



Research Article

Open Access

Davide Ghiglino*, Cesco Willemse, Davide De Tommaso, Francesco Bossi, and Agnieszka Wykowska

At first sight: robots' subtle eye movement parameters affect human attentional engagement, spontaneous attunement and perceived human-likeness

<https://doi.org/10.1515/pjbr-2020-0004>

Received July 23, 2019; accepted December 2, 2019

Abstract: Human-robot interaction research could benefit from knowing how various parameters of robotic eye movement control affect specific cognitive mechanisms of the user, such as attention or perception. In the present study, we systematically teased apart control parameters of Trajectory Time of robot eye movements (rTT) between two joint positions and Fixation Duration (rFD) on each of these positions of the iCub robot. We showed recordings of these behaviors to participants and asked them to rate each video on how human-like the robot's behavior appeared. Additionally, we recorded participants' eye movements to examine whether the different control parameters evoked different effects on cognition and attention. We found that slow but variable robot eye movements yielded relatively higher human-likeness ratings. On the other hand, the eye-tracking data suggest that the human range of rTT is most engaging and evoked spontaneous involvement in joint attention. The pattern observed in subjective ratings was paralleled only by one measure in the implicit objective metrics, namely the frequency of spontaneous attentional following. These findings provide significant clues for controller design to improve the interaction between humans and artificial agents.

Keywords: behavioral sciences, biological control systems, robot control, humanoid robots, eye movements

***Corresponding Author: Davide Ghiglino:** Istituto Italiano di Tecnologia, Via Morego, 30, 16163, Genova, Italy; DIBRIS, Università degli Studi di Genova, Via Opera Pia, 13, 16145, Genova, Italy; E-mail: davide.ghiglino@iit.it

Cesco Willemse, Davide De Tommaso, Agnieszka Wykowska: Istituto Italiano di Tecnologia, Via Morego, 30, 16163, Genova, Italy

Francesco Bossi: Istituto Italiano di Tecnologia, Via Morego, 30, 16163, Genova, Italy; IMT School, Lucca, Italy

1 Introduction

The human eyes play a special role in daily interactions with others. With gaze, we efficiently communicate to others our internal mental states and our interest in the external environment [1, 2]. Sometimes, just by looking at a person's eyes, we are able to infer the other's intentions, emotions, or action plans [3, 4]. During interaction with another human, the eye region typically is the most attended of all facial features and the most used source of information, independent of the specific characteristics of the interaction [5]. In order to improve the quality of the interaction, Pelachaud and Bilvi [6] proposed a communicative model of gaze for embodied conversational agents. The authors claimed that simple variations in the temporal components of gaze behavior might induce different attributions toward the artificial agent. Other authors tested the implementation of gaze behavior in artificial agents during different activities, such as storytelling, conversation and reading (for a review, see [7]) and concluded overall that this implementation requires integration of knowledge from a large number of disciplines, from neuroscience to computer graphics. In order to design such behaviors, a better understanding of human gaze is needed. From a practical point of view, robot designers developed several tools to improve the naturalness of their artificial agents' behaviors. Biologically plausible controllers might facilitate communication during human-robot interaction. Indeed, the implementation of human-like behavior in an artificial agent might encourage the adoption of the same mental models humans spontaneously adopt towards their conspecifics. However, additional studies are needed in order to test this hypothesis and provide designers additional guidelines defining the best design of such controllers. In this context, we present a systematic approach of exploring not only the reliability of theoretical models but also participants' perception of an artificial

agent displaying various types of gaze behaviors. We use objective (eye-tracking) and subjective measures (ratings) in order to obtain a comprehensive picture of how various gaze behaviors are received by the user.

1.1 Aims

In this work, we investigate how different configurations of the same robot controller may affect cognition and attentional engagement of the user, as well as subjective impression of the robot’s human-likeness while maintaining the same task for all the conditions. We aim to provide roboticists with novel methods, grounded in cognitive psychology, for developing customizable controllers and for using effective strategies to configure the existing ones. To address the aims of our study, we filmed an iCub robot [8, 9] that was systematically manipulated to display two specific parameters of eye movements in the iCub controller: Trajectory Time (rTT) and Fixation Duration (rFD). rTT refers to the time required for the robot’s pupil to shift from one fixed position to one other fixed position in space. rFD refers to the amount of time that the robot’s pupil spends on a given target before moving again. We administered a rating scale to examine how these manipulations affected subjective attributions of human-likeness. Furthermore, to tease apart how human cognitive and attentional mechanisms are affected by these manipulations, we tracked participants’ eye movements, as eye movement patterns are closely related to attention and cognition [10]. We expected variations in rTT and rFD to affect subjective attributions of human-likeness as well as characteristic features of attentional engagement and cognitive resources related to observing the robot behavior.

2 Methods

2.1 Participants

Thirty-four participants were recruited for this experiment (mean age = 25; S.D. = 3.9 years; 21 females). All participants reported normal or corrected-to-normal vision and reported no history of psychiatric or neurological diagnosis, substance abuse or psychiatric medication. Our experimental protocols followed the ethical standards laid down in the Declaration of Helsinki and were approved by the local Ethics Committee (Comitato Etico Regione Liguria). Each participant provided written informed consent to participate in the experiment. Participants were not in-

formed regarding the purpose of the study before the experiment but were debriefed upon completion.

2.2 Stimuli

In the current study, we estimated an average human-like TT by applying the model proposed by Baloh et al. [11], using the following formula: $TT = 37ms + 2.7 * |\alpha|$, where α represents the visual angle (in degrees) between two targets. For small angles, the calculated trajectory times would be beyond the iCub physical constraints, thus we selected a visual angle of 60 degrees between the two joint positions. Based on this angle, an average TT of 200 ms was estimated. In the present study, we manipulated the velocity profile and the periodic state of TT. While the first refers to the speed of the eye movement, the latter refers to the variability displayed during the trial. Overall, we considered the implementation of the following behaviors in the iCub robot:

1. A fixed, behavior, showing no variability, calculated as an “average human behavior”.
2. A “human-range variable” behavior based on literature and human’s eye models, showing variability.
3. A “slow-range variable” behavior, designed to be considerably slower than the human-range behavior.

This design allowed us to define the effect of periodic state manipulation by comparing behavior (1) and (2), and the effect of velocity profile by comparing behavior (2) and (3). Consequently, we defined three conditions for rTT: a fixed behavior (F_{TT} during which rTT was constant, a human-range variable (HRV_{TT}) behavior and a slow-range variable behavior (SRV_{TT}). With regard to rFD, previous studies suggested that for humans, the typical pause time between two subsequent eye movements (i.e. fixation duration) is approximately 200 ms [12]. We decided to refer to this lower bound and to explore the same variability range adopted for rTT. Therefore, we adopted the same approach used for rTT, defining three conditions for rFD: a fixed behavior (F_{FD}) during which rFD was constant, a human-range (HRV_{FD}) behavior during which rFD was variable [13], and a slow-range variable behavior (SRV_{FD}). Then, we combined these two factors in a 3 X 3 repeated-measures design, generating 9 final conditions (Table 1).

Based on our experimental conditions, we implemented the nine different behaviors in the iCub robot. The behaviors were filmed using a 4K Handycam FDR-AX53 by Sony (Minato, Tokyo, Japan). Each video started with the robot looking straight-ahead (2 sec). Then, the eyes started moving from the initial position (I) to either posi-

Table 1: Summary of the experimental conditions for rTT (Robot Trajectory Time) and rFD (Robot Fixation Duration).

		rFD		
		300ms	200-400ms	400-600ms
E	200ms	$F_{rTT-F_{rFD}}$	$F_{rTT-HR_{rFD}}$	$F_{rTT-SRV_{rFD}}$
	100-300ms	$HRV_{rTT-F_{rFD}}$	$HRV_{rTT-HRV_{rFD}}$	$HRV_{rTT-SRV_{rFD}}$
	300-500ms	$SRV_{rTT-F_{rFD}}$	$SRV_{rTT-HRV_{rFD}}$	$SRV_{rTT-SRV_{rFD}}$

tion *A* (right) or *B* (left). Then, the eyes moved from one position to the other, for another ten seconds. Immediately after, the robot eyes returned to the *I* position. Subsequently, the head of the robot turned either to the right or to the left with a 70 degrees amplitude. Eventually, the head and the eyes returned to the initial position. The manipulation of occasional head movement was introduced in order to test whether participants would engage in a spontaneous attentional following (measured by fixations that would land laterally with respect to the face, in the same direction as the head movement) and whether their spontaneous attentional following would depend on the robot behavior. Importantly, for the measures of the spontaneous attentional following, the stimuli remained identical across conditions (i.e., the head movement was always the same). Therefore, any differential effects would be due to some sort of “priming” by preceding robot behavior (fixed, variable, human-range or slow-range). Each behavior was filmed twice (one starting from *I* to *A* and one starting from *I* to *B*). Consequently, we filmed 18 videos to be used for the experiment.

2.3 The iCub’s gaze controller

In this study, we used the Cartesian 6-DoF gaze controller developed for the iCub robot [14]. We implemented this to control the trajectories of the neck (T_N) and of the eyes (T_E) of the robot for looking at 3D Cartesian fixation points in space. This controller fits well with our requirements since it allows specifying the point-to-point execution time for the neck (T_N) and the eyes (T_E). Therefore, we implemented the robot’s eye movements simply by tuning the T_E parameter, considering the nine different conditions shown in Table 1. We implemented a Python script interfacing with the IGazeController Yarp class [15]. The functions for controlling both the eyes and the neck are shown in Listing 1. In the moveEyes method, we controlled the robot to move the eyes between two pre-defined fixation points for a total duration of 10 seconds. The fixation points are provided in relative angles (azimuth, elevation,

and vergence) according to the controller’s specifications. Once 10 seconds had elapsed, the moveNeck method was called for shifting the gaze in a pre-defined location (to the left or right). In this case, we released the block on the neck to let the robot rotate the head along the yaw angles.

Listing 1: Python functions for controlling the eyes and the neck of the robot in the different condition of the task.

```

def moveEyes(FD_MIN, FD_MAX, TT_MIN, TT_MAX):
    IGazeControl.blockNeckYaw()
    IGazeControl.blockNeckPitch()
    IGazeControl.blockNeckRoll()
    while True:
        fd = random.uniform(FD_MIN, FD_MAX)
        tt = random.uniform(TT_MIN, TT_MAX)
        fp = gaze_positions.next()
        IGazeControl.setEyesTrajTime(tt)
        IGazeControl.lookAtRelAnglesSync(fp)
        time.sleep(fd)

def moveNeck():
    IGazeControl.blockNeckYaw()
    IGazeControl.lookAtRelAnglesSync(YAW_POS)
    time.sleep(DISTRACTOR_TIME)
    IGazeControl.lookAtAbsAnglesSync(INIT_GP)

```

2.4 Apparatus

The experimental session took place in a dimly-lit room. Stimuli were presented on a 22” LCD screen (resolution: 1366 x 768). A chinrest was mounted on the edge of a table, in order to maintain a distance of 63 cm between participants’ eyes and the screen for the entire duration of the experiment. Consequently, the forward-looking robot’s face subtended 5.5° by 7.1° of visual angle. Whereas a range of eye-tracking techniques has previously been used in robotics research (see [16] for review), infrared / near-infrared videobased eye-tracking was deemed most suitable for our application [17, 18]. Therefore, we used a screen-mounted SMI RED500 eye-tracker by iMotion (Sensomotoric Instruments GmbH, Teltow, Germany) with a sampling rate of 500 Hz and spatial accuracy of 0.4° to record binocular gaze data. The experiment was programmed in and presented with OpenSesame 3.1.8 [19] using the Legacy backend and the PyGaze library [20].

2.5 Procedure

We instructed participants to carefully watch the videos and to evaluate, on a 6-point scale how much the behavior displayed by the robot was human-like (0=not at all, 5=extremely). Each video was repeated six-times across the whole experiment (108 trials in total) in a random order of presentation. The experiment consisted of four blocks of 27 trials to allow for self-paced breaks. During each trial, log messages were sent to the eye-tracker at the video onset, the onset of the robot head movement, and the video offset. For details on the trial structure, see Figure 1. Prior

to the task and before starting the second half of the experiment, a 9-point calibration and 4-point validation thereof were carried out (mean accuracy = 0.89°; S.D. = 0.70°). Additionally, participants were recalibrated when deemed necessary (e.g. when a participant moved their head from the chinrest).

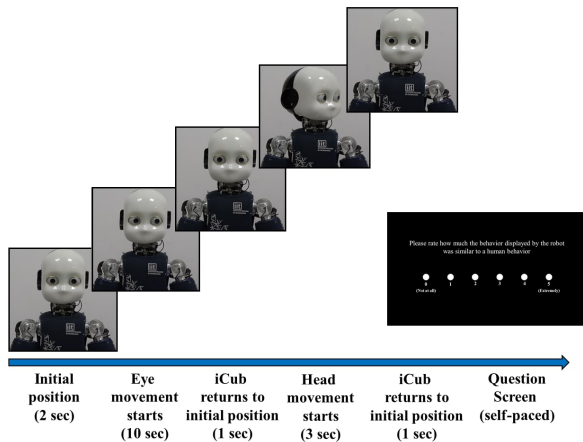


Figure 1: A planar view of the iCub robot used in this study. The picture depicts the typical trial sequence, illustrating the iCub's behavior (staggered panels) and the subsequent human-likeness rating screen (bottom-right). In this example, the robot moves from I to A and turns the head to the right.

2.6 Analyses

In order to explore the potential effects of rTT and rFD on participants' ratings, a mixed model was applied in R Studio [21]. We considered participants' responses as the dependent variable and intercept as a random factor. Then, we computed rTT and rFD as fixed factors of the model. This procedure allowed us to explore the main effects of the single factors and the effect of interaction between the two. Pairwise post hoc comparisons were estimated using the Tukey method. For the purpose of investigating whether the eye movements implemented in the iCub controller affected participants' eye movements during the videos, analyses on the eye-tracking data were performed. Three participants (mean age = 24.67; S.D. = 1.53; 1 female) were excluded from these analyses due to technical errors in the eye-tracker recordings. In order to facilitate processing of the eye-tracking data, we defined an Area of Interest (AoI): the eye region of the robot. (see Figure 2 for details). For each participant, we calculated proportional dwell time (the amount of recorded gaze samples within the AoI regardless of eye movement type) and total fixation

count to investigate where our participants attended. We examined the average fixation duration to underpin the temporal characteristics of these mechanisms on the AoI per condition between the onset of the video and the iCub's head movement. Finally, we investigated whether a fixation lateral to the face occurred within 2500 ms after the robot head movement in each trial as a measure of spontaneous gaze following. We then processed the landing position -the horizontal vector- of the first lateral fixation in the same direction as the head movement. Considering the skewed distribution of our data, we computed these metrics as the dependent variables of different Mixed Models. Each model's output provided us with predicted values of all the metrics input as dependent variables. Such predicted values derive from raw data and were corrected based on the effects taken into account in the Mixed Models. In order to maximize comprehensibility and graphical rendering of the effects, we plotted predictive values instead of the raw data. Subsequent pairwise post hoc comparisons were estimated using the Tukey method.

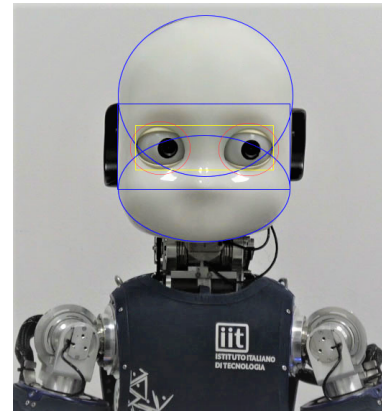


Figure 2: A frame of the trial sequence depicting the iCub robot. The figure indicates the Area of Interest (AoI) used for the analyses, defined as the yellow rectangle.

3 Results

3.1 Subjective reports

For the human-likeness ratings, we found that the slow-range variable (SRV) condition was evaluated as most human-like. We observed a main effect of condition both on rTT ($F=7.67$, $p<0.001$, Figure 3 left) and on rFD ($F=34.61$, $p<0.001$, Figure 3 right). Post hoc comparisons revealed that our participants evaluated the SRV condition signifi-

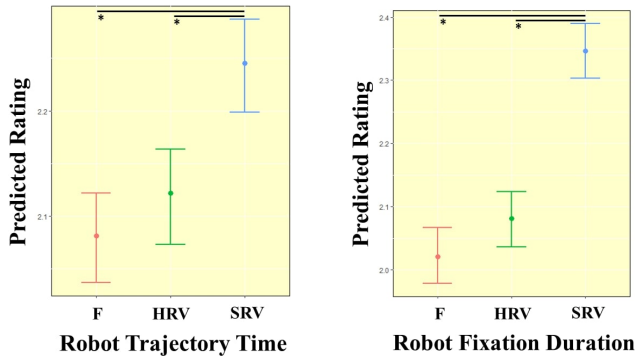


Figure 3: Human-likeness ratings across the study. Predicted values are plotted on the y-axes in order to facilitate the interpretation of results. Vertical bars denote +/- 1 standard error. Horizontal bars denote differences surviving post hoc comparison. Figure on the left shows Robot’s Trajectory Time (rTT) main effect on participants’ ratings (F: Mean=2.08, S.D.=1.30; HRV: Mean=2.12., S.D.=1.29; SRV: Mean=2.25, S.D.=1.32). Figure on the right shows Robot’s Fixation Duration (rFD) main effects on participants’ ratings (F: Mean=2.02, S.D.=1.12; HRV: Mean=2.08., S.D.=1.35; SRV: Mean=2.35, S.D.=1.33).

Table 2: Z ratio and p values resulting from the post hoc comparisons for the human-likeness ratings; asterisks denote the significant results.

	Contrast	z.ratio	p.value
Fixation Duration	F _{rFD} vs HRV _{rFD}	-1.377	0.3528
	F _{rFD} vs SRV _{rFD}	-7.803	<0.0001*
	HRV _{rFD} vs SRV _{rFD}	-6.419	<0.0001*
Trajectory Time	F _{rTT} vs HRV _{rTT}	-0.614	0.8123
	F _{rTT} vs SRV _{rTT}	-3.660	0.0007*
	HRV _{rTT} vs SRV _{rTT}	-3.046	0.0066*

cantly more human-like than the other two conditions. No significant differences were found between the fixed and human-range conditions (see Table 2 for detailed comparisons).

3.2 Objective measures: eye-tracking data

The eye-tracking data is visualized in Figure 4. Heat maps suggested the presence of differences in fixation patterns due to TT manipulation. Our results in the objective implicit eye-tracking measures showed a somewhat different pattern than the subjective explicit reports.

3.2.1 Human-range variability condition (HRV)

Importantly, the eye-tracking data showed that at the implicit level of processing, human-range variability (HRV)

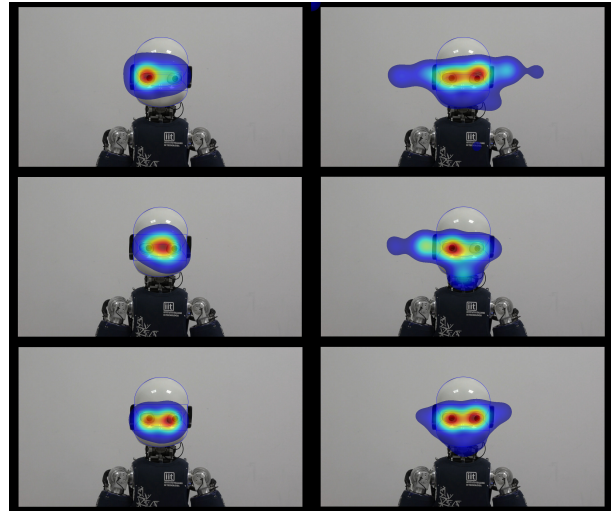


Figure 4: Fixation heat maps for each rTT condition across the experiment. On the left: trial onset until head movement; on the right: head movement until the end of the trial).

engaged participants’ attention more than the slow-range variability (SRV) – a differential effect on fixation durations, and evoked higher degree of spontaneous joint attention than the other two conditions – an effect on the range of lateral fixations and speed of attentional following. In more detail, our results showed that participants fixated on the eye region longer for the HRV condition, as compared to SRV, as evidenced by the significant main effect of rTT for the eye region (F=4.84, p=0.01) in the average fixation duration, and significant difference between HRV and SRV, (z=3.03, p=0.01), planned comparison. No significant differences were found between F and HRV (z=-2.13, p=0.08) or between SRV and F (z=0.93, p=0.63) (Figure 5, left). Furthermore, participants showed a higher degree of spontaneous attentional following in the HRV condition, relative to the other conditions: analyses on lateral fixations recorded after the robot head movement revealed a significant main effect of rFD on the horizontal vector of the first fixation location (F=4.41, p=0.01). Planned comparisons revealed a significant difference between F and HRV condition (t=-2.81, p=0.01). A marginal difference was found between HRV and SRV condition (t=2.24, p=0.07), while no differences were found between F and SRV (t=-0.58, p=0.83) (Figure 5, right).

3.2.2 Fixed-behavior condition (F)

Interestingly, also the fixed behavior condition elicited distinctive gaze patterns, compared to other conditions, which was also – similarly to the HRV condition – not

in line with the subjective ratings. Specifically, we found that participants' gaze was directed toward the eye region significantly more (in terms of dwell times) during the Fixed-behavior condition, than during the other two conditions, as evidenced by a main effect of rTT on participants' proportion of dwell times on the eye region of the iCub ($F=4.01$, $p=0.02$), and the significant differences between SRV ($z=2.48$, $p=0.04$) and HRV ($z=2.43$, $p=0.04$) conditions (Figure 6, left). Analyses did not reveal any effects of rFD on eye region dwell time or of either TT or FD on the dwell time on the whole face. Furthermore, a significant main effect of rTT was found on the number of fixations that occurred in the eye region ($F=4.46$, $p=0.01$). In the Fixed-behavior condition, the number of fixations was larger than in the HRV condition ($z=2.96$, $p=0.01$). No significant differences were found between SRV and F ($z=1.12$, $p=0.50$) or HRV and SRV ($z=-1.82$, $p=0.16$) (Figure 6, right).

3.2.3 Slow-range variable condition (SRV)

The slow-range variable condition elicited a distinctive pattern only in the frequency of instances of attentional following, and this is the only result from the implicit measures that follow the explicit subjective reports. The generalized linear model revealed a significant main effect of rTT on the occurrence of spontaneous gaze following ($\chi^2(2)=6.82$, $p=0.03$). Specifically, planned comparisons revealed a significant difference between SRV and HRV ($z=-2.46$, $p=0.04$). No difference were found between F and either HRV ($z=0.48$, $p=0.88$) and SRV ($z=-1.97$, $p=0.11$) (Figure 7).

4 Discussion

The aim of our study was to examine how various parameters of humanoid eye movements affect the subjective impression of human-likeness and attentional engagement, measured with implicit objective measures (eye-tracking). We manipulated the behavior of the iCub humanoid to display either fixed patterns (fixed trajectory times and fixed fixation durations) or variable trajectory times and fixation durations with a human-range variability (HRV) or a slow-range variability (SRV). Our results showed that the SRV elicited the highest degree of human-like impression, as reported in subjective ratings. Interestingly, when asked to elaborate on their choices, 59% of our sample reported that the "slower" behavior showing variability seemed to be more natural than the oth-

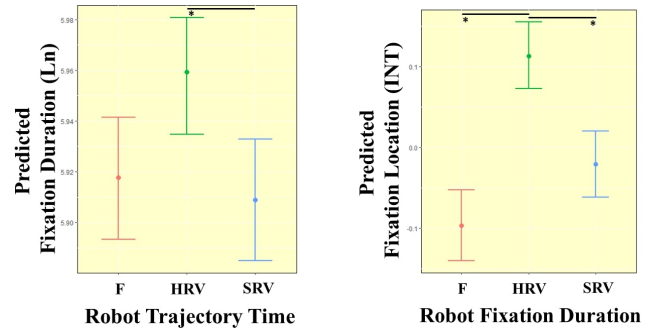


Figure 5: Fixation patterns across the study. Predicted values are plotted on the y-axes in order to facilitate the interpretation of results. Vertical bars denote ± 1 standard error. Horizontal bars denote differences surviving post hoc comparison. Figure on the left shows Robot's Trajectory Time (rTT) main effect on participants' fixation duration (F: Mean=439ms, S.D.=309ms; HRV: Mean=460ms., S.D.=321ms; SRV: Mean=428ms, S.D.=275ms). Log-transformed (Ln) predicted values are plotted on the y-axis in order to facilitate the interpretation of results. Figure on the right shows Robot's Fixation Duration (rFD) main effects on the absolute horizontal vector of the first fixation. (F: Mean=2.6°, S.D.=1.8°; HRV: Mean=4.2°, S.D.=3.0°; SRV: Mean=3.8°, S.D.=2.4°). Location was estimated as the distance from the center of the iCub face. Predicted Blom inverse normal transformed values (INT) are plotted on the y-axis in order to facilitate the interpretation of results.

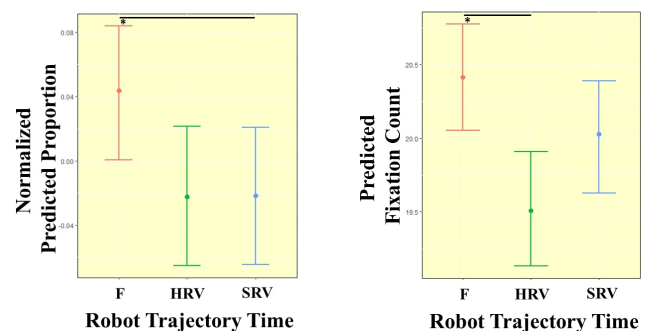


Figure 6: Fixation patterns across the study. Predicted values are plotted on the y-axes in order to facilitate the interpretation of results. Vertical bars denote ± 1 standard error. Horizontal bars denote differences surviving post hoc comparison. Figure on the left shows trajectory Time (rTT) main effect on participants' proportional dwell time (eye region) (F: Mean=0.82, S.D.=0.27; HRV: Mean=0.80., S.D.=0.29; SRV: Mean=0.80, S.D.=0.28). Normalized predicted values of dwell time are plotted on the y-axis in order to facilitate the interpretation of results. Figure on the right shows trajectory Time (rTT) main effects on participants' amount of fixations (F: Mean=20.41, S.D.=9.89; HRV: Mean=19.51, S.D.=10.11; SRV: Mean=20.03, S.D.=9.83). Predicted values of fixation count are plotted on the y-axis in order to facilitate the interpretation of results. Vertical bars denote ± 1 standard error. Horizontal bars denote differences surviving post hoc comparison.

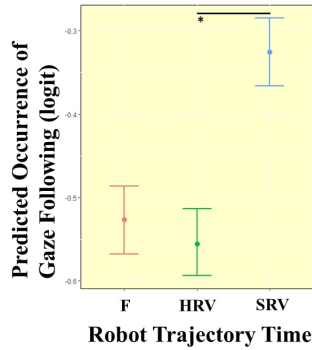


Figure 7: Spontaneous gaze following of participants during the study. Robot Trajectory Time (rTT) main effect on participants' on the proportion of participants' occurrence of gaze following behavior (F: 35.7% of total F events, HRV: 35.7% of total HRV events, SRV: 39.6% of total SRV events). Logit transformed predicted proportion is plotted on the y-axis. Vertical bars denote ± 1 standard error. Horizontal bars denote differences surviving post hoc comparison.

ers were. Some of them reported that this specific behavior seemed to be fluid, while the “faster” behavior seemed “glitchy”. We speculate that when humans approach a robot, they automatically adopt most available strategies to interpret and predict robot behavior (see [22] for a more elaborate argumentation along these lines). These strategies might be influenced by prior assumptions, knowledge¹, expectations (all not necessarily realistic) that participants have regarding how a human-like behavior looks like. Importantly for the purposes of our study, the implicit objective measures showed a different, more informative, pattern. The eye-tracking data indicated that HRV attracted more attention and evoked more attentional engagement, as evidenced by longer fixation durations on the eye region in this condition, compared to the SRV condition. Furthermore, the human-range variability affected joint attention, as participants showed a larger degree of following iCub's directional cues (further location of a lateral fixation elicited by the iCub's head movement), as compared to the other conditions. These results show that participants' implicit (perhaps more automatic) attentional mechanisms became (socially) attuned with the robot behavior when it displayed human-range variability and that this kind of behavior elicits more attentional engagement. On the other hand, the fixed, repetitive, “mechanical” behavior of the robot, although also showed a divergent pattern of results than the explicit subjective

¹ Our participants' education, for example, negatively correlated with the ratings, suggesting that the more a person might be informed about technology, science or research, the more s/he avoids attributing high human-likeness toward an artificial agent ($r=-0.35$, $p=0.04$).

measures, affected the cognitive mechanisms of participants in a different way than the HRV condition. Specifically, it induced a larger number of fixations and visits (proportional dwell times), as compared to the other conditions. This might indicate that participants “scanned” iCub's face more (showed a higher number but shorter fixations) in the mechanistic condition, perhaps because the brain perceived it as unnatural and unfamiliar behavior. This is in line with literature investigating immediateness of biological motion recognition, and suggests the existence of low-level processes that we use to discriminate biological and non-biological motion [23]. We speculate that humans require a higher amount of fixations to scan an agent displaying unnatural behaviors, while fewer fixations are needed when the behavior is biologically plausible. Finally, the pattern of results observed in the subjective ratings was paralleled by only one implicit measure, namely the proportion of instances of attentional following (Figure 7). This indicates that perhaps the general frequency of attentional following was detected at the higher-level of cognitive processes, while other, more implicit and subtle cognitive mechanisms were not. This speculation is based on the following reasoning: explicit measures pinpoint cognitive processes that are accessible to conscious awareness, hence they are higher-level than those that can be captured by implicit measures. In our study, participants reported that the SRV condition appeared most human-like. This might have been a consequence of detecting that at a lower level of processing, they followed the head movements of the robot more when it displayed a slower range of eye movements, relative to faster ranges. In contrast, the other measures (i.e. fixation duration, predicted fixation location, Figure 5) – although clear markers of attentional engagement – were too low-level to reach the conscious (and thereby reportable) level of processing. An alternative explanation might be that at the higher-level of processing, participants' responses were prone to various biases, such as assumptions regarding what constitutes a “human-like” behavior or expectations related to robot behavior. Those biases might have affected conscious reports. As a consequence, the frequency of following the head movements of the robot (Figure 7) was influenced by those higher-level biases, which would be in line with previous literature on top-down biases in attentional following [24–26]. Interestingly, the other mechanisms of attentional engagement (reflected by fixation durations and range of following) were not prone to top-down biases, as they were presumably at a much lower-level of processing. Overall, our results show that explicit subjective reports alone do not provide a comprehensive picture of cognitive mechanisms evoked by observation of (or interaction

with) a robot. Objective measures are necessary to complement subjective reports by addressing specific, and often low-level implicit cognitive processes, an argument put forward in previous literature, due to a dissociation that has been observed between explicit and implicit measures [27]. Related to robot implementations directly, the differences we found in human attentional engagement, as well as a subjective impressions evoked by superficially similarly looking conditions hint, suggests that users' interaction with a robot can be qualitatively affected by subtle differences in its behavioral design. This will be investigated further in our future work. Our findings suggest different strategies to use for the iCub' gaze controller depending on the type of interaction the scenario requires to establish with the user. TE values between 300-500 ms may evoke the impression of more 'naturalness' in the robot's movements. Faster eye movements may appear less smooth (TE values below 300 ms), but if they involve human-range variability (100-300 ms), they should evoke higher attentional engagement. In this context, the TE default value (TE = 250 ms) could be reconsidered to be increased. Importantly, the variability of TE values among different fixations is a parameter which should certainly be considered in robot behavioral design. Our results showed that added variability induces a higher impression of human-likeness, and is more attentionally engaging. A human-like range of trajectory time elicits most attentional engagement, and attunement in the form of spontaneous joint attention.

5 Conclusions

In summary, our results confirmed that both implicit and explicit measures need to be taken into account when evaluating the user's reception of a robot behavioral design. Our data show that at the level of conscious subjective impressions, the variability of behavior (trajectory times in the case of our experiment) creates the most human-like impression. Fixed-time mechanistic trajectories do not only appear as least human-like, but they also induce fragmented, scattered and short "glimpses", which might have a distracting effect on the user and impair smoothness of interaction. Finally, variable robot behavior with human-range of trajectory times attracts attentional focus most, and thereby is most engaging, even though this might not reflect in subjective conscious impressions. Throughout the present study, we proposed an approach that uses research methods from cognitive psychology to test engineering parameters. Combining

such approaches is beneficial for the future of both disciplines, by facilitating the interaction between humans and artificial agents and by improving our knowledge about ourselves.

Acknowledgments: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant awarded to A. Wykowska, titled "InStance: Intentional Stance for Social Attunement. Grant agreement No: 715058).

References

- [1] H. Kobayashi and S. Kohshima, "Unique morphology of the human eye and its adaptive meaning: Comparative studies on external morphology of the primate eye," *J. Hum. Evol.*, vol. 40, no. 5, pp. 419–435, May 2001.
- [2] N. J. Emery, "The eyes have it: The neuroethology, function and evolution of social gaze," *Neurosci. Biobehav. Rev.*, vol. 24, no. 6, pp. 581–604, August 2000.
- [3] A. Marjaninejad and S. Daneshvar, "A low-cost real-time wheelchair navigation system using electrooculography," in *Proceedings of the 22nd Iranian Conference on Electrical Engineering (ICEE)*, IEEE, pp. 1961–1965, May 2014.
- [4] M. Porta, S. Ricotti, and C. J. Perez, "Emotional e-learning through eye tracking," in *Proceedings of the 2012 IEEE Global Engineering Education Conference (EDUCON)*, IEEE, pp. 1–6, April 2012.
- [5] R. J. Itier and M. Batty, "Neural bases of eye and gaze processing: The core of social cognition," *Neurosci. Biobehav. Rev.*, vol. 33, no. 6, pp. 843–863, June 2009.
- [6] C. Pelachaud and M. Bilvi, "Modelling gaze behavior for conversational agents," in *Intelligent Virtual Agents*, T. Rist, R. S. Aylett, D. Ballin, J. Rickel (Eds.), Lecture Notes in Computer Science, vol. 2792, Springer, Berlin, Heidelberg, 2003, pp. 93–100.
- [7] K. Ruhland, C. E. Peters, S. Andrist, J. B. Badler, N. I. Badler, M. Gleicher, et al., "A review of eye gaze in virtual agents, social robotics and HCI: Behaviour generation, user interaction and perception," *Comput. Graph. Forum*, vol. 34, no. 6, pp. 299–326, 2015.
- [8] G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori, "The iCub humanoid robot: an open platform for research in embodied cognition," in *Proceedings of the 8th Workshop on Performance Metrics for Intelligent Systems*, ACM, pp. 50–56, 2008.
- [9] L. Natale, C. Bartolozzi, D. Pucci, A. Wykowska, and G. Metta, "The not-yet-finished story of building a robot child," *Sci. Robot.*, vol. 2, no. 13, eaaq1026, 2017.
- [10] H. Deubel and W. X. Schneider, "Saccade target selection and object recognition: Evidence for a common attentional mechanism," *Vision Res.*, vol. 36, no. 12, pp. 1827–1837, June 1996.
- [11] R. W. Baloh, A. W. Sills, W. E. Kumley, and V. Honrubia, "Quantitative measurement of saccade amplitude, duration, and velocity," *Neurology*, vol. 25, no. 11, pp. 1065–1070, November 1975.

- [12] T. A. Salthouse and C. L. Ellis, "Determinants of eye-fixation duration," *Am. J. Psychol.*, vol. 93, no. 2, pp. 207–234, June 1980.
- [13] T. J. Andrews and D. M. Coppola, "Idiosyncratic characteristics of saccadic eye movements when viewing different visual environments," *Vision Res.*, vol. 39, no. 17, pp. 2947–2953, August 1999.
- [14] A. Roncone, U. Pattacini, G. Metta, and L. Natale, "A cartesian 6-DoF gaze controller for humanoid robots," in *Proceedings of Robotics, Science and Systems*, 2016, pp. 18–22.
- [15] F. Paul, C. Elena, D. Daniele, P. Ali, M. Giorgio, and N. Lorenzo, "A middle way for robotics middleware," *Journal of Software Engineering in Robotics*, vol. 5, no. 2, pp. 42–49, 2014.
- [16] H. Singh and J. Singh, "Human eye tracking and related issues: A review," *International Journal of Scientific and Research Publications*, vol. 2, no. 9, pp. 1–9, 2012.
- [17] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, and H. Jarodzka, J. Van de Weijer, *Eye tracking: A comprehensive guide to methods and measures*, OUP Oxford, 2011.
- [18] L. Bell, J. Vogt, C. Willemsse, T. Routledge, L. T. Butler, and M. Sakaki, "Beyond self-report: A review of physiological and neuroscientific methods to investigate consumer behavior," *Front. Psychol.*, vol. 9, p. 1655, September 2018.
- [19] S. Mathôt, D. Schreij, and J. Theeuwes, "OpenSesame: An open-source, graphical experiment builder for the social sciences," *Behav. Res. Methods*, vol. 44, no. 2, pp. 314–324, June 2012.
- [20] E. S. Dalmaijer, S. Mathôt, and S. Van der Stigchel, "PyGaze: An open-source, cross-platform toolbox for minimal-effort programming of eye tracking experiments," *Behav. Res. Methods*, vol. 46, no. 4, pp. 913–921, December 2014.
- [21] RStudio Team, *RStudio: Integrated Development for R*, RStudio, Inc., Boston, MA, 2015. <http://www.rstudio.com/>
- [22] S. Marchesi, D. Ghiglino, F. Ciardo, J. Perez-Osorio, E. Baykara, and A. Wykowska, "Do we adopt the Intentional Stance towards humanoid robots?" *Front. Psychol.*, vol. 10, p. 450, March 2019.
- [23] W. H. Dittrich, "Seeing biological motion – Is there a role for cognitive strategies?" in *Gesture-Based Communication in Human-Computer Interaction*, A. Braffort, R. Gherbi, S. Gibet, D. Teil, J. Richardson (Eds.), Lecture Notes in Computer Science (Lecture Notes in Artificial Intelligence), vol. 1739, Springer, Berlin, Heidelberg, 1999, pp. 3–22.
- [24] C. Özdem, E. Wiese, A. Wykowska, H. Müller, M. Brass, and F. Van Overwalle, "Believing androids - fMRI activation in the right temporo-parietal junction is modulated by ascribing intentions to non-human agents," *Soc. Neurosci.*, vol. 12, no. 5, pp. 582–593, October 2017.
- [25] E. Wiese, A. Wykowska, J. Zwickel, and H. J. Müller, "I see what you mean: How attentional selection is shaped by ascribing intentions to others," *PLoS One*, vol. 9, no. 4, p. e94339, 2014.
- [26] A. Wykowska, E. Wiese, A. Prosser, and H.J. Müller, "Beliefs about the minds of others influence how we process sensory information," *PLoS One*, vol. 9, no. 4, p. e94339, 2014.
- [27] K. Kompatsiari, F. Ciardo, V. Tikhanoff, G. Metta, and A. Wykowska, "On the role of eye contact in gaze cueing," *Sci. Rep.*, vol. 8, no. 1, p. 17842, December 2018.