**(abstract)**

Recent research provides some compelling evidence that our own motor system plays an important role in predicting the outcome of observed actions by others in our environment. It has been argued that this ability to simulate the actions of others is the key to understanding the mind of others. Indeed, an increasing number of studies have shown that humans spontaneously and unconsciously keep track of the belief states and perspective of others. However, how both systems interact and what the role is of motor simulation in mentalising is still poorly understood. In this study we propose a new paradigm that allows us to study both predictions based on motor cues and on mentalising. Specifically, we looked at anticipatory looking while an agent reached for an object. Crucially, the belief of the agent on the location of this object could be true or false. Results show that participants used motor cues from the agent to predict where the agent would reach for the object. Moreover, we found evidence suggesting that participants also kept into account the beliefs of the agent on the location of the object, modulating the social predictions based purely on motor cues.

**General introduction**

As a social species, coordinating our actions with others is a fundamental part of our day-to-day life. We are constantly cooperating in joint action with other people at work, doing sports, dancing, playing games or having a chat. As such, we constantly use social signals to anticipate the actions of our interaction partners and infer the mental states (such as perceptions, desires and beliefs) underlying these actions (Frith & Frith, 2007). It has been argued that we do so by implementing predictive coding principles in observing the actions of others. We track the movements of our interaction partners using our own motor system (Rizzolatti, Camarda, Fogassi, Gentilucci, Luppino, and Matelli, 1988; Gallese & Goldman, 1998; Rizzolatti, Camarda, Fogassi, Gentilucci, Luppino, & Matelli, 1998; Rizzolatti & Sinigaglia, 2016). Tracking low-level motor cues related to the temporal and spatial properties of observed actions allows us to predict how these actions are likely to unfold, in turn allowing us to derive the either the end state or goal of the action of the action (Blakemore & Decety, 2001; Sartori, Becchio, & Castiello, 2011; Vaziri-Pashkam, Cormiea, & Nakayama, 2017). Moreover, action observation is not a passive process but involves actively sampling the environment for relevant social cues (Gredebäck & Falck-Ytter, 2015; Donnarumma Costantini, Ambrosini, Friston & Pezzulo, 2017). As such, one simultaneously samples information for different cues, such as objects in the environment, eye gaze of the interaction partner, biological movement. This entails that we need to integrate social cues from multiple sources of information, of which some might be ambiguous or even conflicting (Ambrosini, Pezzulo, & Constantini, 2015).

So far the focus has been on situations whereby the action is going to be a successfully achieved unless the world is uncooperative. But there are also situations in which agents have false beliefs about which actions are needed. For example, if you falsely believe that your favourite mug is on the top shelf, you may reach up with the goal of grasping your mug even though achieving this goal would actually require you to reach down to a lower shelf. This creates a challenge for action observation: if we are to predict how an action will unfold, we cannot always rely on how things actually are but must also take into account what the agent believes.

For example, Alice is in the office looking at her monitor. While her eyes are fixed on the screen, we see Alice opening her hand to form a grip and reaching next to her. What is she up to? Merely simulating those movements is insufficient to infer the goal of that action behind it, as this is an action typically related to grasping an object in the environment. We can’t rely on her eye gaze to infer the target object, as she is looking at the screen. Instead, based on low-level motor cues simulated in our own motor system we would derive an approximate target location of her reaching movement. We would therefore make a predictive eye movement toward that location, sampling that location for relevant information. As such, we would infer that the most likely target of her action is her coffee mug based on her grip and movement and understand that she intends to take a sip.

However, reality does not always conform to our prior expectations and sometimes we hold false beliefs on the world around us. Suppose that just before Alice wanted to drink some coffee from her mug, her colleague Dan moved the mug somewhere else, unbeknownst to Alice. She would still reach for the mug although it isn’t there. This creates a challenge for action observation: if we are to predict how an action will unfold, we cannot always rely on how things actually are but must also take into account what the agent believes. It seems that humans have adapted to this challenge through the development of a social-cognitive system that is able to rapidly track different perspectives and belief states, and even when belief-tracking is completely irrelevant for the task at hand (e.g., Samson, Apperly, Braithwaite, Andrews & Scott, 2010; Kovács, Téglás, & Endress, 2010; Grainger, Henry, Naughtin, Comino, & Dux, 2018; Schneider, Bayliss, Becker, & Dux, 2012; Schneider, Slaughter, Becker, & Dux, 2014). Some have argued that the mentalising system is in fact an extension of the mirror-neuron system or is at least heavily dependent on it (e.g., Gallese, 2001; Keysers & Gazzola, 2007). However, the relationship between the action-observation system and the mentalising system is still poorly understood.

As such, this study aims to establish a behavioural paradigm in which we are able to look at both action-observation and mentalising within the same task. There is a considerable body of research indicating that observing a hand pre-shaping before grasping an object elicits fast anticipatory looking to the target matching the hand shape, in line with the notion of efficient behavioural prediction based on motor simulation (see Ambrosini, Constantini, & Sinigaglia, 2011; Ambrosini, Sinigaglia, & Constantini, 2011; Constantini, Ambrosini, & Sinigaglia, 2012; Constantini, Ambrosini, Cardellicchio, & Sinigaglia, 2014). In this study, we expand this paradigm to include to include a case where the agent has a false belief about the location of the target to be grasped. Therefore, we investigated if anticipatory looking behaviour would reflect the integration of motor cues and belief tracking in the case where the observed agent holds a false belief on the target to be grasped. Specifically, we hypothesized that if action observation relies on motor representations (that activate corresponding motor plans in the observer) there should be grip-congruent anticipatory looking behaviour in the true-belief condition. Secondly, if action observation relies on motor representations that can also be modulated by belief tracking, there should be belief-congruent looking behaviour in the false-belief condition.

**Experiment 1**

*1. Methods*

*1.1. Participants*

A total of 43 participants participated in this study (*M* = 19.35, *SD* = 2.36, 29 females). All provided informed consent and had normal to corrected vision, and no history of any neurological pathology. The experimental procedure received ethical approval from the Victoria University of Wellington School of Psychology Human Ethics Committee.

*1.2. Stimuli and Apparatus*

Stimulus presentation was done through videos. All videos were presented in a 1920 x 1080 pixel MP4 format on a 23-inch monitor (510 x 286 mm, 39.91 x 23.09 degrees of visual angle) at 30 FPS. Every video lasted for exactly 698 frames (23.267 seconds). The entire experiment was run through custom Python software, using the Psychopy and Pylink libraries (Peirce & MacAskill, 2018). The monitor was positioned in such a way that the top of the screen was elevated 450mm above the table surface. The headrest was positioned in such a way that the chin was positioned 300mm above the table surface, and the forehead at 450mm above the table surface. The distance between the headrest and the monitor was 700mm. We used an SR Research Eyelink 1000 eyetracker at 500 Hz, which was positioned just underneath the monitor at a distance of 540mm to the headrest. We used this eyetracker to record data from the left eye of the participant. Participants made responses to the primary task using a Psychology Software Tools Serial Response Box.

There were 12 experimental videos in total, crossing belief (true or false) with hand shape (whole hand grip, precision grip, no grip) and counterbalancing each cup’s location (left or right). We also incorporated a visual detection task into each video; the cups could change colour (from the original orange to bright purple) for 10 frames (333ms) once per video and participants were required to respond to the change in colour by pressing the central button on a response box. The sequence of each video unfolded in 5 phases.

To illustrate, consider a false-belief test trial (see Figure 1). In phase 1, the ledge with the cups (locations of small and large cups counterbalanced) was on the ground level. The agent could not visually perceive the cups when the ledge was on the ground. Then the ledge was raised so that the cups were perceptually accessible to the agent, inducing the agent’s belief towards the objects’ locations (<small-cup, left-side> and <large-cup, right-side>). The ledge was lowered back to the ground in phase 2, whereupon the agent was no longer able to perceive the cups. Once lowered, both cups moved to the middle of the screen. The cups could either swap positions resulting in the agent having a false-belief of the objects’ locations or return to their original positions, maintaining the agent’s true-belief of their whereabouts (see Figure 2). At the end of the cups’ movements, a fixation arrow (with accompanying beep sound) appeared on the agent’s resting hand and was presented for 52 frames (1733ms). Participants were instructed to fixate on the arrow for the duration of its presence.

The occurrence of the visual detection task was restricted to very early on in the video sequence to reduce the impact of the visual detection task on the looking behaviour later on in the video and occurred well before the fixation arrow was presented. Specifically, we divided the video sequence before the fixation arrow presentation into 4 segments of equal length. The visual detection task could therefore occur at 118ms, 236ms or 354ms after the start of the video. Timing of the visual detection task was randomised and uniformly distributed over the experiment.

In phase 3, after the disappearance of the fixation arrow, hand-shaping was manipulated. The agent either preshaped her hand in an informative manner (demonstrating a precision grip or a whole hand grip, depending on the target cup) or showed no informative preshaping of her hand (fist remained closed). Times differ slightly between stimuli, but on average, from the end of the arrow being displayed to the hand finishing preshaping and starting to move, this phase lasted for 26.25 frames (787ms).

In phase 4, the hand moves upward along a path that branches off to the left and right. The occluder dropped as soon as the agent’s hand reached the crossroad at the top of the Y-shaped path, effectively masking the agent’s entire arm from view. We were inspired by other relevant studies (e.g., Kochukhova and Gredebäck, 2007; Paulus, Hunnius, van Wijngaard, Vrins, van Rooij, & Bekkering, 2011) in introducing an occluder to facilitate anticipatory eye movements to one of path’s exits. On average, the phase ranging from the start of hand movement to the occluder fully covering the hand lasted for 55.25 frames (1657ms).

Split this up to get the TOI, along with average length of TOI

In phase 5, the hand appeared in one of the exits and contacted the cup on that side. In the true-belief condition, the agent contacted the cup that was congruent with the preshaping of her hand. In the false-belief condition, the agent wound up contacting the cup that was incongruent with the preshaping of her hand. For example, as per Figure 1, the agent possessed a false-belief that the large and small cups were on the right-side and left-side, respectively, and her goal to contact the large cup (as suggested by the whole-hand grip) was modulated by her false-belief of the cups’ spatial whereabouts. In this case, the agent’s hand appeared at the exit where she believed the large cup to be and therefore wound up contacting the small cup instead. Where there was no informative hand preshaping, the agent simply touched the large cup with her fist in 50% of the trials and she simply touched the small cup with her fist in the remaining 50% of no-preshaping trials.

*1.3. Procedure*

Participants were asked to give informed consent at the start of the experiment. After that, they were familiarised with the events in the videos. We paid special attention to explaining the familiarisation videos. We did this through presenting the participants with a model displaying how the videos were filmed. We informed the participant that they would be observing video stimuli in the experiment and that some background as to how the videos were made was required. We informed them of the different camera angles (see below) and the inability of the agent to observe the cups if the ledge was lowered. Insert photo of setup Participants were instructed to stabilise the head in the headrest, after which the familiarisation trials were presented. These familiarisation trials where only 240 frames (8 seconds) long were and didn't feature the primary visual detection task. Notably, the familiarisation videos had two side by side panels: one panel corresponds to the top-down perspective featured in the experimental trials, the other panel was the same scene shot from a frontal angle. A total of 8 familiarisation videos were presented. For the whole hand preshaping condition and the precision grip preshaping condition we used a true-belief situation, in which the hand contacted the cup congruent with hand preshape. We also used both cup locations (<small-left, large-right> and <small-right, large-left>), resulting in a total of 4 videos (2 cup locations *x* 2 hand preshaping conditions). For the no-preshape condition, we used both possible cup locations and both possible locations of contact (again, 2 *x* 2), to ensure the no-preshaping condition is not associated with a spatial bias.

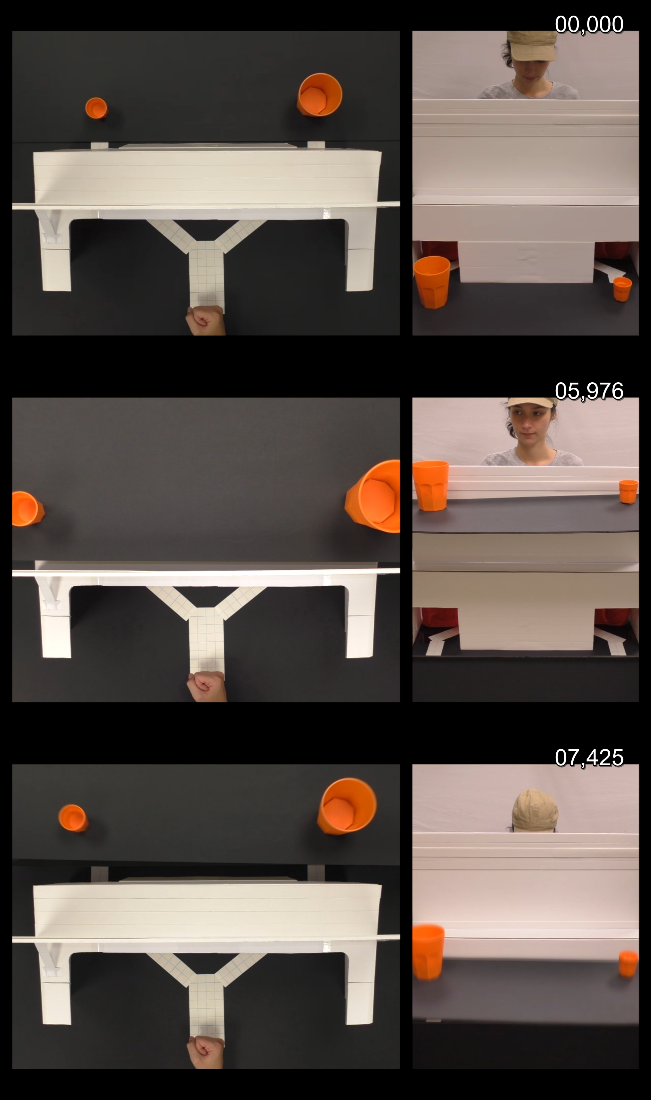


Figure X. Time line of a familiarisation trial. The occluder starts lowered and is then raised up. At this point, the agent looks at both objects before the ledge gets lowered again, breaking eye contact with the cups on the ledge. In appendix?

The main goal of the familiarisation trials was to ensure that participants had accurate information about the line of sight between the agent and the cups (i.e., to ensure that participants were exposed to the notion that the agent was unable to see the cups when the ledge was lowered but was able to see the cups when the ledge was raised). The videos ended as soon as the cups returned to the lower position during Phase 2. The instruction during familiarisation was to merely observe the videos (verbally communicated and visually presented on the computer screen).

After familiarisation, the participants were introduced to the Eyelink eyetracker. Instructions were presented on the screen and verbally communicated by the experimenter. When instructions were clear, participants completed a first calibration and validation run using a standard 9-point grid (a 3x3 grid uniformly distributed over the screen). After calibration, participants were presented with a test-trial (random selection of a no-grip video) to ensure they correctly responded to the visual detection secondary task, as well as correctly fixated on the fixation stimulus during Phase 2 of the experimental videos.

A total of 72 experimental trials were presented in 4 blocks of 18 videos. The number of false- versus true-belief trials was balanced per block. The different hand preshaping conditions (whole hand grip, precision grip, closed fist) were also balanced per block. After each block, the Eyelink eyetracker was recalibrated using the same procedure as outlined above.

After the stimulus presentation, participants were required to complete a short debriefing. The debriefing consisted of two binary response questions (yes/no): When the cups are raised, is the agent able to see the cups?" and "When the cups are lowered, is the agent able to see the cups?"

A third open-ended question asked participants if they could discern any stable behavioural pattern with the agent in the video. This was done to make sure the participant did not explicitly reason about the belief state of the agent. I.e., if a participant gave and answer indicating that the participant was tracking the agent’s beliefs the data from this participant would be withheld from the analysis.

*1.4. Data-analysis*

Raw eyetracking data and other experimental data were imported into R for analysis. We withheld any subject from the final analysis that either failed to answer the debriefing questions correctly (see above), had 75% of trials removed (see below), or otherwise indicated that they were explicitly tracking the belief state of the agent. This resulted in a final sample of N = 34.

On trial-level, we withheld all trials that had no reaction time on the visual detection task, failed to fixate on the fixation arrow during presentation of fixation arrow, or failed to fixate on any of the two AOIs during Phase 4. This resulted in a total of 1979 trials being analysed from the original 2448 trials recorded from 34 participants. Our analyses focused on the fixation patterns of participants, since in this paradigm participants were free to look anywhere at all times during the trial (with exception of the presentation of the fixation arrow). All fixations were defined in terms of a central (average) pixel location with x and y coordinate in a 1920 x 1080 field, with a specific starting time and duration.

The Areas of Interest (AOIs) were constructed as follows. For the arrow, we drew a box with the height and width of arrow, with an extra 50 pixels to account for eyetracking inaccuracies. For the cup AOIs, a more complex procedure was required. As all cups vary slightly in position over different videos, we calculated the central point for both cups in every video as an x - and y-coordinate (in pixels). We then averaged these points to a single x-coordinate for the left AOI, and another central x–coordinate for the right AOI. We averaged y-coordinates over both AOIs so we ended up with a single central y-coordinate for both sides. The lower y-boundary of the AOIs were based on the location of the box, plus 50 pixels of padding to account for measurement error. For the left and right x -boundaries as well as the upper y-boundary we based the AOI size on the maximal variation between cup locations and the size of the large cup (plus 50 pixels of padding to account for measurement errors).

Include excerpts from code in an Appendix. See Appendix X for more details

The Time of Interest (TOI) was defined as follows: for each different video we determined the precise frame at which the video transferred from Phase 3 to 4, and from Phase 4 to 5 (resp. the occlude obscuring the hand and the hand appearing in one of the exits). These frames were used as trigger to derive a precise timestamp as to when the TOI (Phase 4) started and finished in each trial. All fixations that started at a time within this TOI were considered for further analysis. As mentioned above, trials that did not show a single fixation within either AOI during the TOI were not considered for analysis.

We used the following measures as dependent variables:

1) the number of fixations within an AOI during the TOI. We aggregated these numbers within participant per condition to arrive at average number of fixations per trial as function of hand-preshaping condition.

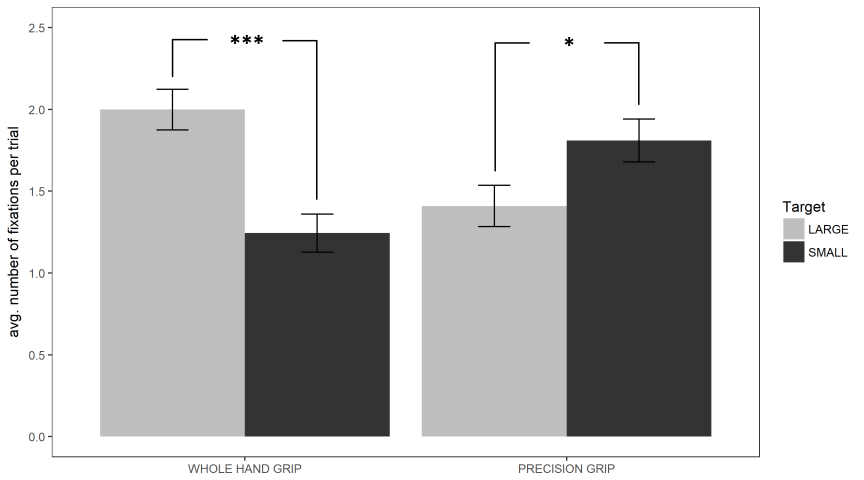
2) total dwell time, or the sum of all fixation durations within an AOI during the TOI. However, since the duration of the TOI varied slightly over videos we divided total dwell time by the total length of the TOI (for that specific video). As such, the second dependent variable is the total dwell time per AOI (within the TOI), proportional to the length of the TOI. Aggregating across the trials demonstrating each type of grip shape, we calculated participants’ total dwell time scores (expressed in percentage) on each AOI.

The data for both measures did not meet the assumptions of a paired-sample t-test due to being lower-bounded by zero in both cases, as well as being non-normally distributed. Hence, we used the Wilcoxon rank sum test with continuity correction in all the analyses below as a non-parametric alternative. The paired samples were the left and right AOIs in this case, interpreted as a within-subjects measure. All analyses were based on a .05 significance level (α = 0.05).

*2. Results*

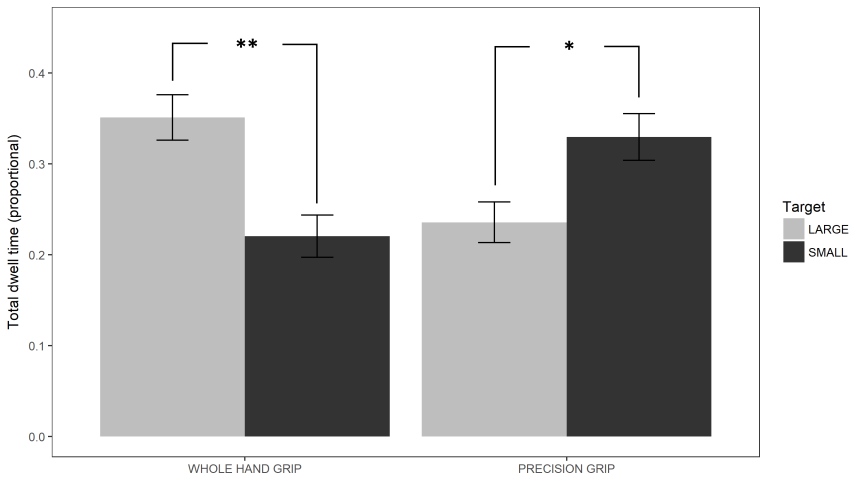
*2.1. Analysis of true belief condition*

*Number of fixations*. There was no significant difference between the number of fixations (large: M = 1.610; small: M = 1.624) towards the large and small cups in the closed-fist condition [W = 574.5, p = 0.971, 95% CI = [-0.348; 0.369]]. There was, however, significantly more fixations to the large cup (M = 1.999 fixations) than to the small cup (M = 1.244 fixations) in the whole-hand-grip condition [W = 900, p <0.0001, 95% CI = [0.378; 1.091]], indicating a bias towards the motor congruent AOI. There was also a similar motor congruency effect whereby participants produced significantly more fixations to the small cup (M = 1.810 fixations) than to the large cup (M = 1.410 fixations) in the precision-grip condition [W = 742, p = 0.045, 95% CI = [5.519e-05; 0.818]].



Insert Figure. Average number fixations per trial in the true-belief condition as a function of grip shaping. Error bars represent standard error of the mean.

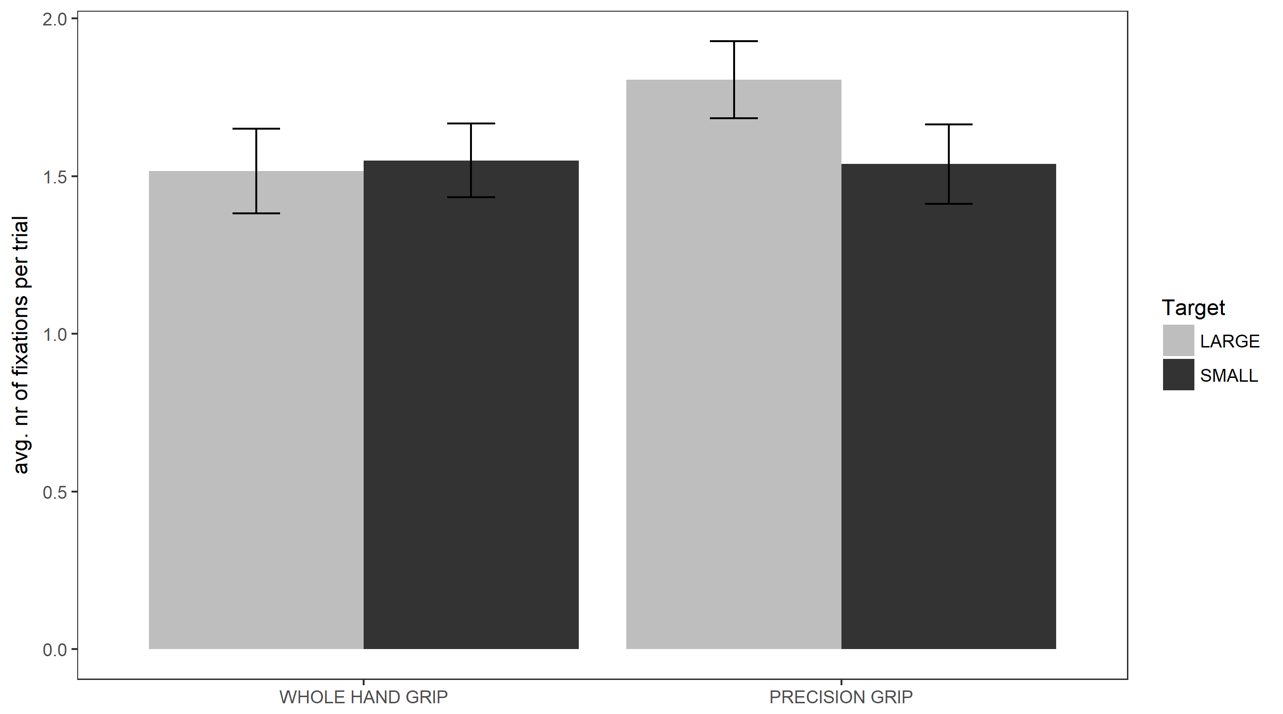
*Total dwell time, proportional to TOI length*. There was no significant difference between total dwell time on the large (M = 0.246) and small (M = 0.234) targets in the no-grip condition [W = 609, p = 0.708, 95% CI = [-0.037;0.053]]. Dwell time on the large target (M = 0.351) was significantly greater than dwell time on the small target (M = 0.220) for the whole hand grip [W = 883, p <0.001, 95% CI = [0.061;0.192]]. And conversely, dwell time on the small target (M = 0.330) was significantly greater than dwell time on the large target (M = 0.236) for the precision grip [W = 773, p = 0.017, 95% CI = [0.021;0.157]].



Insert Figure. Total dwell time per trial in the true-belief condition. Error bars represent standard error of the mean.

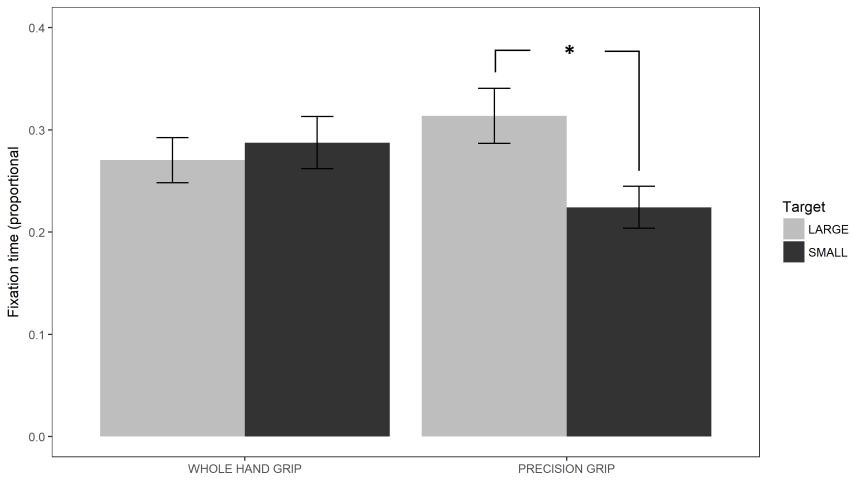
*2.3. Analysis of false-belief condition*

*Number of fixations.* There was no significant difference between the number of fixations to the large target (M = 1.816 fixations) and the small target (M = 1.508 fixations) in the no-grip condition, [W = 727.5, p = 0.066, 95% CI = [-0.0152; 0.636]]. There was also no significant difference between the number of fixations to the large (M = 1.516 fixations) and small targets (M = 1.550 fixations) in the whole-hand grip condition [W = 591, p = 0.878, 95% CI = [-0.364; 0.399]] or between the large targets (M = 1.806 fixations) and small targets (M = 1.529 fixations) in the precision-grip condition [W = 698, p = 0.142, 95% CI = [-0.100; 0.666]]. (Report means in the above sentences, even when not significant).



Insert figure. Average number fixations per trial in the false-belief condition. Error bars represent standard error of the mean.

*Total dwell time, proportional to TOI length*. There was no significant difference between total dwell time on the large (M = 0.282) and small (M = 0.222) targets in the no-grip condition [W = 734, p = 0.056, 95% CI = [-0.001; 0.103]]. There was also no significant difference between total dwell time on the large and small targets (large: M = 0.270; small: M = 0.287) in the whole-hand-grip condition [W = 610, p = 0.699, 95% CI = [-0.056; 0.077]]. However, we did find that total dwell time on the large target (M = 0.314) was significantly greater than total dwell time on the small target (M = 0.224) in the precision-grip condition, suggesting a bias for belief tracking to modulate motor processing [W = 759, p = 0.027, 95% CI = [0.012; 0.148]].



Insert Figure. Total dwell time per trial in the false-belief condition. Error bars represent standard error of the mean.

*3. Discussion*

Based on the results in the true-belief condition, we can conclude that there is strong evidence for implicit motor-driven anticipatory looking. On both measures (number of fixations and total dwell time) we found convincing evidence that hand pre-shaping is driving the gaze towards the size-congruent target. Specifically, participants showed a clear bias towards the cup that matched hand pre-shaping in size. That is, when the hand formed a large (whole hand) grip participants looked more often and longer to the large cup, and looked more and longer to the small cup in case a small (precision) grip was observed. This strongly suggests that participants inferred the target cup based on their own motor representation of a grasping movement in relationship to the object. This is in line with the findings of Ambrosini and colleagues (2011) and our results provide a conceptual replication of this study and extend it by applying different methods and measures. As such, this study supports the body of literature on implicit behavioural prediction based on motor simulation in action observation.

There is also mixed evidence for action observation being underpinned by belief tracking modulating or interacting with motor representations in the false belief condition. That is, we found that anticipatory looking was biased towards the target that was in line with the beliefs of the agent, but was opposite to actual state of affairs. Specifically, considering total dwell time in the false belief - precision grip condition, it seems that the gaze was pulled in the opposite direction: away from the motor-congruent location towards the belief-congruent location. These findings could potentially be explained through the conflicting information conveyed by the motor representations (“a whole hand grip indicates the intention to grasp the large cup, so that is where the hand will go”) and the outcome of low level mental state computation (“the agent believes the large cup is still on the left side”) as these are per definition leading to opposite predictions in the false belief condition. Interestingly, the hand or the location of the cups was completely irrelevant for the task at hand and there was no need for the participants to track the intentions of the agent at all. As such, current results are in line with other research on mentalising: without conscious awareness we anticipate the actions of other agents and this is reflected in our behaviour, in this case anticipatory looking.

In an active inference framework one could explain these results as the outcome of two competing hypotheses, respectively the hand reaching either to the small or the large cup. Under this framework, the observer establishes a generative model that makes predictions on future sensory input (Donnarumma et al., 2017; Friston, Adams, Perrinet & Breakspear, 2012). Here, this model would be based on prior events indicating that the agent has a false belief of the situation. That is, the observer establishes a predictive model of the environment based on the state of affairs as observed, including the position of the cups and the perspective of the agent. As such, the observer expects future actions of the agent to unfold according to his or her incorrect beliefs. Hence, when the agent preshapes the hand the observer is able to infer that the intention is to grasp a specific cup (large or small) based on the motor simulation of the hand shape, while the generative model informs the observer that the agent is likely to reach for the wrong cup as the agent is unaware of the cups having swapped. Thus, under the framework of active inference, the observer is most likely to sample information from the cup location that conforms with this prediction. That is, that specific location becomes more salient in function of the generative model active at the time (Donnarumma et al., 2017). However, the results only show this to be the case in the precision grip condition. Moreover, the effects are not as strong in the false belief condition as compared to the true belief condition. It is possible that the results are due to the different properties of motor simulation and mentalising. Motor simulation is a fast (in Ambrosini et al. 2011: on average 193ms) and automatic process, requiring little if any cognitive resources (Spunt & Lieberman, 2013) while implicit mentalising has been found to rely on cognitive resources to some extent (Schneider, Lam, Bayliss & Dux, 2012). It is therefore possible that mentalising is a slower and more demanding process, resulting in a suboptimal resolution of the conflicting hypotheses related to both processes that is still very much influenced by the incorrect predictions based on motor simulation. It is entirely possible that such temporal difference between both processes might be reflected in timing of anticipatory looking: a fast initial response in line with motor predictions that is subsequently corrected based upon the output of the slower and more demanding mentalising process. Our design did not allow for such temporal distinction (see below) but this notion is a tantalising avenue for future research.

However, we found substantial but non-significant differences in the no-preshaping condition. Specifically, there seems to be a bias towards the large object. This could indicate that there is a saliency effect of the large cup; when no information is available on where the hand might go the looking behaviour might default to the most salient object nearby. Note that similar saliency effects have been described in Ambrosini et al. (2011), where looking behaviour was significantly influenced by the size of the target object and were faster to fixate the large object when no pre-shaping was performed. However, no similar bias was found in the true belief condition, despite being identical to the false belief condition with the exception of the cups not swapping position. It is possible that the effects of saliency interact with belief reasoning. One could argue that due to implicit false belief processing relying at least minimally on executive cognitive resources (Schneider, Lam, Bayliss & Dux, 2012), in turn leaving visual attention more susceptible to distraction (Lavie, 2005) in function of object salience (Desimone & Duncan, 1995). In our case, this would mean that implicitly tracking the false beliefs of the agent caused the participant to be more inclined to rely on object salience (size of the object in this case, as the only distinguishing feature between both cups) to drive visual attention. As such, the significant difference in looking behaviour in the precision grip condition needs to be interpreted with caution as a saliency effect might explain a portion of this difference: in the precision grip condition saliency might account for increased looking times at the large target, while in the whole hand grip condition saliency could drive looking behaviour to a null difference. Nonetheless, the disappearance of the bias in anticipatory looking towards the motor-congruent object in the false belief condition in itself argues for the ability of spontaneous mentalising to modulate the outcome of behavioural predictions based on motor simulation.

A drawback of this study is that unlike the paradigm of Ambrosini et al. (2011) where participants were fixating a specific fixation cross before the measurement period, our study uses a free-looking paradigm in which the gaze of the participant is free to roam anywhere, only being instructed to watch the video. This free-look paradigm was largely necessitated by the type of stimuli we used. Although we did use a fixation cross, there was still a period lasting 2,44 seconds (on average) before the measurement period during which the gaze could move freely. As such, we couldn’t measure the temporal properties of anticipatory looking behaviour during the measurement period. This could be very interesting, because such data could possibly distinguish temporal differences between predictions based on motor simulation and mentalising and hence shed new light on the effects found in the present study. Additionally, we opted to compare both AOIs (large and small) rather than considering a single AOI (as per Ambrosini et al., 2011). This limited us significantly in our options for data-analysis and modelling the data. However, this method did allow us to directly compare looking behaviour at both AOIs. Future research could potentially disentangle the effects of target size and belief by modelling the outcome of these variables on a set of continuous measures. The paradigm could also be adapted to allow for different measurements of implicit mentalising. A prime candidate is mouse tracking (see van der Wel, Sebanz & Knoblich, 2014) as this is a continuous measure that allows us to track the decision-making process online.

In conclusion, the present study found substantive evidence that we use our own motor system to predict the outcome of observed actions by others. Moreover, we found tentative evidence suggesting that the outcome of these motor predictions is modulated through spontaneous computation of the belief state of the agent being observed.

**Experiment 2**

In the previous experiment we found tentative evidence that participants implicitly and rapidly engaged in mentalising to predict the outcome of an observed goal-directed motion. However, this study does allow for alternative interpretations of the results that do not rely on motor simulation (e.g. perceptual matching, statistical learning). Therefore we aimed to test the hypothesis that the motor system is causally implicated in determining the outcome pattern in predictive looking behaviour. This was done in analogy with Ambrosini, Constantini and Sinigaglia (2011). In this study, the researchers found that tying participants’ hands behind their backs (hence restricting their motor affordances) resulted in a significant reduction in speed of predictive looking behaviour. In the context of our implicit Theory of Mind framework, we hypothesised that restricting the hand movements of participants would result in a similar reduction in proactive looking behaviour. Specifically, we hypothesised that restricting the same motor affordances required to perform the observed action would result in a) a significant reduction in the discrepancy of number of fixations towards the belief-congruent location versus the opposite location and b) a similar reduction in total looking time discrepancy.

Do we hypothesis a greater effect of restriction in the false belief condition, and if yes, why?

*1. Methods*

*1.1. Participants*

A total of 125 participants participated in this study (*M* = 19.3, *SD* = 1.78, 85 females). All provided informed consent and had normal to corrected vision, and no history of any neurological pathology. The experimental procedure received ethical approval from the Victoria University of Wellington School of Psychology Human Ethics Committee.

*1.2. Stimuli and apparatus*

Analogous to experiment 1. Hand movement in the ‘restricted condition’ (see Procedure) was achieved by asking participants to hold on tight to a soft elastic band behind their back. We instructed participants that this was required to correct posture and prevent slouching during the eye-tracking part of the experiment.

*1.3. Procedure*

Analogous to Experiment 1. Added was a between-groups condition: movement restriction. Participants were randomly assigned to either the ‘unrestricted group’ (note that this group was under the exact same procedure as the participants in Experiment 1) or the ‘restricted group’.

Participants in the ‘restricted condition’ were instructed to apply the restricting elastic band after the familiarisation trials, and prior to the first calibration of the eye-tracker.

*1.4. Data-analysis*

Analogous to Experiment 1. In all within-subject comparisons we applied a two-sided paired sample t-test if the assumptions of normality and homogeneity of variance were met. If not, we used a Wilcoxon signed rank test. A similar approach was used in the between-subjects condition (movement restriction). We compared the unrestricted and the restricted groups by calculating a congruency score for each grip condition and comparing these between groups. The congruency score for number of fixations in each grip condition was calculated as the difference between number of fixations to the target location (where the hand would reach for the cup) and the non-target locations, per participant. The same procedure was followed for dwell time. We did not use a post-hoc correction on p-values as we only concerned ourselves with a limited number of pre-planned tests.

*2. Results*

*2.1. Baseline condition (unconstrained)*

*2.1.1. True belief condition*

*Number of fixations.* There was no difference between the number of fixations to the large target (M = 1.711) and the small target (M = 1.557) in the no-grip condition [t = -1.234, p = 0.31, CI = [-0.404;0.097]]. There was a significant difference in the whole-hand grip condition between the number of fixations to the large target (M = 2.744) and the small target (M = 0.476) [W = 1275, p < 0.0001, CI = [2.036;2.456], Cliff’s *d* = 0.999]. There was also a significant difference between the number of fixations to the large (M = 2.775) and small (M = 0.413) target in the precision-grip condition [W = 1275, p < 0.0001, CI = [2.208, 2.568], Cliff’s *d* = 1].

*Total dwell time, proportional to TOI.* On average participants looked longer at the small target (M = 0.329) as compared to the large target (M = 0.273) in the no-grip condition [W = 985, p < 0.001, CI = [0.029;0.097], Cliff’s *d* = 0.406]. Participants looked significantly longer at the large target (M = 0.558) as compared to the small target (M = 0.063) in the whole-hand grip condition [W = 1275, p < 0.0001, CI = [0.459;0534], Cliff’s *d* = 1]. Similarly, participants looked longer at the small target (M = 0.568) as compared to the large target (M = 0.0541) [W = 1275, p < 0.0001, CI = [0.486;0.547], Cliff’s *d* = 1].

*2.1.2. False belief condition*

*Number of fixations.* We found no difference in the number of fixations between the large target (M = 1.767) and the small target (M = 1.743) in the no-grip condition [t = - 0.226, p = 0.822, CI = [-0.235;0.187]]. In the whole-hand grip condition, participants fixated more often on the large target (M = 2.759) as compared to the small target (M = 0.455) [W = 1275, p > 0.0001, CI = [2.192;2.467], Cliff’s *d* = 1]. In the precision-grip condition participants were looking more frequently at the large target (M = 2.985) as compared to the small target (M = 0.594) [W = 1274, p < 0.0001, CI = [2.187;2.486], Cliff’s *d* = 0.996).

*Total dwell time, proportional to TOI.* We did not find a difference in looking times between the large target (M = 0.307) and the small target (M = 0.332) in the no-grip condition [t = 1.161, p = 0.251, CI = [-0.019;0.069]]. In the whole-hand grip condition, participants looked significantly longer at the small target (M = 0.066) as compared to the small object (M = 0.559) [W = 1275, p <0.0001, CI = [0.473;0.526], Cliff’s *d* = 0.999]. In the precision-grip condition, participants looked significantly longer at the large target (M = 0.53) as compared to the small target (M = 0.079) [W = 1275, p < 0.0001, CI = [0.419;0.499], Cliff’s *d* = 0.999].

*2.2. Restricted condition*

*2.2.1. True belief condition*

*Number of fixations.* We found no difference between the number of fixations on the large target (M = 1.628) and the small target (M = 1.569) [W = 448.5, p = 0.104, CI = [-0.455;0.083]]. In the whole-hand grip condition, participants looked more frequently at the large target (M = 2.737) versus the small target (M = 0.304) [W = 1275, p < 0.0001, CI = [2.208;2.583], Cliff’s *d* = 0.997]. Similarly, participants looked more often at the small target (M = 2.619) as compared to the large target (M = 0.348) in the precision grip condition [W = 1275, p < 0.001, CI = [2.077;2.416], Cliff’s *d =* 1].

*Total dwell time, proportional to TOI.* No difference was found between the large target (M = 0.298) and the small target (M = 0.346) in the no-grip condition [W = 871, p = 0.052, CI = [-0.001;0.084]]. Participants looked longer at the large target (M = 0.598) than the small target (M = 0.038) in the whole-hand grip condition [W = 1275, p < 0.0001, CI = [0.534;0.594], Cliff’s *d* = 1]. In the precision-grip condition, participants looked longer at the small target (M = 0.048) as compared to the large target (M = 0.611) [W = 1275, p < 0.0001, CI = [0.539;0.592], Cliff’s *d* = 1].

*2.2.2. False belief condition*

*Number of fixations.* There was no difference in the no-grip condition: participants did not look more frequently at the large object (M = 1.671) or the small object (M = 1.635) [W = 468.5, p = 0.77, CI = [-0.258;0.179]]. In the whole-hand grip condition, however, participants looked more frequently at the small target (M = 2.803) as compared to the large target (M = 0.316) [W = 1275, p < 0.0001, CI = [2.250;2.649], Cliff’s *d* = ]. In contrast, participants looked more frequently at the large target (M = 2.865) as compared to the small target (M = 0.438) in the precision grip condition [W = 1275, p < 0.0001, CI = [2.166;2.625], Cliff’s *d =* 0.994].

*Total dwell time, proportional to TOI.* There was no difference between the large target (M = 0.323) and the small target (M = 0.329) in the no-grip condition [W = 720, p = 0.429, CI = [-0.020;0.042]]. However, participants looked longer at the small target (M = 0.600) as compared to the large target (M = 0.049) in the whole-hand grip condition [W = 1275, p < 0.0001, CI = [0.520;0.580], Cliff’s *d* = 1]. Similarly, participants looked longer at the larger target (M = 0.564) as compared to the small target (M = 0.061) in the precision grip condition [W = 1275, p < 0.0001, CI = [0.476;0.548], Cliff’s *d* = 0.968].

*2.3. Free versus Restricted comparison*

*2.3.1. True belief condition*

*Number of fixations.* We found no difference in congruency scores between the unrestricted (M = 2.29) and the restricted condition (M = 2.266) in the precision grip condition [W = 1963.5, p = 0.378, CI = [-0.143;0.333]]. Nor did we find a difference between the unrestricted (M = 2.293) and the restricted group (M = 2.423) in the whole-hand grip condition [W = 1576, p = 0.389, CI = [-0.375;0.167]].

*Dwell time, proportional to TOI.* We found a significant difference between the unrestricted (M = 0.514) and the restricted (M = 0.571) groups in the precision grip condition [W = 1377, p = 0.028, CI = [-0.089;-0.004], Cliff’s *d* = -0.233]. Similarly, we found a significant effect between the unrestricted group (M = 0.504) and the restricted group (M = 0.567) in the whole-hand grip condition (t = -3.121, p = 0.002, CI = [-0.104;-0.023], Cohen’s Delta = 0.579].

*2.3.2. False belief condition*

*Number of fixations.* There was no difference in the precision grip condition between the unrestricted (M = 2.408) and the restricted group (M = 2.423) [W = 1813, p = 0.929, CI = [-0.250;0.249]]. Similarly, there was no difference between the unrestricted (M = 2.250) and restricted group (M = 2.453) in the whole-hand grip condition [W = 1579.5, p = 0.257, CI = [-0.357;0.083]].

*Dwell time, proportional to TOI.* We found no difference between the unrestricted (M = 0.464) group and the restricted (M = 0.509) group in the precision grip condition [W = 1448, p = 0.068, CI = [-0.087;0.003]]. However, we did find a significant difference between the unrestricted group (M = 0.495) and the restricted group (M = 0.553) in the whole-hand grip condition [W = 1280, p = 0.007, CI = [-0.089;-0.141], Cliff’s *d* = -0.287].

Define fixations as 100ms and over?

Different results with shorter TOI?

3. Discussion

The results of Experiment 2 provided a high-power replication of the results of Experiment 1. When unrestricted, participants looked more frequently and longer to the target that was congruent with the beliefs of the agent. That is, in the true belief condition participants showed increased proactive looking behaviour towards the target that was compatible with the hand pre-shaping of the agent. Conversely, in the false belief condition, participants showed significantly increased anticipatory looking behaviour towards the target that matched the agent’s beliefs and not the target matching the hand pre-shaping of the agent. That is, looking behaviour of participants was in line with one would expect if participants based predictions on the mental state of the observed agent, as opposed to an egocentric bias reflecting one’s own mental model of the environment. As such, the results form a strong theoretical replication of earlier work done on the subject (e.g. Ambrosini et al. 2011, Kovács et al., 2010).

However, the group assigned to the restricted-movement condition (i.e. had restrained hand movement) showed a similar pattern of results. A comparison between both groups showed that there was either no difference, or even a small increase in gaze proactivity in the restricted-movement group. As such, our results fail to provide a theoretical replication of the study of Ambrosini, Constantini and Sinigaglia (2011). That is, participants that had their hand movements restricted and were hence momentarily unable to perform the actions they observed, did not show a decrease in proactive gaze behaviour in line with the mental state of the agent, counter to what a motor-simulation account of mental state processing. However, there a number of potential explanations for this disparity. importantly, unlike the experimental procedure of Ambrosini et al. (2011), we implemented a free-looking paradigm. A such, we are not able to look at the trajectory and speed of the first fixation, from a specific initial point of fixation. As such, our results reflect the total looking behaviour over a prolonged timeframe. It might be the case that restricting the observer’s motor repertoire only impedes the accuracy of early looking responses but is subsequently corrected. That is, it is possible that the involvement of motor representations in driving anticipatory gaze behaviour decreases over time. Tentative evidence for this notion can be found in Maranesi, Ugolotti, Serventi, Bruni, Bimbi, Fogassi & Bonini (2013). In this study, the researchers found that gaze timing predicted the magnitude and timing of peak activity in the mirror neuron system, with earlier gaze onsets being linked to increased neural discharge, after which the signal decays. The researchers suggest that motor representations are crucial in automatically driving the initial gaze shift towards an observed action, further sampling that location for information, feeding back to the representations of that movement in the motor system. This interpretation is reinforced by the findings of Elsner, D’Ausilio, Gredebäck, Falck-Ytter & Fadiga (2013), who found that Transcranial Magnetic Stimulation (TMS) of the motor cortex hand area resulted in a delayed gaze shift towards to target of an observed movement, implying that motor representations related to a specific movement play a crucial role in directing predictive eye movement towards the target of an observed motion. It is therefore likely that restricting motor affordances will have a larger impact on the initial gaze shift towards a target location than any subsequent appraisals.

Based on the current results, we suggest that future studies could expand this paradigm by including a temporal component. It could be the case that early onset gaze shifts are primarily driven by motor simulation and are subsequently corrected by top-down modulation based on mentalising processes.

Run analysis of First Fixation, and time to first fixation (TTFF)?

Experiment 3: fixate eyes until lid is fully closed?

4. General Conclusion

In this study, we aimed to investigate the interaction between our capabilities for mentalising and motor-based prediction during action observation. While previous research found strong evidence that people tend to shift their gaze towards the target of an observed motion based on simulation in their own motor cortex (mirror neurons), it remained unclear if this gaze shift is driven by observers’ own mental model or a mental model based of the agent. That is, does an observer base predictive eye-movement in relation to an observed action purely on motor cues, or is this kind of information sampling modulated by mentalising processes? Is predictive eye movement driven by the observer’s implicit estimation of the belief state of the agent, who can have false information on the situation? In a first experiment, we found that predictive eye movement took advantage of motor cues, with limited evidence for modulation based on mentalising processes. However, in a second experiment with larger statistical power we found strong evidence for modulation based on a simulated mental state of the observer. That is, if the agent had false information on the location of the target of its movement, participants looked more frequently and longer at the location where the agent would look based on their (false) mental model of the target’s location. We also tested if this effect would diminish or disappear if the participant’s movements were restricted and they themselves would be unable to perform the observed movement. Whereas previous studies found that this impairs predictive eye movements towards a target location, we found no effect of movement restriction.

**References**

Ambrosini, E., Costantini, M., & Sinigaglia, C. (2011). Grasping with the eyes. *Journal of neurophysiology*, *106*(3), 1437-1442.

Ambrosini, E., Pezzulo, G., & Costantini, M. (2015). The eye in hand: Predicting others' behavior by integrating multiple sources of information. *Journal of neurophysiology*, *113*(7), 2271-2279.

Ambrosini, E., Sinigaglia, C., & Costantini, M. (2012). Tie my hands, tie my eyes. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(2), 263.

Blakemore, S. J., & Decety, J. (2001). From the perception of action to the understanding of intention. *Nature reviews neuroscience*, *2*(8), 561.

Costantini, M., Ambrosini, E., & Sinigaglia, C. (2012). Out of your hand's reach, out of my eyes' reach.

Costantini, M., Ambrosini, E., Cardellicchio, P., & Sinigaglia, C. (2013). How your hand drives my eyes. *Social cognitive and affective neuroscience*, *9*(5), 705-711.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience*, *18*(1), 193-222.

Donnarumma, F., Costantini, M., Ambrosini, E., Friston, K., & Pezzulo, G. (2017). Action perception as hypothesis testing. *Cortex*, *89*, 45-60.

Friston, K., Adams, R., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: saccades as experiments. *Frontiers in psychology*, *3*, 151.

Frith, C. D., & Frith, U. (2007). Social cognition in humans. *Current Biology*, *17*(16), R724-R732.

Gallese, V. (2001). The 'shared manifold 'hypothesis. From mirror neurons to empathy. *Journal of consciousness studies*, *8*(5-6), 33-50.

Gallese, V., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in cognitive sciences*, *2*(12), 493-501.

Grainger, S. A., Henry, J. D., Naughtin, C. K., Comino, M. S., & Dux, P. E. (2018). Implicit false belief tracking is preserved in late adulthood. *Quarterly Journal of Experimental Psychology*, *71*(9), 1980-1987.

Gredebäck, G., & Falck-Ytter, T. (2015). Eye movements during action observation. *Perspectives on Psychological Science*, *10*(5), 591-598.

Keysers, C., & Gazzola, V. (2007). Integrating simulation and theory of mind: from self to social cognition. *Trends in cognitive sciences*, *11*(5), 194-196.

Kochukhova, O., & Gredebäck, G. (2007). Learning about occlusion: Initial assumptions and rapid adjustments. *Cognition*, *105*(1), 26-46.

Kovács, Á. M., Téglás, E., & Endress, A. D. (2010). The social sense: Susceptibility to others’ beliefs in human infants and adults. *Science*, *330*(6012), 1830-1834.

Lavie, N. (2005). Distracted and confused? Selective attention under load. *Trends in cognitive sciences*, *9*(2), 75-82.

Paulus, M., Hunnius, S., van Wijngaarden, C., Vrins, S., van Rooij, I., & Bekkering, H. (2011). The role of frequency information and teleological reasoning in infants' and adults' action prediction. *Developmental psychology*, *47*(4), 976.

Peirce, J. W., & MacAskill, M. R. (2018). *Building Experiments in PsychoPy*. London: Sage.

Rizzolatti, G., & Sinigaglia, C. (2016). The mirror mechanism: a basic principle of brain function. *Nature Reviews Neuroscience*, *17*(12), 757.

Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G., & Matelli, M. (1988). Functional organization of inferior area 6 in the macaque monkey. *Experimental brain research*, *71*(3), 491-507.

Samson, D., Apperly, I. A., Braithwaite, J. J., Andrews, B. J., & Bodley Scott, S. E. (2010). Seeing it their way: evidence for rapid and involuntary computation of what other people see. *Journal of Experimental Psychology: Human Perception and Performance*, *36*(5), 1255.

Sartori, L., Becchio, C., & Castiello, U. (2011). Cues to intention: the role of movement information. *Cognition*, *119*(2), 242-252.

Schneider, D., Bayliss, A. P., Becker, S. I., & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of experimental psychology: general*, *141*(3), 433.

Schneider, D., Lam, R., Bayliss, A. P., & Dux, P. E. (2012). Cognitive load disrupts implicit theory-of-mind processing. *Psychological science*, *23*(8), 842-847.

Schneider, D., Slaughter, V. P., Becker, S. I., & Dux, P. E. (2014). Implicit false-belief processing in the human brain. *NeuroImage*, *101*, 268-275.

Spunt, R. P., & Lieberman, M. D. (2013). The busy social brain: evidence for automaticity and control in the neural systems supporting social cognition and action understanding. *Psychological science*, *24*(1), 80-86.

van der Wel, R. P., Sebanz, N., & Knoblich, G. (2014). Do people automatically track others’ beliefs? Evidence from a continuous measure. *Cognition*, *130*(1), 128-133.

Vaziri-Pashkam, M., Cormiea, S., & Nakayama, K. (2017). Predicting actions from subtle preparatory movements. *Cognition*, *168*, 65-75.

New:

Maranesi, M., Ugolotti Serventi, F., Bruni, S., Bimbi, M., Fogassi, L., & Bonini, L. (2013). Monkey gaze behaviour during action observation and its relationship to mirror neuron activity. *European Journal of Neuroscience*. <https://doi.org/10.1111/ejn.12376>

Elsner, C., D’Ausilio, A., Gredebäck, G., Falck-Ytter, T., & Fadiga, L. (2013). The motor cortex is causally related to predictive eye movements during action observation. *Neuropsychologia*. https://doi.org/10.1016/j.neuropsychologia.2012.12.007

Appendix X

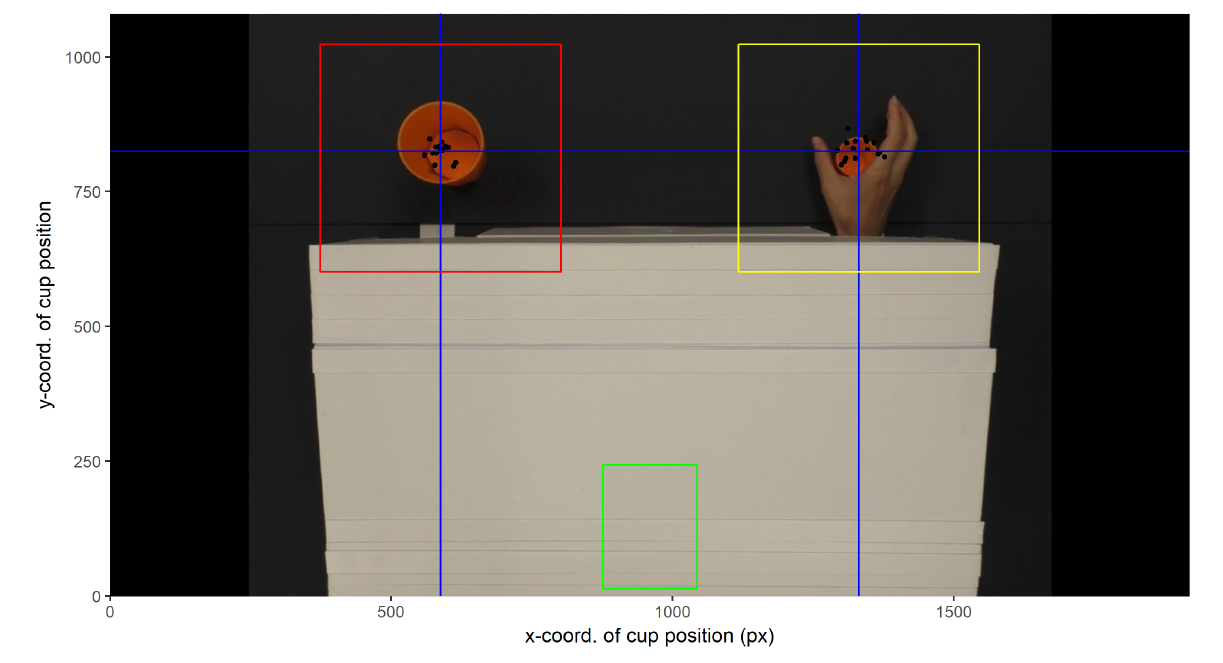
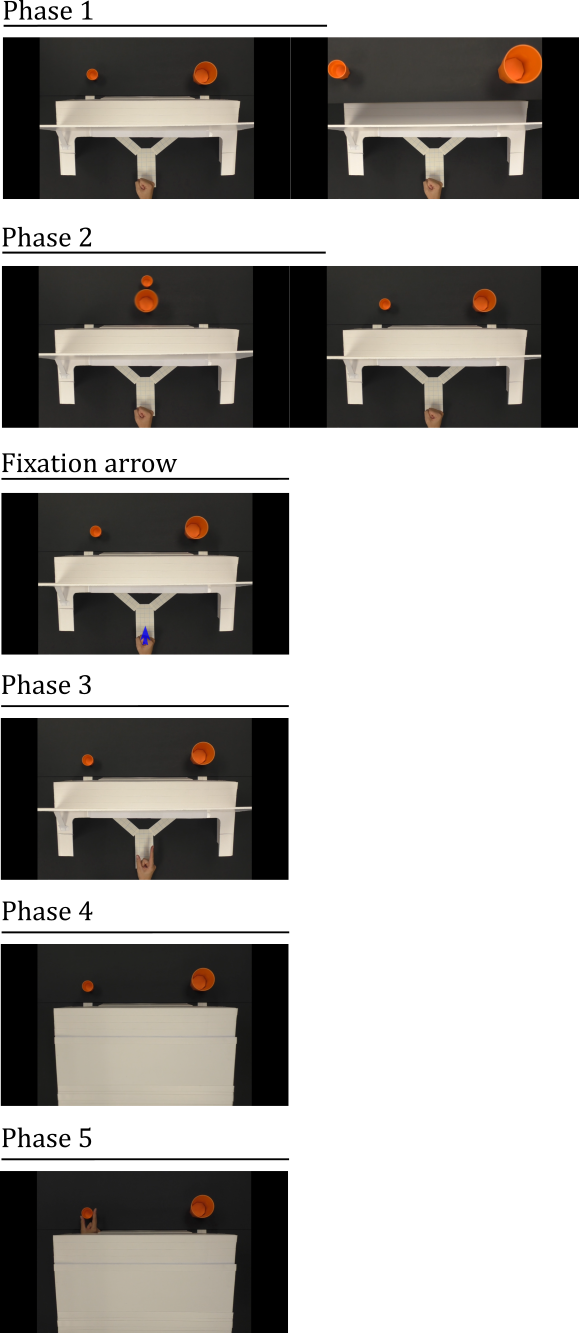
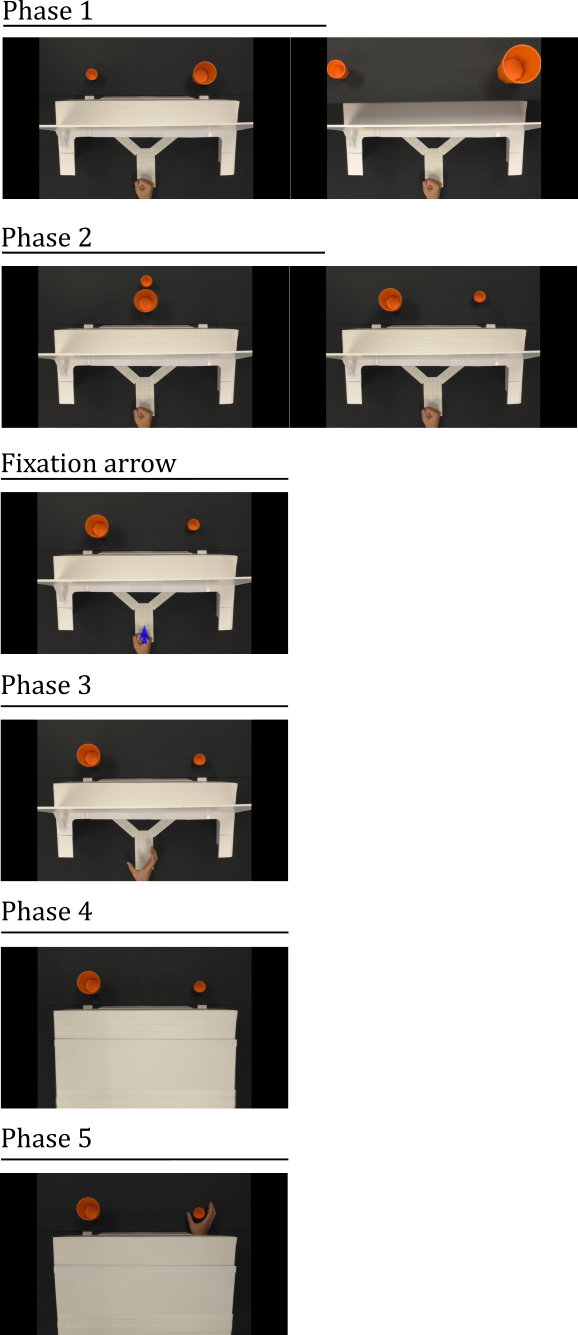


Figure 3: Construction of Areas of Interest. Black dots indicate cup locations per video. Blue lines indicate the means used to determine the centers of the red and yellow AOIs (covering the left cup and right cup, respectively). The green rectangle represents the AOI associated with the fixation arrow.

Appendix X





Appendix X.

### Processing Eyelink data ###

# custom functions

source('pixel\_conversion.R')

# global variables

mov.dur.f <- 698 #frames

fps.std <- 30

screen.W <- 1920

screen.H <- 1080

# Import cup positions

file <- read.csv(pos.input)

# Getting AOI definitions

# IMPORTANT: the eyetracker counts pixels from top-left, not bottom-left, so we need to mirror the Y axis

# properties of the arrow

arrow.H <- mm\_to\_px(51-17)

arrow.W <- mm\_to\_px(263-245)

arrow.X <- mm\_to\_px(254)

arrow.Y <- mm\_to\_px(34)

arrow.padding <- 0.75

# defining AOI limits

a.xmin=(arrow.X - (arrow.W \* arrow.padding))

a.xmax=(arrow.X + (arrow.W \* arrow.padding))

a.ymax=screen.H - (arrow.Y - (arrow.H \* arrow.padding))

a.ymin=screen.H - (arrow.Y + (arrow.H \* arrow.padding))

# properties of cup positions

# Note: in the wrong scale... multiply with a scaling constant

scale <- 1.406

file$leftX <- mm\_to\_px(file$leftX)\*scale

file$leftY <- screen.H - (mm\_to\_px(file$leftY)\*scale)

file$rightX <- mm\_to\_px(file$rightX)\*scale

file$rightY <- screen.H - (mm\_to\_px(file$rightY)\*scale)

avg.leftX <- mean(file$leftX)

avg.leftY <- mean(file$leftY)

avg.rightX <- mean(file$rightX)

avg.rightY <- mean(file$rightY)

avg.Y <- mean(avg.leftY,avg.rightY)

#size of large cup

cup.size <- 160.63

# size of padding

padding <- 50

# calculate padding

diff.x.left <- max(file$leftX) - min(file$leftX)

diff.x.right <- max(file$rightX) - min(file$rightX)

X.padding <- max(diff.x.left,diff.x.right)

diff.y.left <- max(file$leftY) - min(file$leftY)

diff.y.right <- max(file$rightY) - min(file$rightY)

Y.padding <- max(diff.y.left,diff.y.right)

# Defining left cup AOI

l.xmin=(avg.leftX - X.padding - padding)

l.xmax=(avg.leftX + X.padding + padding)

l.ymin=(avg.Y - Y.padding - padding)

l.ymax=(avg.Y + Y.padding + padding)

# Defining right cup AOI

r.xmin=(avg.rightX - X.padding - padding)

r.xmax=(avg.rightX + X.padding + padding)

r.ymin=(avg.Y - Y.padding - padding)

r.ymax=(avg.Y + Y.padding + padding)