***Chapter 4: Experiment 1***

***Hi Pieter – do not fuse Tables and Figures – they are traditionally separated. Look at a range of journals – you will see that they keep figures separated from tables, or else they try to have a figure do all the work with numbers and stats shown (or stats provided in text or the caption). Fusing them can make your thesis readers nit pick – remember to make the reading experience easy and streamlined.***

Status: reviewed, edited, awaiting final feedback

***4.1 Introduction***

As discussed in previous chapters, a major caveat in the current body of research on social prediction making is the interaction between motor-driven behavioural simulation and representations of mental states. To that extent, new paradigms are needed to investigate mentalising operations in conjunction with action observation of goal-directed action. This chapter proposes a paradigm that combines a classic action observation paradigm but adds a requirement for implicitly tracking false beliefs in order to make correct behavioural predictions. In this action observation paradigm of Ambrosini, Constantini and Sinigaglia (2011), an agent reaches for one of two target objects: a large cup or a small cup. The agent’s hand could match grip aperture to the target object in either a precision grip or whole hand grip, or display keep the hand closed. Just like these researchers, we were interested in predictive gaze behaviour towards either target. The crucial addition of this study, however, is that prior to the agent performing a reaching motion the cups could swap position. This swap occurs unbeknownst to the agent, rendering her belief true or false. The agent, of course, acts in accordance with her beliefs on the location of the targets.

As such, this study investigates the case in which both systems would in principle yield different behavioural predictions if operating in isolation. As there are two possible targets for action, prediction of the target of the reaching action requires integration of both goal-tracking and belief-tracking abilities. If the participant were to respond purely based on motor resonance, her eyes would be attracted to the cup that matches hand aperture, as this object is more salient due to shared motor representations with the agent, as per Donnarumma, Constantini, Ambrosini, Friston & Pezzulo (2017)*.* However, such predictions would be incorrect if the target cups swapped position. To arrive at the correct prediction, a participant would have to 1) derive the target of the grasping motion based on grip aperture, and 2) modulate the resulting target prediction based on implicitly tracking the registrations of the agent.

Our specific hypotheses were as follows:

1) We expect a significant effect of pre-shaping, in line with previous findings using this paradigm. That is, when motor cues are sufficient to predict the unfolding of future actions (i.e., in the true belief condition), we expect to find that anticipatory gaze is significantly biased towards the target that matches hand aperture.

2) If motor-based predictions of action goals interface with representations of agent-object registrations, we expect to find that anticipatory looking times are significantly biased towards the action target that is congruent with the agent’s belief when motor cues are insufficient to predict the outcome of the grasping action (i.e., in the false-belief condition).

As the focus of this investigation is the role of the motor system in *implicit* mentalising processes, steps were undertaken to ensure the implicit nature of the task and the underlying process to the best of our abilities. To that extent we masked the task to appear as a stimulus detection task and made no references to the mental states of the agent. In fact, the agent had no relevance to the task at all. As such, the task demands were implicit, in the sense that belief-tracking was implicitly measured through anticipatory looking patterns. Furthermore, we attempted to ensure that the underlying was implicit by means of a debriefing session. Participants that reported to have been explicitly reasoning on the mental state of the agent (e.g., “*if the cups were swapped, she would think it was still in the old location and reach for the cup there*”) were eliminated for further analysis.

*PIETER, when it comes to formatting, is there something you can do to get rid of all these redundant white spaces? MSWORD is horrible at this – but perhaps you will be a formatting genius to make the spacing all look visually stunning and pleasant to the eye!*

***4.2 Methods***

*4.2.1. Participants*

A total of 43 participants participated in this study (*M* = 19.35, *SD* = 2.36, 29 females). These participants were first year undergraduate students made available by Victoria University of Wellington’s Introduction to Psychology Research Programme. After data cleaning (see below), the final sample size was N = 35. All provided informed consent and had normal to corrected vision. The experimental procedure received ethical approval from the Victoria University of Wellington School of Psychology Human Ethics Committee.

To do: power analysis

*4.2.2. 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. The visual detection task served two purposes. First, it served as a disguise for the real purpose of the experiment as this was the only task participants needed to perform. Secondly, it served as attention check.

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 & 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).

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 X, 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.

Insert Diagram here

*4.2.3. 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. 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. See Appendix X for a visual example of a familiarisation trial.

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).

*4.2.4 Data analysis*

All experimental data (eye-tracking data and behavioural data) were imported into R for analysis. At the participant level, we withheld participants (N = 8) from analysis that had 1) a wrong answer on the essential debrief questions (see above) or 2) had an incomplete number of data points (2 belief conditions *x* 3 grip conditions, for a total of 6 data points) (N = 1) or 3) self-reported to have been explicitly mentalising (N = 4) or 4) had severe issues with correct calibration of the eye-tracker (N = 3). On a trial level, we removed all trials from the original total of 3075 recorded trials that had 1) video display issues, or a framerate lower than 59.6 FPS (0 trials), 2) were not completed, 3) had timing issues due to eye-tracker and video streams being mismatched (7 trials), 4) had no fixation on the arrow during presentation of arrow (133 trials), or 5) had no fixations during our Time of Interest (TOI) in either one of the Areas of Interest (AOI) (383 trials). After data cleaning on both participant level, 35 participants remained totalling 2125 trials, or 69.1% of recorded trials.

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). The AOI in which the agent reached for the cup at the end of the video was defined as the ‘target’ AOI. The opposite AOI was designated as the ‘distractor’ AOI henceforth.

Include image of AOIs here

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 (respectively, the occluder 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.

For each trial, we calculated a Differential Looking Time Score to represent overall gaze bias. This method was preferred over using “location” as a predictor variable as this method increases model complexity. Furthermore, the distribution of our data reveals that implementing this method tends to lead to heavily skewed data distributions, violating the assumption of normality of residuals in ANOVA tests.

Consequently, we adopted the following approach. First of all, we calculated the sum of the durations of all fixations within each of the AOIs, defined as the ‘target’ AOI and the ‘distractor’ AOI, per trial. DLTS was then defined, per trial, as: the total looking time for the target AOI, minus total looking time for the distractor AOI, divided by the total length of the TOI.

*Where:*

*t = total looking time per AOI during TOI*

*l = total duration of TOI*

This results in a score between -1 and 1, where -1 equals the full length of TOI spent fixating on the distractor AOI, and 1 equals the full length of the TOI spent looking at the target. 0 would hence indicate no overall bias towards either AOI during the TOI. One could also interpret these as the differential proportion of preferential looking times.

We choose to deviate here from the classic definition of DLTS that is frequently used in anticipatory looking in implicit mentalising studies (e.g., Hayashi et al., 2020; Senju, Southgate, White & Frith, 2009; Kulke & Rakoczy, 2019; Kulke, Reiß, Krist & Rakoczy, 2019; Kulke, von Duhn, Schneider & Rakoczy, 2018; Burnside, Ruel, Azar, Poulin-Dubois, 2018). In the classic definition, DLTS is defined by dividing by the sum of fixations towards target and distractor. The issue here is twofold. First of all, since any trials that only have fixations towards one AOI and not the other will always result in a score of either -1 or 1, the result is an extremely tail-heavy distribution of trial-wise DLTS. Furthermore, the information contained in duration of fixations is lost as only the relative difference of both AOIs is accounted for, and not the actual duration of the fixations in relation to the Time of Interest. As such, a fixation that lasts for 100 ms weighs equally as a fixation of 1000 ms, for example. Dividing the differential looking times by the length of the Time of Interest not only preserves the information contained in the actual looking times, but also results in differential looking times that are more in line with a normal distribution (see Figure X).

Chart, histogram

Description automatically generated

*Figure X. Histogram representing distribution of raw data of Experiment 1, using our modified definition of differential looking time scores (DLTS). Red line indicates a theoretical normal distribution with the same mean and standard deviation as our sample, for reference.*

In order to test our hypotheses on the effects of our experimental variables on both measures, we modelled our data using Linear Mixed-effects Models (LMM) using the *lme4* package in R (bates, Mächler, Bolker & Walker, 2015). There are a number of benefits to using this method. First and foremost, traditional ANOVA methods with repeated measures with multiple datapoints per cell (as is the case here) will often prompt us to aggregate the data, potentially resulting in false positives (so-called Type I errors), reduced power or yielding inaccurate perceptions of robustness of effects (Meteyard & Davies, 2020). Linear mixed effect models, in contrast, captures more complex data structures such as data clustered within-participant, by explicitly including these data clusters in our model. This altogether eliminates the necessity to aggregate data, enabling us to use our entire datasetwhen modelling our results and testing hypotheses.

The main characteristic of *mixed models* is that they make a distinction between fixed effects and random effects. Fixed effects in psychology experiments usually consist of our a priori experimental conditions, i.e., the predictor variables which we suspect to have a specific relationship to an outcome variable. In contrast, random effects are effects that come from outside factors related to our experimental design. Often, in psychological experiments, this will be the factor ‘participant’. By explicitly modelling ‘participant’ as a random effect (a random sample from a population) we basically model how our fixed effects yield varying results over units (people) of that population and account for how people individually deviate from an average trend. This offers a much more powerful way of handling individual data and looking at trends on the level of both the individual and the population.

Belief and grip were within-subjects fixed categorical factors. Participants, as experimental units, were treated as a random factor and thus included in the model. We further extended the model by including block number to account for learning effects. Additionally, we expanded the model to include the main effect of target size, i.e., if the target of movement is the large cup or the small cup. This was done for two reasons. First of all, a look at DLTS scores per video makes us suspect that especially in the no-grip condition, scores are driven by the saliency of the large cup. Secondly, Ambrosini et al. (2011) found higher accuracy, fewer saccades and faster arrival times for the large object, even when no hand pre-shaping was present. As such, we were interested to see if participant looked longer at the large cup (as a more salient object) than the small cup.

Where possible, we opted to use Least Square means instead of observed means as this method accounts for imbalances in design. Additionally, it takes advantage of the strengths of our linear mixed regression models to represent the data.

We tested the effect of variables included in our model by using Type III Analyses of Variance with Satterthwaite's method (Satterthwaite, 1946). Any post-hoc contrast tests were done using the *lsmeans* package in R (Length, 2016). Degrees of freedom were estimated using the Kenward-Roger method (Kenward & Roger, 1997). All reported t-tests on least square estimates had p-values adjusted for multiple comparisons alpha-inflation using the Holm method (Holm, 1979).

We checked if our models met the assumptions of linear mixed models: linearity, homogeneity of variance and normality of residuals. All these conditions were met

*Pieter – too much white space here – you will need to format more effectively the visual layout.*

***4.3 Results***

*4.3.1 Basic model*

Here we modelled the data using the main effects of belief and grip, plus their interaction effect. We found a significant main effect of belief, *F*(2,2052) = 4.854, *p* = .028. There was also a significant main effect of grip, *F*(1,2053) = 10.319, *p*<.0001. The interaction between belief and grip was not significant, *F*(2,2054) = 1.766, *p* = .171.

A pairwise post-hoc test on the main effect of belief revealed that participants had a significantly higher DLTS in true belief trials [LSM = .069, SE = .02, 95% CI = [.028;.11]] as compared to the false belief condition [LSM = .035, SE = .02, 95% CI = [-.006;.076]], *t =* -2.203, p = .027.

A pairwise post-hoc test (see Table X and Figure X) with Kenward-Roger estimation of degrees of freedom and Holm-corrected p-values, found that scores in both the whole hand grip condition, *t =* 5.019, p<.0001, and the precision grip condition, *t =* 5.84, p<.0001, differed significantly from the no grip condition, in line with our hypotheses. There was no difference in LS mean score between the whole hand grip condition and precision grip condition, *t =* .792, *p* = .428.

|  |  |  |  |
| --- | --- | --- | --- |
| *Grip* | *LS Mean* | *SE* | *95% CI* |
| Whole hand | .07 | .022 | .026;.113 |
| No grip | .003 | .022 | -.041;.047 |
| Precision | .084 | .022 | .041;.127 |

*Table X. DLTS per grip condition, model-based Least Square means including standard error and 95% confidence interval.*

Chart, box and whisker chart

Description automatically generated

*Figure X. Main effect of grip, model-based LS means. Error bars represent standard error of the mean, stars represent significant mean differences.*

We also had a closer look at the different levels of belief and grip to test our a priori hypotheses. First, we found that DLTS was significantly higher than baseline (i.e., participants look longer at the correct action target) in the pre-shaped true belief conditions (TB-WHG and TB-PG), signalling that participants looked significantly longer at the motor-congruent target when the agent has a correct belief on the location of the action target that matches her hand aperture. Moreover, in the precision grip + false belief condition (FB-PG), we found that DLTS was significantly higher than chance level as well, indicating that in this condition participants correctly looked longer at the *belief*-congruent action target in anticipation of the agent’s actions. See Table X below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Belief-grip* | *LS Mean* | *SE* | *95% CI* | *t* | *df* | *p* |
| FB-WHG | .037 | .026 | -.014 ; .088 | 1.449 | 115 | .3 |
| FB-NG | -.018 | .026 | -.069 ; .033 | -.693 | 117 | .49 |
| FB-PG | .086 | .026 | .036 ; .137 | 3.369 | 112 | .003\*\* |
| TB-WHG | .102 | .025 | .052 ; .153 | 4.019 | 111 | <.001\*\*\* |
| TB-NG | .026 | .026 | -.028 ; .075 | -.91 | 118 | .365 |
| TB-PG | .082 | .025 | .032 ; .132 | 3.248 | 107 | .003\*\* |

*Table X. DLTS per belief and grip condition, model-based Least Square means including standard error and 95% confidence interval, as well as t-tests of the LS means in relation to baseline 0. Stars indicate significance levels of t-tests of least-square means against baseline 0.*

We also tested for the difference between false belief and true belief conditions in each level of grip. We found that while this difference was significant in the whole hand grip condition (FB-WHG vs TB-WHG), *t* = 2.478, *p* = .039, DLTS in the no grip, *t* = -.656, *p* = 1.0, and precision grip, *t =* -.159, *p* = 1.0, conditions did not differ between levels of belief.

Chart, box and whisker chart

Description automatically generated

*Figure X. DLTS per level of belief and grip, model-based LS means. Error bars represent standard error of the mean. Stars above bars represent significant mean differences. Stars on the bar represent significance of LS mean difference versus baseline 0. Results show a significant effect of hand preshaping in the true belief condition. In the false belief condition, we see significant belief-congruent looking times in the precision grip condition. These results are in line with out a priori hypotheses.*

TODO: include scatter plot of responses in FB to check for implicit strategies

*4.3.2 Extended model*

We also investigated the impact of adding crucial covariates to the model. First, we added block number to account for learning effects, including both a main effect and three-way interaction effects with belief and grip. This revealed that while there was a significant main effect of block number, *F*(3,2040) = 3.964, *p* = .008, this variable did not interact with belief, *F*(3,2035) = 1.592, *p* = .189, nor with grip, *F*(6,2035) = .943, *p* = .463, nor was there a three-way interaction effect, *F*(6,2035) = .265, *p* = .953. Other effects remained unaffected by adding in block number as covariate. An overview of the main effect of block number can be found below (Table X, Figure X).

However, we found that adding “block” as predictor variable results in no significant reduction in model deviance when comparing this model to the basic model, χ2(18) = 24.052, *p* = .153, signalling that adding this variable does not significantly improve the fit of the model to the observed data.

|  |  |  |
| --- | --- | --- |
| *Contrast* | *t* | *p* |
| 1-2 | -2.302 | .086 |
| 1-3 | -2.55 | .054 |
| 1-4 | -3.251 | .007\*\* |
| 2-3 | -.21 | 1.00 |
| 2-4 | -.919 | 1.00 |
| 3-4 | -.722 | 1.00 |

Chart, box and whisker chart

Description automatically generated

*Figure X. Main effect of block number, model-based LS means. Error bars represent standard error of the mean, stars represent significance of LS means difference.*

*Table X. Overview of post-hoc t-tests on LS means between blocks.*

An overview of LS means per level of block, belief and grip can be found in Appendix X-X.

Secondly, we extended the model to include target size to account for a saliency effect towards the larger object. However, interpretation and testing of this the effect of target size in relation to DLTS and the interaction effects of target size with other variables is complicated by the relationship between these variables. That is, in the whole hand grip condition and precision grip conditions, target size is directly tied to the belief condition. Specifically, in the true belief condition target size matches grip: a whole hand grip matches with the large cup, while a precision grip matches with the small cup. However, in the false belief condition, the belief-congruent target of a whole hand grip is the small cup, while the target of a precision grip is the large cup, since both cups were swapped without the agent knowing. Fully crossing ‘hand pre-shape’ and ‘target size’ is impossible as no condition is included in which the hand reaches for the *opposite sized* cup (e.g., precision grip to large cup) in the true belief condition, or likewise, where the agent reaches for the *matching sized* cup (e.g., precision grip to small cap) in the false-belief condition. As such, we can’t directly look at the interaction effect between target size and grip due to inevitable rank deficiencies in the model. We therefore included ‘target size’ as a main effect only. It is important to note that this main effect is estimated purely based on the variance that can be explained through target size in *in the no-grip condition only.*  Implementing this main effect across belief and grip conditions assumes that this saliency effect will be uniform over these conditions. Stated otherwise, the extended model accounts for the effect of target size under the assumption that this size effect does not interact with belief and/or grip. This assumption is fundamentally untestable, as per the reasons stated above.

There are, however, a couple of checks we can do that might give us some indication. First of all, if size matters outside of the no-grip condition, we should see that: 1) in the true belief condition, DLTS will be higher in the whole hand grip condition (as the target is large cup) as compared to the precision grip condition (target is the small cup), 2) in the false belief condition, DLTS will be higher in the precision grip condition (target = large) as compared to the whole hand grip condition (target = small). Testing this assumption, we find that there was no difference between the TB-WHG condition and the TB-PG condition, *t =* .78, *p* = .436. In the false belief condition, we only found a marginal difference between the FB-WHG condition and the FB-PG condition, *t =* -1.822, *p* = .069. This suggests that the size of the target object might not play as big a role in the preshaped conditions as in the no-grip condition. We must therefore be cautious with interpreting the results of the extended model too strongly, as target size might interact with belief and/or grip. If this is the case, we probably overestimate the effect of target size in the pre-shape grip conditions if we base this estimate on results in the no-grip condition.

An ANOVA on this model revealed that there indeed was a significant effect of size, *F*(1,2035) = 7.26, *p* = .007. The main effects of block number, belief, grip, and their interaction effects, remained unchanged by the addition of this covariate. A post-hoc t-test showed that participants tended to look longer at the large object [LSM = .089, SE = .023, 95% CI = [.043;.135]] as compared to the small object [LSM = .015, SE = .023, 95% CI = [-.031;.062]], *t =* 2.694, *p* = .007.

Finally, we compared our basic model with the extended model including the covariates block number and the main effect of target size. This revealed that the extended model accounts for significantly more variance than the basic model, χ2(19) = 31.384, *p* = .036. We are therefore inclined to use the more prudent extended model, as saliency effects might partially account for the results found under the basic model. Learning effects could also present, but do not seem very impactful.

We tested our main hypotheses again under this extended model correcting for learning and saliency effects, see Table X below. As we can see, DLTS was significantly higher than baseline in the true belief condition + whole hand grip (TB-WHG) and true belief condition + precision grip (TB-PG), analogous to results in the basic model. In the false belief condition, DLTS is significantly higher than baseline in the whole hand grip condition (FB-WHG) only, hence this belief-tracking effect is now found in the other preshaped grip condition as compared to the basic model. This makes sense, as in the FB-PG condition, the belief-congruent target is the large cup, and hence looking times will be boosted by any saliency effects due to size. Conversely, in the FB-WHG condition the saliency effect of size must be *overcome*, as here the belief-congruent target is small. Therefore, accounting for size effects boosts scores in this condition.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Belief-grip* | *LS Mean* | *SE* | *95% CI* | *t* | *df* | *p* |
| FB-WHG | .073 | .029 | .016 ; .131 | 2.523 | 185 | .037\* |
| FB-NG | -.019 | .026 | -.07 ; .032 | -.753 | 117 | .453 |
| FB-PG | .049 | .029 | -.008 ; .106 | 1.705 | 182 | .18 |
| TB-WHG | .066 | .029 | .009 ; .123 | 2.297 | 180 | .046\* |
| TB-NG | .024 | .026 | -.027 ; .076 | .944 | 118 | .347 |
| TB-PG | .119 | .029 | .062 ; .175 | 4.153 | 175 | <.001\*\*\* |

*Table X. DLTS per belief and grip condition, model-based Least Square means under the extended model. Included are standard error and 95% confidence interval, as well as t-tests of the LS means in relation to baseline 0.*

We also tested for the difference between false belief and true belief conditions in each level of grip. We found that this difference was not significant in either the whole hand grip condition (FB-WHG vs TB-WHG), *t* = -.187, *p* = .852, DLTS in the no grip, *t* = 1.613, *p* = .21, and precision grip, *t =* 1.834, *p* = .20.

Chart, box and whisker chart

Description automatically generated

*Figure X. DLTS per level of belief and grip, model-based LS means, corrected for learning effect and saliency effect. Error bars represent standard error of the mean. Stars above bars represent significant mean differences. Stars on the bar represent significance of LS mean difference versus baseline 0. The effects of hand preshaping is largely preserved in the true belief condition, and we still find evidence for belief-based modulation of looking times, in line with out a priori hypotheses*

TO DO:

* See if we can hint to power analysis: see Meteyard & Davies
* Make a table for ANOVAs, including slopes, SE and 95% CI, include random effects

***4.4 Discussion***

In this experiment, we implemented a completely novel combination of action observation and false belief reasoning in an experimental paradigm. Although this study is highly exploratory, we had a couple of firm a priori hypotheses based on previous research looking at action observation (e.g. Ambrosini, Constantini, & Sinigaglia, 2011) or implicit mentalising (e.g. Kovács, Téglás, & Endress, 2010) in isolation.

First of all, our results did confirm our first hypothesis, which states that there would be a significant preshaping effect, essentially providing a conceptual replication of the work of Ambrosini and colleagues (2011). Indeed, we found that gaze was significantly more proactive in the conditions where the hand formed a shape (either a whole hand grip or a precision grip) as compared to the condition in which the hand did not form a shape. That is, participants tended to look longer at the object that matched the hand shape if this information was available. If no motor cue was present (the no-grip condition), the gaze behaviour of participants closely resembled chance level: they were equally likely to look at the target or the distractor object.

Secondly, we did find that belief played a role: predictive gaze behaviour was significantly larger in the true belief condition as compared to the false belief condition. This effect is not surprising, as directing implicit gaze behaviour in the true belief condition can be reliably based on a single cue (hand preshaping), whereas belief-congruent predictive visual sampling requires integrating information from two sources (hand preshaping plus belief state computation) in the false belief condition.

Thirdly, and more critically, we found evidence that when participants were presented with a case in which the agents had a false belief, their predictive gaze behaviour reflected that false belief. That is, in the false belief condition the gaze was biased towards the object that was *belief-congruent* but *motor-incongruent*. This is a rather novel finding. For the first time, we found predictive gaze behaviour as function of the integration of motor cues (hand preshaping) and implicit belief-reasoning processes. That is, if participants were to solely rely on motor cues (i.e., displayed motor-congruent gaze behaviour), their differential looking time scores (DLTS) would fall well below chance level as we scored belief-congruent fixations as positive. However, it seems that participants took hand preshaping as a cue signalling the agent’s action goal, and modulated the resulting predictions based on the false belief state of the agent.

Interpretation here is somewhat complicated in function of the used model. Under the basic model, we find this belief-congruency effect in the precision grip condition only, whereas we find this effect in the whole hand grip under the extended model. This is to be expected, as in the false-belief condition the target object for the precision grip is the large cup. As such, this target is more likely to attract the attention of the participant, and this saliency effect will boost looking behaviour towards this location. Conversely, when the agent has a false belief, the target of a whole-hand grip is the small cup, and the distractor target is the large cup. Therefore, predictive gaze behaviour towards the target location would be diminished by the saliency effect as attention needs to be shifted away from the more salient larger object. Therefore, accounting for the saliency effect within our extended model will attenuate belief-congruent looking bias in the precision grip condition but boost belief-congruent looking bias in the whole-hand grip condition

It bears repeating, however, that the extended model is likely to overestimate the saliency effect by a serious margin. This is because a) the main effect of size is by necessity estimated purely based on the effect of size in the no-grip condition, discounting any potential interaction effects of this variable with grip and/or belief, and b) we found no differences directly attributable to target size in either the true or the false belief conditions. We are therefore likely to be erring on the side of caution if we interpret the results from the perspective of the extended model accounting for target size. Regardless of what model is used, however, our conclusion is that in the false-belief condition, anticipatory looking was significantly biased towards the belief-congruent target and not the motor-congruent target.

Statistical modelling indicated that there was a potential learning effect over time: participants tended to have higher DLTS as the experiment progressed. One possibility for such a learning effect is that participants implicitly learned a behavioural rule due to statistical commonalities between different trials. That is, participants could have arrived at a rule as “when the cups remain in place, the hand goes to the matching cup” in the true belief condition, and “the hand goes to the *opposite* cup when the cups swap”. While at face value this might be a signal that predictive behaviour simply reflects a learned relationship between video events and outcomes, there are some strong arguments against explaining away the results as a mere side-product of statistical learning (see the General Discussion in Chapter 7 for further discussion).

***4.5 Conclusion***

Our results yielded compelling evidence in favour of our hypotheses. Specifically, we found that gaze was proactive when the agent’s grip matched the size of a specific target. No such proactivity was found when there was no grip information. These results provide a direct conceptual replication of previous work on the role of motor representations in predictive eye movement. However, our results expanded upon these by showing that this predictive gaze behaviour was modulated by the false belief state of the agent. This suggests that behavioural predictions based on motor cues were integrated with belief-tracking processes. Although previous research suggest that target size can influence anticipatory looking, we were able to control for such effects. More important is evidence suggesting learning effects, which could indicate that the results may be due to statistical learning rather than belief tracking. This issue will be addressed in Experiment 3.

***Chapter 5: Experiment 2***

Status: reviewed, not edited

***5.1 Introduction***

After Experiment 1, the research goals was to check for the involvement of motor representations. While we have theoretical reasons to expect the involvement of the motor system in estimating the target based on hand aperture of the agent, this is still very much an open empirical question. To tackle this issue, we took the same approach as Ambrosini, Sinigaglia & Constantini (2012). Recall from Chapter 2 that in this study, the research found that temporarily restricting the motor affordances of the participants, so they were unable to physically execute the observed action, resulted in reduced anticipatory looking towards the target of that action as compared to the group that was able to move freely. Following the same reasoning, we postulated that temporarily restricting the motor affordances of the participants in such a way that they would be unable to grasp a cup should reduce the activity in those motor representations related to grasping. As such, we anticipated a reduction in proactive gazing towards the target of the grasping motions of an agent. To that extent, one group of participants would watch the same videos as in Experiment 1, while the hands were constrained. A second group without any such restrictions would do the same, serving as a baseline condition. Furthermore, this baseline group could provide a replication of the results of our first study.

Our specific hypothesis for this experiment was as follows: we expect to find a significant main effect of group, and a significant interaction effect of group with grip condition. Specifically, we expect to find that the factor group has an effect on differential looking times scores in the whole hand grip and precision grip conditions only. While in analogy to Ambrosini and colleagues (20120) we expect that previous effects of motor cues (true belief condition) and of implicit false-belief tracking (false belief condition) will still be present, we expect to find that these effects will be significantly reduced in the movement-restricted group as compared to the baseline (unrestricted) group, resulting in significantly lower differential looking time scores.

***5.2 Methods***

*5.2.1 Participants*

A total of 125 participants participated in this study (Mean age= 19.3, *SD* = 1.78, 85 females). These participants were first year undergraduate students made available by Victoria University of Wellington’s Introduction to Psychology Research Programme. After data cleaning (details can be found below), the final sample size was N = 99*.* Groups were more or less equally matched, with 49 participants in the baseline group and 50 participants in the restricted group. All provided informed consent and had normal to corrected vision. The experimental procedure received ethical approval from the Victoria University of Wellington School of Psychology Human Ethics Committee.

To do: power analysis

*5.2.2 Stimuli and Apparatus*

Stimuli and apparatus were analogous to Experiment 1. The only addition was a group that had motor affordances restricted. This was done by means of a soft elastic band. Participants were asked to hold on firmly to the band, clutching their hands around the band behind their backs, during every block of trials. Participants were told that this would correct posture during the eye-tracker sessions as it straightened out the back.

*5.2.3 Procedure*

The procedure followed was analogous to that of Experiment 1, with the exception that participants in the movement-restricted group were required to hold on to the elastic band at the start of every experimental block.

*5.2.4 Data-analysis*

The approach to data-analysis was analogous to Experiment 1. Of the original 125 participants, 9 participants were excluded for failing the debrief questions, 14 for explicitly mentalising, and 2 for subpar data quality. Of the 8043 recorded trials, the participant exclusion left 6664 trials. The trial-based exclusion criteria further reduced this number: 748 trials were removed for failing the attention check, 1221 trials were removed for showing no fixation towards either target in the time of interest, and 29 trials were removed for being incomplete, resulting in an overall exclusion rate of 41.9%.

***5.3 Results***

*5.3.1 Basic model*

First, we implemented a full model that included the factors group, belief, grip and their interaction effects. There was no main effect of group, *F*(1,99.4) = .122, *p* = .728. There was no significant main effect of grip, *F*(2,4581.8) = 2.821, *p* = .06. There was, however, a significant main effect of belief, *F*(1,4581.3) = 11.209, *p* < .001 (see Figure X, table X). Moreover, the interaction effect between belief and grip was also significant, *F*(2,4578.6), *p* = .032. The group factor showed no interaction effect with either belief, *F*(1,4581.3) = .089, *p* = .766, nor with grip, *F*(2,4581.8), *p* = .605. There was no significant three-way interaction effect between group, belief and grip, *F*(2,4578.6) = 2.925, *p* = .054.

|  |  |  |  |
| --- | --- | --- | --- |
| *Belief* | *Mean (LS)* | *SE* | *95% CI* |
| FB | -.0002 | .008 | -.016 ; .015 |
| TB | .028 | .008 | -.012 ; .043 |

Chart, box and whisker chart

Description automatically generated

*Figure X, table X. Table represents the Least Square means for the main effect of belief, as well as standard error of the LS mean and 95% confidence interval. These means are visually represented on the figure, with stars indicating the level of significance of the mean difference. Error bars represent standard error of the mean. This main effect of belief condition is not expected under our hypotheses and diverges from results found in Experiment 1.*

Taking a closer look at the significant factors in our model through Holm-corrected pairwise tests, we found that DLTS scores are significantly higher in the true belief condition as compared to the false belief condition, *t* = -3.348, *p*<.001.

We also performed a series of follow up tests on the interaction effect between belief and grip, as per our a priori hypotheses. An overview of the Least Squares Means of the interaction effect between belief and grip can be found in Table X and Figure X. Also included are significance test statistics of every model mean with the baseline 0.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Belief - grip* | *LS Mean* | *SE* | *95% CI* | *t* | *p* |
| FB - WHG | -.026 | .012 | -.049 ; -.004 | *-2.3* | *.066* |
| FB - NG | .006 | .011 | -.016 ; .029 | *.56* | *.574* |
| FB - PG | .019 | .011 | -.003 ; .042 | *1.727* | *.169* |
| TB - WHG | .032 | .011 | .01 ; .054 | *2.842* | *.009\*\** |
| TB - NG | .017 | .012 | -.006 ; .04 | *1.457* | *.145* |
| TB - PG | .034 | .011 | .011 ; .056 | *2.959* | *.009\*\** |

*Table X. The model-based Least Square Means of every level of the interaction effect between belief and grip conditions, as well as their standard error and 95% confidence interval. Also included is the significance test (t-tests) of each LS mean against 0. All reported p-values are Holm-corrected for multiple comparisons.*

In the true belief condition, we found that DLTS was significantly higher than baseline in both the whole hand grip, *t* = 2.828, *p* = .009, and the precision grip conditions, *t* = 1.975, *p* = .009. This is in line with our hypotheses and confirm results of Experiment 1. In the false belief conditions, participants’ gaze was not significantly biased towards one side or the other (all *p*>.05). This result is at odds with our hypotheses and our previous findings in Experiment 1.

We also tested if there was an effect of belief within each level of grip. We found that in the whole hand grip condition (FB-WHG versus TB-WHG), participants looked significantly longer at the target in the true belief condition than the false belief condition, *t* = 4.078, *p*<.001. In the other grip conditions, the true belief versus false belief contrast was not significant (all *p*>.05).

Chart, box and whisker chart

Description automatically generated

*Figure X. Model-based Least Square means in function of belief and grip conditions. Error bars represent the standard error of the mean. Stars on the bars represent the significance of each LS mean compared to baseline 0. Stars above the bars represent the significance levels of the difference between means. The effect of preshaping was significant in the true belief condition. In the false belief condition, no significant deviations from baseline 0 were observed.*

Contrary to our hypotheses, we did not find a significant main effect of group, nor a significant interaction effect with grip, nor a significant three-way interaction effect of group, grip and belief. We did test the crucial contrasts as per our a priori hypotheses regardless. An overview can be found in Table X. However, none of these contrasts were significant.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Belief x grip* | *Contrast: group* | | *t* | *p* |
| FB-WHG | baseline – restricted | 1.538 | | .498 |
| TB-WHG | baseline – restricted | -1.143 | | .76 |
| FB-PG | baseline – restricted | -.777 | | .875 |
| TB-PG | baseline – restricted | .494 | | .875 |

*Table X. Contrast tests of model-based LS means between groups, per level of belief and grip, pre-shaping grip conditions only.*

A visual representation of the different levels of belief and grip in function of group can be found in Appendix X.

*5.3.3 Extended model*

We extended our model to include the main effect of block number to check for learning effects. Indeed, we found that there was a significant effect of block number, with DLTS tending to be higher as trial number increases, *F*(3,4638.7) = 2.623, *p* = .047. Adding this variable to the model had no impact on other main effects and interaction effects. Post-hoc t-tests (see Table X, Figure X) found that DLTS is significantly higher in the 4th block as compared to the 1st block, *t* = -2.799, *p* = .031. None of the other contrasts reached significance.

Chart, box and whisker chart

Description automatically generated*Table X. Overview of post-hoc t-tests on LS means between blocks.*

|  |  |  |
| --- | --- | --- |
| *Contrast* | *t* | *p* |
| 1-2 | -1.218 | .561 |
| 1-3 | -1.528 | .507 |
| 1-4 | -2.799 | .031\* |
| 2-3 | -.334 | .738 |
| 2-4 | -1.656 | .489 |
| 3-4 | -1.32 | .561 |

*Figure X. Main effect of block number, model-based LS means. Error bars represent standard error of the mean, stars represent significance of LS means difference.*

An overview of LS means per level of block, belief and grip can be found in Appendix X-X. Note that increments are similar over all levels of belief and grip as we included only the main effect of block number as to not overcomplicate the model.

We further expanded the model to include the main effect of target size, as per Experiment 1. Indeed, there is a significant effect of target size on differential looking time scores, *F*(1,4589) = 19.919, *p* < .001. This was confirmed by a follow-up t-test, showing that DLTS is significantly larger when the target is the large cup, as compared to when the target is the small cup, *t* = 4.463, *p* < .001 (see Table X, Figure X).

We checked for other indicators of the effect of target size as well. In the true belief condition, we compared the whole hand grip condition to the precision grip condition. This showed no significant DLTS bias towards the large object (TB-WHG), *t* = -.117, *p* = .907. In the false belief condition, we did find that DLTS was significantly increased when the belief-congruent target was large (FB-PG) as compared to when it was small (FB-WHG), *t =* 3.204, *p* = .001. This signals that target size might have played no role in the true belief condition, but did affect results in the false belief condition.

|  |  |  |  |
| --- | --- | --- | --- |
| *T size* | *Mean (LS)* | *SE* | *95% CI* |
| Large | .047 | .01 | -.028 ; .066 |
| Small | -.018 | .01 | -.037 ; .002 |

Chart, box and whisker chart

Description automatically generated

*Figure X, table X. Table represents the Least Square means for the main effect of target size, as well as standard error of the LS means and 95% confidence interval. These means are visually represented on the figure, with stars indicating the level of significance of the mean difference. Error bars represent standard error of the mean.*

Note that it is still impossible to check any interaction effects with belief and grip conditions as target size is directly tied to belief and grip combinations. As such, we only included a single slope for target size in our model, de facto assuming that the effect of target size is uniform over levels of belief and grip. It follows that this slope (*b =* -.0645) is estimated based on the effect of target size in the no-grip condition, since this is the only grip condition in which there is no direct link between grip and target size.

It is obviously of great interest to see what impact the significant main effect of target size has on the results and what the estimated means look like when accounting for the target size effect. Below, we provide an overview of the model-based Least Square means for a model that includes: the three-way interaction effect between group, belief and grip, the main effects of these three variables, as well as the main effects of block number and target size.

A Type III Analysis of Variance Table with Satterthwaite's method of the final model yields the following results (see below, Table X). Note that the main effect of belief and the interaction effect between belief and grip remain unchanged as compared with the basic model (see above).

|  |  |  |  |
| --- | --- | --- | --- |
| *Fixed effects* | *df* | *F* | *p* |
| Target size | 1, 4587 | 19.919 | <.001\*\*\* |
| Block | 3, 4638 | 2.8 | .038\* |
| Group | 1, 100 | .131 | .717 |
| Belief | 1, 4578 | 11.44 | <.001\*\*\* |
| Grip | 2, 4579 | 2.88 | .056 |
| Group × belief | 1, 4578 | .092 | .762 |
| Group × grip | 2, 4579 | .512 | .597 |
| Belief × grip | 2, 4580 | 3.896 | .021\* |
| Group × belief × grip | 2, 4575 | 2.936 | .053 |

*Table X. Overview of the significance tests of each of our fixed effects and interaction effects.*

Below we also look at our main hypotheses, testing on the interaction effect between belief and grip under a model that controls for a potential size effect.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Belief - grip* | *LS Mean* | *SE* | *95% CI* | *t* | *p* |
| FB - WHG | .008 | .014 | -.019 ; .034 | *.556* | *1.00* |
| FB - NG | .006 | .011 | -.016 ; .028 | *.564* | *1.00* |
| FB - PG | -.011 | .013 | -.037 ; .015 | *-.843* | *1.00* |
| TB - WHG | .001 | .013 | -.025; .028 | *.085* | *.932* |
| TB - NG | .018 | .012 | -.004 ; .041 | *1.574* | *.232* |
| TB - PG | .068 | .014 | .041 ; .094 | *4.997* | *<.0001\*\*\** |

*Table X. The model-based Least Square Means of every level of the interaction effect between belief and grip conditions, as well as their standard error and 95% confidence interval. Also included is the significance test (t-tests) of each LS mean against 0. All reported p-values are Holm-corrected for multiple comparisons.*

As we can see in the Table X, with ‘target size’ controlled for, differential looking time scores were only biased towards the target in the TB-PG condition, *t =* 4.997, *p* < .001. In all the other conditions, participants’ gaze did not differ significantly from baseline 0, that is no bias towards either the target or the distractor.

When testing for the effect of belief per grip condition, we found that in the precision grip condition, DLT scores were significantly higher in the true belief condition as compared to the false belief condition, *t* = 3.875, *p* <.001. In both the whole hand grip condition, *t =* -.314, *p* = .835, and the no-grip conditions, *t* = .811, *p* = .835, DLT scores were no different when comparing the true and the false belief condition.

Chart, box and whisker chart

Description automatically generated

*Figure X. Model-based Least Square means in function of belief and grip conditions, for a full model accounting for the effect of target size. Error bars represent the standard error of the mean. Stars on the bars represent the significance of each LS mean compared to baseline 0. Stars above the bars represent the significance levels of the difference between means.*

Finally, we used a series of log-likelihood test to make a stepwise comparison between the basic model and models including block number and grip effect as covariates. This revealed that the inclusion of block number results in a model that is a significantly better fit than the basic model, χ2 (3) = 7.989, *p* = .046. Including the main effect of target size yields an even better fit, χ2 (1) = 19.953, *p* < .001. We therefore conclude that it is more prudent and correct to consider the results as per the extended model that accounts for learning effects and the effect of target size, at least for the false belief condition. In the true belief condition, it is likely that the target size effect estimated based on the no-grip condition misrepresents the data.

***5.4 Discussion***

First and foremost, this experiment served as an attempt to - at least conceptually - replicate the findings of Ambrosini et al. (2012). Using a similar design as in the first experiment, we tested if performance in a group that was movement-restricted would yield different proactive looking patterns as compared to a baseline group. To that effect, participants in the movement restricted group held on to a soft elastic band behind their backs.

However, results indicated that motor restriction yielded no significant group differences: we found no effect of group, and post-hoc comparisons of conditions within this group with their counterparts in the baseline (unrestricted) group. This was unexpected. One potential explanation for the lack of effect of motor restriction is the comparatively long measurement window (M = 4.264 s, SD = .503 s) and the specific measure used. That is, we used differential looking times over a longer time period as dependent variable. In contrast, Ambrosini and colleagues (2011) reported differences due to motor restriction on the timing (onset time and arrival time) of gaze shifts towards a target. Similarly, in Low et al. (2020) differences due to motor restriction were reported on reaction times. It may therefore be that effect of motor restriction will only be readily apparent on more fine-grained measures such as reaction time and gaze shifting, but not in overall attentional bias over longer periods of time. A further, perhaps practical, explanation might be that having participants clutching tightly onto the elastic band as we did here is potentially less disruptive to motor processing and resulted in a reduced level of impairment on motor affordances. Ambrosini and colleagues (2011) kept participants’ hands rigidly tied up behind their backs and their participants also had to keep their wrists crossed and keep their fists clenched throughout the experiment. It must be pointed out, however, that motor restriction does not always produce the hypothesised attentional shifts expected under a mirror neuron account (Shaw, 2014; McMannus & Thomas, 2018). Indeed, recall from Chapter 2 that Vannuscorps and Caramazza (2016; 2017) found that proactive gaze in tasks akin to Ambrosini and colleagues (2011) was preserved in patients without upper limbs. This is an indication that the ability to perform an action is not a prerequisite to understand that action and predict its outcome. More research is needed to clarify when and how motor restriction impacts object-related attentional biases.

Secondly, this study was intended to serve as a high-power replication attempt of our findings in the first experiment. However, while the power was almost double, the experiment also encountered technical difficulties, mainly in the form of software issues leading to reduced equipment performance: low framerate, high rate of frames dropped and as a result unreliable timings for recorded fixations. Nevertheless, we did find similar result patterns as in the first experiment. Similar to our first experiment, we found that there was a significant effect of belief, with a higher rate of target-oriented predictive gaze behaviour in the true belief condition as compared to the false belief condition. However, we found no main effect of grip in this study. Instead, grip interacted with belief.

Post-hoc analyses under the basic model revealed that in the true belief condition, results followed a similar pattern as in the first experiment: participants tended to look significantly more often towards the cup that matched the grip aperture of the agent. In the no-grip condition, participant did not show a bias towards either AOI. In the false belief condition, however, results start to diverge from the first experiment. That is, participant gaze was neither significantly biased towards the belief-congruent nor the motor-congruent object. However, it could be argued that gaze behaviour that is not significantly biased towards the motor-congruent object is in itself evidence for belief-based modulation of predictive gaze behaviour, as will be discussed in the General Discussion in Chapter 7.

Thirdly, here again we checked for the influence of two important variables that can potentially be confounders: learning effect and a general saliency effect for the larger cup. Current results are in line with those in the first study: there seems to be a learning effect where participants will show increasing target-congruent gaze biases as the experiment progresses. Moreover, there was a significant effect of target size, with increased looking times towards the larger object. In this case, however, we did find that DLTS in the false belief condition was biased towards the large target. In the true belief condition, no difference attributable to target size was found. This is tentative evidence in favour of a target size *x* belief interaction effect. That is, it is unlikely that target size has an impact on gaze in the true belief condition, while it is likely that target size impacts looking time patterns in the false belief condition. This means that the disappearance of the preshaping effect in the true belief condition under the extended model could very well be due to an overestimation of the target size effect. In this case, we might be better off to interpret the results under the basic model. However, it is advisable to look at the results of the false belief condition with the target size effect in mind, as per the extended model.

To this effect we extended the model with these covariates and checked our hypotheses again. This, however, changes little to our interpretation of the data: in the false belief condition, we still see DLTS patterns that are not biased towards either the belief-congruent, nor the motor-congruent object. In the true belief condition, we still see clear evidence for a motor-based preshaping effect.

It has to be noted that this study shows differential looking time scores that are much more trending towards the baseline zero as compared to the first experiment. We do not have a direct explanation for this finding, other than increased signal noise due to technical difficulties.

***5.5 Conclusion***

Overall, our results are a partial replication of our first study. In the true belief condition, we again saw a clear effect of hand preshaping on anticipatory looking times. In the false belief condition, we saw that anticipatory looking times generally sat around chance level. Although less clear-cut than results in the first experiment, this still entails that overall looking scores were not completely driven by hand preshaping, at least suggesting an impact of the opposing belief state of the agent. Again, we have to be cognizant of the confounding effect of object saliency due to object size as well as potential learning effects. Our crucial manipulation in this experiment was the addition of a group that was motor-restricted and therefore unable to grasp for a cup, which should in principle attenuate motor simulation and action prediction. This, however, was not the case. We attribute this null-effect to differences in our paradigm in relation to previous studies, although our results could also add to recent studies (McMannus & Thomas, 2018; Vannuscorps & Caramazza, 2016; 2017) calling into question the impact of motor restriction.

***Chapter 5: Experiment 3***

Status: reviewed, not edited

***5.1 Introduction***

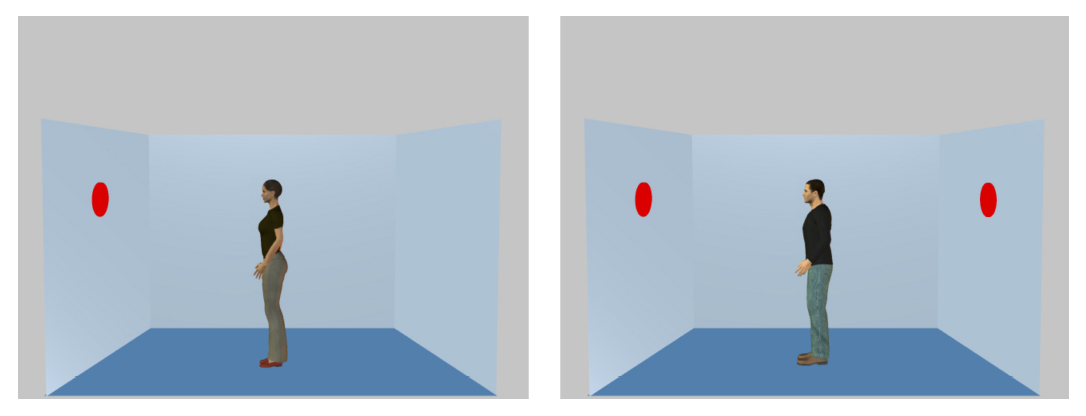
The final study under our research plan was primarily aimed at investigating to what extent the processes under investigation are function of cognitive resources. Furthermore, we extended the paradigm to include an explicit response component to investigate the differential impact of cognitive resources on implicit and explicit prediction making. This allows us to investigate the correlation between implicit eyetracking measures we have used thus far and explicit prediction making. Lastly, we expanded our debriefing procedure slightly to tap into qualitative aspects of our study. As there now was an explicit component to the task, what did the participants themselves think drove their prediction-making process?

The addition of cognitive load followed the reasoning that demands on cognitive resources would a have different impact in the different processes under scrutiny. More specifically, we expect that cognitive load should have small to non-existent effects on the processes we considered to be implicit, and therefore *efficient* and minimally relying on cognitive resources. Conversely, we expect to find an effect of secondary task load on the outcome of the explicit mentalising process, as measured by explicit predictions of the participant on the behaviour of the agent.

Firstly, purely motor-driven implicit responses, i.e., looking times in the true belief condition, should be relatively immune to secondary task load. This hypothesis is derived from key tenets of the direct-matching hypothesis and ideomotor theory of action and action observation: that observing an action unfold *automatically* activates matching action plans based upon the bi-directional connectivity between representations of that action and its typical sensory consequences through so-called mirror neurons (e.g., Rizolatti, Fadiga, Gallese & Fogassi, 1996, see Chapter 2, Section 2.2). A strong line of evidence for this notion comes from the literature on *automatic imitation*, which shows that observing an action facilitates performance of that action, even if the observed action is irrelevant to the current task (e.g., Brass, Bekkering, Wohlschläger & Prinz, 2000; Brass, Bekkering & Prinz, 2001, See Chapter 3, Section 3.2). Automatic imitation is generally taken as a reliable proxy for processes within the mirroring system. The automaticity of imitative processes is well supported by empirical work, in the sense that this process is fast, occurs independently of intentions and persists even under high attentional load (Heyes, 2011; Cracco et al., 2018; Ramsey, Darda & Downing, 2019). Although attentional factors and added cognitive load significantly modulate the strength of imitative responses (Chong, Cunnington, Williams & Mattingley, 2008; 2009; Puglisi, Leonetti, Cerri & Borroni, 2018), the automatic imitation response generally does not disappear altogether (Cracco et al., 2018). We therefore expect that cognitive load could modulate but not nullify the effect of hand preshaping in the true belief condition, as the current consensus on the mirroring system holds that mirroring processes function sufficiently automatic to occur even under cognitive load.

Secondly, under a dual-process of mentalising, we would expect little to no effect of cognitive load on implicit mentalising measures, i.e., looking times (Butterfill & Apperly, 2013). This position is perhaps more contentious given previous research. In a study by Schneider, Lam, Bayliss & Dux (2012), participants watched videos that in which an agent observed a puppet move a ball between two boxes in front of her, following a classic false-belief scenario that had the puppet move the ball from one box to the other in the absence of the agent, creating a false-belief condition. Crucially, the researchers manipulated task load. One group of participants was subjected to a secondary task with a relatively low cognitive load in the form of paying attention to an auditory stream of letters. A second group had a significantly harder task: they were to count the number of “2-back” repetitions in the auditory stream. A third baseline group simply watched the videos. Crucially, in the no-ball location participants showed significantly higher looking times in the false belief condition as compared to the true belief condition, which is taken as index for implicit false belief processing. This index effect disappeared under both low cognitive load and high cognitive load, as looking times were in line with an egocentric bias (the actual location of the ball). However, to the best of our knowledge this study has not been replicated and can therefore not be considered conclusive. With this scarcity of evidence in mind, it is informative to look at evidence for the relative automaticity of related concepts.

More evidence has accumulated on the automaticity of *level 1 visual perspective taking*. This refers to the ability to infer whether someone else can see a specific stimulus. This ability is held as a critical component of mentalising, especially implicit forms of mentalising. Researchers found that participants implicitly represented the task-irrelevant perspective of an avatar in a dot-counting task, resulting in an altercentric interference effect (Samson, Apperly, Braithwaite, Andrews & Scott, 2010).



*Figure X. Original stimuli from the experiment by Samson et al. (2010), adapted from O’Grady, Scott-Phillips, Lavelle & Smith (2020). In this task, one is required to make a judgement on the number of dots based on the self-perspective or the other-perspective. The judgement itself consist of a button press if a digit on the screen (e.g., “1” or “2”) matches the number of dots seen from this perspective. The panel on the left represents a “congruent” trial, as the number of dots from the perspective of the participant is equal to that of the avatar. The panel on the right is an “incongruent” case, as the participant sees two dots while the avatar sees only one. Typically, such incongruent trials result in a slower response time in the self-perspective, indicating that an implicitly held representation of the avatar’s perspective interferes with judgements made from one’s own perspective.*

Qureshi, Apperly and Samson (2010) had a group of participants perform this task while simultaneously performing a task that had participants perform a tapping task, requiring to give an opposite number of key presses to a set number of tones (e.g., one press for two tones heard, and vice versa). This task placed additional demands on executive function due to the necessity to inhibit a congruent response. Results indicated that the secondary task *increased* altercentric and egocentric biases in the visual perspective taking task, suggesting that the additional executive demands affected perspective selection but not perspective calculation, bolstering the case for the relatively automatic representation of other’s visual perspective. Similarly, Todd, Cameron and Simpson (2017) showed that these altercentric intrusion effects are relatively immune to added time pressure. Recent work, however, has found limitations to the automaticity of Level 1 perspective taking. When combining the dot perspective task above with a secondary task that requires participants to hold a series of letters in working memory, Qureshi and Monk (2018) found that this additional working memory demand diminished altercentric intrusions, suggesting that working memory is required for maintaining a representation of someone’s visual perspective. At present the emerging picture is still unclear, but a brief summary would be that executive functioning is involved in inhibiting the egocentric perspective and selecting the altercentric perspective. Moreover, while the altercentric perspective seems to be calculated in a relatively automatic and efficient manner, working memory could be required to *maintain* a representation of the altercentric perspective (Martin, Perceval, Davies, Su, Huang & Meizer, 2019).

Given the similarity between both concepts (i.e., automatic representation of another’s visual perspective and automatic representation of another’s object registrations), we considered that we would also find similar results on implicit measures of belief-tracking. That is, as we expect implicit tracking of agent registrations to be a relatively efficient and automatic process; we expect to find that additional demands placed on executive functioning could diminish the previously found belief-tracking effect but will not nullify it, provided that this secondary task places little to no demands on working memory.

We do expect, however, that cognitive load will have an impact on the accuracy of any explicit judgements made on the future behaviour of the agent. There is precedent for this expectation. In a study by Lin, Keysar & Epley (2010), for example, it was found that participants under working memory load showed a reduced capacity to apply mentalising processes to a communication game, indicating that explicit mentalising is an effortful process that requires attentional resources to function. These results were replicated by Cane, Ferguson and Apperly (2017), who furthermore showed that this effect is modulated by motivation. Finally, a recent longitudinal study of Kloo, Kristen-Antonow & Sodian (2020) found that explicit but not implicit mentalising capabilities are related to the development of executive functioning in preschool children, in line with a dual-process account of mentalising. Taken together, these studies suggest that the ability of the participant to explicitly predict the future behaviour of the agent based explicit on mentalising computations should be selectively reduced by added secondary task load and subsequent reduction in attentional resources.

The secondary task we implemented in this study is adapted from a recent study by Michael, Wolf, Letesson, Butterfill, Skewes & Hohwy (2017). In this study, researchers were interested assessing whether a secondary task posing additional executive demands would preserve Level 1 perspective-taking in the dot perspective task. The task consisted of an auditory tone monitoring task, where participants had to judge if two rapidly presented tones had the same frequency or a different one (more details in Section 5.2.2). Michael and colleagues (2017) found that this secondary task indeed impacted spatial cueing but not visual perspective-taking. As this task requires attentional shifting to the auditory stimuli and discriminating them rather maintaining a representation of them in working memory, this task should be taxing on attentional processes but not working memory, rendering this task theoretically capable of distinguishing between the hypothesised implicit and explicit mentalising processes under investigation in this thesis. In our experiment, participants had to remember their answer and respond at the end of the trial, purely for practical reasons. We expect that this modification will place little to no additional demands on working memory as only one element needs to be recalled at the end of the trial (“same tone” or “different tone”). We therefore expect that in the group of participants under this secondary task load the proactive gaze behaviour effects will largely be preserved, in both the true and the false belief conditions, indicative of mirroring and implicit mentalising processes, respectively. We do, however, expect that the reduction in attentional resources will severely impact explicit predictions, as measured by overt responses, resulting in a significant group effect on this measure.

In order to measure any explicit judgements, we cut the ending of the videos so that the participant never saw the hand actually reaching for a cup. Instead, we asked the participants to make an explicit statement of where the agent would reach: left or right. Additionally, this change to the stimuli means that we can exclude any event-outcome learning rules, as discussed in Chapter 3. Furthermore, this will allow us to correlate implicit and explicit measures for mentalising. This is of theoretical importance as recent studies have failed to find to replicate numerous anticipatory looking implicit mentalising paradigms (see Chapter 2). In light of these non-replications it has been questioned if anticipatory looking is a reliable method of measuring mentalising processes (Kulke & Hinrichs, 2021; Dörrenberg, Rakoczy & Liszkowski, 2018). It is therefore of interest to find convergent validity between explicit and implicit measures to support the notion that differential looking behaviour is indeed a reflection of behaviour anticipation due to mentalising processes.

In sum, the third experiment in this thesis implements a modified version of our action-observation task, but with an added group condition. A first group will perform the task with an added task demand aimed at selectively impairing attentional shifting but not working memory. The second group would not receive any additional task demands, serving as a baseline condition. We expect that in this baseline condition we will replicate previously found effects of motor cueing (true belief condition) and implicit belief-tracking (false-belief condition) in anticipatory looking times, and that these effects will extend to the accuracy of explicit predictions as well. Furthermore, in the true belief condition, where there is no need for explicit mentalising, we expect that the secondary task will largely preserve these effects in anticipatory looking times as these are taken as reflective of the outcome of relatively automatic and efficient mirroring and implicit mentalising processes. In the false belief condition, however, we expect that these effects will be preserved in anticipatory looking times but *not* in explicit predictions, which should be function of effortful and cognitively demanding explicit mentalising processes. We expect to find a significant positive correlation between implicit anticipatory looking times and the accuracy of explicit predictions. This correlation should be weaker in the false-belief trials in the group under secondary task load, however.

***5.2 Methods***

*5.2.1 Participants*

After data cleaning (see below, 1.4), the total number of participants was *N* = 48 (Mean age = 19.4, SD = 1.976, 37 female, 11 male, 0 other). These participants were first year undergraduate students made available by Victoria University of Wellington’s Introduction to Psychology Research Programme. Due to external circumstances (COVID-19) the groups were slightly unbalanced, with 20 participants in the baseline group and 28 participants in the secondary task group. Our statistical methods are robust to such group imbalances, however (Brown, 2021). All provided informed consent and had normal to corrected vision. The experimental procedure received ethical approval from the Victoria University of Wellington School of Psychology Human Ethics Committee.

*5.2.2 Stimuli and Apparatus*

The eye-tracking setup was analogous to previous experiments. We also used the same video stimuli as in previous experiments. However, in contrast to previous experiments the videos were cut off right before the hand appeared out of the occluder. This was done by ending the video loop 8 frames prior to the “hand appearing” event in each video.

Unlike previous experiments, this experiment featured a question prompt (“on which side of the screen will the hand appear?”) after each trial upon participants were asked to respond vocally with “left” or “right”. This was the case for both experimental groups. Answers were recorded usingan external microphone, as well as noted down manually by the experimenter. The voice records were stored and analysed off-line after the main experiment sequence was completed using Google voice recognition software.

In the mind-tied group, this experiment also included a secondary task. The point of this task was to exert attentional load upon the participants during the measurement period of the experimental trials. The task consisted of an auditory discrimination task (see Section 5.1). Tones consisted of two 30 ms segments of a simple sine wave sampled at either 800 or 1200 Hz. Tones were offset with 266 ms. The tones were selected at random, in counter-balanced fashion. I.e., in 50% of cases the tones would match, in the other 50% of cases they would be different. The first tone was presented 100 ms after the fixation arrow disappeared.

*5.2.3 Procedure*

Up until first calibration of the Eyetracker, the experimental procedure was the same as in previous experiments. However, after the first calibration participants were presented with the instructions for the main body of the experiment, which also introduced the explicit prediction task (see below). In the mind-tied condition, another instruction was added including the tone detection task. That is, participants were instructed to memorise if they thought the tones they heard during the video were the same tone (i.e., the same frequency) or different from each other. We then tested if the participants understood the instructions by having them complete a random test-trial. This test trial would be repeated (each time with a new random video) until the participants showed sufficient understanding of the tasks.

A total of 60 trials were displayed in the main section of the experiment. These were divided over 5 blocks of 12 trials each. Per block, we displayed 4 false belief trials with 2 whole hand grip and 2 precision grip trials, 4 true belief trials with 2 whole hand grip and 2 precision grip videos, and 4 no-grip trials, of which 2 were false belief and 2 were true belief (although this is only nominal in this condition as there is no motor information to signal intention of the agent). Between blocks the participant had an opportunity to disengage from the eye-tracker setup and take a break. The participant could choose when to resume the experiment. The eye-tracker was recalibrated before a new block of experimental trials was commenced.

Each trial featured the same videos as in previous experiments, with the notable exception that the video display loops were cut off 8 frames before the hand appeared out of the occluder. As such, the participants never at any point saw the hand actually reaching for a cup. During phases 1 and 2, there was no difference with previous experiments. However, in the mind-tied condition the tones were presented after the fixation arrow disappeared. After the video was over, a prompt appeared with the explicit prediction question (see above). Participants had 3 seconds to respond with a vocal “left” or “right”, at normal conversational level. In the mind-tied group, this prompt was followed up with a prompt that asked if the tones they heard were the same or different (“were the two tones the same or were they different?”). Participants responded to this question by a keypress on the keyboard: “S” for “same” and “D” for “different”.

After the final block, we had a short debriefing session. We had the same two binary response questions (yes/no) as in previous experiments. Specifically, “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?".

We also included a third open ended question: “What made you decide in picking left or right on the prediction task?”. If the participant indicated that they based their answer on hand preshaping, we would ask the follow up question: “did you take into account that the agent couldn’t see the cups being swapped?”. Answers were coded by the experimenter based on similarities in participants’ responses. We were of course most interested in responses that indicated that the participants used either the hand preshaping, or the hand preshaping in combination with the agent’s belief state (i.e., if they were aware that the agent could not see the cups being swapped.

*5.2.4 Data-analysis*

To a large extent, the methodology we used was analogous to that in previous experiments, with a few exceptions due to differences in design. All experimental data (eye-tracking data and behavioural data) were imported into R for analysis. We withheld participants (N = 13) from either group condition from analysis that had a wrong answer on the essential debrief questions (5 participants) or had an incomplete number of data points for all combinations of within-participant conditions (2 belief conditions *x* 3 grip conditions, for 6 data points per participant; 8 participants withheld). On a trial level, we removed all trials that had 1) a framerate lower than 59.6 FPS (52 trials), 2) were not completed (0 trials), 3) had timing issues due to eye-tracker and video streams being mismatched (185 trials), 4) failed the attention checl (169 trials), or 5) had no fixations during our Time of Interest (TOI) in either one of the areas of interest (1003 trials). This left 1757 trials, for an overall exclusion rate of 45.5%.

AOIs were constructed in analogy with previous experiments. TOI was defined as the interval between the disappearance of the fixation arrow and the end of the video.

In analogy with how differential looking time scores (DLTS) are assigned, we scored explicit responses as correct if they are in line with were the hand would reach if the videos were shown completely. That is, in the true belief condition, the correct side is the side where the cup size matches hand pre-shaping. In the false belief condition, the correct answer reflects the side where cup size is *opposite* to hand pre-shaping, as the cups have been swapped, rendering the belief of the agent to be false. Therefore, a correct answer in the false belief condition reflects reasoning in line with the agent’s beliefs.

***5.3 Results***

*5.3.1 Differential Looking Time Scores*

*5.3.1.1 Basic model*

First of all, we implemented a basic model that looks at the main effects of group, belief and grip, as well as their interaction effects. An overview of the hypothesis tests on effect significance of model parameters can be found below in Table X.

|  |  |  |  |
| --- | --- | --- | --- |
| *Fixed effects* | *F* | *df* | *p* |
| Group | 1.791 | 1, 47.26 | .187 |
| Belief | 128.852 | 1, 1724.8 | <.0001\*\*\* |
| Grip | 2.619 | 2, 1721 | .073 |
| Group × belief | .914 | 2, 1724.8 | .34 |
| Group × grip | .803 | 2, 1721 | .448 |
| Belief × grip | 31.681 | 2, 1717.16 | <.0001\*\*\* |
| Group × belief × grip | .461 | 2, 1717.16 | .63 |

*Table X. Results of Type III Analysis of Variance with Satterthwaite's method of the basic model.*

Contrary to our hypothesis, we did not find a significant effect of group, nor did this fixed effect interact with any of the other fixed effects. A three-way interaction was also non-significant. The main effect of belief was significant, however, as was the interaction effect of belief and grip.

A post-hoc follow up test revealed that DLTS was, over levels of group and grip, significantly higher in the true belief condition as compared to the false belief condition, *t =* -11.348, p<.0001.

|  |  |  |  |
| --- | --- | --- | --- |
| *Belief* | *Mean (LS)* | *SE* | *95% CI* |
| FB | -.046 | .007 | -.06 ; -.033 |
| TB | .049 | .007 | .036 ; .063 |

Chart, box and whisker chart

Description automatically generated

*Figure X, Table X. An overview of the model-based LS means for DLTS in function of belief, including standard error and 95% confidence interval. These means are visually represented on the figure, with stars indicating the level of significance of the mean difference. Error bars represent standard error of the mean.*

We also performed a series of follow up tests on the interaction effect between belief and grip, over different levels of group, as per our a priori hypotheses. An overview of the Least Squares Means of the interaction effect between belief and grip can be found in table X and Figure X. In the true belief condition, DLTS was significantly higher than baseline, reflecting anticipatory looking towards the grip-congruent cup, in line with the findings of previous experiment. Importantly, we found that in the false belief condition the mean DLTS was significantly below baseline 0. In other words, contrary to previous experiments and our a priori hypotheses, participants looked significantly longer at the *motor-congruent* and *belief-incongruent* target. In the no-grip condition, results did not differ from baseline in either belief condition.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Belief - grip* | *LS Mean* | *SE* | *95% CI* | *t* | *p* |
| FB - WHG | -.074 | .011 | -.095 ; -.053 | *-6.85* | *<.0001\*\*\** |
| FB - NG | .015 | .011 | -.037 ; .007 | *-1.339* | *.181* |
| FB - PG | -.05 | .011 | -.071 ; -.029 | *-4.669* | *<.0001\*\*\** |
| TB - WHG | .096 | .011 | .075 ; .117 | *8.911* | *<.0001\*\*\** |
| TB - NG | -.009 | .011 | -.031 ; .013 | *-.796* | *.426* |
| TB - PG | .061 | .011 | .04 ; .082 | *5.696* | *<.0001\*\** |

*Table X. The model-based Least Square Means of every level of the interaction effect between belief and grip conditions, as well as their standard error and 95% confidence interval. Also included is the significance test (t-tests) of each LS mean against 0. All reported p-values are Holm-corrected for multiple comparisons.*

Chart, box and whisker chart

Description automatically generated

*Figure X. Model-based Least Square means in function of belief and grip conditions. Error bars represent the standard error of the mean. Stars inside the bars represent the significance of each LS mean compared to baseline 0. Stars above the bars represent the significance levels of the difference between means.*

Furthermore, we found that in the whole hand grip condition (FB-WHG versus TB-WHG), looking times were significantly larger for the true belief condition as compared to the false belief condition, *t =* 11.73, *p* < .001. Similarly, DLTS was higher in the true belief versus the false belief condition in the precision grip condition (FB-PG versus TB-PG), *t =* 7.729, *p* < .001. No such difference was observed in the false belief condition, *t =* .412, *p* = .68.

Although we did not find a significant main effect of group, nor a significant interaction effect with grip or belief, nor a three-way interaction effect (see above), we still tested the contrasts outlined by our hypotheses. An overview can be found in Table X. However, none of these contrasts were significant.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Belief x grip* | *Contrast: group* | | *t* | *p* |
| FB-WHG | baseline – restricted | -1.692 | | .365 |
| TB-WHG | baseline – restricted | -.069 | | .945 |
| FB-PG | baseline – restricted | -1.063 | | .633 |
| TB-PG | baseline – restricted | -1.252 | | .633 |

*Table X. Contrast tests of model-based LS means between groups, per level of belief and grip, pre-shaping grip conditions only.*

A visual representation of the different levels of belief and grip in function of group can be found in Appendix X.

*5.3.1.2 Extended model*

We further expended the model by taking into account block number, as per previous experiments. This revealed, however, contrary to previous experiments, no significant main effect of block, *F*(4,1735) = .168, *p* = .954. Other effects and interaction effects remained unchanged. A visual representation can be found in Appendix X-X.

As per previous experiments, we also check for a main effect of target size. We found, however, no significant main effect of target size, *F*(1,1722) = .273, *p* = .601.

As neither the addition of block number nor target size into the model as fixed effects seems to explain any additional variance as compared to the basic model, χ2 (5) = .949, *p* = .967, we therefore adopt the more basic model, solely concerned with the main effects of group, belief and grip, plus their interaction effects and three-way interaction effect.

*5.3.2 Accuracy of explicit responses*

*5.3.2.1 Basic model*

We implemented a general linear mixed effects model using a binomial distribution with logit link as answers on the explicit task were binary (correct or false). Due to the complexity of a three-way interaction in this type of distribution, a model including a three-way interaction effect between group, belief and grip failed to converge. As such, only the main effect of group is included in the model, along the main effects and interaction effect of belief and grip.

Furthermore, as the *lme4* package no longer provides p-values based on F- or t-tests in the case of generalised linear mixed models and also no longer supports the Markov Chain Monte Carlo sampling method, we estimated the effect based on stepwise likelihood ratio tests (i.e., testing a model minus a fixed effect against the full model). An overview of the results can be found below (Table X). Main effects of belief and grip cannot be tested directly using this method due to involvement in interaction effects (see Chapter 4, Section 4.3.2) and are hence not represented in this table. The main effects of belief and grip are, however, included in the statistical model.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Fixed effect parameter* | *F* | χ2 | *df* | *p* |
| Group | .045 | .136 | 1 | .712 |
| Belief × grip | 49.335 | 100.727 | 2 | <.0001\*\*\* |

*Table X. Estimated effects of model parameters based on stepwise likelihood ratio tests.*

We conducted follow-up tests on the model-based least-square means within belief and grip conditions. This revealed that accuracy was significantly higher in the true belief condition [LSM = .714, SE = .022, 95% CI = [.669;.756]] as compared to the false belief condition [LSM = .404, SE = .025, 95% CI = [.357;.453]], across levels of group and grip, *z* = -11.905, *p* < .001 (see Figure X). Concerning the differences within levels of grip, we found that accuracy was significantly higher in the whole hand grip condition [LSM = .64, SE = .029, 95% CI = [.581;.695]] as compared to the no grip condition [LSM = .475, SE = .028, 95% CI = [.42;.531]], *z* = 4.989, *p* < . 001, and in the precision grip condition [LSM = .579, SE = .028, 95% CI = [.522;.633]] as compared to the no grip condition, *z* = -3.3, *p* = .003.

Chart, bar chart

Description automatically generated

*Figure X. The main effects of belief and grip on accuracy of explicit responses, model-based Least Mean Squares. Error bars represent the standard error of the mean. Stars above the bars represent the significance levels of the difference between means. TODO: significance*

Every level of combinations of belief and grip conditions was tested against chance level, in this case 50%. An overview of this analysis can be found in Table X below. Overall, while accuracy in true belief preshaped grip conditions (TB-PG and TB-WHG) was significantly above chance level (all *p’*s < .001), accuracy in the preshaped false belief conditions (FB-PG and FB-WHG) was significantly *lower* than baseline (all *p’*s < .001). In the conditions without preshaping, accuracy did not differ from chance level (all *p’*s > .05).

We also tested differences within levels of grip. For the whole hand grip condition, accuracy was significantly higher for the true belief condition (TB-WHG) versus the false belief condition (FB-WHG], *z* = 11.473, *p* < .001. Likewise, within the precision grip condition accuracy was significantly higher in the true belief condition (TB-PG) versus the false belief condition (FB-PG), *z* = 9.113, *p* < .001. No such difference was found for the no-grip condition (TB-NG versus FB-NG, *p* > .05). Figure X provides an overview of these results.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Belief - grip* | *LS Mean* | *SE* | *95% CI* | *z* | *p* |
| FB - WHG | .351 | .033 | .289 ; .418 | *-4.262* | *<.0001\*\*\** |
| FB - NG | .492 | .036 | .422 ; .562 | *-.231* | *.817* |
| FB - PG | .373 | .033 | .31 ; .44 | *-3.664* | *<.001\*\*\** |
| TB - WHG | .854 | .023 | .804 ; .893 | *9.804* | *<.0001\*\*\** |
| TB - NG | .458 | .036 | .39 ; .528 | *-1.174* | *.24* |
| TB - PG | .76 | .028 | .701 ; .811 | *7.455* | *<.0001\*\** |

*Table X. The model-based Least Square Means of every level of the interaction effect between belief and grip conditions, as well as their standard error and 95% confidence interval. Also included is the significance test (t-tests) of each LS mean against 0. All reported p-values are Holm-corrected for multiple comparisons.*

Chart, bar chart

Description automatically generated

*Figure X. The interaction effect between belief and grip in function of accuracy of explicit responses, model-based Least Mean Squares. Error bars represent the standard error of the mean. Stars above the bars represent the significance levels of the difference between means.*

To test our a priori hypothesis, we compared between levels of group within levels of belief and grip. A visual representation can be found in Appendix X-X. None of the contrasts reached significance.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| *Belief x grip* | *Contrast: group* | | *z* | *p* |
| FB-WHG | baseline – restricted | -.371 | | .999 |
| TB-WHG | baseline – restricted | -.371 | | .999 |
| FB-PG | baseline – restricted | -.371 | | .999 |
| TB-PG | baseline – restricted | 1.667 | | .382 |

*Table X. Contrast tests of model-based LS means between groups, per level of belief and grip, pre-shaping grip conditions only. P-values are Holm-corrected.*

*5.3.2.2 Extended model*

As with DLTS, we also investigated the effect of block number (learning effect) and target size (saliency effect). This revealed that for accuracy of explicit responses, there was a significant effect of block number, χ2 (4) = 13.791, p = .008. A detailed account can be found below (Table X, Figure X).

|  |  |  |
| --- | --- | --- |
| *Contrast* | *z* | *p* |
| 1-2 | -1.13 | .791 |
| 1-3 | -2.62 | .067 |
| 1-4 | -3.052 | .019\* |
| 1-5 | -2.722 | .051 |
| 2-3 | -1.481 | .575 |
| 2-4 | -1.9 | .317 |
| 2-5 | -1.625 | .481 |
| 3-4 | -.411 | .99 |
| 3-5 | -.186 | .99 |
| 4-5 | .213 | .99 |

Chart, bar chart

Description automatically generated

*Figure X. Main effect of block number, model-based LS means. Error bars represent standard error of the mean, stars represent significance of LS means difference. TODO: significance*

*Table X. Overview of post-hoc t-tests on LS means between blocks.*

We also tested the effect of target size. However, we found no main effect, χ2 (1) = 1.364, p = .243.

*5.3.3 Correlation analysis between DLTS and accuracy of explicit responses*

We were also interested in comparing looking time bias (as measured by DLTS) and the accuracy of explicit responses as to the intention of the agent to reach for either the target cup or the distractor cup. Specifically, we looked at the correlation patterns between DLTS and response accuracy. In order to do so, we had to aggregate a score DLTS and accuracy per participant, over levels of group, belief and grip. This resulted in 6 data points per participant for each measure.

First we looked at the correlation between DLTS and accuracy of explicit responses across all group, belief and grip conditions. Since at this level neither DLTS nor accuracy was normally distributed (as per Shapiro-Wilks normality tests) as well as bounded between respectively -1 and 1, and 0 and 1, we performed a Spearman’s rank correlation analysis. The results show a strong significant positive correlation between DLTS and accuracy, *rs*= .694, *p* < .001, 95% CI = [.619;.767]. Note that the confidence interval is based on a 1000 replications bootstrap resampling method.

*Chart, scatter chart

Description automatically generated*

*Figure X. Correlation between DLTS and accuracy of explicit responses. The dashed blue line represents a fitted linear model, while the red line represents a LOESS (non-parametric local regression) fit. Based on this graph, we can assume linearity or at least a monotonic relationship.*

Next, we looked at the correlation between DLTS and accuracy per group using a similar methodology. This correlation was significant in the baseline condition, *rs*= .672, *p* < .001, 95% CI = [.541;.774], as well as in the secondary task group, *rs*= .7, *p* < .001, 95% CI = [.582;.797].

Chart

Description automatically generated

*Figure X. Correlation between DLTS and accuracy of explicit responses, per group. The dashed blue line represents a fitted linear model, while the red line represents a LOESS (non-parametric local regression) fit.*

TODO: different titles

We investigated the correlation patterns for every belief-grip combination as well. The results of this analysis can be found in the table below, along with a visual representation of the relationship between DLTS and accuracy. Overall, DLTS was significantly correlated with the accuracy of explicit predictions, across all combinations of grip and belief conditions. While this correlation between our implicit and explicit measure is moderate to strong in the preshaped conditions (PG and WHG), the correlation is considered weak in the no-grip conditions.

|  |  |  |  |
| --- | --- | --- | --- |
| *Condition* | *rs* | *95% CI* | *p* |
| TB: Whole Hand Grip | .64 | .428;.794 | <.0001\*\*\* |
| TB: No Grip | .349 | .019;.624 | .015\* |
| TB: Precision Grip | .763 | .578;.865 | <.0001 \*\*\* |
| FB: Whole Hand Grip | .627 | .405;.779 | <.0001 \*\*\* |
| FB: No Grip | .438 | .155;.653 | .002\*\* |
| FB: Precision Grip | .639 | .415;.828 | <.0001 \*\*\* |

*Table X. Spearman’s rank correlation coefficient per level of belief and grip, including 95% confidence interval and p-values*

*Pieter – too much white space here – you will need to format more effectively the visual layout!.*

Chart, scatter chart

Description automatically generated

*Figure X. DLTS and accuracy of explicit responses plotted against each other, per belief and grip combination, for the baseline group only. Blue lines represent the linear model fit for each plot.*

*5.3.5 Accuracy of secondary tone task*

We also had a look at performance on the secondary tone detection task in the secondary task group. We found a global accuracy of 63.22%, and a mean accuracy aggregated per participant of 65.29%, standard deviation = 17.31%. Mean accuracy was significantly higher than chance level (50%), as per a Wilcoxon signed rank test, V = 343.5, *p* < .001.

We tested if performance on the tone detection task was correlated to performance on the explicit mentalising task. There was, however, no significant correlation between both variables, *rs* = .067*, p* = .733, 95% CI = [-.326;.427].

We also tested if performance on the tone detection task was correlated to Differential Looking Time Scores (DLTS). However, there was no linear or monotonic relationship between these variables either, *rs* = .177*, p* = .367, 95% CI = [-.208;.551].

We also tested a general linear mixed effect model with binomial distribution and logit link, but found that neither belief, χ2 (1) = .114, *p* = .736, grip, χ2 (2) = 3.368, *p* = .186, nor their interaction, χ2 (2) = 5.356, *p* = .069, had a significant effect on tone task accuracy. See Figure X for an overview of secondary task accuracy in function of belief and grip conditions.

Chart, bar chart

Description automatically generated

*Figure X. Accuracy on the tone detection task in the secondary task group in function of belief and grip conditions. Error bars represent the standard error of the mean.*

*5.3.6 Method of prediction*

*Optional: redo with no-grip conditions?*

During the debriefing session, we asked participants what motivated their decision on the explicit response task. Of the total of 48 participants, N = 20 indicated that they kept the beliefs of the agent into account, N = 16 decided purely based on the shape of the hand, N = 4 said they picked at random, N= 3 reported that they estimated based on perceived motions of the hand, N = 1 purely went for the large cup, and N = 4 had answers that were non-conclusive. For the sake of simplicity, we distinguished between mentalising (belief-based) methods and non-mentalising (other) methods only, omitting non-conclusive answers for further analysis. An overview of the distribution of participant’s methods can be found in the table below.

|  |  |  |
| --- | --- | --- |
| *Method* | *Baseline group* | *Mind-tied group* |
| Belief | 11 | 9 |
| Other | 9 | 15 |

*Table X. Two-way frequency table of self-reported method of prediction and experimental group condition.*

We were interested to see if experimental group (“baseline” versus “secondary task”) had an impact on what method for explicit prediction was implemented by the participants. However, a chi-squared test with Yates’ continuity correction revealed that there was no association between these two variables, χ2 (1) = .734, *p* = .392.

*5.3.6.1 Method of prediction and Differential Looking Time Scores (DLTS)*

We were interested in the relationship between method of responding to the explicit response task and looking behaviour, as measured by DLTS. To do so, we added “method of prediction” as a covariate into linear mixed effect model. Specifically, we investigated the main effects of method, belief and grip, plus their interaction effects and three-way interaction effect. We omitted no-grip trials from this analysis as they are of no relevance to these analyses.

Our conjecture was that in the true belief condition, the implemented strategy should not make a difference on predictive eye gaze as participants should arrive at the same conclusion, given that motor cues and mental state of the agent are congruent. As such, we expect to find DLTS above baseline in both method groups, and no difference in DLTS between method groups. In the false belief condition, however, we expect to find DLTS scores significantly above baseline in the mentalising group, as DLTS was defined as positive if gaze was congruent with a mentalising-based method of prediction of the outcome. However, if participants reasoned from a strictly motor-based perspective, they should score significantly below baseline. Recall that if the hand makes a precision grip, the motor-congruent target object is the small object. This is, however, incorrect as the cups have been swapped unbeknownst to the agent. In this case, the large object would be target of movement, as this is where the agent believes the small object is. Hence, a purely motor-based method of prediction will consistently result in a negative DLTS, while a mentalising method will yield a positive DLTS.

An ANOVA of our model revealed that while there was no main effect of method, *F*(1,44) = 3.135, *p* = .084, method did show a significant interaction with belief, *F*(1,1094) = 11.421, *p* < .001. In Table X and Figure X below, we report the least-square means for the method *x* belief interaction effect.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Method* | *Belief* | *LS Mean* | *SE* | *95% CI* | *t* | *p* |
| Belief | FB | -.026 | .014 | -.055 ; .002 | *-1.821* | *.072* |
| other | FB | -.089 | .013 | -.116 ; -.064 | *-6.985* | *<.0001\*\*\** |
| belief | TB | .074 | .014 | .046 ; .102 | *5.203* | *<.0001\*\*\** |
| other | TB | .08 | .013 | .054 ; .106 | *6.096* | *<.0001\*\*\** |

*Table X. The model-based Least Square Means of every level of the interaction effect between belief and method, as well as their standard error and 95% confidence interval. Also included is the significance test (t-tests) of each LS mean against 0. All reported p-values are Holm-corrected for multiple comparisons.*

To investigate the interaction effect between belief and grip, we also tested for the difference in DLTS between groups per belief condition. These contrasts indicated that there was no difference between mentalising and other strategies in the true belief condition, *t =* -.302, *p* = .764; however, in the false belief condition participants that used mentalising strategies had significantly higher DLTS as compared to participants using other strategies, *t*=3.293, *p* = .003. This is unsurprising, as a mentalising strategy would yield a positive DLTS in this analysis.

Chart, box and whisker chart

Description automatically generated

*Figure X. Model-based Least Square means in function of self-reported method of prediction and belief conditions. Error bars represent the standard error of the mean. Stars on the bars represent the significance of each LS mean compared to baseline 0. Stars above the bars represent the significance levels of the difference between means.*

*5.3.6.2 Method of prediction and accuracy of explicit responses*

Of course, one would expect participants that reported using a belief-based strategy, answering the question “on which side of the screen will the hand appear?”, to have a better task performance. Again, we expected that this effect would only be prevalent in the false belief condition, as here different strategies (belief-based, or motor-based and other) lead to different predictions.

To test this, we again added “method of prediction” as a predictor variable in a model that tests for main effects and interaction effects with belief, for the pre-shaping grip conditions only. A general linear mixed effects model with a binomial outcome variable and logit link was used. Specifically, this model looked at accuracy of explicit responses in function of method, belief, and the interaction effect between method and belief. To simplify the model and since we are not interested in scores in the no-grip condition, grip was excluded as a factor here.

We used a stepwise likelihood ratio test to test the interaction effect of method and belief, as main effects of factors in an interaction effect are not testable with this method. This revealed a significant interaction effect between method and belief, χ2 (1) = 18.283, *p* < .001. We took a further look at the interaction effect between method and belief. Again, we expected to find higher accuracy in the group that implemented a belief-based strategy. An overview can be found in Table X and Figure X below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Method* | *Belief* | *LS Mean* | *SE* | *95% CI* | *z* | *p* |
| Belief | FB | .542 | .055 | .434 ; .647 | *.764* | *.445* |
| other | FB | .216 | .036 | .154 ; .293 | *-6.137* | *<.0001\*\*\** |
| belief | TB | .83 | .034 | .752 ; .887 | *6.532* | *<.0001\*\*\** |
| other | TB | .805 | .034 | .73 ; .863 | *6.589* | *<.0001\*\*\** |

*Table X. The model-based Least Square Means of every level of the interaction effect between belief and method, as well as their standard error and 95% confidence interval. Also included is the significance test (z-tests) of each LS mean against 0. All reported p-values are Holm-corrected for multiple comparisons.*

We also had a look at the differences between groups within level of belief. This showed that accuracy was significantly higher in the mentalising group as compared to the group implementing other strategies, *z* = 4.779, *p* < .001, for the false belief condition. In the true belief condition, no such group difference was found, *z* = .518, *p* = .604.

*Chart, bar chart

Description automatically generated*

*Figure X. Bar plot representing model-based least-square means for accuracy on the explicit response task, per method of responding. Error bars represent the standard error of the mean. Stars on the bar represent significance level of the mean difference as compared to chance level (50%). Stars above the bars represent the significance levels of the difference between means. Note that chance level is represented by the dashed horizontal line.*

Pieter – remember to try to get rid of so much white space. Do not underestimate the importance of a clean visual presentation – make this thesis appear like a real book in formatting! And remember to start filling in the Figure numbers and Table numbers (and do the power analyses) – the little things can be super important!

***5.4 Discussion***

The purpose of this experiment was to investigate the impact of a secondary task load and subsequent diminished attentional resources on implicit and explicit mentalising within our experimental framework. In order to do so, we made a number of adjustments to the paradigm. Firsly, the final phase of the videos (the hand reaches out of the occluder and grasps a cup) were cut off and replaced by a prompt asking which side of the box the hand would appear. Secondly, participants in the secondary task group were required to perform an auditory discrimination task while the hand was preshaping. As this study included multiple different facets, we will discuss the results per topic below.

*5.4.1 Secondary task load*

As expected, placing the participants under secondary task load seemed to have no impact on implicit measures. That is, anticipatory looking, which we consider as proxy for the outcome of motor resonance processes in the true belief condition, and of combined motor resonance and implicit belief-tracking processes in the false belief condition, was unaffected by additional task load. Contrary to our hypotheses, however, added secondary task load had no impact on participants’ explicit predictions of the agent’s reaching movement either. Furthermore, performance on the secondary tone detection task was not correlated with either implicit looking bias or accuracy of explicit responses, suggesting that there was no trade-off between allocation of attentional resources towards either the secondary tone discrimination task or behavioural prediction making, be it implicit or explicit.

In first instance, it would be tempting to attribute this null result on the explicit prediction measure to task demands of the tone discrimination task. That is, one could argue that the task was too easy and did not place sufficient demands on cognitive resources. Or, conversely, one could state that the tone task was too demanding, causing participants to guess at random. The data do not support either interpretation, however, as accuracy was both significantly higher than chance level while at 63.22% being far from ceiling level. It is interesting to note that this overall accuracy is lower than the overall accuracy of 78.48% reported by Michael and colleagues (2017). This could potentially be due to higher perceptual demands of our video stimuli of the primary task, in comparison to the static images used in the perspective taking task. As such, it could be that the primary task took precedence over the secondary task, reducing accuracy on this tone task.

An alternative explanation for the lack of effect of the secondary task could be the timing of the tone detection task. A similar logic here could apply for the null effect of the secondary tone task as for the absence of effects of motor restriction in Experiment 2. That is, the tone presentation phase lasted for 326 milliseconds in a time-of-interest window that lasted on average 4.26 seconds. One could state that this short task interval did not impact attentional processes during the entire measurement window sufficiently to produce any effects. That is, our measures are not sensitive enough to pick up conflicting processes.

For comparison, we can look back at the two studies mentioned in the introduction of this chapter. Both Lin, Keysar & Epley (2010) and Cane, Ferguson & Apperly (2017) used a director task to measure explicit mentalising capabilities. In this task, participants interact with a human confederate or computerised “director” to move different objects within a grid of boxes. Crucially, the participant and director are on opposite sides of the boxes and hence have different perspectives. The content of some of the boxes is occluded from the view of the director by opaque walls. Furthermore, objects come in sets, where objects are similar but differ in a single feature, e.g., “the toy car with blue stripes” and “the toy car with the red dot”. The director then references objects and asks the participant to move these around the grid. Successful performance on this task requires taking into account the perspective of the director (which differs due to the occluders) and the ambiguity due to object similarity. The measures used in these studies were the timing and accuracy of participant responses, assumed to be function of explicit mentalising processes. Lin and colleagues (2010) also included “number of fixations to a distractor” as dependent measure. During this task, participants performed a secondary working memory task consisting of memorising digits.

*5.4.2 Mirroring and mentalising effects*

The results yielded a strong and convincing preshaping effect. That is, implicit gaze behaviour was significantly biased towards the grip-congruent cup in the true belief condition. Moreover, this effect was clearly present in the explicit predictions of the participants as well. This is a clear indication that participants capitalised on motor cues to predict future behaviour of the agent, and that this effect permeates into both indirect and direct measures of prediction-making. This result is especially striking in this experiment, as there was never an outcome phase. That is, participants never saw the hand actually grasping any of the cups. These results solidify the case that motor cues tend to strongly drive implicit and explicit prediction making in action observation (Donnarumma et al., 2017). Moreover, as this effect was not affected by secondary task load, the current results bolster the case for the cognitive efficiency of observation-induced activation of motor representations (see Cracco et al., 2018).

In the false belief condition, an entirely different pattern of results emerged when compared to previous studies. That is, we found that participants looked significantly more often and longer at the distractor (motor-congruent) cup and not to the belief-congruent cup. Interestingly, this effect was found in explicit responses as well, with participants considering it more likely that the agent would reach for the motor-congruent cup, despite having a false belief on the cup’s location. This runs directly counter to the results of our previous experiments: whereas before we found evidence for either belief-congruent looking biases, or at least an absence of a clear motor-based bias, we find that participants here, on average at least, seemed to discount the agent’s mental state when making implicit and explicit predictions.

Importantly, this effect occurred in both groups, and was hence unrelated to the secondary task load. Indeed, no effect was found between experimental group and self-reported method. As such, our experiment is perhaps not informative the effects of a secondary task load on implicit or explicit mentalising processes, as on average, our results do not suggest any mentalising processes to have been taking place at all. We must therefore look at other changes in our experimental paradigm that could potentially account for these vastly differing result patterns. It could be that the explicit response question prompt itself changed the way participants acted. Whereas in previous experiments event sequences in the video were completely irrelevant to the task at hand (that is, reacting the colour change of the cups and the fixation arrow), in this experiment we directly asked participants to make a prediction on how the rest of the video would hypothetically unfold. The prompting question was phrased in such a way that it made no reference to the mental state of the agent, but the intended effect was to have the participant engage in overt prediction making. Since the videos being cut at the end, there was never an observable relationship between events in the video and the hypothetical reaching actions of the agent. Hence the participant had to settle on a response strategy that would be best to answer the overt question. The lack of feedback may have discouraged participants from using the more tentative strategy of basing predictions on mentalising processes, instead favouring the clear motor cue of grip to drive predictions, as motor cues tend to exert a strong effect, which can even override opposing cues from different sources (Ambrosini, Pezzzulo & Constantini, 2015; See Section 5.4.3). This latter possibility is corroborated by the absence of a learning effect on looking times, in contrast to the findings of our previous experiments. Note that on the explicit mentalising measure, however, we did find a learning effect, possibly reflecting that participants only settled on a given strategy after a few blocks. We will discuss the notion of response strategies further in Section 5.4.3.

Note that in this experiment, we found no effect on target size on both looking times and explicit answers, indicating that in the no-grip condition participants did not use object saliency as a cue for either looking behaviour or making explicit predictions on outcome. This can tentatively be explained by the absence of the outcome phase: whereas in previous experiments basing your predictions on target size would yield a 50% accuracy rate, rendering this strategy somewhat effective, no such link could -implicitly or otherwise- be established in this experiment. As such, participants may have picked purely at random in this particular grip condition.

Interestingly, this study points out that our implicit measure (differential looking times score) is strongly correlated with overt, explicit behavioural predictions. Furthermore, this correlation was equally strong in both groups. Unsurprisingly, measures were more strongly correlated in the preshape conditions, and was weaker in when there was no information from motor cues. Notably, the correlation between explicit and implicit measures was still significant in the no-grip condition, pointing out that even here implicit sampling of information was to extent in line with overt prediction making. We will discuss this topic further in Chapter 7.

*5.4.3 Individual differences in response strategy*

Breaking away from the previous experiments reported in this thesis, we used self-reported prediction strategy as a covariate rather than excluding explicitly mentalising participants. That is, we asked participants directly what strategy they used to predict what cup the hand would reach for, and then incorporated this into a statistical model of the results. Interestingly, the mentalising strategy was the most common in this study (42% of participants). The second largest group (33.33% of participants) said they based their responses purely on hand preshape. To simplify modelling, we compared mentalising participants with other strategies (mainly comprised of participants using a motor-based strategy). Analyses revealed that while the implemented strategy did not have a direct influence on looking times or explicit predictions, it did interact with belief condition. Specifically, we found that while both strategies showed a clear preshaping effect in the true belief condition, these groups differed vastly in the false belief condition. While mentalising participants had an accuracy at baseline for both implicit looking times and explicit accuracy, participants implementing a different strategy scored significantly below baseline on both measures. That is, their responses were consistently in line with a motor-based strategy. This should not be surprising as the most common self-reported strategy in this group was basing explicit responses on hand preshape.

Moreover, the difference between mentalising and mirroring participants in the false belief condition was significant, on both implicit and explicit measures. Since we had no a priori hypotheses on the effect of grip and this was not a planned comparison, we did not pursue this analysis further, but it points out a number of important considerations. Firstly, it is quite likely that the globally motor-congruent results on both measures are largely driven by the large group of participants that based their predictions purely on hand preshape. This begs the question why such strategy was prevalent. Secondly, while the largest group reported to implement a mentalising strategy, this group only performed at chance level on both implicit and explicit measures of behavioural prediction making. This could signal that participants were not very confident in this approach. However, the results of this group, both in terms of looking times and overt predictions, are in line with the looking times of participants in Experiment 2. As we discussed there, the absence of a clear motor-based effect in the false belief condition lends some degree of credibility to the notion that implicit belief-based inferences at least modulated prediction making.

This relative prevalence of a pure motor-based strategy as self-reported by participants can potentially be explained through the results of Ambrosini, Pezzulo & Constantini (2015). In this study the researchers used similar stimuli as in Ambrosini et al. (2011) discussed in Chapter 2, measuring proactive eye movement in combination with videos in which a hand reached for either a large or small tomato, preshaping the hand to match the size of the tomato. Crucially, in this experiment, the hand preshaping could be misleading and only matched the size of the tomato in 50% of trials. Moreover, the experimenters included the face of the actor, whose eyes would be directed to one of the tomatoes prior to reaching. Again, actor gaze could be misleading and only matched the reaching target in 50% of trials. In this way, both gaze and preshape could be congruent, both gaze and preshape could be incongruent, gaze could be congruent while preshape was incongruent, and gaze could be incongruent while preshape was congruent. The results showed that while actor gaze influenced anticipatory looking, the motor information in hand preshaping exerted a stronger effect. Moreover, while participants relied less on actor information over time, reflecting a learning effect, no such learning effect was found for preshaping, despite both actor gaze and preshape being equally unreliable. This indicates that participants automatically take advantage of motor cues for behavioural prediction, even overriding other potential sources of information. In analogy, it could be that in our experiment, the absence of feedback resulted in hesitancy over the reliability of predictive strategies, prompting some participants to rely solely on preshaping information as a more salient, automatically processed source of information.

The overall explanation would then be that in previous experiments (Experiments 1 and 2, see Chapters 4 and 5) the feedback provided at the end of the videos could implicitly reinforce a mentalising strategy based upon statistical learning (e.g., Ruffman, 2014; the impact of statistical learning will be further discussed in Chapter 7). This potentially increased the incidence of an implicit mentalising strategy in Experiments 1 and 2 relative to Experiment 3. Although speculative, this account is in line with the learning effects found in Experiments 1 and 2, but not Experiment 3. In Experiment 3, however, due to the absence of any sort of feedback, participants would be unable to explicitly tune their strategy to relations between events in the video and the outcome of the video. A relatively larger number of participants could therefore have defaulted to a motor-based prediction strategy due to the strong cueing effect exerted by automatically processed motor information. Participants that did use a mentalising strategy could have done so only hesitantly, as there was no clear indication that their strategy yielded correct predictions, resulting in no clear bias towards either the motor-congruent or belief-congruent cup in the false belief condition. However, at present these remain open questions for future research.

***5.5 Conclusion***

In this experiment we implemented secondary task load and added overt prediction-making as response variable. Results deviated quite significantly from previous experiments. Generally, the responses were strongly in line with hand preshaping but were generally incongruent with the agent’s belief, indicating that participants did not track the mental states of the agent on neither implicit nor explicit measures. This effect was not modulated by secondary task load, which runs counter to previous studies on both implicit and explicit forms of mentalising. Arguably the most crucial finding is that participants tended to gravitate towards a specific response strategy, most primarily basing their responses on either mentalising or motor cues. The absence of belief-based modulation of implicit and explicit mentalising seemed largely driven by the large group of participants using only motor cues. Mentalising participants, however, scored largely around baseline level, in line with Experiment 2. A number of potential explanations are suggested, but these remain questions for future studies. Finally, we found a strong positive correlation between anticipatory looking scores, which we took as implicit measure of behavioural prediction, and explicit behavioural predictions. This suggests that in at least in the paradigm outlined in present thesis, anticipatory looking times seem to reflect ongoing predictive processes.