Directing Human Attention with Pointing

Xun Wang[1], Mary-Anne Williams[1], Peter Gärdenfors[2], Jonathan Vitale[1],
Shaukat Abidi[1], Benjamin Johnston[1], Benjamin Kuipers[3] and Alan Huang[4]

Abstract— **Pointing is a typical means of directing a human's attention to a specific object or event. Robot pointing behaviours that direct the attention of humans are critical for human-robot interaction, communication and collaboration. In this paper, we describe an experiment undertaken to investigate human comprehension of a humanoid robot's pointing behaviour. We programmed a NAO robot to point to markers on a large screen and asked untrained human subjects to identify the target of the robots pointing gesture. We found that humans are able to identify robot pointing gestures. Human subjects achieved higher levels of comprehension when the robot pointed at objects closer to the gesturing arm and when they stood behind the robot. In addition, we found that subjects performance improved with each assessment task. These new results can be used to guide the design of effective robot pointing behaviours that enable more effective robot to human communication and improve human-robot collaborative performance.**

I. INTRODUCTION

Attention is typically understood to be the cognitive process of selectively concentrating on a specific aspect of perceptions, normally the most relevant aspect, while ignoring other aspects at the same time. It is a critically important cognitive activity in humans [1], [2], [3], [4] for reacting to stimuli, achieving goals, problem solving, communicating and collaborating with others. Attention involves the allocation of processing resources [5], [2]. Humans have limited computational resources and must manage them judiciously. Paying attention not only has a computational cost, but it creates opportunity costs as well, since not paying attention to the most important object/event can have undesirable consequences, e.g. not paying attention to traffic while crossing the road can be deadly.

The ability to direct attention is critical for problem solving and the ability to direct other peoples attention plays a crucial role in facilitating collaboration. People use eye movement and gestures to direct the attention of others to objects and events. Gestures such as pointing can be used to control and influence another agents behaviour by directing their attention. Young children can point, and apes can be taught to point and often do so to signal to adults to give them food that is out of reach [6], [7], [3]. Pointing can have many meanings; it can be used to inform, request, identify, show

something to share and indicate preferences [8]. Tomasello [9], [10] argues that intention is a fundamental prerequisite for pointing and for understanding pointing [11], [12], [13].

As robots increasingly interact and collaborate with humans, they will require skills that allow them to direct peoples attention to important and relevant objects/events[1]. Determining how and when to direct peoples attention is a critically important human-robot interaction skill for robots. A whimsical robot that distracts people from important activities and/or the task at hand, will not be appreciated or, worse, could be dangerous to interact with. Pointing can be an efficient and effective tool for robots to direct peoples attention so that they can work fluently and safely with people.

This paper investigates how people comprehend a robot's pointing behaviours. It turns out that there are two widely accepted kinds of pointing behaviours: imperative and declarative pointing behaviours. Imperative pointing is basic request to give an object out of reach to the robot. Imperative pointing is sometimes viewed as an extension of power, as well as a form of communication e.g. "I want that". Declarative pointing, on the other hand, allows a robot to direct the attention of a human towards a specific object/event. Declarative pointing is a complex behaviour because it requires a robot to confirm that the human has comprehended their pointing behaviour successfully, e.g. it is looking at the object/event that the robot is trying to direct their attention towards and in some cases achieve joint attention [17]. Joint attention has been identified as fundamental to human relationships that rely on the sharing of experience and knowledge [7] and plays a critical role in collaboration.

We present a quantitative experiment in which people are asked to determine the target of a robots pointing gesture. It provides new evidence and insights that can be used to design efficient and effective pointing gestures in robots. Section 2 describes our human-robot interaction experiment. The empirical results are presented in Section 3 and analysed in Section 4. The final discussion, design guidelines and ideas for future human-robot interaction experiments are given in Section 5.

[1]Xun Wang *et al.* are from Social Robotics Studio, Centre of Quantum Computation and Intelligent Systems, University of Technology, Sydney, Australia xun.wang-1@uts.edu.au

[2]Peter Gärdenfors is with Cognitive Sciences, Lund University Peter.Gardenfors@lucs.lu.se

[3]Benjamin Kuipers is with Intelligent Robotics Laboratory, University of Michigan kuipers@umich.edu

[4]Alan Huang is now with School of Mathematics and Physics, The University of Queensland alan.huang@uq.edu.au

[1]Robots can comprehend human pointing by design[14]. Abidi *et al.* [15] developed a system that responds to human pointing behaviours by interpreting the angle of a human's arm; whereas Rich *et al.* [16] used a visual hand tracking system to interpret human pointing.

II. MEASURING HUMAN COMPREHENSION OF POINTING

There is an important body of work studying the comprehension of gestures such as pointing in human-human interactions [11], [2] and this work is critical to understanding human comprehension of robot gestures. In human pointing experiments, the comprehension criterion is typically measured by extending or pivoting a body part such as the hand or head, towards an object/event of reference [2], [18], asking subjects to identify the referenced object/event and then measuring accuracy. Comprehension of pointing depends on whether it is proximal or distal [19], [3]. Proximal pointing occurs when the target is touched by the pointing device (e.g. stick, finger) and, as such, presents little opportunity for ambiguity. Distal pointing occurs when the target is remotely situated in relation to the pointing device [19]. The comprehension of distal pointing can be measured by the so-called detection accuracy that subjects exhibit when they locate the target of a pointing gesture [20]. This spatial location task requires a subject to visually discriminate and identify the target of a pointing gesture [5]. It turns out that detection accuracy is well known for eye gaze, but Butterworth and Itakura [20] found detection accuracy to be low for distal pointing gestures.

In a human-robot scenario, St. Clair, Mead and Mataric [21] undertook an experiment to investigate the effect of visual saliency in pointing. They used a humanoid robot making deictic gestures in front of a subject. A transparent panel was placed between the robot and the experiment participant. The experimenters asked to the subjects to identify the point on the transparent screen referenced by the robot pointing in two scenarios: an empty panel (no salient objects), and a panel with some randomly placed round marker (presence of salient objects). They found that environmental saliency, in this specific form of human-robot interaction, leads only to minimal improvements on the detection of the target.

In contrast to the work of St. Clair, Mead and Mataric [21], other studies have focused more on measuring the human comprehension of robot pointing in open space. Hato *et al.* [22] conducted an experimental investigation into the deictic interactions that reference spacial regions. They found that people use two kinds of pointing models for referring to regions in the space: using the index finger maintaining stable the arm in a direction, and using the index finger reproducing a circle through the movement of the arm in a circular way. The authors used the observed human behaviour for the robot used in the experiment, in order to let the robot refer to specific areas in an open space. The results show that a robot with region-cognition, namely a limited well known list of the possible regions in the space, can express the referred region better than without this region-cognition of the considered space. Furthermore, they found that the dimension of the considered region has an impact on the detection performance, namely a large size region is recognised better than a small one. Williams *et al.* [23] undertook an experiment using a PR2 robot in a static

pointing pose in a public building to study the way humans interpret robot pointing. This experiment was designed to test whether people interpret robot pointing in the same way as they interpret people pointing reported in Wnuczko and Kennedy [24]. They found that the majority of people use the arm angle to determine the pointing target and that the head angle of the PR2 robot plays a significant role in influencing how people determine the target of a static PR2 robot pointing posture.

These previous experiments have established a number of important aspects of human comprehension of robot pointing. In the following section we describe a new laboratory based human-robot interaction experiment. This experiment draws inspiration from the human-human interaction experiment proposed by Bangerter [19] but we collect more precise data using an iPad interface. Since Williams *et al.* [23] found the elongated head shape of the PR2 influenced human comprehension our experiment uses a NAO robot whose head has a circular side profile rather than an elongated rectangular side profile like the PR2 head. Furthermore, contrary to Bangerters experiment using a single line of targets, the visual stimuli are placed in a two-dimensional grid introducing more complexity and realism to the task [19].

III. THE HUMAN-ROBOT INTERACTION EXPERIMENT

A. The Objective of the Experiment

The objective of the experiment is to investigate how people interpret dynamic distal robot pointing behaviours for the purpose of designing efficient and effective robot pointing behaviours in human-robot interaction. The experiment used a NAO robot that, starting from a neutral stance, dynamically pointed towards a target randomly chosen from a grid of 12 possible targets. Each subject participated in 10 trials, attempting to identify the target of the pointing gesture with each trial.

B. The Robot Platform used in the Experiment

A NAO humanoid robot was used as the robot platform in the experiment. The NAO is an inexpensive and programmable humanoid robotic platform that stands approximately 60cm tall (see Figure 1). It has 25 degrees of freedom in total and four degrees of freedom in each arm. We used the NAO RoboCup edition, meaning that it does not have the ability to control individual fingers. We used Aldebaran's proprietary middleware, NAOqi, to develop a pointing system that points towards targets at known positions in 3D space, and then communicates gesture and timing information over a wireless network to an iPad user interface.

C. The Experimental Hypotheses

The experiment was designed to test four hypotheses regarding human comprehension of dynamic robot distal pointing gestures using the method of measuring detection accuracy of human subjects. The first hypothesis provides an important baseline for human interpretation of dynamic

robot pointing gestures. We would like to establish that pointer gesture can be an effective tool for directing people's attention.

H1: Humanoid robot pointing gestures can direct human's attention and humans can interpret the gesture without additional cues.

The second hypothesis concerns the best position for a person to stand when trying to comprehend the target of a robots pointing gesture. This hypothesis investigates the importance of where a person trying to comprehend a robot pointing gesture should stand relative to the robot. Testing this hypothesis will inform the design of robot positioning in robot pointing behaviours.

H2: Humans identify the robot target object more accurately when standing on the same side of the robot as the arm that the robot uses to point.

The third and fourth hypotheses focus on which arm is more effective in pointing to different sides of the body and testing them will inform which arm a robot should use when directing human attention. The third hypothesis tests comprehension accuracy and the fourth comprehension speed. The third hypothesis is used to test if higher levels of detection accuracy can be achieved when the robot uses the arm on the same side as the target. The fourth hypothesis is used to test if subjects can determine the target of robot pointing gestures faster when the robot uses the arm on the same side as the target.

H3: Humans find the robot target object more accurately when it lies on the same side as the arm the robot uses to point.

H4: Humans find the robot target object more quickly when it lies on the same side as the arm the robot uses to point.

With the fifth hypothesis, we seek to confirm our expectation that there is a familiarity effect: that observers improve their comprehension of pointing gestures, even if the system does not provide feedback on their accuracy.

H5: Human comprehension of robot pointing gestures improves in less than 10 observations.

D. Experimental Design

The experiment was conducted over the course of a university open day, in which prospective students, their families and other interested members of the public were invited to learn about the university's course offerings and research. As such, our experiment participants were drawn from a broad cross-section of society but that would be biased towards individuals and families interested in the university's research and offerings. That is, our sample population was biased towards young adult age group. The participants were indiscriminately recruited based on solely their willingness to participate with the experiment.

In the experiment, 12 (3x4) green circles labeled with the letters A to L were projected onto the wall (Fig.2). These labeled green circles were shown in the same configuration on an iPad, used as the input device for the human participants. The NAO robot was placed at a fixed position in front of the screen. A human subject was instructed to stand in one of three fixed positions (Left, Right or Behind) next to the robot (see Fig.1). The subject was then briefed on the procedure for the experiment and the operation of the iPad interface.

The experiment started with the robot randomly selecting and pointing at one of the green circles on the screen with its right arm. The human subjects then identified the target circle that the robot was pointing towards. The participants were able to answer the question by tapping one of the circles displayed on the iPad. If a participant could not identify the target within the maximum allowed time period of ten seconds, we interpret this at the subject unable to identify the target. We provided audio and visual cues to the participants to indicate the beginning and the ending of answering session. The response time of the participant was recorded (measured in seconds from the end of the visual cue to the selection of a green circle on the screen of the iPad by the subject). Once the answer was recorded, the robot moved its arm to a rest position next to its leg, pointing downwards. The above stated steps were repeated ten times for each human subject.

At the end of the experiment we asked the participants to answer a brief questionnaire about their gender and age group. We considered five age groups: 5 to 15, 16 to 25, 26 to 35, 36 to 55 and above 55. As our interest was oriented towards understanding the relationships between the robot pointing behaviour and human ease of comprehension, we asked the subjects if they had any free comments about the difficulty found in the proposed task. Furthermore, we asked for free comments about the perceived level of confidence when determining the target of the robot pointing. The entire experiment required about 5 to 10 minutes per participant.

Fig.1, shows the spatial configuration of the experiment. This particular spatial setup is constrained by the physical parameters of the room used for the experiment and the angles that could be achieved by the robots arm since the robot had to point towards all 12 targets projected on the wall.

Fig. 2, shows the relative size and separation of the graphical objects displayed on the projector screen. The robot was pre-programed to randomly point at one of these marks from its fixed position.

IV. EMPIRICAL RESULTS

In this section we report the experimental results for 36 untrained experiment subjects. The gender, age and location relative to the robot of the 36 subjects was noted. There were 21 male subjects and 15 female subjects across four age groups: 5 subjects in the 5-15 years of age range; 26 subjects in 16-25 years: 2 subjects in 26-35: 3 subjects in 36-55 year range. Each subject was asked to stand in one of three positions. The total number of subjects in each position was: Left: 15; Behind: 13; Right: 8.

The detection accuracy was measured for each subject. On average, the subjects scored 31% accuracy over ten trials (three out of ten correct answers) with a standard deviation of
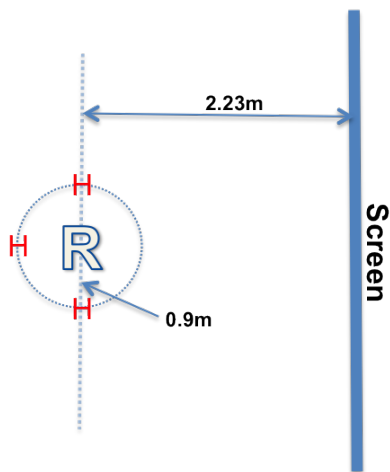
Fig. 1. The NAO robot was positioned at **R**. 2.23m from the screen on a wall where the targets were projected. Human participants were placed to the left, right and behind the robot (H), and asked to determine the target of the pointing gesture.
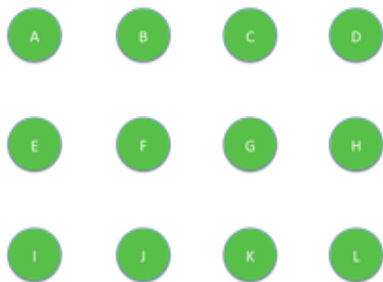

Fig. 2. The configuration of the target objects that the robot directed human subjects attention toward.

Fig. 3. The experimental set up: (a) NAO humanoid robot in a neutral stance with subject holding iPad standing behind the robot, and (b) NAO humanoid robot in pointing stance with subject holding iPad standing to the left of the robot. (Photos by Pramod Parajuli).

1.5. Grouped by gender the accuracy was 32% for males and 30% for females: 30%. Grouped by age, the accuracy for 5-15 year olds was 26%, for 16-25 years olds was 34%, for 26-35 year olds was 15% and for 36-55 year olds the accuracy was 33%. Grouped by location relative to the robot: the accuracy was 32% for subjects placed to the left of the robot, 34% for subjects behind the robot, and 26% for subjects to the right. The highest number of correct targets identified by a subject was 6 out of 10 (3 out 36 subjects achieved this level of success). Every subject made at least one correct answer. Across all 360 pointing events, the subjects were unable to determine any target on only 3 occasions. The subjects detection accuracy of robot pointing towards the 12 green circle targets was measured during the experiment and given in Table I, below.

TABLE I
SUBJECT DETECTION ACCURACY TASK FOR EACH TARGET.

A: 0.16	B: 0.25	C: 0.50	D: 0.90
E: 0.13	F: 0.10	G: 0.22	H: 0.44
I: 0.00	J: 0.04	K: 0.34	L: 0.55

The time taken for each subject during the detection accuracy task was measured in seconds and recorded. The overall average time to respond was 2.51s. Grouped by gender: the average response time for males was 2.68s, and for females was 2.72s. Grouped by age: the average response time for subjects between 5-15 was 2.17s; for the 16-25 age group the average response time was 2.37s; for the 26-35 age group the average response time was 2.51s; for the 36-55 age group the average response time was 4.20s. Grouped by location relative to the robot: Left: 2.68s; Behind: 2.72s; Right: 2.19s. The average time taken for the detection accuracy task for each of the 12 green circle targets was measured during the experiment and given in Table II, below.

TABLE II
LENGTH OF TIME TAKEN BY SUBJECTS DURING THE DETECTION
ACCURACY TASK.

A: 2.60	B: 2.84	C: 2.30	D: 1.57
E: 2.70	F: 2.55	G: 3.10	H: 1.92
I: 2.41	J: 2.77	K: 3.16	L: 2.07

V. EXPERIMENT DATA ANALYSIS

The experimental results show that people can interpret pointing gestures because 93% of all targets identified by subjects during the experiment were deemed correct or directly adjacent to the target (vertical, horizontal and diagonally) and 31% of answers were precisely correct. Less than 1% (3 out of 360 targets took more than 10 seconds to determine) of the robot pointing behaviours were incomprehensible to the subjects. This provides strong support for hypothesis **H1** that humanoid robot pointing

can direct human's attention and human *can* interpret the pointing gesture, especially where the required gesturing precision is similar to or greater than the degree separation between adjacent circles used in the experiment. The less than expected accuracy of the human interpretation is due to the small physical size of the robot and the fact that our RoboCup NAO does not have movable fingers and can only point using its fists. Low accuracy has little impact on the validity of **H1** since accuracy was not tested for **H1**, only the ability of subjects to comprehend a robot's pointing gesture when it raises its arm and extends in a direction.

For hypothesis **H2**, **H3** and **H5**, the subjects pointing interpretation accuracy was analysed using a statistically based binomial mixed model and allowed for correlation within each participant. For observation from participant we model the probability of correct detection using the following equation,

$$Prob(Correct_{ij}) = \frac{e^{X_{ij}^T\beta+\mu_j}}{1+e^{X_{ij}^T\beta+\mu_j}}, i=1,2,...,10, j=1,2,...,36$$
$$\mu_j \sim N(0,\sigma_U^2), j=1,2,...,36.$$

The model is fit in R environment[2] with the glmer function in the lme4 package using a model that incorporates fixed effects gender, age, position, target, accuracy, ordering, subjective self-evaluations and an individual effect factor. If any estimated effect has a p-value between 0.1 and 0.05, we say that the effect is marginally significant. If any estimated effect has a p-value of less than 0.05, we say that the effect is "significant".

Our major findings are as follows: (i) there were no significant differences in gender in terms of interpretation accuracy, (ii) accuracy rates for each subject increased as the experiment went on (significant effect), thereby confirming hypothesis **H5**, and (iii) the subject location with the best detection accuracy was behind the robot; being right of robot was worst and the difference is marginally significant. Thus, we failed to confirm **H3** and have weak evidence for the converse (i.e. that a more distant observation position is more effective for interpreting robot gestures.

We found significant evidence to support **H2**, that humans identify the robot target object more accurately when standing on the same side of the robot as the arm that the robot uses to point. In particular, the most easily identified target was D located at the top right corner of the screen with a success ratio of 0.9. The least identifiable target was I located at the bottom left corner of the screen, no subject identified it correctly. Subjects made more accurate interpretations when the robot pointed to a target that is on the same side as its pointing arm, i.e. as the robot used the right arm to point, participants gained higher performance when identifying targets on the right side of the screen. The targets in order of easiest to correctly detect are: D,L,C,H,K,B,G,A,E,F,J,I.

8	6	3	1
9	10	7	4
12	11	5	2

This analysis shows that subjects exhibited higher levels of interpretation accuracy for targets that lie on the same side as the robots pointing arm. Interestingly, an additional observation we made was that after adjusting for gender, age, position, total questionnaire scores, time order and target circle, the person-to-person variability is quite small. That is, recognition accuracy tends not to be an individual characteristic, but something that can be predicted from demographic information of a participant.

To evaluate **H4**, that humans find the robot target object more quickly when it lies on the same side as the arm the robot uses to point, we constructed a linear mixed model analogously to that used previously for **H2**, **H3** and **H5**, that also allowing correlations within each participant. That is, for observation from participant we model the probability of correct pointing interpretation by the following formula,

$$Time_{ij} = N(X_{ij}^T\beta+\mu_j,\sigma^2), i=1,2,...,10, j=1,2,...,36$$
$$\mu_j \sim N(0,\sigma_U^2), j=1,2,...,36.$$

Again, this model is fit in R with the glmer function using a model that incorporates fixed effects gender, age, position, target, answer time, ordering, subjective self-evaluations and an individual effect factor.

Major findings are as follows: (i) there were no significant differences in gender for time taken to complete the interpretation task, (ii) the 36-55 age group took the longest time to answer, while the 5-15 age group took shortest time and the difference was marginally significant, (iii) there was no significant differences in response time for the three different positions, (iv) for each subject, the time to respond shortened with each successive detection task with an average 0.63 second reduction in response between the first and last detection tasks. The targets in order of shortest response time are: D,L,H,C,I,E,K,J,B,A,F,G. This order is presented in Table IV, below, and is similar to that for interpretation accuracy.

10	9	4	1
6	11	12	3
5	8	7	2

The difference in recognition time between the targets on the left and right hand side of the grid is significant, lending support for **H4**. Interestingly, after adjusting for gender,

age, position, total questionnaire scores, target circle and correctness of target, the person-to-person variability is quite large; some people just take longer (or shorter) to answer!

VI. DISCUSSION

This paper presents the first quantitative experiment that investigating whether a NAO robot can direct human's attention through dynamic-distal pointing behaviours, how a human interpret the pointing behaviour and factors that influence human interpretation. The experimental results support our initial hypotheses, **H1**, **H3**, **H4** and **H5**. The experiment confirmed that NAO robot can direct human's attention through pointing and human can interpret the pointing behaviours, since the human subjects were able to identify the targets correctly or directly adjacent to the correct targets. Only in 3 out of 360 trials, i.e. less than 1% human subjects were unable to comprehend the robot pointing. However, the average accuracy of identifying correct targets is still low. We believe the small size and limited pointing capability of the NAO robot are two main factors for this low accuracy. In the next experiment, we plan to use a human size robot to verify this hypothesis.

The experiment showed that humans can comprehend robot pointing both more accurately and more quickly when the robot uses the arm closest to the target. This result has practical use in helping human to better understand humanoid robot pointing. That is, robot pointing behaviours should be designed to use the arm that lies on the same side of the robot body as the target to help human comprehension of robot pointing.

The experiment also revealed a familiarity effect: humans tend to get better at recognising pointing gestures without any feedback on their performance. This improvement in human interpretation of robot pointing appear in very limited exposure time to the robot. This means, humans can quickly self-adapt to understand robot pointing and this provides further evidences for robot pointing as a viable communication tool for human-robot interaction.

Finally, the experiment did not find any support for **H2**. Instead, our analysis shows that the most effective place to stand when determining the pointing target of a robot is behind the pointing arm. This discovery can be exploited in robot pointing behaviour design and it has significant impact in face-to-face human-robot interaction when humans and robots are in close proximity and the robot is attempting to direct a humans attention.

The experiment presented in this paper is our initial experiment on understanding human's comprehension of humanoid robot pointing. Our data sample population was relatively small and may be biased towards particular age groups. In the future, we will conduct larger scale experiments with better experimental controls. We will conduct geometric analysis of the spatial relationship between human participants, robot and markers and how this relationship may effect human interpretation of robot pointing. We may use different visual markers and introducing audio cues to study their effects.

Finally, we plan to carry out comparison studies on different types of humanoid robots.

REFERENCES

[1] L. Hay, "Accuracy of children on an open-loop pointing task," *Perceptual and Motor Skills*, vol. 47, no. 3f, pp. 1079–1082x, 1978.

[2] A. Kendon, "Do gestures communicate? a review," *Research on language and social interaction*, vol. 27, no. 3, pp. 175–200, 1994.

[3] D. J. Povinelli, J. E. Reaux, D. T. Bierschwale, A. D. Allain, and B. B. Simon, "Exploitation of pointing as a referential gesture in young children, but not adolescent chimpanzees," *Cognitive Development*, vol. 12, no. 4, pp. 423–461, 1997.

[4] R. D. Wright and L. M. Ward, *Orienting of attention.* Oxford University Press, 2008.

[5] J. R. Anderson, *Cognitive psychology and its implications.* Macmillan, 2005.

[6] Y. Bruinsma, R. L. Koegel, and L. K. Koegel, "Joint attention and children with autism: A review of the literature," *Mental retardation and developmental disabilities research reviews*, vol. 10, no. 3, pp. 169–175, 2004.

[7] A. C. MacPherson and C. Moore, "Attentional control by gaze cues in infancy," *Gaze-following: Its development and significance*, pp. 53–75, 2007.

[8] L. Schilbach, M. Wilms, S. B. Eickhoff, S. Romanzetti, R. Tepest, G. Bente, N. J. Shah, G. R. Fink, and K. Vogeley, "Minds made for sharing: initiating joint attention recruits reward-related neurocircuitry," *Journal of Cognitive Neuroscience*, vol. 22, no. 12, pp. 2702–2715, 2010.

[9] M. Tomasello, "Why dont apes point?" in *Roots of Human Sociality: Culture, Cognition and Interaction*, N. J. Enfield and S. C. Levinson, Eds., 2006, pp. 506–524.

[10] ——, *Origins of human communication.* MIT press Cambridge, 2008.

[11] I. Brinck, "Attention and the evolution of intentional communication," *Pragmatics & Cognition*, vol. 9, no. 2, pp. 259–277, 2001.

[12] P. R. Cohen and H. J. Levesque, "Intention is choice with commitment," *Artificial intelligence*, vol. 42, no. 2, pp. 213–261, 1990.

[13] L. Kopp and P. Gärdenfors, *Attention as a minimal criterion of intentionality in robots.* Lund University, 2001.

[14] K. Nickel and R. Stiefelhagen, "Visual recognition of pointing gestures for human-robot interaction," *Image and Vision Computing*, vol. 25, no. 12, pp. 1875–1884, 2007.

[15] S. S. Raza Abidi, M. Williams, and B. Johnston, "Human pointing as a robot directive," in *Proceedings of the 8th ACM/IEEE international conference on Human-robot interaction.* IEEE Press, 2013, pp. 67–68.

[16] C. Rich, B. Ponsler, A. Holroyd, and C. L. Sidner, "Recognizing engagement in human-robot interaction," in *Human-Robot Interaction (HRI), 2010 5th ACM/IEEE International Conference on.* IEEE, 2010, pp. 375–382.

[17] N. Eilan, C. Hoerl, T. McCormack, and J. Roessler, *Joint Attention: Communication and Other Minds: Issues in Philosophy and Psychology: Issues in Philosophy and Psychology.* Oxford University Press, 2005.

[18] D. McNeill, *Gesture and thought.* University of Chicago Press, 2008.

[19] A. Bangerter and D. M. Oppenheimer, "Accuracy in detecting referents of pointing gestures unaccompanied by language," *Gesture*, vol. 6, no. 1, pp. 85–102, 2006.

[20] G. Butterworth and S. Itakura, "How the eyes, head and hand serve definite reference," *British Journal of Developmental Psychology*, vol. 18, no. 1, pp. 25–50, 2000.

[21] A. St Clair, R. Mead, and M. J. Mataric, "Investigating the effects of visual saliency on deictic gesture production by a humanoid robot," in *RO-MAN, 2011 IEEE.* IEEE, 2011, pp. 210–216.

[22] Y. Hato, S. Satake, T. Kanda, M. Imai, and N. Hagita, "Pointing to space: modeling of deictic interaction referring to regions," in *Proceedings of the 5th ACM/IEEE international conference on Human-robot interaction.* IEEE Press, 2010, pp. 301–308.

[23] M.-A. Williams, S. Abidi, P. Gärdenfors, X. Wang, B. Kuipers, and B. Johnston, "Interpreting robot pointing behavior," in *Social Robotics.* Springer, 2013, pp. 148–159.

[24] M. Wnuczko and J. M. Kennedy, "Pivots for pointing: Visually-monitored pointing has higher arm elevations than pointing blindfolded." *Journal of Experimental Psychology: Human Perception and Performance*, vol. 37, no. 5, p. 1485, 2011.