)$) as a function of the four experimental conditions, for all effects that reached significance. As long as verbal behaviour is concerned, only when the participant was speaking were there significant effects of experimental factors on the probability to observe an electrodermal response. There was a significant effect of the nature of the agent ($F(1,9) = 8.42$, $p = 0.03$, $\eta^2_{partial} = 0.58$) and of the bidirectionality ($F(1,9) = 6.01$, $p = 0.05$, $\eta^2_{partial} = 0.50$), while agent by bidirectionality interaction didn't reach significance ($F(1,9) < 0.01$, $p = 0.97$). As predicted, the probability of having a physiological event was greater when the subject was speaking to a human compared to an artificial discussant, and during live compared to videoed conditions.

Relation with Oculomotor Behaviours

Next oculomotor behaviours were analysed (Figure 5). Interestingly, the pattern of significant posterior probabilities differs depending on whether the gaze was on the face, the eyes or the mouth. For the former, only the bidirectionality of the interaction reached significance ($F(1,9) = 5.33$, $p = 0.05$, $\eta^2_{partial} = 0.37$; nature of agent: $F(1,9) = 0.12$, $p = 0.74$; agent by bidirectionality $F(1,9) = 0.64$, $p = 0.44$). Probability of electrodermal event when the face was gazed at increased in the live compared to the videoed condition. When the eyes were being looked at, only the nature of the agent significantly influenced the posterior probability ($F(1,9) = 3.86$, $p = 0.05$, $\eta^2_{partial} = 0.49$; bidirectionality of the interaction $F(1,9) = 1.69$, $p = 0.20$; agent by bidirectionality interaction $F(1,9) = 0.02$, $p = 0.88$). Probability increased when the discussant was the human fellow. Finally, considering the mouth being watched, the effect of the nature of the agent didn't reach significance ($F(1,9) = 0.08$, $p = 0.78$), while the bidirectionality of the interaction ($F(1,9) = 7.49$, $p = 0.02$, $\eta^2_{partial} = 0.45$) and agent by bidirectionality interaction ($F(1,9) = 5.00$, $p = 0.05$, $\eta^2_{partial} = 0.36$) were both significant. The important feature of the significant interaction was that posterior probability increased significantly (from 0.043 to 0.069) when the mouth of a human was observed live compared to videoed, but no effect of the bidirectionality of the interaction was found for the artificial agent.

-- insert Figure 5 around here --

Discussion

The objective of this paper is to propose a new experimental approach to investigate the physiological and behavioural correlates of natural conversations in order to elucidate changes associated with adopting the intentional stance. The demonstration is still preliminary at this stage - several required technical improvements have been identified during data collection, the sample of participants for this proof of concept of the experiment is limited (n=10), and a number of complementary analyses will be run. Nevertheless, the finding of significant effects in line with expectations validates the experimental approach that could benefit to several research communities - social cognitive neuroscience, psychiatry, linguistics, social embodied conversational agents and human-robot interactions in particular.

Current Findings

Only linguistic and gaze behaviours were analyzed at this stage. A first issue concerned the time scale of the probabilistic temporal associations under scrutiny. It is clear, given the deconvolution of electrodermal activity and the physiological delays, that synchronicity at the frequency used for data preprocessing (30 Hz, meaning co-occurrence of events within 33 ms windows) was improbable. An exploratory approach was adopted. The same posterior probabilities were calculated with time windows of 100, 167, 200, 333 and 500 ms. While results were quite consistent when comparing the time windows, the scale of 200 ms always provided, when significant, the lowest *p* value. It is interesting to compare this to the conclusion of Laming (1968) that a simple reaction time to a visual stimulus, when no other task is required, is around 220 ms, as it strongly supports that physiological responses are automatically triggered in response to perceptual event.

When investigated further, it should be noted that the significant effects are not similar across the behaviours investigated. Probability of observing physiological responses when the participant is speaking is influenced by the two factors independently: it increases for the human compared to the artificial agent and for live compared to videoed condition, implying that this probability is affected similarly (in both case the delta is 1% of probability) to the degradation of the social competence of the interaction. Increase in this probability could represent the participant's engagement in the conversation and therefore provide a good marker of the social competence of the interacting agent.

The result when gaze is directed to the mouth shows a significant effect of live vs videoed for the human, but not the artificial agent. It is possible that the imperfect rendering of mouth movements for speech with the version of GRETA used in this experiment makes the use of lip reading useless and therefore dissociated from physiological responses. In contrast, looking at the mouth is used to help the understanding of the discourse, that is useful for live but not videoed conversation with humans as only in the former case is new information provided. In that case, the physiological effects associated with gaze directed to the mouth would correspond to increased attention to the content communicated through speech. Further investigations focusing on the content of the discourse could confirm this interpretation.

The case of the eyes strongly confirms the importance of eye contact to provide a sense of social presence (Senju & Johnson, 2009). It is associated with physiological responses when interacting with the human irrespective of the nature of the interaction, meaning that even in videoed interactions (practically, when watching a movie), the fact of observing human eyes is associated with an increased probability of having a electrodermal response. This is even more surprising as, because of technical limitations of the current set-up (see Further Developments), only the artificial agent provided the feeling of direct gaze, while the human discussant gaze was directed downward. The difference of probability between the human and artificial agents represents an objective measure of the sense of presence elicited by mutual gaze (Senju & Johnson, 2009).

When the face is watched, the probability of observing a physiological response is increased for live versus videoed conditions, with no effect of or interaction with the nature of the agent. This could be surprising as recognizing live from videoed conditions was more complicated for the artificial than the human discussant (see Further Development). But errors are always the same, live conditions reported as videoed, so that repetition of a previous interactions (ie video) was always correctly recognized. The current result is therefore likely to represent reduced surprise or attention when a previously experienced discussion is repeated.

These significant findings with a limited dataset argue in favour of the validity of the experimental approach proposed here, namely recording behavioural and physiological data during a natural discussion when varying the social presence through two factors, the nature of the agent being interacted with, a natural or an artificial agent, and the bidirectionality of the interaction itself, live or videoed. The core of the approach is then to compare the natural social interaction condition, Human/Live, with control conditions, either a bidirectional discussion but with an artificial agent, for which we don't adopt an intentional stance (Artificial/Live), or the same video and audio input from the same human but not bidirectional (Human/Videoed).

Further Developments

Firstly, several limitations of the existing experimental setup have been identified in this proof-of-concept phase and are currently being addressed. The absence of direct eye contact in video conferencing is critical to investigate natural social interactions, for which mutual gaze is a central element eliciting a sense of presence and an engagement (Senju & Johnson, 2009). Recently, technical solutions to this limitation have been proposed, based on online video correction (Kuster et al., 2012). An alternative option is the use of a device based on a semi-transparent mirror system. The image of the other agent is on a screen located behind the mirror angled 45° from the horizontal, and a camera located above the mirror records the image of the speaker on the mirror, giving the impression of real eye contact. A second issue in the current set-up is the delay introduced by the Wizard of Oz program used to control the artificial agent, so that participants frequently reported the Artificial/Live condition to be Artificial/Videoed. This is a computer engineering and programming issue, but it could have impacted some of the present results. Finally, the presence of the experimenters in the same room as the participant could also have given rise to a form of reactivity effect (French and Sutton, 2010). Both the participant and discussant should be isolated in a their respective room.

Secondly, a number of analyses have to be developed from the corpus of behavioral and physiological data already acquired. The other recorded physiological measures is blood pressure, but heart rate variability can't be analysed using the same approach as skin conductance, and a different preprocessing has to be developed to extract events from raw data. Behavioral analyses focused on speaker turn-taking and direction of gaze of the participant but other variables can be extracted from the raw data. Head movements - translations and rotations - are extracted from the videos and should be used to analyze mimicry of these oscillatory movements. It will allow us to investigate, for example, whether or not physiological responses are more likely when speakers' movements are coordinated. A frame by frame distribution of 22 emotions can be extracted and used to investigate propagation of positive emotions (smiles and laughter) and their correlation with physiological markers of engagement. This is also the case of the semantic content of the conversation, as it is possible that physiological responses will be more likely for infrequent than frequent words as rare words cause surprise in the perceiver. Another direction for improvement is time series analyses. More complex statistical approaches, for example incorporating multiple factors and temporal causality like Granger causality or cross-recurrence analyses, also present promising developments.

Finally, the objective of this experimental approach is to be extended to neurophysiological investigations, in particular using functional magnetic resonance imaging. This methodology will be helpful to investigate the neural bases of abnormal social behaviour in autism spectrum disorders (Rolison et al., 2015) as well as to assess the potential of artificial agents as interacting partners in this population (Chaminade et al., 2015). Most required devices are already available, and the few remaining technical difficulties are all tractable. fMRI data can be investigated using a similar procedure than for the electrodermal response presented in the current report. Preprocessing involves deconvoluting fMRI time series in brain regions devoted to well-characterized dimensions of social cognition into boxcar (for sustained activity) or delta (for single events) functions time series, that will then be analysed as the electrodermal activity to provide evidence for temporal relations between these activities and various aspects of behaviour. Altogether the current results validate the experimental approach proposed here to investigate the physiological bases of a natural social behaviour, a discussion between two agents, which can be extended to neurophysiological investigations.

Acknowledgements

The experimental approach described in this manuscript would not have been possible without a large number of collaborators: Christine Deruelle and Professor Da Fonseca at the INT (Institut de Neurosciences de la Timone, Aix-Marseille Université [AMU] & Centre National de la Recherche Scientifique [CNRS]

UMR 7289, Marseille) in the discussions of the experimental paradigm, Catherine Pelachaud and Magalie Ochs at the LTCI (Laboratoire Traitement et Communication de l'Information, mixed Télécom ParisTech & CNRS UMR 5141, Paris) provided the embodied conversational agent GRETA, technical help was obtained by INT support team (Joël Baurberg and Xavier Degiovanni); master student Louise Merly developed the first version of the experimental setup, psychiatry intern Raphaël Curti was responsible for data recording (with support from Farah Wolfe) as well as the coding of personality questionnaire; Laurent Prévot and master student Léo Baiocchi, from the LPL (Laboratoire Parole et Langage , AMU & CNRS UMR 7309, Aix-en-Provence), extracted speech data from the audio recording of the experiment, company Picxel contributed to the extraction of the face tracking data from the video recordings

References

Bach, D. R., & Friston, K. J. (2013). Model-based analysis of skin conductance responses: Towards causal models in psychophysiology. *Psychophysiology*, *50*(1), 15–22.

Benedek, M., & Kaernbach, C. (2010). Decomposition of skin conductance data by means of nonnegative deconvolution. *Psychophysiology,* 47(4), 647-658.

Bigi, B., Watanabe, T. & Prévot, L. (2014). Representing Multimodal Linguistics Annotated Data, *9th International conference on Language Resources and Evaluation (LREC)*, Reykjavik (Iceland).

Chaminade, T., Hodgins, J., & Kawato, M. (2007). Anthropomorphism influences perception of computer-animated characters' actions. *Soc Cogn Affect Neurosci, 2*(3), 206-216.

Chaminade, T., & Cheng, G. (2009). Social cognitive neuroscience and humanoid robotics. *Journal Of Physiology-Paris*, *103*(3-5), 286–295.

Chaminade, T., Rosset, D., Fonseca, D. D., Nazarian, B., Lutcher, E., Cheng, G., & Deruelle, C. (2012). How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Frontiers In Human Neuroscience*, *6*.

Chaminade, T., Fonseca, D., Rosset, D., Cheng, G., & Deruelle, C. (2015). Atypical modulation of hypothalamic activity by social context in ASD. *Research In Autism Spectrum Disorders*, *10*, 41–50.

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of personality and social psychology*, *76*(6), 893.

Dawson, M.E., Schell, A.M., & Filion, D.L. (2007). The electrodermal system. In J.T. Cacioppo, L.G. Tassinary, & G.G. Berntson (Eds.). *Handbook of Psychophysiology* (3rd edition; pp. 159-181). Cambridge, UK: Cambridge University Press.

Dennett, D. C. (1996), *The Intentional Stance (6th printing)*, Cambridge (MA, USA): The MIT Press.

Diehl, J. J., Schmitt, L. M., Villano, M., & Crowell, C. R. (2012). The clinical use of robots for individuals with Autism Spectrum Disorders: A critical review. *Research In Autism Spectrum Disorders*, *6*(1), 249–262.

French, D. P., & Sutton, S. (2010). Reactivity of measurement in health psychology: How much of a problem is it? What can be done about it? *British Journal Of Health Psychology*, *15*(3), 453–468.

Frith, C. D., & Allen, H. A. (1983). The skin conductance orienting response as an index of attention. *Biological psychology*, *17*(1), 27-39.

Hauber, J., Regenbrecht, H., Hills, A., Cockburn, A., & Billinghurst, M. (2005). Social presence in two-and three-dimensional videoconferencing. *Proceedings of 8th Annual International Workshop on Presence*, London (UK), 189-198.

Khalfa, S., Isabelle, P., Jean-Pierre, B., & Manon, R. (2002). Event-related skin conductance responses to musical emotions in humans. *Neuroscience letters*, *328*(2), 145-149.

Kilpatrick, D. G. (1972). Differential responsiveness of two electrodermal indices to psychological stress and performance of a complex cognitive task. *Psychophysiology*, *9*(2), 218-226.

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., & Kircher, T. (2008). Can Machines Think? Interaction and Perspective Taking with Robots Investigated via fMRI. *PLoS ONE*, *3*(7), e2597.

Kuster, C., Popa, T., Bazin, J.-C., Gotsman, C., & Gross, M. (2012). Gaze correction for home video conferencing. *ACM Transactions On Graphics*, *31*(6), 1.

Laming, D. R. J. (1968). *Information Theory of Choice-Reaction Times*. Academic Press, London.

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3), 437-454

Ochs, M., Niewiadomski, R., Brunet, P., & Pelachaud, C. (2012). Smiling virtual agent in social context. *Cognitive Processing*, *13*(S2), 519–532.

Pelachaud, C. (2009). Modelling multimodal expression of emotion in a virtual agent. *Philosophical Transactions Of the Royal Society B: Biological Sciences*, *364*(1535), 3539–3548.

Pfungst, O. (1911). *Clever Hans (The horse of Mr. von Osten): A contribution to experimental animal and human psychology* (Trans. C. L. Rahn). New York: Henry Holt.

Riek, L.D. (2012). Wizard of oz studies in hri: a systematic review and new reporting guidelines. *Journal of Human-Robot Interaction* **1**.

Rolison, M. J., Naples, A. J., & McPartland, J. C. (2015). Interactive Social Neuroscience to Study Autism Spectrum Disorder. The Yale Journal of Biology and Medicine, *88*(1), 17–24.

Roth, W. T., Dawson, M. E., & Filion, D. L. (2012). Publication recommendations for electrodermal measurements. *Psychophysiology*, *49*, 1017-1034.

Schilbach, L. (2010). A second-person approach to other minds. *Nature Reviews Neuroscience*, *11*(6), 449–449.

Schilbach, L., Timmermans, B., Reddy, V., Costall, A., Bente, G., Schlicht, T., & Vogeley, K. (2013). Toward a second-person neuroscience. *Behavioral And Brain Sciences*, *36*(04), 393–414.

Senju, A., & Johnson, M. H. (2009). The eye contact effect: mechanisms and development. *Trends In Cognitive Sciences*, *13*(3), 127–134.

Turing, A. (1950), Computing machinery and intelligence. *Mind,* 59(236), 433-460.

Wykowska, A., Chaminade, T. & Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. *Phil. Trans. R. Soc. B, 371*(1693), 20150375.

Xiong, X. & De la Torre, F. (2013). Supervised Descent Method and its Application to Face Alignment. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

Figure Legend

Figure 1: The two series of three images were used as support for discussion in the cover stories. They were chosen to respond to a number of criteria: first, to actually represent an homogenous advertisement message; second that different interpretations of this message are possible; third, to avoid real humans or social interactions, and fourth, to still be interpretable in social terms. Anthropomorphized fruits and vegetables represented superheroes in series 1 and rotten fruits in series 2.

Figure 2: Experimental setup (explanations provided in experimental setup section of the main text).

Figure 3: Combination of face and gaze tracking on one frame. Blue dots represent features tracked by the face tracking program. Circles indicate regions of interest on the Discussant based on face tracking, the green dot the direction of the Participant's head and yellow dot the direction of Participant's gaze. In this specific frame, [IsFace] and [IsMouth] are equal to 1 for both discussants, [IsEyes] equals 0.

Figure 4: Probability that the effect of interest (Agent, Bidirectionality and Agent by Bidirectionality interaction) significantly affects the posterior probabilities of obtaining a physiological response given an observed behaviour, at the five different sampling frequencies used for the analysis (thick grey line: $p < 0.05$).

Figure 5: Posterior probability of observing a physiological event given a behavior as a function of the four experimental conditions defined by the nature of the agent (Human, Artificial) and the bidirectionality of the interaction (Live, Videoed). Error bars indicate standard error.