

Interacting with a Social Robot Affects Visual Perception of Space

Carlo Mazzola
CONTACT & RBCS,
Istituto Italiano di
Tecnologia
& DIBRIS,
Università di Genova
Genova, Italy
carlo.mazzola@iit.it

Alexander Mois
Aroyo
CONTACT
Istituto Italiano di
Tecnologia
Genova, Italy
alexander.aroyo@iit.it

Francesco Rea
RBCS
Istituto Italiano di
Tecnologia
Genova, Italy
francesco.rea@iit.it

Alessandra Sciutti
CONTACT
Istituto Italiano di
Tecnologia
Genova, Italy
alessandra.sciutti@iit.it

ABSTRACT

Human partners are very effective at coordinating in space and time. Such ability is particularly remarkable considering that visual perception of space is a complex inferential process, which is affected by individual prior experience (e.g. the history of previous stimuli). As a result, two partners might perceive differently the same stimulus. Yet, they find a way to align their perception, as demonstrated by the high degree of coordination observed in sports or even in everyday gestures as shaking hands. Robots would need a similar ability to align with their partner's perception. However, to date there is no knowledge of how the inferential mechanism supporting visual perception operates during social interaction. In the current work, we use a humanoid robot to address this question. We replicate a standard protocol for the quantification of perceptual inference in a HRI setting. Participants estimated the length of a set of segments presented by the humanoid robot iCub. The robot behaved in one condition as a mechanical arm driven by a computer and in another condition as an interactive, social partner. Even if the stimuli presented were the same in the two conditions, length perception was different when the robot was judged as an interactive agent rather than a mechanical tool. When playing with the social robot, participants relied significantly less on stimulus history. This result suggests that the brain changes optimization strategies during interaction and lays the foundations to design human-aware robot visual perception.

CCS CONCEPTS

• Human-centered computing-Interaction design • Human-centered computing-Collaborative and social computing devices

KEYWORDS

Space perception; central tendency effect; shared perception; interaction; alignment; bayesian models; iCub; humanoid

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.
HRI '20, March 23–26, 2020, Cambridge, United Kingdom.
© 2020 Copyright is held by the owner/author(s).
ACM ISBN 978-1-4503-6746-2/20/03.
DOI: <https://doi.org/10.1145/3319502.3374819>

ACM Reference format:

Carlo Mazzola, Alexander Mois Aroyo, Francesco Rea and Alessandra Sciutti. 2020. Interacting with a Social Robot Affects Visual Perception of Space. In *Proceedings of 2020 ACM/IEEE International Conference on Human Robot Interaction (HRI'20)*, 23-26 March 2020, Cambridge (UK). ACM, New York, NY, USA, IX pages. <https://doi.org/10.1145/3319502.3374819>

1 Introduction

Humans are very good at coordinating with each other in space and time. We see it in everyday actions, as passing an object or shaking hands, and it is even more evident in sports, dance, music and complex collaborative activities. This is the kind of coordination ability that we would desire in future robots interacting with us.

This human skill is all the more remarkable, given that the two partners might perceive the world differently. Indeed, our perception of the world does not only depend on the incoming sensory information, but is also influenced by our prior knowledge, associated with the statistical properties of our world. This phenomenon traces back to Helmholtz, who introduced the concept of unconscious inference, which helps to shape vision [1]. Incorporating previous experience, or priors (e.g. the history of the previous stimuli), into the current percept helps the brain to cope with the uncertainty resulting from sensory and neural noise and ambiguity [2].

An example is the inferential process of perception called *central tendency* in which previous perceptual experience (prior) modifies the current perception. In general, judgments of time, weights, forces, extent of movements, length, etc. show the same tendency to gravitate toward a mean magnitude. For instance, when you are asked to judge the length of a set of short segments and, then, you are shown the next one, you will perceive it as shorter than it actually is because you will use the average (short) length of the previous segments as prior. Perception is therefore affected by the mean value of the stimuli previously experienced [1, 3-4]. Recent Bayesian models of the phenomenon [1, 4] have shown that the central-tendency strategy (i.e. taking into account the statistics of previous stimuli rather than just the current stimulation in a perceptual judgment) optimizes perception. In fact, this strategy leads to an increased reliability associated with

an overall reduction in the error magnitude [3], [4]. As a result, the estimate of one stimulus differs depending on the distribution from which it is drawn, *i.e.* its statistical context. The very same physical stimulus can be then perceived differently by two observers, if they have different previous experience [5].

However, to collaborate efficiently with others we need to share similar representations of the world and, in particular, of its temporal and spatial properties. In a cooperative situation, a mechanism like central tendency that sacrifices the accurate perception of the current stimulus could not represent the best strategy. In fact, humans somehow find a way to align their perception notwithstanding this inferential process. Different interactive purposes require different levels of alignment. Indeed, although in certain contexts extreme precision in the coordination is required – as in sports or in surgical procedures – in everyday life a “sufficient” alignment can support effective interaction, even in presence of different sensorimotor experiences and different prior experiences. Thus, a successful shared perception would not always require to completely put aside prior experience. According to the context, alignment might require just a minor reliance on our own prior in order to give more importance to what the partner perceives.

Enabling robots to be aware of the partner’s perception is important to establish sufficient alignment, as a function of what is needed for a successful interaction. Sharing a similar perception of the world has been deemed as determinant in supporting intuitive mutual understanding between agents of any nature [6]. Conversely, in robotics, perception mainly aims at reaching high levels of accuracy in the detection of the physical properties of the environment, rather than trying to accommodate or replicate potential distortions in the way the human partners sense it. There have been some attempts at modeling human peculiarities in the perception of space, for instance in the context of replicating visual perspective taking mechanisms [7], but the majority of the robotic platforms is designed to maximize the physical accuracy of environment perception, also through the use of a variety of non-biologically inspired sensors (as RGB-D cameras or lasers).

We claim that it would be important to design robots who are aware of these “distortions” and consider them when interacting with human partners, to guarantee the necessary alignment for interaction. In other words, we are advocating the need for a *human-aware perception* to support collaborative planning, manipulation, navigation and action in general. To do that we have to endow robots with an accurate model of how humans modulate their perception and inferential process during interaction.

Recently a high degree of interest has emerged in the effect of recent experience on visual perception, highlighting the substantial relevance of the use of priors has on individual percepts [3], [8], [9]. However, there is **no knowledge of how perceptual inference shapes visual perception during social interaction**. This is, indeed, a crucial question. During interaction, the brain faces two potentially conflicting goals: maximizing individual perceptual stability by using central tendency or maximizing perceptual alignment with the partner to

facilitate coordination, by limiting the reliance on individual previous experience.

In this work, we propose a first step toward understanding how human perceive space *during interaction*, to be able to model the same perception mechanism in robots and enable perceptual alignment. In the current research, robots become a fundamental tool of investigation. One of the crucial limits to the study of perception during interaction has so far been the impossibility of maintaining rigorous control on the stimulation, while allowing for a direct involvement in a dynamic exchange with another agent. The robotic platform allows porting the stimuli used in perceptual investigations to the domain of online collaboration, bringing controllability and repeatability to an embodied and interactive context.

Mounting evidence has shown that being involved in an interaction is very different than just observing it for many behavioral and neurophysiological aspects [10]. For instance, Schillibach and colleagues have demonstrated that in a task requiring participants to fixate an object on a screen, the underlying brain activity differed significantly when the participant was doing this ‘together’ with a virtual other, by following his gaze, or alone [10]. During action planning, Kourtis et al. [11] compared the preparation of a joint action measured through EEG with the planning of individual actions. They found a significant increase of the neural activation in the first case, suggesting the inclusion of the representation of the other’s action. The presence of a social rather than individual context has also behavioral consequences. For instance, when reaching to grasp the very same object, the kinematics of the action varies significantly if it is driven by a communicative rather than an individual intention [12] and even if the social context is competitive rather than collaborative [13].

It is however unknown how social involvement might affect the very basis of the inference mechanism supporting visual perception. A hypothesis is that the brain might aim for the emergence of shared perception, even if this implies selecting a different solution if compared to individual strategy and sacrificing robustness to sensory noise. The current experiment verifies whether interacting with a social or a non-social agent affects visual perception of spatial stimuli.

In particular, we are replicating a standard protocol for the quantification of central tendency in space perception [14], porting it into an interactive, HRI context. The task consisted in reproducing the length of a series of segments shown on a touchscreen. The assessment of the nature of the errors in the reproduction of the lengths allowed measuring the degree of central tendency for each participant in different experimental sessions. In the main experimental validation, the iCub humanoid robot [15] acted as stimulus demonstrator, by physically indicating the extremes of the segments to be reproduced with its hand. We manipulated the interactivity of the experimental context by realizing two conditions that we presented as *mechanical* and *social*. In the **mechanical condition**, the humanoid robot iCub behaved as a mechanical arm showing a neutral facial expression and gazing at a fixed point in the scene. In the **social condition**, the robot presented the same stimuli, but

it established a mutual gaze giving an implicit idea of turn taking with a friendly expression. Our hypotheses are as follows: (i) in the *social* condition, the central tendency phenomenon will be reduced, as the brain will favor giving more relevance to current sensory stimulation rather than stimulus history, in order to facilitate instantaneous alignment of perception with an interaction partner; (ii) the effect will be stronger the more social the robot will be perceived during the interaction.

2 Methods

In this study, we evaluated whether space perception changes when the perceptual task is not performed in isolation but with another agent. We focused on the central tendency strategy and we assessed whether, given the very same distances to be reproduced, they were reproduced differently when the robot – who acted as stimulus demonstrator – exhibited a different level of sociality (see Fig. 1.1 and Fig. 1.3).

2.1 Participants

For this study, 30 participants were recruited (15 males, 15 females) from 19 to 46 years old ($M=28$, $SD=6$). For 3 participants the robot could not complete the task due to technical failures and they were therefore excluded from the analysis, leaving a sample of 13 males, 14 females, of age $M=29$, $SD=6$. Participants were students or workers in different areas with a background in Humanities, Engineering, Computer Science, Architecture, Medicine, Statistics, and Politics and were chosen from the ones who answered to a recruiting e-mail sent to a local mailing list. Most participants lived in the region where the experiment was run. 11 participants (4 males, 7 females) had already previously done experiments with iCub robot. All the participants gave their written informed consent before participating. The regional ethical committee approved the study (Comitato Etico Regione Liguria). All participants were compensated with a sum of 15 € for their time.

2.2 Design

The experiment comprised 4 different sessions. In two of them, the participants pursued an individual task using a touchscreen. The other two conditions involved the interaction with the robot

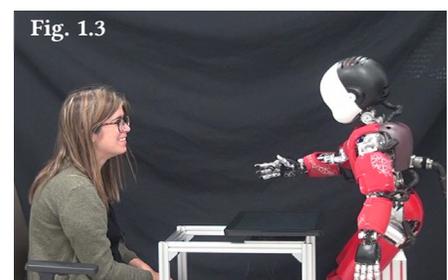
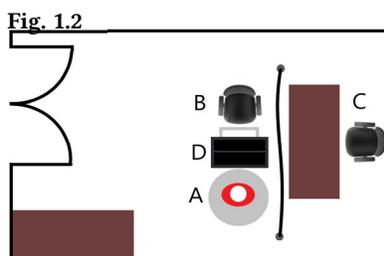
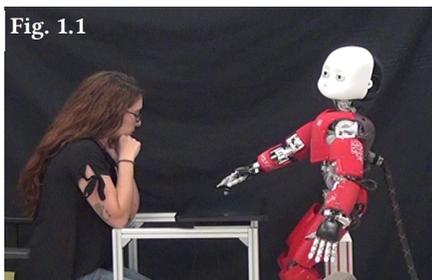
employing the same touchscreen. The order of conditions was counterbalanced. Participants could start either with the individual sessions or with the robot. Moreover, also the order of the two sessions with the robot – the *mechanical* and the *social* condition – was counterbalanced. The experiment lasted approximately 90 minutes. Participants had been previously informed of the duration of the experiment.

For the experimental set-up (see Fig. 1.2), participants were sitting in an office chair without wheels. The chair faced a structure of the height of 75 cm on which an ELO 2002L 20" touchscreen monitor was mounted. For the sessions performed with the robot, the robot was placed on the other part of the touchscreen at 20 cm from it on a fixed platform. For the individual sessions, instead, the robot was hidden behind a black curtain. The blinds were closed and the room was lit up with artificial light in order to ensure the same lighting conditions for all participants. Another curtain hid the experimenter's station with a table and the computers connected to the touchscreen and to the robot. After having explained the task to the participants, the experimenter sat behind the curtain. In the experimental room there were two cameras to record the experiment and a Tobii pro glasses 2 device was used as eye-tracker. The touchscreen was programmed with MATLAB 2019a and Psychtoolbox on a Windows 10 pc that recorded the responses of participants.

2.2.1 Individual Sessions.

2.2.1.1 Individual Length Discrimination.

The first individual session consisted in a *length discrimination* task aiming to test the perceptual acuity of the participants. Three red disks of 1 cm diameter appeared for 0.4 s in sequence with an interval of 1.5 s on a white straight line crossing the screen at its central height. After stimulus disappearance, subjects had to judge whether the longest segment was the first, delimited by the first and the second disk, or the second one, delimited by the second and the third disk, by typing respectively "1" or "2" on a keyboard located between them and the touchscreen. Participants performed this task for 66 trials. One of the two distances (*standard*) measured always 10 cm, while the other (the *comparison*) changed from trial to trial according to a QUEST adaptive procedure [14]. This design represents a very simple measure of length discrimination, where priors do not influence performance. The proportion of times in which the *comparison* interval was judged longer than the *standard* was plotted as a



Panel 1. Pictures and schema of the set up. Fig. 1.1 The *mechanical* condition. The robot is performing the task gazing at a fixed point. Fig. 1.2 The set-up of the experimental room for sessions to be performed with the robot. The robot (A) in front of the participant (B) with the touchscreen between them (D), and the experimenter (C) behind the black curtain. For the individual tasks, the set-up was the same except for the robot that was hidden behind another black curtain. Fig. 1.3 The robot in the *social* condition gazing at the screen and at the participant.

function of *comparison* amplitude and fit by a cumulative Gaussian distribution. The standard deviation of the fitted Gaussian represents the perceptual threshold, which is the minimal difference between two lengths that the participant can reliably distinguish.

2.2.1.2 Individual Length Reproduction. The second individual session consisted in a *length reproduction task* (see Fig. 2 A). The set-up was the same as the previous session but in this case, for each trial only two consecutive red disks of 1 cm diameter appeared on the white line. The duration of the stimulus measured 0.6 s. The first disk was positioned at a variable distance from the left border of the screen (0.5–3.5 cm, randomly selected). On its disappearance, a second disk appeared after 2 s at a variable distance to the right of it, disappearing again after 0.6 s. For each participant, 11 different lengths were presented 6 times, randomly, in a range of lengths from 6 cm up to 14 cm increasing each 0.8 cm. After the stimulus, participants had to touch the screen on the white line on the right of the second disk in order to reproduce that distance. A red disk came on to indicate where they had touched, but no feedback about the correctness of the response was provided. Participants had 3 practice trials after which they performed the task for 66 trials. This individual session served to compare our results with traditional studies on central tendency.

2.2.2 Sessions with the robot. In the other two sessions, on the other part of the touchscreen with respect to participants there was a robot acting as stimulus demonstrator on the touchscreen. Except for the use of iCub robot, the set-up of the experimental room was the same as the individual sessions.

In both sessions, the robot pressed two different points in sequence on the touchscreen along a visible straight white line (as above). Then participants had to press a third point at equal distance from the last shown by the robot. Unlike the individual session, the touchscreen did not show any red disk in the points where iCub or the participant touched. iCub was programmed to show 11 different lengths 6 times randomly. The programmed lengths were equal to those of the individual session (from 6 cm to 14 cm, increasing each 0.8 cm). The interval between the first and the second press of the robot was on average of 5.45 s (*SD* 0.4 s) depending on the length of the stimulus, while each trial lasted on average of 15.76 s (*SD* 2.2 s). The task of the participant was the same for both the sessions. We scheduled 66 trials for each session with 3 additional practice trials at the beginning.

For the two sessions, we employed the same robot, but its behavior differed. In one condition, iCub appeared as a social partner of the task, while in the other one it did not show any kind of social behavior. Participants were explicitly informed of the difference of the behavior of iCub.

2.2.2.1 Length Reproduction with mechanical robot. In the *mechanical* condition (see Fig. 1.1), the robot behaved only as a mechanical stimulus. To this aim, the iCub face showed a fixed neutral facial expression: the LEDs of its face were switched on but iCub did not smile or show any emotion. Its gaze was programmed to be directed to a fixed point in the environment in a manner that the robot did not seem to gaze at participants or to show any awareness of the environment. We instructed

participants about the task with these words: “Now I turned off iCub’s social intelligence. The computer is just driving its hand motions in a predefined pattern. It will be touching two positions on the touch screen. Please reproduce the distance between these two points, by pressing a third one at equal distance from the last one”.

2.2.2.2 Length Reproduction with social robot. In the *social* condition (see Fig. 1.3), the eye-cameras of the robot were switched on and iCub changed its facial expression according to its action. When iCub looked at participants, before the session and between trials, it smiled showing a friendly expression. When iCub was busy in touching the screen, it showed a focused expression. To enhance the impression of animacy, for the entire duration of the interactive condition, the eyelids were blinking. Using a face-tracker module, iCub was able to establish mutual gaze with the participants, looking at their face at the beginning of the session, before and after each stimulus demonstration, giving an implicit idea of turn taking. Moreover, while performing its task, it exhibited a natural oculo-manual coordination, anticipatorily directing its gaze toward the point it was going to press. Finally, to enhance its social behavior, the robot told fixed sentences to participants. During speech, the mouth-LEDs simulated the lips movement in coordination with the words iCub was saying. At the beginning of the session, iCub welcomed participants and explained them the task with these words: “Hello, I’m iCub! Now we will play together. I will touch the screen twice and you will touch the screen a third time to replicate the distance. Are you ready?”. At the end of the session, it said goodbye to them: “Thank you for having played with me! It has taken a bit of a long time, but you are helping us a lot! See you next time!” It also incited twice participants to play well the game from time to time, saying: “Well done! Keep it up!” and “Come on, there are only a few more trials left, keep focused”. The condition was also framed as social by the experimenter, who introduced the session saying, “Now iCub is fully working, with its social intelligence on and its cameras are switched on to look at you and at the screen. It will be showing you two positions on the touch screen. Please reproduce the distance between these two points, by pressing a third one at equal distance from the last shown by the robot”.

2.2.3 Questionnaires. Additionally, we collected data from a set of questionnaires through Google Form. The first questionnaire was compiled before coming in the laboratory and included some questions about participants’ previous experience with robots, the Italian version of TIPI test on participants’ personality [16], the Autism-spectrum Quotient test (AQ test) [17], [18] to measure the degree to which adults with normal intelligence have the traits associated with the autistic spectrum, and the NARS questionnaire, to evaluate the attitudes of participants towards robots [19].

Then, immediately after each session with the robot, a set of additional questionnaires was proposed. First, participants had to fill the Inclusion of other in the self-scale (IOS) questionnaire [20], by which it is possible to check how close the person interviewed feel to another subject (in our case the robot) in a range from 1 to 7. Furthermore, it was asked to participants to answer to the subscales *Anthropomorphism*, *Animacy*, *Likeability*, and *Perceived Intelligence* of the Godspeed questionnaire [21] and to the sub-

scales *Mind experience* and *Mind agency* of a Mind perception test [22][23]. These aimed to obtain from the participants their impressions of the robot after each condition so that it was possible to compare them. When we proposed these questionnaires, we employed a 7 points likert scale, from 1 to 7. In order to make results more readable, we put results of all these sub-scales in proportion with 1, where 1 is the highest possible result (e.g. Anthropomorphism has 5 items in a 7 points likert scale: its highest possible result would be 35). At the end of the experiment, a final questionnaire was given in order to collect participants' opinions about the experiment, to get their feedback about the behavior of iCub and to know which was the difference between the sessions with the robot according to them.

2.3 The Robot

To generate controlled stimulation in an interactive context we used the humanoid robot iCub [15]. Such a complex agent endowed with sensors and actuators allows generating controlled and precise actions that enable the experimenter to replicate the rigorous perceptual stimuli traditionally adopted in the standard perceptual investigations. However, it also generates complex behaviors, which could entail the coordinated use of multiple effectors (e.g., eye and hand or two arms) and the timely and adaptive response to a participant's behavior. The robot is endowed with multiple sensors (two cameras, a skin over its body, composed of multiple tactile sensors, microphones in its head) whose input can be processed almost online and used to modify robot behavior according to participants' actions. For the purpose of this experiment, it was sufficient employing only the left eye-camera using a Face Tracker module to make iCub look at participants' face at the end of each trial and the iKinGazeCtrl [24] module to make it direct its gaze toward the point on the touchscreen. The head of iCub is provided with LED lights to enable it to show different facial expressions. The lights were fixed on a neutral face for the *mechanical* condition, while for the *social* condition iCub was programmed to change facial expression according to its actions.

Moreover, to design robot behaviors we made use of the CartesianController module [25] already designed by the iCub community, which produces human-like minimum jerk movements with an average hand speed of about 0.1 m/s. Through this module, the robot presented the stimuli on the touchscreen moving the torso and the right arm for all its degrees of freedom, except for the fingers that were fixed. During all the experiment, iCub showed a biological motion. Due to the combination of the motor imprecision intrinsic to the robot actuators, the presented positions could slightly differ from the programmed ones. For this reason, we submitted the sets of stimuli presented by the robot in the two conditions to a Kolmogorov-Smirnov test. Results showed that the distributions were not significantly different (for all participants $p > 0.93$) providing evidence that the statistical context was comparable in the *mechanical* and the *social* conditions. The robot was programmed to perform the task with pre-defined movements and speech. The experimenter just started the execution of the routines. Except for the face-tracking behaviour in the *social* condition, the robot was not responsive to

stimuli from participants. In the experimental room, there was only one experimenter, who was a researcher of the laboratory and was aware of the experimental hypothesis. During the experiment, he checked the correct realization of the experiment without any control on the robot.

2.4 Data Analysis

To assess the degree of central tendency in spatial perception, we followed a well-established approach [14], [26]. As a direct measure of central tendency, we computed the regression index for each of the reproduction tasks (*individual*, *with mechanical robot*, *with social robot*) as the difference in slope between the best linear fit of the reproduced values and the identity line. The regression indexes were compared among different conditions through either t-tests or ANOVAs, according to the design. When a data distribution was not normal (checked with Lilliefors test), the corresponding non-parametric tests were performed. The data resulting from questionnaires were also used as independent variables in linear regressions. The details of each analysis are reported in the results. From the *length discrimination* task, we derived the spatial acuity of each participant, i.e., the minimal difference between two lengths that the participant could reliably discriminate. This corresponds to standard deviation of the cumulative Gaussian fitting the proportion of responses "the comparison stimulus is longer" (more details in [14]).

We considered outlier any participant who exhibited results exceeding the interval $[Mean - 2.5 SD; Mean + 2.5 SD]$.

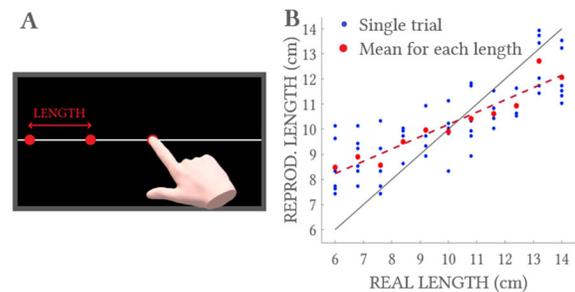


Figure 2. A) Sketch of the individual length reproduction task. B) Example of results of a length reproduction task. Reproduced values are plotted against the corresponding stimuli. The regression index is computed as the difference between the slope of identity line (1, gray) and the slope of the best linear fit of the data (red).

3 Results

In this experiment we wanted to assess whether being involved in an interaction with another social agent could modify the perception of the length of a segment it showed, if compared when the same stimulus was shown by a mechanical stimulus generator.

3.1 Baseline analysis

First, we checked that all participants had a visual acuity that enabled them to perceive reliably the difference in length of the

proposed stimuli. Indeed, the phenomenon of central tendency is highly influenced by perceptual acuity [4] and an abnormal perceptual threshold would have determined abnormal performances in the regression tasks. The average perceptual acuity was of about 2.39 cm ($SD=0.62$ cm) well compatible with the task, with no participant exhibiting outliers values.

To evaluate the central tendency we plotted the reproduced lengths against the presented length (see Fig. 2 B) and measured the regression index, defined as the difference in slope between the best linear fit of the data and the identity line. An index close to 0 corresponds to a reduced central tendency, whereas values close to 1 are associated to a strong use of the previous stimulus history.

We repeated the same procedure for all the sessions (*individual, with mechanical robot and with social robot*). Before advancing with the analyses, we checked these results for the presence of outlier values. This led to the exclusion of 2 participants (1 male and 1 female), whose regression index was larger than the average plus 2.5 SD s, one in the *individual* condition and the other in the *social* condition. Such result suggests that they were not paying attention to the task and were not reproducing lengths reliably. In fact, they did not even differentiate in their reproduction the two extremes of the range presented (ca. 6 cm - 14 cm). The sample on which we performed the full analysis was then composed of 12 males, 13 female, age $M=29$, $SD=6$, of which 10 had already interacted with iCub for other experiments.

In the *individual* condition, participants exhibited a clear central tendency, with an average regression index of 0.446 ($SD=0.13$). Through an one-sample t-test, this value resulted significantly different from 0 ($t(24)=16.81$, $p<0.01$, Cohen's $d=3.469$) and in line with the results obtained in previous experiments on central tendency [14]. This confirmed that participants leveraged on the inferential mechanism to perform the reproduction task.

3.2 Manipulation check

To verify that participants perceived the *social* and the *mechanical* conditions differently, we compared their responses to the Godspeed questionnaires reported after the two conditions. We expected in particular to find a significant increase in the *Animacy* subscale of the questionnaire, as we were aiming to manipulate the perception of the robot being a live agent versus a non-alive machine and as this scale encompasses questions as the level of perceived interactiveness and of responsiveness to stimuli. A paired-sample t-test showed that in the *social* condition the robot was perceived with a significantly higher level of *Animacy* ($M=0.62$, $SD=0.15$) than in the *mechanical* condition ($M=0.37$, $SD=0.15$): $t(24)=-7.56$, $p<0.01$, Cohen's $d_z=1.60$. This is evident also from Fig. 3, where all evaluations of animacy performed after the *social* condition result higher than the evaluations the same individual provided after the *mechanical* condition (all points lying above the identity line).

The manipulation had a significant effect on the general perception of the robot. A Wilcoxon Signed-Ranks Test indicated that for Anthropomorphism values referred to the *social* condition

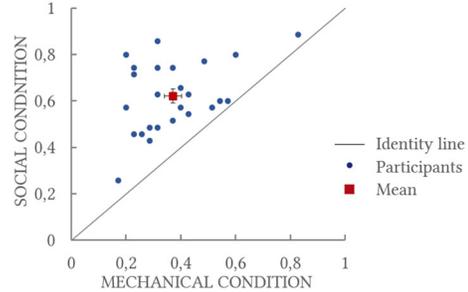


Figure 3. Plot of the values of Godspeed subscale–Animacy for each participant in both *mechanical* and *social* condition.

($M=0.55$, $SD=0.17$) are significantly higher than the ones of the *mechanical* condition ($M=0.37$, $SD=0.14$): $Z=-3.85$, $p<0.01$, Cohen's $d_z=1.03$. A paired sample t-test revealed the same result in *Likeability* (Soc. C. $M=0.82$, $SD=0.18$; Mech. C. $M=0.63$, $SD=0.21$; $t(24)=-5.03$, $p<0.01$, Cohen's $d_z=0.97$) and in *Perceived Intelligence* (Soc. C. $M=0.70$, $SD=0.11$; Mech. C. $M=0.62$, $SD=0.14$; $t(24)=-3.91$, $p<0.01$, Cohen's $d_z=0.76$). Also considering different dimensions of the mind perception, we observed a significant increase in both the *Mind experience* subscale (Soc. C. $M=0.44$, $SD=0.27$; Mech. C. $M=0.25$, $SD=0.17$; Wilcoxon Signed-Ranks Test: $Z=-2.75$, $p<0.01$, Cohen's $d_z=0.76$) and the *Mind agency* subscale (Soc. C. $M=0.60$, $SD=0.25$; Mech. C. $M=0.42$, $SD=0.20$; paired sample t-test: $t(24)=-4$, $p<0.01$, Cohen's $d_z=0.83$) and the same regarding the "Inclusion of other in the self-scale" questionnaire (Soc. C. $M=0.63$, $SD=0.20$; Mech. C. $M=0.41$, $SD=0.22$; Wilcoxon Signed-Ranks Test: $Z=-3.59$, $p<0.01$, Cohen's $d_z=1.21$).

3.3 Perception during interaction

In the reproduction tasks involving the robot as demonstrator, the average regression index was 0.26 ($SD=0.17$) a value significantly lower than the one measured in the individual condition ($M=0.45$, $SD=0.13$) (paired sample t-test: $t(24)=4.66$, $p<0.01$, Cohen's $d_z=0.96$). This general decrease can be partially due to the difference in the type of stimulation: just two brief red disks represented the length to be reproduced in the *individual* condition; while in both robotic sessions the whole arm motion was visible, providing a richer information. According to the Bayesian models described in [27], the presence of less sensory noise would yield to less central tendency.

However, comparing the two different robot-based conditions a significant decrease in regression index emerged, even though the length stimuli were the same. The regression index was on average 0.29 ($SD=0.19$) in the *mechanical* condition and 0.23 ($SD=0.17$) in the *social* condition. A mixed model ANOVA with "order of conditions" as between factor and "condition" as within factor confirmed that the regression index was significantly larger in the *mechanical* condition ($F(1,23)=7.22$, $p=0.013$, $\eta_p^2=0.24$) (see Fig. 4 B). On the contrary, there was no significant effect of order ($F(1,23)=0.32$, $p=0.58$) nor of the interaction between the two.

Given the more complex oculo-motor behaviour of iCub in the *social* condition, we evaluated whether the time interval between

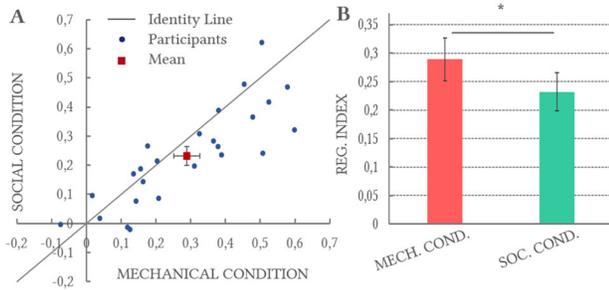


Figure 4. A) Individual regression index and B) Mean regression index in the two robot conditions.

two consecutive trials was longer than in the mechanical condition. The average inter trial interval in the mechanical condition ($M=15.46$ s, $SD=0.6$ s) resulted indeed significantly shorter than in the social condition ($M=15.88$ s, $SD=0.62$ s), of about 0.4 s (paired sample t-test: $t(24)=-2.36$, $p=0.027$, Cohen's $d_z=0.47$). To ensure that this difference was not causing the change in regression index observed, we ran a linear regression between the variation (ΔRI) in regression index and the difference in inter trial interval between the conditions (ΔT). No correlation was found as ΔRI resulted to be approximately constant among different values of ΔT ($F=0.01$, $R^2=0.03$, $p=0.93$).

To verify whether previous experience interacting with iCub impacted on the results, we divided all participants in two groups (if they had already performed experiments with a iCub or not). A Mixed Model ANOVA with “condition” as within factor and “previous experience” as between factor did not reveal any significant effect of previous experience ($F(1,23)=0.34$, $p=0.57$) on the significant variation of regression index between conditions.

Therefore, results indicate that on average participants exhibited less central tendency when they were involved in an interactive context than when they were playing alone with a computer or with a mechanical device showing them the stimuli.

To further explore the factors modulating the observed results we performed a multiple linear regression to test the effect of the changes in the evaluations of the Godspeed subscales on the amount of individual changes in regression index between the two conditions (ΔRI). The results of the regression indicated that the model explained 32% of the variance ($F(4,20)=2.36$, $p=0.088$, $R^2=0.32$). The only significant predictor was “Anthropomorphism” ($F(1,20)=7.74$, $p=0.01$, $\beta=-0.74$, Cohen's $f^2=0.39$), whereas no other regression reached significance (see Fig. 5). We did not find any significant effect in a multiple linear regression between ΔRI and the two scales *Mind experience* and *Mind agency* ($F(2,22)=0.91$, $p=0.42$, $R^2=0.08$) nor in a simple regression with respect to the change in IOS values ($F(1,23)=0.06$, $p=0.81$, $R^2=0.002$).

For none of the results of the questionnaires investigating the personality of participants (TIPI, AQ, NARS), it was found a significant linear regression with the variation of the regression index: TIPI: multiple linear regression ($F(1,19)=0.81$, $p=0.56$, $R^2=0.18$), AQ test: simple linear regression ($F(1,23)=1.24$, $R^2=0.05$, $p=0.28$), NARS questionnaire simple linear regression ($F(1,23)=0.19$, $R^2=0.01$, $p=0.67$).

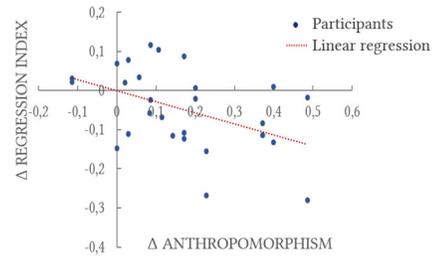


Figure 5. Individual variations of the regression index in the two robot conditions as a function of the variations of the results from the Godspeed scale *Anthropomorphism*.

4 Discussion

When interacting with others, it is often necessary that the two partners coordinate in space and time their behaviors, both with respect to each other and with respect to external stimuli. For instance, while dancing, the two partners have to adjust to the other's moves while at the same time perceiving the correct timing from the musical input. Hence, a perception based on an inferential process that responds to a need for optimality in the formation of individual perceptual decisions poses a crucial question when considering joint actions. Precisely, how does the brain deal with the conflicting goals of maximizing perceptual stability on the one hand and aligning with the perception of the partner on the other?

The hypothesis driving this work is that the brain puts in place mechanisms that might favor the emergence of shared perception, even at the expenses of selecting a sub-optimal solution, if compared with individual strategy. Our results show that we favor accurate estimation of a physical stimulus – if embedded in an interaction – rather than a stable, though less veridical perception, as the one normally derived by optimization in individual situations produced by the central tendency mechanism. In other words, in interactive scenarios, accuracy becomes more important than robustness to perceptual noise to allow for the successful completion of a cooperative effort. Therefore, the current perceived stimulation (e.g. the length of a movement) becomes less biased by the stimulus history (i.e., by the average of the lengths previously observed) and there is a minor effect of the central tendency strategy during interaction.

If we assume that the Bayesian model of the central tendency is preserved also in interactive conditions, this change could be explained in terms of a reweighting between the relevance of the statistical context (or *prior*) and the current stimulation (*likelihood*), with a decrease of the relative importance of the former with respect to the latter. This would imply that the participant –when involved in an interaction– relies more on the (variable) instantaneous stimulation than during individual tasks, where the story of previous stimuli plays a stronger role. There is wide evidence of the inferential process being dynamic, with the relative weight given to the prior varying as a function of sensory precision and becoming higher when the uncertainty of the sensory input increase, as predicted by the Bayesian framework

[4]. Our current new results suggest that the interactivity of the context could represent an additional factor modulating the integration of prior knowledge and incoming sensory stimulation.

The use of a humanoid as a controllable interactive agent, whose perceptual and motor decisions can be fully under experimental control, provided us the unique possibility to have an active, embodied interaction with a partner that could be assimilated to a machine or to a living agent, while presenting the exact same stimuli across conditions. We operated the manipulation both explicitly and implicitly. On the one hand, the experimenter framed the robot's role in the two conditions differently: as a "tool" for the computer to show stimuli, or as a social agent. On the other hand, the robot exhibited either a passive, fixed state or a series of social and intelligent behaviors (in terms of looking patterns, facial expressions and speech). The combination of these two types of manipulation was on average quite effective, with judgments for anthropomorphism, animacy, likability, perceived intelligence and even mind attribution significantly higher in the *social* vs. the *mechanical* condition. The current study cannot tear apart the relative contributions of the explicit framing and the implicit social signals embedded in robot behavior. Since mounting evidence has highlighted the power of communication of robot gaze, e.g. as a mean to unconsciously govern turn taking in interaction [28][29] or to reveal goals [30], we believe that implicit gazing behavior could have alone a significant impact on the selection of the perceptual strategy. Future research is needed to verify this hypothesis.

Getting back to the question about what triggered the change in perceptual strategy, the results suggest that the change in perceived "anthropomorphism" of the iCub between the two conditions played an important role. Indeed, the more the robot was judged as having increased its anthropomorphism in the *social* condition, the less the perception in that condition was influenced by the statistical context, with respect to the *mechanical* one. So it seems that the more human-like the partner was perceived to be by participants, the less their perceptual strategy considered the previous stimuli, in favor of the current one. Since the robot was in both conditions a humanoid platform, moving its arm and torso according to biological motion rules, it was the combination of gazing, facial expressions and speech, together with the experimental framing, that drove this change in judgment, with no change in robot shape or its motion kinematics.

Future research will aim to clarify whether interaction with an anthropomorphic and social agent leads to different perceptual strategies as a function of it exhibiting a pro-social versus an anti-social behavior toward the human partner. As a potential confound it might be suggested that in the *social* condition the gaze of the robot was providing and additional information about the location of the two extremes of the length to be reproduced, providing a richer stimulation, which in turn would have produced less need to rely on the statistical context to improve perception. However, the strong linear dependency of the change in regression index on the attributed anthropomorphism of the robot seems to exclude this possibility. If the stimuli in the social condition were simply richer, the increased information would have been reasonably the same for all participants.

What do these results tell to the HRI community? In everyday activities we often have to coordinate with spatial-temporal dynamic inputs deriving from the environment – e.g., to catch an object before it falls – and with the partner, whose perceptual and motor abilities might differ from ours. Aligning with the partner's perception implies understanding what will he or she perceive (e.g. with respect to the timing and location of the falling objects) and adapt our action accordingly. This ability is necessary also in robots aiming at interacting with us. Understanding how humans perceive the world while interacting will lay the foundations to make autonomous technology adaptive to each user's needs.

A robot aware of the sensory distortion of a partner could have important applications in the context of assistive and rehabilitation robotics. This could be particularly relevant for the elderly people since it has been proposed that their perception might be more strongly distorted due to an over-reliance on internal priors [31] that could play a role in fostering detachment and isolation. Novel robots, with adaptive perception, will be able to predict potential distortions in their partners and adapt to them – i.e., by tailoring their actions to complement those of the human in time, force and space, e.g. in an object passage. Alternatively, robots might also correct this kind of distortions, e.g., by systematically warning patients of the inaccuracy and correcting the misperception. In particular, often rehabilitation requires patients to repeat multiple time the same set of actions, which might induce them to regress to their average behavior. A robot used in such type of rehabilitation could use a measure of central tendency to actively counteract it and push for improvement. In the process of achieving this long-term goal, human-robot interaction has once more proven to be a fundamental tool to investigate the bases of human social abilities [32],[33] and the very mechanisms of visual perception. Therefore, controllability, repeatability and reliability of robots allow to lead the research in real-world scenarios, and further investigate how interaction with social agents affects human perceptual and cognitive processes.

Additionally, these results provide an innovative way of measuring how the human partner unconsciously perceives "being" with another agent. Indeed, by measuring how the interaction with different agents affects individual perception of a physical property of a series of stimuli (as its length here, but the very same phenomenon of central tendency applies to almost all magnitudes, including time [27]) we can gain an indirect and unconscious measure of how *interactive* the context is perceived to be. This would represent a novel measure complementing the existing measures of involvement in joint actions (as the Joint Simon Effect [34] or the automatic Visual Perspective Taking [35]), enabling to assess through a purely perceptual task whether a human-robot interaction is implicitly processed as social or not.

ACKNOWLEDGMENTS

We thank Matilde Antonj and Joshua Zonca for their precious help. This work was supported by a Starting Grant from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme. G.A. No 804388, wHiSPER.

REFERENCES

- [1] H. Helmholtz, *Handbuch der physiologischen Optik*. Leipzig, 1866.
- [2] D. C. Knill and A. Pouget, "The Bayesian brain: the role of uncertainty in neural coding and computation," *Trends Neurosci.*, vol. 27, no. 12, pp. 712–9, Dec. 2004.
- [3] M. Jazayeri and M. N. Shadlen, "Temporal context calibrates interval timing," *Nat. Neurosci.*, vol. 13, no. 8, pp. 1020–6, Aug. 2010.
- [4] G. M. Cicchini, R. Arrighi, L. Cecchetti, M. Giusti, and D. C. Burr, "Optimal encoding of interval timing in expert percussionists," *J. Neurosci.*, vol. 32, no. 3, pp. 1056–60, Jan. 2012.
- [5] N. W. Roach, P. V. McGraw, D. J. Whitaker, and J. Heron, "Generalization of prior information for rapid Bayesian time estimation," *Proc. Natl. Acad. Sci.*, vol. 114, no. 2, pp. 412–417, 2017.
- [6] K. Fong, T., Nourbakhsh, L., & Dautenhahn, "A Survey of Socially Interactive Robots: Concepts, Design, and Applications," *Rob. Auton. Syst.*, vol. 42, no. 3–4, pp. 143–166, 2003.
- [7] T. Fischer and Y. Demiris, "Markerless Perspective Taking for Humanoid Robots in Unconstrained Environments," *2016 IEEE Int. Conf. Robot. Autom.*, pp. 3309–3316, 2016.
- [8] J. Fischer and D. Whitney, "Serial dependence in visual perception," *Nat. Neurosci.*, vol. 17, no. 5, pp. 738–743, 2014.
- [9] J. Taubert, D. Alais, and D. Burr, "Different coding strategies for the perception of stable and changeable facial attributes," *Sci. Rep.*, vol. 6, pp. 2–8, 2016.
- [10] L. Schilbach *et al.*, "Toward a second-person neuroscience," *Behav. Brain Sci.*, vol. 36, no. 4, pp. 393–414, 2013.
- [11] D. Kourtis, N. Sebanz, and G. Knoblich, "Predictive representation of other people's actions in joint action planning: An EEG study," *Soc. Neurosci.*, vol. 8, no. August 2013, pp. 1–12, 2012.
- [12] L. Sartori, C. Becchio, B. G. Bara, and U. Castiello, "Does the intention to communicate affect action kinematics?," *Conscious. Cogn.*, vol. 18, no. 3, pp. 766–772, 2009.
- [13] I. Georgiou, C. Becchio, S. Glover, and U. Castiello, "Different action patterns for cooperative and competitive behaviour," *Cognition*, vol. 102, pp. 415–433, 2007.
- [14] A. Sciutti, D. Burr, A. Saracco, G. Sandini, and M. Gori, "Development of context dependency in human space perception," *Exp. Brain Res.*, vol. 232, no. 12, pp. 3965–3976, 2014.
- [15] G. Metta *et al.*, "The iCub humanoid robot: An open-systems platform for research in cognitive development," *Proc. 8th Work. Perform. metrics Intell. Syst.*, pp. 50–56, 2008.
- [16] C. Chiorri, F. Bracco, T. Piccinno, C. Modafferi, and V. Battini, "Psychometric Properties of a Revised Version of the Ten Item Personality Inventory," *Eur. J. Psychol. Assess.*, 2014.
- [17] S. Baron-cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley, "The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome / High-Functioning Autism, Males and Females, Scientists and Mathematicians," *J. Autism Dev. Disord.*, vol. 31, no. 1, 2001.
- [18] L. Ruta, D. Mazzone, and L. Mazzone, "The Autism-Spectrum Quotient – Italian Version: A Cross-Cultural Confirmation of the Broader Autism Phenotype," *no. 123*, pp. 625–633, 2012.
- [19] D. S. Syrdal, K. Dautenhahn, K. L. Koay, and M. L. Walters, "The Negative Attitudes towards Robots Scale and Reactions to Robot Behaviour in a Live Human-Robot Interaction Study," 2003.
- [20] D. Aron, A., Aron, E. N., & Smollan, "Inclusion of other in the self scale and the structure of interpersonal closeness," *J. Personal. Soc. Psychol.*, vol. 63(4), 596, 1992.
- [21] C. Bartneck, D. Kuli, and E. Croft, "Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots," *Int. J. Soc. Robot.*, vol. 1, no. 1, pp. 71–81, 2009.
- [22] F. Ferrari, M. Paola, and P. Jolanda, "Blurring Human – Machine Distinctions: Anthropomorphic Appearance in Social Robots as a Threat to Human Distinctiveness," *Int. J. Soc. Robot.*, vol. 8, no. 2, pp. 287–302, 2016.
- [23] H. M. Gray, K. Gray, and D. M. Wegner, "Dimensions of Mind Perception," *Science (80-)*, vol. 315, no. 5812, pp. 619–619, 2007.
- [24] A. Roncone, U. Pattacini, G. Metta, and L. Natale, "A Cartesian 6-DoF Gaze Controller for Humanoid Robots," *Robot. Sci. Syst.*, vol. 2016, 2016.
- [25] U. Pattacini, F. Nori, L. Natale, G. Metta, and G. Sandini, "An experimental evaluation of a novel minimum-jerk Cartesian controller for humanoid robots," in *IEEE/RSJ 2010 International Conference on Intelligent Robots and Systems, IROS 2010 - Conference Proceedings*, 2010, pp. 1668–1674.
- [26] G. M. Cicchini, R. Arrighi, L. Cecchetti, M. Giusti, D. C. Burr, and V. Moruzzi, "Optimal encoding of interval timing in expert percussionists," *J. Neurosci.*, vol. 32, no. 3, pp. 1056–60, 2012.
- [27] M. Jazayeri and M. N. Shadlen, "Temporal context calibrates interval timing," *Nat. Neurosci.*, vol. 13, no. 8, pp. 1020–1026, 2011.
- [28] K. Ruhland *et al.*, "A Review of Eye Gaze in Virtual Agents, Social Robotics and HCI: Behaviour Generation, User Interaction and Perception," vol. 00, no. 0, pp. 1–28, 2015.
- [29] O. Palinko, A. Sciutti, L. Schillingmann, Y. Nagai, and G. Sandini, "Gaze Contingency in Turn-Taking for Human Robot Interaction: Advantages and Drawbacks," *2015 24th IEEE Int. Symp. Robot Hum. Interact. Commun.*, pp. 369–374, 2015.
- [30] O. Palinko, F. Rea, G. Sandini, and A. Sciutti, "A Robot Reading Human Gaze: Why Eye Tracking Is Better Than Head Tracking for Human - Robot Collaboration," *2016 IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 5048–5054, 2016.
- [31] M. Turgeon, C. Lustig, and W. H. Meck, "Cognitive Aging and Time Perception: Roles of Bayesian Optimization and Degeneracy," *Front. Aging Neurosci.*, vol. 8, no. 102, 2016.
- [32] T. Chaminade and G. Cheng, "Social cognitive neuroscience and humanoid robotics," *J. Physiol.*, vol. 103, pp. 286–295, 2009.
- [33] A. Sciutti and G. Sandini, "Interacting With Robots to Investigate the Bases of Social Interaction," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 12, pp. 2295–2304, 2017.
- [34] G. Knoblich and N. Sebanz, "The Social Nature of Perception and Action," *Curr. Dir. Psychol. Sci.*, vol. 15, no. 3, pp. 99–104, Jun. 2006.
- [35] D. Samson, I. A. Apperly, J. J. Braithwaite, B. J. Andrews, and S. E. B. Scott, "Seeing it their way: Evidence for rapid and involuntary computation of what other people see," *J. of Experimental Psychol. Hum. Percept. Perform.*, vol. 36 (5), no. 1255, pp. 1–39, 2010.