**When Your Beliefs Guide My Hands:**

**Visibly Constraining an Agent Modulates Observers’ Automatic Belief Tracking**

**Jason Low1\*, Katheryn Edwards1, Stephen A. Butterfill2**

1School of Psychology, Victoria University of Wellington, New Zealand

2Department of Philosophy, University of Warwick, United Kingdom

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\* Corresponding author at: School of Psychology, Victoria University of Wellington, PO Box 600, Wellington, New Zealand 6140.

Email addresses: [Jason.Low@vuw.ac.nz](mailto:Jason.Low@vuw.ac.nz) (J. Low), [katheryn@lolux.net](mailto:katheryn@lolux.net) (K. Edwards), [s.butterfill@warwick.ac.uk](mailto:s.butterfill@warwick.ac.uk) (S. Butterfill)

**1. Introduction**

Tracking what is likely to happen in dynamic social situations is inherently challenging, and— because observing others activates our own motor systems—one view is that motor representations facilitate action and social understanding. Converging evidence revealed through brain imaging, transcranial magnetic stimulation (TMS), and reaction times suggests that observation of an agent’s action leads to activation of a corresponding motor representation in the observer (as reviewed by Rizzolatti & Sinigaglia, 2016). Neuroimaging studies support that the observation of actions done by others triggers activity in the motor execution-related brain areas that includes the ventral premotor and inferior parietal cortices (Cattaneo & Rizzolatti, 2009; Rizzolatti, Fogassi & Gallese, 2009). Consistent with imaging data, studies applying magnetic stimulation to the motor cortex during observation of reach-to-grasp actions reveal increased excitability in those parts of the motor cortex that correspond to the observed movement. For example, when adults watch another individual grasp an object, motor evoked potentials triggered by the stimulation are elevated in the very same muscles that would be active if the observers were making that grasping action (Fadiga, Fogassi, Pavesi, & Rizzolatti, 1995; Urgesi, Moro, Candidi, & Aglioti, 2006). Most importantly for the present study, behavioural studies profiling reaction-time benefits on motor priming tasks indicate that movement observation (a task-irrelevant stimulus showing an agent lifting her index finger, for instance) can accelerate our execution of the same action and decelerate our execution of an incompatible action (Brass, Bekkering, Wohlschlager, & Prinz, 2000; Cracco, Bardi, Desmet, Genschow, Rigoni, De Coster, et al., 2018).

Tracking others’ behaviour involves being sensitive not just to an action itself but also to the context of an action. There is evidence suggesting that our motor system can code the context in which specific actions are embedded. Iacoboni, Molnar-Szakacs, Gallese, Buccino, and Massiota (2005) monitored adults’ brain activations as participants watched three types of movies: clips of different background contexts (cup, teapot and plate of food arranged as if someone was about to have breakfast, or had finished); clips of a hand physically executing different grips (precision or whole-hand) to grasp a cup without context; and clips of a hand physically executing different grips to grasp a cup in the different contexts. The imaging data showed that the condition where participants witnessed grasping actions in embedded contexts, compared to the other conditions, elicited higher cortical motor activations. The results suggest motor representations can code more complex goal outcomes (grasp for drinking, grasp for putting away) of which the immediate action is a part.

Motor representations can carry information about the features of a situation even in the absence of any movement towards a target object. Costantini, Ambrosini, Tieri, Sinigaglia and Committeri (2010) required participants to execute a grip with either their left or right hand upon the presentation of a task-irrelevant go-signal (a virtual mug with its handle aligned to the left or right). The researchers found that elicitation of a spatial alignment effect—faster responding when the required hand to execute the motor act was compatible with the orientation of the mug handle—depended on the apparent possibility for participants to interact with the virtual object. The spatial alignment effect was elicited when the virtual mug appeared in near reachable space where it could be apparently acted upon by participants. In a follow-up study, Costantini, Committeri and Sinigaglia (2011) found that a bystander’s space constraints for potential action could also modulate participants’ own motor behaviour; the spatial alignment effect occurred also when the mug was presented outside the reaching space of the participants but within the reaching space of a computer avatar. The findings suggest that motor activations can be deployed even without watching movements being performed, such that anticipating someone else’s potential action can excite planning and control of actions in the observer to influence task performance.

Costantini and colleagues’ (2011) findings that graspable objects potentially ready to an agent’s hand prime the execution of participants’ own hand movements might suggest that some motor programmes can be relatively effector specific. That said, Costantini and colleagues remind readers that there can be different levels of motor coding during action observation (see also Grafton & Hamilton, 2007; Rizzolatti & Sinigaglia, 2010). Whilst some motor representations can map the low-level parameters that compose action (patterns of bodily configurations, joint displacements or muscle activations; Fadiga, Craighero, & Oliver, 2005), others can selectively carry information relating to the goal of an agent’s action (i.e., the anticipated or observed outcomes to which another’s purposive behaviour may be directed). There are TMS data indicating that motor representations code the goal of an agent’s tool-mediated behaviour and not necessarily the muscular pattern required to perform the movement, and that virtual lesions to the ventral premotor cortex impair judgments about the outcomes that bodily actions are directed to but preserves judgements about which body parts are being observed (Cattaneo, Caruana, Jezzini, & Rizzolatti, 2009). Likewise, when we witness an agent perform an action goal using a *non*-typical effector (kicking a ball with her finger or grasping a pencil with her foot, for instance), there is motor facilitation in the effector muscles that we typically use to achieve that action goal (i.e., the leg or the hand, respectively) (Betti, Castiello, & Sartori, 2015; Senna, Bolognini, & Maravita, 2014). Similarly, Triberti, Repetto, Costantini, Riva and Sinigaglia (2016) found that adults, after being trained to use their feet to grasp a ball, started to respond more as if they would typically use their feet to operate with other objects such as hammers and cups, which typically involve hand movements. Overall, the context of an action may be coded at multiple levels, whereby some motor representations that support social understanding may be relatively abstract for tracking the goal—and perhaps even the belief-informed goal—that an agent’s potential or overt action is directed to (Rizzolatti & Sinigaglia, 2016; Rizzolatti, Fogassi & Gallese, 2001).

If tracking the outcome-directedness of potential actions is an extension of motor processing, then constraining an agent’s opportunity to act should significantly disrupt motor representations concerning another’s realm of agency. Liepelt, Ullsperger, Obst, Spengler, von Cramon and Brass (2009) instructed participants to lift their index or middle finger in response to a number stimulus presented between the index and middle finger of a photograph of an agent’s static hand. There was a marked slowing of participants’ reaction time in the condition where the observed agent’s corresponding index and middle fingers were tied to the table with metal clamps as compared to the conditions where the agent’s fingers were unrestrained or where the agent’s non-corresponding fingers (thumb and ring finger) were restrained. The results highlight that the physical body of an agent is constitutively relevant to certain kinds of cognition. The fact that another person’s restraint leads to a finger-specific slowing of reaction time in the observer, even when the restraint was response irrelevant, suggests that our motor system automatically takes on the task-relevant circumstances governing someone else’s potentially upcoming goal-directed action. Dovetailing with Liepelt et al.’s findings, Costantini et al. (2011) found that priming of participants’ own motor system to react more quickly (as demonstrated with the spatial alignment effect paradigm) was obliterated when the observed agent’s apparent possibility to interact with an object was temporarily restricted (a transparent panel was placed between the computer avatar and the handled mug).

The available evidence that motor representations enable us to track the goals of others’ actions, although broad, does not entail that motor representations underpin the primary ways in which human beings engage in social cognition to play a role in the tracking of beliefs which specify someone’s reasons for action (Rizzolatti & Sinigaglia, 2010, p. 271). Going beyond this evidence, our hypothesis is that information about beliefs can feed into motor predictions concerning an agent’s potential actions. Our research experiment therefore aims to test whether, and to what extent, visibly constraining an agent from potentially interacting with an object modulates observers’ abilities to automatically track that agent’s belief. Our experiment bears on two broader issues.

The first issue concerns how observers could ever distinguish between motor representations that are self-triggered and other-triggered. This issue arises given that witnessing someone else’s action leads to a direct activation of corresponding motor processes in the observer. The challenge can be acute. For example, patients with frontal lobe injury overly imitate someone else’s interfering finger movements that are not response relevant, and patients with Parkinson’s disease find it difficult to control automatic imitation of others’ postural instability as induced by motor contagion (Brass, Derfuss, Matthes-von Cramon, & von Cramon, 2003; Pelosin, Bisio, Pozzo, Lagravinese, Crisafulli, Marchese, et al., 2018). Neurotypical individuals, however, can enforce their own internally planned movement against an externally perceived movement. Bardi and Brass (2016) suggest that a deeper understanding of control of concurrently activated goal-directed motor representations may be gained by investigating its embedding in theory-of-mind (ToM) abilities. Consistent with this suggestion, studies show that there are overlapping brain activations when adults perform mentalising and motor-priming tasks (Spengler, von Cramon, & Brass, 2009). Moreover, training on a motor task that makes salient differences between the one’s own motor action and someone else’s motor action can transfer to improvements in spontaneously tracking another’s visual perspective on a ToM task (Santiesteban et al., 2012). This raises the question of whether impairing abilities to represent actions motorically might impair performance on a ToM task, which our experiment is designed to answer.

Second, in many social situations, successfully tracking an agent’s motor goal that specifies a target object also involves tracking the correctness of the agent’s belief about that object. Consider, for example, that Maxine is going to perform an action the goal of which is to grasp her ball. She falsely believes that her ball is in the right-side box, whereas actually it is inside the left-side box. If we ignore the Maxine’s false belief, then fixing only upon grasping the ball as the goal of her action would generate the wrong expectation of how it would unfold – incorrectly predicting that she would reach to the left-side box to grasp her ball (Butterfill & Apperly, 2016). We need to track Maxine's belief in order to correctly identify the potential motor outcome of her action, which is that she would reach to the empty right-side box to grasp her ball. This creates a challenge for action observation: if we are to track how an agent’s action will potentially unfold, we cannot always rely on how things are but must also take her belief or belief-like state into account. Could the motor representations critical for tracking how the agent’s actions will potentially unfold also play a role in tracking her beliefs? No study, until now, has demonstrated that motor representations also matter for successfully tracking of an agent’s belief.

Bardi and Brass (2016) conjectured that a functional connection between control of motor representations and belief tracking may be studied “in implicit ToM tasks where participants’ performance depends on concurrent activations of different representations of the environment” (p. 162). We explored Bardi and Brass’ conjecture by taking advantage of Kovacs and colleagues’ (2010) object-detection task, which yields a critical effect showing that adults’ rapid and automatic tracking of a bystander agent’s belief has an impact on observers’ own actions. The object-detection task involves adults observing a ball rolling behind a wall on a table and then rolling from behind the wall to off the table. The video sequence also includes an agent who was present for some of the ball’s movements but not others and, consequently, could have a true or a false belief about the ball’s location. In the outcome phase, the agent returns, the wall is lowered, and participants must use their hand to press a button as fast as possible if the ball is present behind the wall. The task is considered to target automatic belief tracking because no reference is made to the agent’s belief about the ball’s location, and the agent does not perform any overt action when the wall is lowered. The critical finding, which has been well-documented, is that, compared to a baseline situation in which neither the participant nor agent expected the ball to be present (P-A-), participants are faster to respond when only the agent expected the ball to be present (P-A+), implying that the agent’s belief regarding the ball’s location is automatically encoded (Bardi, Six, & Brass, 2017; Bardi, Desmet, & Brass, 2018; Deschrijver, Bardi, Wiersema, & Brass, 2016; Edwards & Low, 2019; El Kaddouri, Bardi, De Bremaeker, Brass, & Wiersema, 2019; Kovács, Teglas, & Endress, 2010; Nijhof, Brass, Bardi, & Wiersema, 2016).

**1.1. Current study**

The primary aim of our study was to determine whether and to what extent visibly constraining an agent from potentially interacting with an object modulates observers’ abilities to automatically track that agent’s belief. We measured elicitation of susceptibility to an agent’s belief (as indicated by the P−A+ < P−A− effect) using the ball-detection task. We adapted this task to create three distinct version which differed in the constraints upon the agent’s potential movement to act on the ball in the outcome phase (see Figure 1). In the Free-Agent version of the ball-detection task, the agent returned at the outcome phase wearing the same clothing that he wore during the first phase of the video; the agent was visibly free to move and, potentially, to act on the ball. In the Constrained-Agent version of the ball-detection task, the agent returned at the outcome phase with his upper body and limbs bound up tightly in a white linen sheet (“mummified”); the agent was visibly constrained from using his upper body and limbs to potentially act on the ball. Extrapolating from research evidence showing that physical constraints on others’ action possibilities are mapped onto our own representations of the environment (Costantini et al., 2011; Liepelt et al., 2009), one possibility may be that the effect of the agent’s belief speeding up participants’ own action performances is a consequence of the way that information about beliefs feeds into motor predictions concerning the agent’s potential actions, and the way that those motor predictions would then facilitate response times. With respect to our primary research aim, we predicted that the P−A+ < P−A− effect would be elicited in the free-agent context but obstructed in the constrained-agent context.

A screenshot of a cell phone

Description automatically generated

***Figure 1.*** *Schematic showing logical structure of the study. There were three versions of the ball-detection task, presenting distinct constraints on the agent’s ability to move and, potentially, to act upon the ball when he returned to the scene (between-subjects manipulation). Participants experienced the Free-Agent version (no sheet), or the Constrained-Agent version (sheet restricted potential upper body movement) or the Loose-Sheet version (sheet allowed potential upper body movement). Each group of participants completed two forms of the ball-detection task (within-subjects manipulation): in the Hand-Placement form of the task, the agent placed the ball on the table with his hand; in the Foot-Placement form of the task, the agent placed the ball on the table with his foot.*

One might argue that the obstruction of the critical effect in the Constrained-Agent version as compared to the effect’s elicitation in the Free-Agent version stems from the perceptual novelty of the agent’s (“mummified”) appearance in the former version’s outcome phase. To rule out this low-level possibility, we also studied elicitation of automatic belief tracking in a third version of the ball-detection task. In the Loose-Sheet version, the agent returned at the outcome phase with a white linen sheet (same yardage as in the Constrained-Agent version) placed loosely (like a tunic) over the clothing he wore during the first phase of the video (see Figure 1). Consequently, in the Loose-Sheet version, the agent also returned with a novel appearance, but he remained visibly able to move his upper body and limbs and, potentially, to act upon the ball. If perceptual novelty modulates belief tracking, then the P−A+ < P−A− effect should be obstructed in the Loose-Sheet and the Constrained-Agent versions. But if motor representations of action context modulate belief tracking, as per our primary prediction, then the P−A+ < P−A− effect should be elicited in the Free-Agent and Loose-Sheet versions but obstructed in the Agent-Constrained version. The contrast between our predictions over the elicitation of the P−A+ < P−A− effect (in the Free-Agent, Constrained-Agent, and Loose-Sheet versions of the ball-detection task) with the predictions based on perceptual novelty is summarized in Table 1.

***Table 1.*** *Contrasting predictions about the ball-detection task where visibly constraining an agent from potentially interacting with an object modulates observers’ belief tracking.*

|  |  |  |  |
| --- | --- | --- | --- |
| Is the P−A+ < P−A− effect predicted in this version of the ball-detection task? | | | |
|  | Free-Agent version | Constrained-Agent version | Loose-Sheet version |
| Our prediction | Yes | No | Yes |
| Perceptual novelty prediction | Yes | No | No |

An exploratory aspect of our study was to learn about whether the P−A+ < P−A− effect might depend on there being a match between the effector used by the agent to act and the effector used by participants to respond. Considering the hierarchical structure of motor representations wherein information can be instantiated from effector-specific to effector-general levels (Rizzolatti & Sinigaglia, 2010; Sartori et al., 2015), any indication on whether the P−A+ < P−A− effect depends on matching agent’s and participants’ effectors would bear on whether, if motor representations of action do indeed drive the P−A+ < P−A− effect, those motor representations specify which effector the agent uses to act. Would the effect of belief tracking facilitating participants’ own representations of the environment be elicited only when the agent used his hand to grasp and position the ball? Or might the P−A+ < P−A− effect also be elicited if the agent were to use his foot to interact with the ball? Each of the three versions of the object-detection task (Free-Agent, Constrained-Agent, and Loose-Sheet versions; between-subjects manipulation) came in two forms. The Hand-Placement form of the task began with the agent grasping and placing a ball on the table using his right hand; the Foot-Placement form of the task begin with the agent grasping and placing the ball on the table using his right foot (within-subjects manipulation and counterbalanced). If motor representations modulating belief tracking are relatively effector-specific, then the P−A+ < P−A− effect should be elicited only, or more strongly, in the Free-Agent and Loose-Sheet versions where the agent had used his hand (rather than foot) to place the ball on the table. However, if motor representations specify goals but are neutral on the effector used by the agent, then the P−A+ < P−A− effect should be elicited in the Free-Agent and Loose-Sheet versions and elicited regardless of whether the agent had used his hand or his foot to act upon the ball.

**2. Method**

**2.1. Participants**

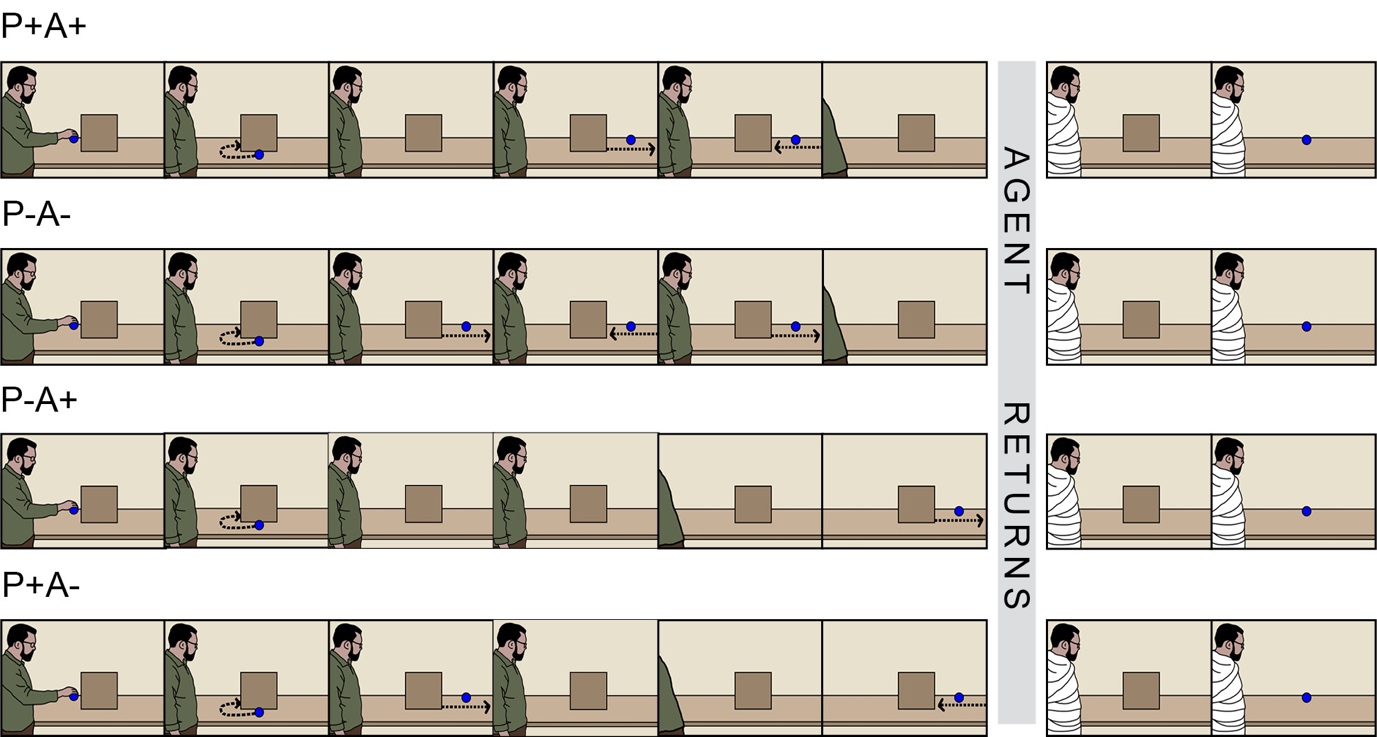
A total of 144 adult participants, made available by the Victoria University of Wellington’s Introduction to Psychology Research Programme, signed up to take part in the study. Of these, 48 participants were randomly allocated to each of three versions of the object-detection task. An a priori analysis using G\*Power (Faul, Erdfelder, Lang, & Buchner, 2007) (input parameters: α = .05, power =.8) determined that a sample size of at least 33 participants was required to detect the standardised effect size in each version. While not a direct replication of Kovács et al. (2010), our sample size was based on the standardised effect size of their critical effect (*r* =.45). This was calculated using the formula, *r2* = *t2*/ (*t2* + *df*), where *t* = reported t-test statistic of Kovács et al.’s critical effect = 2.42, and *df* = 23. Having a larger number of individuals safeguarded against participant dropout, and other factors affecting data collection such as experimenter error or computer malfunction. All participants signed informed consent forms prior to participation and were debriefed orally at the end of the session. Of the 48 participants originally allocated to the Free-Agent version, one participant was removed due to a technical fault and one participant did not perform above a 75% accuracy threshold. As a result, analysis was undertaken on the data of 46 participants (mean age, 18.76 years; range, 17-22; male to female ratio, 10/36). Of the 48 participants allotted to the Constrained-Agent version, two participants did not perform above a 75% accuracy threshold meaning that analysis was undertaken on the data of 46 participants (mean age, 18.98 years; range, 18-29; male to female ratio, 13/33). Of the 48 participants allotted to the Loose-Sheet version, three participants did not perform above a 75% accuracy threshold so that analysis was undertaken on the data of 45 participants (mean age, 19.29 years; range, 17-44; male to female ratio, 10/35). In total, analysis was undertaken on the data of 137 participants. The ratio of females to males was 104/33 and the mean age was 19.01 years (range 17 to 44). The study was approved by Victoria University of Wellington’s Human Ethics Committee.

**2.2. Materials**

All stimuli and instructions to the participant were presented via E-Prime 2.0. In each of the three experiments individuals watched a total of 80 videos in an object-detection paradigm. The on-screen video dimensions were 38cm x 21cm; all videos had a frame rate of 25 frames per second (fps), a 720 x 576 resolution, and were 21 seconds in length. Participants experienced two forms of their respective version of the object-detection task (adapted from Edwards & Low, 2019); these forms differed according to whether the agent placed an object on a table with his hand or his foot (order counterbalanced). There were 40 videos involving hand placements and 40 videos involving foot placements. Given that the primary aim of our study was to discover whether and to what extent visibly constraining an agent from potentially interacting with an object modulates observers’ abilities to automatically track that agent’s belief, we shall illustrate events in the videos of the object-detection task by way of the Constrained-Agent version.

**2.2.1. Constrained-Agent version**

For the Constrained-Agent version involving hand placements, each video began with an agent (a real human actor) standing in front of a table with a wooden wall. The agent held a ball in his outstretched hand. The agent placed the ball in front of the wall and a series of events ensued in which the ball moved around the table in a self-propelled manner. In the ball’s first movement, it moved behind the wall so that it could not be seen by either the participant or agent. Following this movement, the events in the videos varied to create four belief-induction conditions (see Figure 2).



***Figure 2.*** *Schematic showing main events of the four belief-inducing conditions in the Constrained-Agent version (ball present trials). In the P+A+ condition, both participant and agent expect the ball to be present; in the P*−*A*− *condition neither participant nor agent expect the ball to be present. In the P+A*− *condition only the participant expects the ball to be present, and in the P*−*A+ condition only the agent expects the ball to be present. The events for the Loose-Sheet and Free-Agent versions were identical except that in the former the agent returned with a sheet loosely draped over his upper body, and in the latter the agent returns with no sheet (appearance unchanged).*

In the belief induction phase, expectations were induced by manipulating the movement of the ball and by varying the time that the agent left the scene. The belief-induction conditions differed according to whether the participant expected a ball to be present (P+) or absent (P−) in the outcome phase and, further, whether the agent expected a ball to be present (A+) or absent (A−) in the outcome phase. For clarity and brevity, we outlined here the four belief-inducing conditions (P+A+, P−A−, P+A−, P−A+) when paired with a ball present outcome. Let us first consider the P+A+ and P−A− conditions which induced consistent expectations between the participant and agent. The events in the P+A+ condition led both the participant and the agent to believe there was a ball behind the wall; both saw the ball move behind the wall in its final movement. Likewise, in the P−A− condition, both participant and agent witnessed the ball ultimately exit the scene, so that neither were led to believe that there was a ball behind the wall. The P+A− and P−A+ conditions induced inconsistent beliefs. In the P+A− condition, the participant and agent saw the ball leave the scene. However, the agent was absent when the ball returned to rest behind the wall. In this case, the participant, but not the agent, was led to believe the ball was behind the wall. Finally, in the P−A+ condition, the agent was present when the ball left the scene but did not witness the ball’s return. Again, the agent’s and participant’s expectations were inconsistent as only agent had reason to believe that the ball was behind the wall.

In the outcome phase, the agent returned to the same location as at the start of the video, but the agent was tightly bound in a white linen sheet (i.e., “mummified”) which restricted his upper body and hand movements. His presence signalled the final event of the video which depicted one of two outcomes: 1) when the wall rapidly fell away, the ball was present (ball-present outcome); or 2) when the wall rapidly fell away, the ball was not present (ball-absent outcome). As such, participants experienced 8 trial types, comprised of four belief-induction conditions paired with one of two possible outcomes.

The videos showing foot placements by the agent were identical to those showing hand placements by the agent, except that each video commenced with the agent placing the ball on the table with his right foot rather than his right hand.

**2.2.2. Free-Agent version**

The videos for the Free-Agent version (whether the agent used his hand or his foot to place the ball on the table) were identical to those of the Constrained-Agent version except that the agent returned to the scene in the same guise and he had his arms free by his sides (see Figure 1).

**2.2.3. Loose-Sheet version**

The videos for the free-agent movement context (whether the agent used his hand or his foot to place the ball on the table) were identical to those of the constrained-agent context except that when the agent returned to the scene in the outcome phase, a loose white linen sheet hung over his shoulders so that arm and hand movements were unrestricted (see Figure 1).

**2.3. Procedure**

Participants were tested individually in a quiet laboratory setting. Each person sat at a Dell Optiplex 9020 desktop with a 23” screen (16:9 aspect ratio). They were guided through the ball-detection task via on-screen directions which explained the format of the test trials and provided the correct procedure for responding. The initial screen stated, “This is an object-detection task. Your job is to press a key as quickly as you can when you see something appear behind a wall”. Participants then completed the Hand-Placement and Foot-Placement forms of the task (counterbalanced). The instructions for both forms were identical except for the information in brackets: “In the first (second) half of the experiment you will see 40 videos, lasting a total of about 15 minutes. They will look like this (relevant frame of video provided). In each video, the person will leave the scene, then return. Press the ‘Q’ key with your left hand as soon as the person has completely left the scene. When the wall disappears do one of the following with your right hand: Press the ‘N’ key if the ball is present; Press the ‘M’ key if the ball is absent”.

In both forms of the said version of the ball-detection task, each trial consisted of an initial fixation cross (1000ms), then a short video. During each video, the participant had to make two responses: an attention check (pressing a key within 2000ms of the agent leaving the scene), and an object detection (selecting whether or not the ball was revealed when the wall dropped away). The timings of each trial’s events were the same across tasks (see Supplementary Figure S1 for example of event timings). For each form, 40 test trials were presented in a pseudorandom order in two blocks. The first block contained 24 trials comprising three cycles of four different conditions with a ball-present or ball-absent outcome. After a student-led break the participants experienced another block of 16 trials (two cycles of four different belief-inducing conditions with either a ball-present or ball-absent outcome). The participants were then instructed that they were to undertake a second round of the ball-detection task. Thus, across the hand placement and foot placement forms of the task, participants experienced 80 trials in total.

At the start of each form of the task, a training phase exposed the participants to 4 practice trials with response time feedback. No performance feedback was given during the test phase to minimize trial time and distraction. The entire experiment took approximately 35 minutes in total. Following Edwards and Low’s protocol (2019), on completion of the experiment participants were asked to complete a form purportedly surveying their experience of how easy it was to sign up for laboratory experiments in exchange for partial course credits (e.g., “Have you found it easy to find suitable timeslots?”). The final question, “What was the experimenter testing?” sought to determine whether the participants were primed to consider the bystander’s belief. Although not a funnelled debriefing protocol we were confident from survey answers that mental state attribution was not deemed to be the target of our research; all participants’ answers referred to the measuring of attention or reaction times in the pursuit of object detection.

**3. Results**

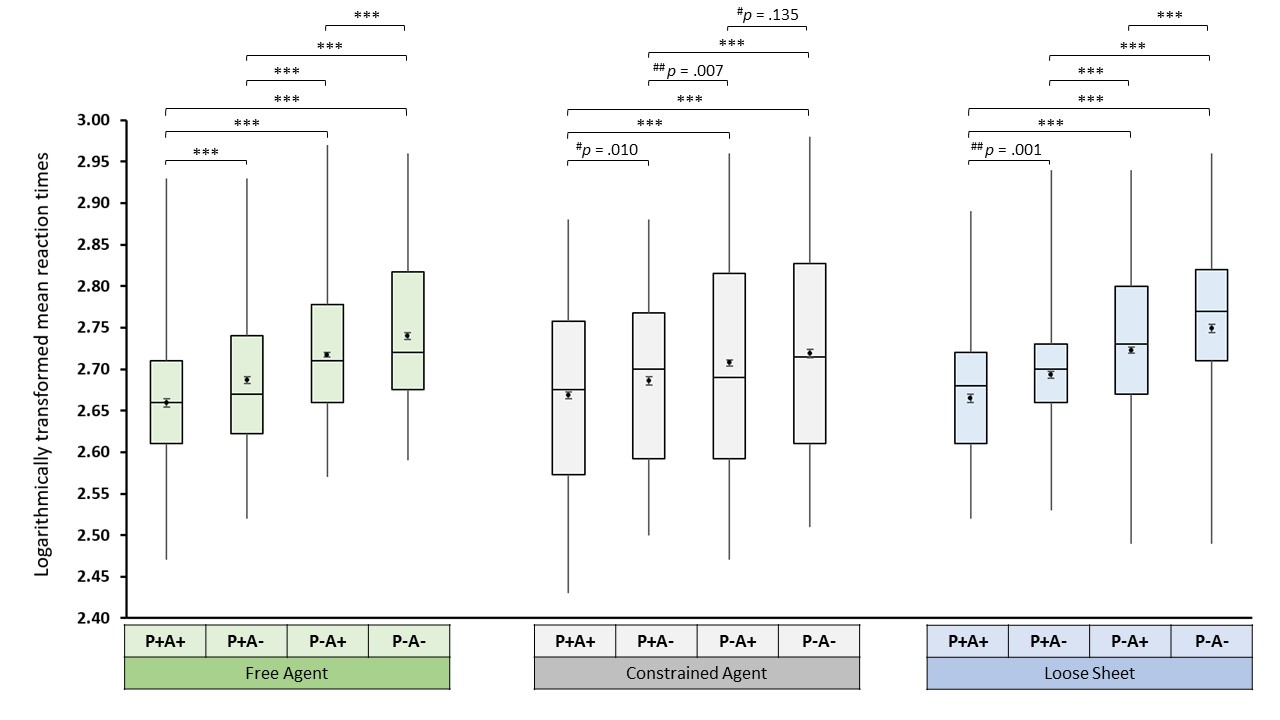
**3.1. Data screening**

All statistical analyses were conducted with IBM SPSS Statistics 23 (SPSS Inc., Chicago, IL, USA). Analysis was undertaken on correct responses, defined as those in which the participant accurately detected the presence or absence of the ball. All statistical tests were two-tailed. Reaction times for trials in which participants failed to respond to an attention check were excluded (1.5% of trials in the Free-Agent version, 2.3% of trials in the Constrained-Agent version, and 2.7% of trials in the Loose-Sheet version). Following an outlier analysis, all data points greater than 3 standard deviations above or below the participant’s overall mean in each task was removed. As a result, 63 individual RTs were omitted from the Free-Agent version (1.7% of total), 51 from the Constrained-Agent version (1.4% of total), and 60 from the Loose-Sheet version (1.7% of total). Tests for normality revealed a positive skew in reaction times and error rates. A logarithmic transformation of the reaction time data was performed to fit the assumptions of an ANOVA before proceeding with further statistical analyses. As such, where means are reported in the main text, the means describe logarithmically transformed data. The extent of the positive skew for the error data necessitated non-parametric testing. Greenhouse Geisser corrections were used whenever the assumption of sphericity was violated. Finally, only trials in which there was a ball to be detected (i.e., ball-present trials) were included in formal analyses (Bardi, Desmet, & Brass, 2019).

**3.2. Response times and errors**

We performed a 3 (Version: Free-Agent; Constrained-Agent; Loose-Sheet) x 2 (Form: Hand-Placement; Foot-Placement) x 4 (Condition: P+A−; P+A−; P−A+; P−A−) mixed model ANOVA. The findings were straightforward. There was only a significant main effect of Condition (*F*(2.734, 366.406) = 113.09, *p* = .000, ηp2 = 0.46) that was qualified by a significant interaction effect of Version x Condition (*F*(5.469, 366.406) = 2.58, *p* = .022, ηp2 = 0.04). The form of the effector (Hand-placement or Foot Placement) used by the agent to place the ball on the table did not make any difference to participants’ responding; the Form x Version x Condition interaction was not significant (*F*(5.620, 376.568) = 0.70, *p* = .641, ηp2 = 0.01). All other effects were also not significant. We carried out a one-way ANOVA for each version of the ball-detection task to unpack the significant Version x Condition interaction.

The main effect of Condition was significant in the Free-Agent version (*F*(3, 135) = 54.30, *p* < .001, ηp2 = 0.55), in the Constrained-Agent version (*F*(2.62, 117.67) = 20.18, *p* < .001, ηp2 = 0.31), and in the Loose-Sheet version (*F*(2.49, 109.41) = 43.09, *p* < .001, ηp2 = 0.50). Pairwise t-tests were performed to interpret the significant main effects, with Bonferroni correction (with significance threshold at *p* < .0083) to account for multiple comparisons. Supporting our primary prediction, response times were significantly faster in the P−A+ condition than in the P−A− condition in the Free-Agent version (*t*(45) = 4.435, *p* < .001; *r* = .55) and in the Loose-Sheet version (*t*(44) = 3.796, *p* < .001; *r* = .50). Also confirming our primary prediction, there was no difference between the P−A+ and P−A− conditions in the Constrained-Agent version (*t*(45) = 1.521, *p* = .135; *r* = .22). The pattern of responding is shown in Figure 3. We then compared response times for the other conditions. The pairwise comparisons of response times for the other conditions (see Table 2 for overview) indicated that, in all three versions of the ball-detection task, participants were fastest to respond in the P+A+ condition and slowest to respond in the P−A− condition; in addition, their reaction times in the P+A− condition were significantly faster than in the P­−A+ condition (readers interested in pairwise comparisons involving ball-absent data of conditions in each version may refer to Supplementary Table S1).



***Figure 3.*** *Box plots presenting logarithmically transformed mean response times for the four conditions in each of the different versions of the ball-detection task. Means are represented by dot markers; associated error bars represent the standard error of the mean. Comparisons marked with \*\*\* (denoting p < .001) and ## (denoting p < .0083) survived Bonferroni correction for multiple comparisons. Comparisons marked with # (denoting p > .0083) did not survived Bonferroni correction for multiple comparisons.*

***Table 2.*** *Overview of pairwise comparisons in each version of the ball-detection task.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Version** | **Comparison** | **Paired Differences** | | ***t*** | ***p*** |
| ***m*** | ***sd*** |
| **Free Agent** | P-A- *versus* P-A+ | .023 | .035 | 4.435 | <.001a |
|  | P-A- *versus* P+A- | .054 | .045 | 8.062 | <.001a |
|  | P-A- *versus* P+A+ | .081 | .051 | 10.756 | <.001a |
|  | P-A+ *versus* P+A- | .031 | .045 | 4.654 | <.001a |
|  | P-A+ *versus* P+A+ | .059 | .047 | 8.518 | <.001a |
|  | P+A- *versus* P+A+ | .027 | .052 | 3.561 | =.001a |
| **Constrained Agent** | P-A- *versus* P-A+ | .010 | .045 | 1.521 | =.135b |
|  | P-A- *versus* P+A- | .032 | .055 | 3.937 | <.000a |
|  | P-A- *versus* P+A+ | .049 | .050 | 6.743 | <.001a |
|  | P-A+ *versus* P+A- | .022 | .052 | 2.839 | =.007a |
|  | P-A+ *versus* P+A+ | .039 | .036 | 7.418 | <.001a |
|  | P+A- *versus* P+A+ | .018 | .045 | 2.694 | =.010b |
| **Loose Sheet** | P-A- *versus* P-A+ | .026 | .046 | 3.796 | <.001a |
|  | P-A- *versus* P+A- | .056 | .049 | 7.645 | <.001a |
|  | P-A- *versus* P+A+ | .084 | .066 | 8.539 | <.001a |
|  | P-A+ *versus* P+A- | .030 | .051 | 3.966 | <.001a |
|  | P-A+ *versus* P+A+ | .058 | .052 | 7.545 | <.001a |
|  | P+A- *versus* P+A+ | .028 | .050 | 3.746 | =.001a |

*Note:*

*aComparison remained significant after Bonferroni correction; bComparison did not survive Bonferroni correction.*

We also undertook an orthogonal analysis of the ball-present data to explore the influence of the agent’s belief and the participant’s belief. We performed a 3 (Version: Free-Agent; Constrained-Agent; Loose-Sheet) x 2 (Form: Hand-Placement; Foot-Placement) x 2 (Belief Holder: P, A) x 2 (Belief: +, −) mixed model ANOVA. We followed Edwards and Low’s (2019) approach by computing four scores, P+, P−, A+ and A−. The A+ score captured the scenario where the agent was led to expect the outcome ([P+A+] + [P−A+] /2), whereas the A- score captured the scenario where the outcome was unexpected by the agent ([P+A−] + [P−A−] /2). The P+ score captured the scenario where the participant expected the outcome (([P+A+] + [P+A−]) / 2), whereas the P- score captured the scenario where the participant did not expect the outcome (([P−A+] + [P−A−]) / 2). The mixed model ANOVA revealed two significant interactions, which are explained below.

The significant main effect of Belief Holder (*F*(1, 134) = 20.10, *p* < .001; ηp2 = .13) was qualified by a significant Belief Holder x Belief interaction (*F*(1, 134) = 43.26. *p* < .001; ηp2 = 0.24). The interaction (which was also reported by Edwards and Low, 2019) can be explained by the observation that the effect of belief (+ or −) was stronger when it was the participants themselves that held the mental state than when it was the agent that held the mental state. In order words, the difference between P+ (*M* = 2.68, *SD* = 0.09) and P− (*M* = 2.73, *SD* = 0.11) responding was larger than that between A+ (*M* = 2.69, *SD* = 0.10) and A− (*M* = 2.71, *SD* = 0.10) responding.

The significant main effect of Belief (*F*(1, 134) = 218.63, *p* < .001; ηp2 = .62) was qualified by a significant Belief x Version interaction (*F*(2, 134) = 4.74, *p* = .010; ηp2 = .07). The interaction can be explained as follows. The difference between a ball-expected-absent (−) and a ball-expected-present (+) belief on task responding was larger in the Free-Agent version (*M*Difference = 0.04, *SD* = 0.03) than compared to the Constrained-Agent version (*M*Difference = 0.03, *SD* = 0.03) ((*t*(90) = 2.81, *p* = .006; *r* = .28). Similarly, the difference between a ball-expected-absent (−) and a ball-expected-present (+) belief on task responding was larger in the Loose-Sheet version (*M*Difference = 0.04, *SD* = 0.03) than compared to the Constrained-Agent version ((*t*(82.73) = 2.64, *p* = .010; *r* = .28). The belief effect on task responding did not differ between the Loose-Sheet and Free-Agent versions ((*t*(82.33) = 0.24, *p* = .809, *r* = .03).

Finally, to finish our analyses, we confirmed that participants showed a high level of accuracy (regardless of either the version or form of the ball-detection task), revealed by low mean error proportions across the different trial conditions. Tests for normality revealed that the error data was positively skewed. Pooling across the different versions, a Friedman test revealed no statistically significant differences in mean error proportions across the four trial conditions (P+A+ = 0.06; P+A− = 0.07; P−A+ = 0.07; P−A− = 0.07; Χ2(3) = 5.95, *p* = 0.114).

**4. Discussion**

We aimed to determine whether and to what extent visibly constraining an agent from potentially interacting with an object modulates observers’ abilities to automatically track that agent’s belief. The P−A+ < P−A− effect was elicited in the Free-Agent version of the ball-detection task where the agent was visibly free to move and to potentially act on the ball. Importantly, confirming our primary prediction, the P−A+ < P−A− effect was not elicited in the Constrained-Agent version where the agent was visibly constrained from using his upper body and limbs to potentially act on the ball. These findings suggest that motor representations of the action context can modulate belief tracking. However, it is possible that participants may not have shown reaction time benefits in the Constrained-Agent version of the ball-detection task because the agent’s “mummified” appearance in the outcome phase was perceptually novel. Our Loose-Sheet version of the ball-detection task provided an important control that helps to rule out a low-level interpretation based on perceptual novelty. In the Loose-Sheet version, the agent also returned with a novel appearance, but he remained visibly able to move his upper body and limbs and, potentially, to act upon the ball. In further confirmation of our primary prediction, we found that the P−A+ < P−A− effect was elicited in the Free-Agent *and* Loose-Sheet versions but obstructed in the Agent-Constrained version.

[More to add in Discussion]

* The classical interpretation of the P+A- < P-A- effect is that the reaction-time difference is a consequence of belief tracking on an object detection task (Kovacs et al., 2010; Edwards & Low, 2019). In other words, adults automatically computed and maintained the agent’s belief about the ball’s actual location, and tracking of the agent’s belief facilitated participants’ perceptual task of detecting the presence of the ball even when participants’ themselves did not expect the ball to be there. Our discovery—that intervening on how participants represent someone else’s action motorically also affects participants’ belief tracking—adds a twist to the interpretation of the P+A- < P-A- effect. The P+A- < P-A- effect may be a consequence of the way that information about beliefs feeds into motor predictions of the agent and those motor predictions then facilitate (or not) response times.
* One possibility is that the hand-placement versus foot-placement occurred too early in the event sequence for the manipulation to make much difference participants’ detection of the ball. With a different paradigm, it may turn out that belief tracking might be effector specific. With respect to our current findings, our cautious interpretation is that we regard response times on the object-tracking tasks as reflecting a sort of motor process that is influenced by belief tracking, but the motor process is unlikely to mainly take into account the specific effector involved.

SUPPLEMENTARY/APPENDIX

***Table S1.*** *Overview of pairwise comparisons in each version of the ball-detection task for ball-absent trials.*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Version** | **Comparison** | **Paired Differences** | | ***t*** | ***p*** |
| ***m*** | ***sd*** |
| **Free Agent** | P-A- *versus* P-A+ | .016 | .049 | 2.206 | =.033b |
|  | P-A- *versus* P+A- | -.037 | .053 | -4.744 | <.001a |
|  | P-A- *versus* P+A+ | -.049 | .059 | -5.635 | <.001a |
|  | P-A+ *versus* P+A- | -.053 | .054 | -6.639 | <.001a |
|  | P-A+ *versus* P+A+ | -.065 | .062 | -7.212 | <.001a |
|  | P+A- *versus* P+A+ | -.012 | .041 | -2.011 | =.050b |
| **Constrained Agent** | P-A- *versus* P-A+ | .007 | .037 | 1.291 | =.203b |
|  | P-A- *versus* P+A- | -.031 | .052 | -3.985 | <.000a |
|  | P-A- *versus* P+A+ | -.041 | .062 | -4.465 | <.001a |
|  | P-A+ *versus* P+A- | -.038 | .057 | -4.488 | <.001a |
|  | P-A+ *versus* P+A+ | -.048 | .064 | -5.100 | <.001a |
|  | P+A- *versus* P+A+ | -.010 | .051 | -1.389 | =.172b |
| **Loose Sheet** | P-A- *versus* P-A+ | .016 | .061 | 1.721 | =.092b |
|  | P-A- *versus* P+A- | -.041 | .075 | -3.615 | =.001a |
|  | P-A- *versus* P+A+ | -.051 | .067 | -5.108 | <.001a |
|  | P-A+ *versus* P+A- | -.056 | .068 | -5.583 | <.001a |
|  | P-A+ *versus* P+A+ | -.066 | .064 | -6.926 | <.001a |
|  | P+A- *versus* P+A+ | -.010 | .045 | -1.531 | =.133a |

*Note:*

*aeffect remained significant after Bonferroni correction; beffect did not survive Bonferroni correction.*

**A screenshot of a cell phone

Description automatically generated**

**Figure S1.** Timings of the main events of the four conditions (illustrated by way of the Constrained-Agent version of the ball-detection task) (timing are the same in the other task versions). Two main events are highlighted: (1) the attention check triggered by the exit of the agent (hatched border); (2) removal of the screens (bold border).