Separate timescales for spatial and anatomical information processing of body stimuli

Baptiste M. Waltzing[1] *, Siobhan McAteer [1], Marcos Moreno-Verdú [1], Elise E. Van Caenegem[1], Yue Du[2], Robert M. Hardwick[1]

[1]Institute of Neurosciences, UC Louvain, Belgium Avenue Mounier 54, 1200, Bruxelles, Belgium

[2]Department of Neurology, Johns Hopkins University, Baltimore, Maryland, United States

*Correspondence should be addressed to Baptiste Waltzing (baptiste.waltzing@uclouvain.be)

Declaration of Interest: The authors declare no competing interest.

Abstract

Observing different body stimuli can influence the speed and accuracy of our responses. Prior work indicates this effect is influenced by factors such as spatial congruence and perspective. We hypothesized that the influence of these factors would vary depending on the amount of time that participants had to process visual stimuli. Experiment 1 was a reaction time task (n=29) with stimuli varying in spatial congruence (congruent, incongruent, neutral), perspective (first- or third-person) and stimulus type (body or control). Experiment 2 (n=50) used the same stimuli in a "Forced Response" paradigm, which controlled the time participants had to prepare a response. This allowed us to assess responses as a function of preparation time. Experiment 1 showed effects of spatial congruence, with longer reaction times and more errors for spatially incongruent stimuli. This effect was greater for body stimuli. Experiment 2 showed that spatial information was processed faster than anatomical information, inducing incorrect responses at short preparation times for spatially incongruent body stimuli. There was little-to-no corresponding effect for control stimuli. Both experiments also showed weak-to-no effects of perspective, which appear to have been driven by spatial congruence. Our results indicate that spatial information is processed faster than anatomical information during observation of body stimuli. These data are consistent with the dual visual streams hypothesis, whereby spatial information would be processed rapidly via the dorsal stream, while anatomical processing would occur later via the ventral stream. These data also indicate differences in processing between body and control stimuli.

Key words: Spatial congruence, Perspective, Stimuli processing

Public significance statements: This study provides novel insight into the time-course of information processing, showing that spatial information is processed faster than anatomical

information for body stimuli. The results also challenge the established view that visual perspective is critical to process body stimuli, demonstrating that this may instead result from lower-level effects of spatial congruence.

Introduction

Action observation (i.e. observing a movement being performed) plays a central role in our everyday lives. We observe the movements of others in many contexts, including to imitate them in order to learn new movements (Gatti et al., 2013), to respond with a complementary action such as shaking hands (Curioni et al., 2020; Faber et al., 2016), or without any specific intention. In recent years, action observation has also been used in the medical field, notably for the rehabilitation of movement-related disorders (Caligiore et al., 2017; López et al., 2019; Machado et al., 2019; Nicholson et al., 2019). Despite this widespread interest, there are still fundamental questions regarding how stimuli presenting actions are processed; addressing these questions is a key challenge in optimizing the use of action observation in order to use it as effectively as possible in applied situations.

In non-human primates, the effects of action observation are mediated by mirror neurons. These neurons, first discovered in the F5 region of the brains of macaques, are activated both during the observation of an action, and during the execution of the same action (di Pellegrino et al., 1992; Rizzolatti et al., 1996). Neurons with similar properties have been identified in premotor and parietal areas, and are thought to form a “mirror neuron system” that could mediate the effects of action observation (Blakemore & Frith, 2005). In humans, action observation leads to the recruitment of functionally similar premotor and parietal areas that are considered to form an “action observation system” (Fadiga et al., 1995; Hardwick et al., 2018). These discoveries provide a potential neural basis for behavioural effects whereby observing the actions of another person can influence our own movements.

Research has demonstrated that observing actions performed by other people can prime us to perform similar movements (Edwards et al., 2003; Hardwick & Edwards, 2011, 2012; Press

et al., 2011). In particular, previous work indicates that reaction times are reduced following the observation of an action compatible with the required response (Brass et al., 2001, Edwards et al. , 2003). Subsequent studies have investigated factors that influence such priming effects (for a review see (Kemmerer, 2021)). In this study, we focused on three of them in particular. The first is the nature of the stimulus, which may be body-related, or control. Studies have shown that the observation of body-related stimuli (e.g. a hand (Gowen et al., 2010)) can induce a priming effect if the requested movement is congruent with the observed movement. There is mixed evidence that this priming effect is smaller or absent during the observation of control stimuli (e.g. geometric shapes (Biermann-Ruben et al., 2008)) and is therefore argued that there would be a particular processing for body stimuli (for a review see (Gowen & Poliakoff, 2012)). This greater priming effect could be explained by the pre-activation of motor areas when the action is observed (Jeannerod, 2001). Many different control stimuli have been used in past studies, e.g. squares, pens, dots, crosses, etc; for review see (Gowen & Poliakoff, 2012). While numerous possible control stimuli were possible, each with respective strengths and limitations, our choice in the present study was conditioned upon certain requirements. A primary point was that we required control stimuli that were asymmetrical, so that they could undergo vertical and horizontal symmetries equivalent to a change in laterality or perspective of the observed hand. We also required the stimuli to be easy to associate with the required response, so that extensive training would not be necessary. Some of the control stimuli previously used were superimposed on hand images (e.g. X superposed on finger)(Brass et al., 2001). As we wanted stimuli that didn't include a body element, this choice was not possible. We therefore chose to use the letters of the keyboard buttons that participants would press to act as the control stimuli in the present task, as they fulfilled the criterion noted above. A second factor is perspective. It has been

shown that observing an action from a first-person perspective (consistent with observing our own movements through our own eyes) had a greater priming effect and was therefore associated with faster reaction times than actions observed from a third-person perspective (observing an action from the vantage point of another person) (Mibu et al., 2020). This effect could be explained by the fact that some studies have shown greater activation of areas of the action observation system when observing actions from a first-person perspective than from a third-person perspective (Angelini et al., 2021; Ge et al., 2018; Oosterhof et al., 2012). The third factor we were interested in was the spatial congruence of the observed action. Studies have shown that observing an action that is spatially congruent with the required action (e.g. a right hand in first-person perspective for a response with the right hand) (fig 1A) has a greater priming effect than observing an action that is spatially incongruent (e.g. a left hand in first-person perspective for a response with the right hand) (Avikainen et al., 2003; Bekkering et al., 2000; Chiavarino et al., 2007; Wohlschläger & Bekkering, 2002). This could be explained by greater activation of the contralateral hemisphere of the hand observed in the first-person perspective but greater activation of the ipsilateral hemisphere in the third-person perspective (Shmuelof & Zohary, 2008).

We therefore examined how these three factors influenced action observation. Experiment 1 used a classical reaction time task to study possible interactions between the factors of stimulus type, perspective and spatial congruence. We predicted that although perspective may have an effect, the most important point would be spatial congruence, with shorter reaction times for spatially congruent stimuli than for spatially incongruent stimuli. This hypothesis is inspired by the work of Simon et al. (Simon & Wolf, 1963) who showed that the spatial position of the observed stimuli played a key role in their processing, with shorter reaction times when they were spatially congruent. While traditional reaction time-based

studies provide general information about stimulus processing, they are limited in their ability to examine the precise time-course of information processing. To address these limitations, Experiment 2 used a forced response paradigm (Haith et al., 2016; Hardwick et al., 2019; Schouten & Bekker, 1967; Vleugels et al., 2020). This procedure limits the processing time available to participants, providing a measurement of accuracy as a function of the response preparation time available, allowing us to obtain a novel insight into how the stimulus is processed. We expect to observe differences in processing between body and control stimuli. We therefore hypothesize that, due to the activation of the action observation system, which includes motor-related areas specifically following the observation of body stimuli (Hardwick et al., 2018) body-related stimuli would be more susceptible to changes in spatial congruence. Similar to our prediction for Experiment 1, we also hypothesised that spatial congruence may interact with perspective during this task.

Experiment 1

Experiment 1 Methods

In this first experiment, we investigated how action observation could be influenced by factors including the nature of the stimulus, perspective, and the spatial congruence of the stimulus with the participant's responses. To do this, we performed a reaction time task to examine variations in both reaction time and errors. All procedures were reported in accordance to the 'Action Observation' part of the "Guidelines for Reporting Action Simulation Studies" checklist (Moreno-Verdú et al., 2024). This study was not preregistered.

Participants

A group of 30 healthy young participants were recruited via the Prolific online platform (https://www.prolific.com/). A single participant was excluded for an extremely high number of unanswered or incorrect trials and very short reaction times (see section 1.1 of appendices for more details). This left 29 participants aged between 18 and 34 years (mean = 25.3 years, SD = 4.2 years). All participants were right handed, 10 were women, 19 were men. The study by Catmur and Heyes, 2011 (Catmur & Heyes, 2011), suggests a moderate effect size (dz = 0.71)(Cohen, 1977) for the influence of spatial congruence on reaction time. A sample size calculation with Gpower 3.1 for a paired two tails t-test with alpha = 0.05, power = 0.8 and a moderate effect size (dz = 0.71) suggests that a sample size of more than 18 participants would be sufficient. To ensure sufficient power in case of additional variability, the number of participants was increased to 30. They were all naive about the purpose of the experiment. All participants gave their consent before the start of the experiment and were financially compensated (3£) for their time. The study was approved by the IPSY ethics committee of UCLouvain.

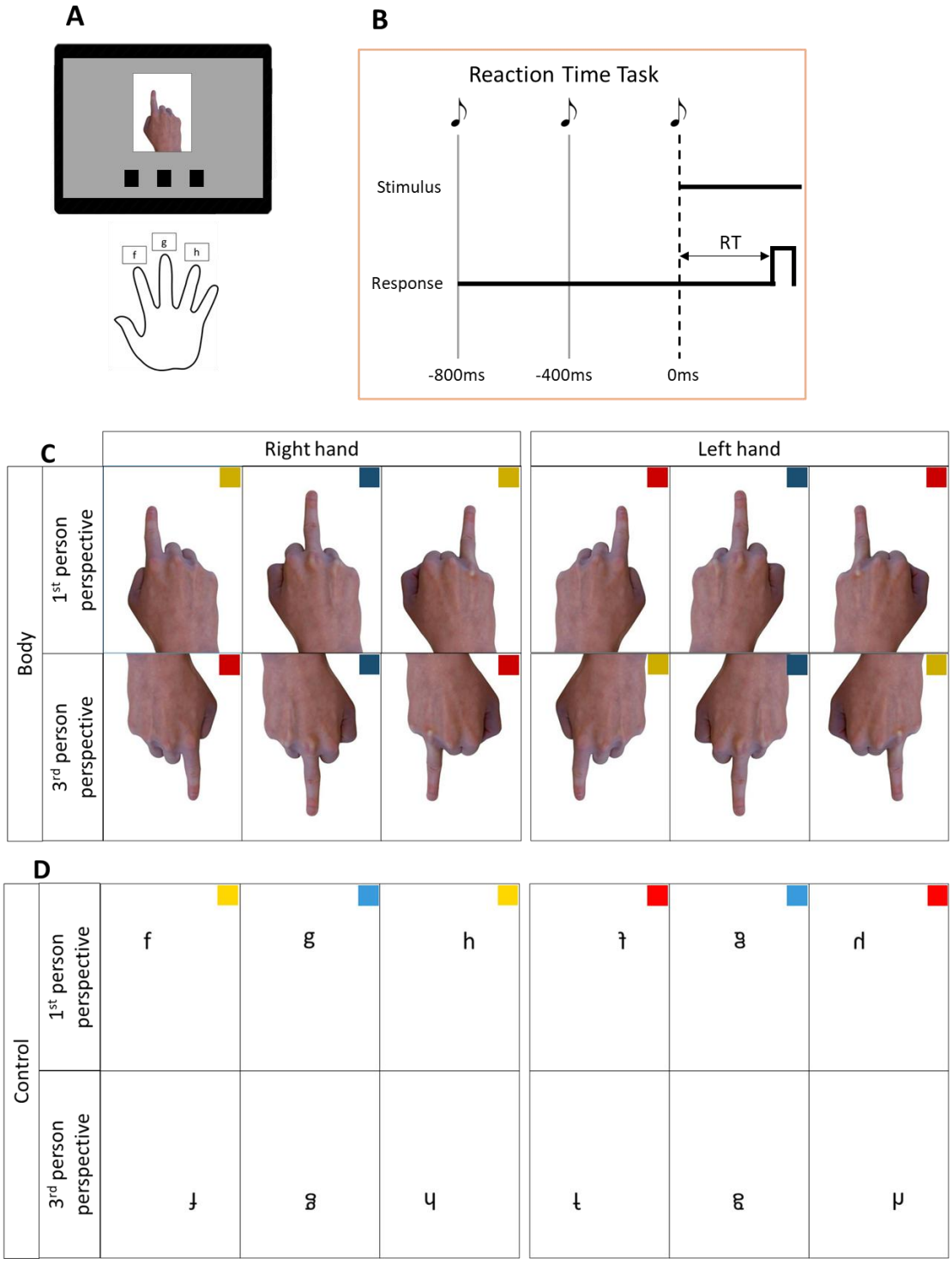

Figure 1. A. General experimental task set up. Participants positioned their right hand with the index, middle and ring fingers resting on the f, g and h keys respectively. The stimulus appeared in the centre of the screen, indicating which key should be pressed. Underneath were 3 black squares, each corresponding to a key and providing visual feedback. The square of the key pressed turned green if the response was correct and red if the response was incorrect. B. Reaction time paradigm. In each trial the participant heard a sequence of three

equally spaced tones, and a stimulus always appeared at the same time as the third tone. Participants were instructed to respond as quickly and accurately as possible once the stimulus appeared. C. Body stimuli. The different body stimuli are presented, which can show a right or left hand, first- or third-person perspective, with index, middle or ring finger raised. The coloured squares are not part of the stimuli, but indicate the spatial congruence: yellow for spatially congruent stimuli, red for spatially incongruent stimuli and blue for spatially neutral stimuli. A darker shade is used for body stimuli and a lighter shade for control stimuli. This colour code is reused in the results plots. D. Control stimuli. These stimuli were presented in the same position as the body stimuli, whose low-level characteristics they mimic.

General Task Procedures

The task was created in PsychoPy2 version 2023.2.2 (Peirce et al., 2019) and run online via PsychoPy2's associated webpage Pavlovia (https://pavlovia.org/) (Bridges et al., 2020).

This reaction time task consisted of three different conditions: the familiarisation condition, the body stimuli condition, and the control stimuli condition. The task was the same for all three conditions, but each condition was characterised by different stimuli. Participants sat in front of a computer keyboard with the index, middle, and ring fingers of their right hand positioned on the f, g, and h keys of the keyboard, respectively (fig 1A). For each trial, the participants heard three short tones (50ms duration) spaced 400ms apart (fig 1B). To make it easier to distinguish between the different tones, the first two tones had a lower pitch than the final tone (i.e. notes "La" and "Si", respectively). A visual stimulus was presented at the same time as the third tone. Depending on the stimulus that appeared, the participants had to respond by pressing on the keyboard with their index, middle, or ring finger. Participants were instructed to answer as quickly and as accurately as possible once the stimulus appeared on the screen. They were given a maximum of 1s to respond, or the trial was considered to be 'time out'. Once they had answered, visual feedback was displayed on the screen to indicate

whether or not the answer was correct. The feedback consisted of three small black squares at the bottom of the screen, each corresponding to one of the three keyboard buttons. The box corresponding to the button that the participant responded with changed to provide feedback, turning green if the response was correct, or red if it was incorrect. The trial had a fixed duration of 1 second regardless of the participant's reaction time, and was followed by a 200ms interval in which no stimulus was shown on screen. In all conditions the order of trials was pseudo-randomised such that the same response was presented on a maximum of two consecutive trials. There was a break between each block (this could be as long as the participant desired, with a minimum of at least 10s), during which the instructions were displayed on the screen. Each block began with a short check in which the participant had to press the three keys, in order to prompt the participants to ensure their fingers were placed on the correct keys.

Familiarisation condition. The familiarization block was used to help participants gain some practice in completing the task. For this condition, one of three different stimuli could be displayed on the screen: an arrow pointing to the left, upwards, or to the right. For these stimuli, participants had to respond with their index, middle and ring fingers respectively. This block consisted of 20 trials for each stimulus. Data from this condition was not included in later analyses.

Body stimuli condition. For this condition, the stimuli displayed were images of a hand with one finger raised (fig 1C). Participants were instructed to respond using the finger that appeared raised on the displayed stimulus (i.e. if the stimulus presented an index finger, the participant had to respond with their own index finger by pressing on the f key). The hand could be presented in first- or third-person perspective, be a right or left hand, and have the

200 index, middle, or ring finger raised, for a total of 12 different stimuli (2 perspectives * 2 hands
201 * 3 fingers). The same three images of the right hand in first-person perspective, one for each
202 finger, were used for all stimuli, with mirroring and/or rotation to create left hand/third-
203 person perspective stimuli. The position of the stimulus relative to the participant's own hand,
204 and thus the spatial congruence of the stimulus relative to the required response, was
205 therefore influenced by the perspective and the hand observed. For example, when the
206 stimulus presented the index finger of a right hand in the first-person perspective, the position
207 of the finger was spatially congruent with that of the participant's own index finger; by
208 contrast, if the stimulus presented the index finger on a *left* hand in first-person perspective,
209 the position of the finger was spatially incongruent with the response required by the
210 participant. While these manipulations affected the positions of the index and ring fingers,
211 being centred, the stimuli with the middle finger were always spatially congruent regardless
212 of the perspective and the hand observed. For this condition, there were two blocks of 120
213 trials, with each stimulus being presented 10 times per block. The order of the trials was
214 pseudorandomised to limit the number of consecutive responses with the same finger to two.
215 Participants therefore completed a total of 20 trials with each stimulus, and a total of 240
216 trials in this condition.

217 *Control stimuli condition*. Previous studies have suggested differences in processing for
218 body compared to control stimuli (Gowen & Poliakoff, 2012). Therefore, the aim of this
219 condition was to have control stimuli that mimicked the low-level characteristics of body
220 stimuli, to see whether the effects observed were specific to action observation or not. The
221 stimuli were the lowercase letters "f", "g" or "h" (fig 1D) in Calibri font. Participants were asked
222 to respond to these stimuli with their index, middle and ring fingers respectively. To obtain
223 the 12 different stimuli equivalent to the body stimuli, the letter stimuli were first placed at

the respective locations of the index, middle and ring fingers of the right hand stimulus in first-person perspective. The same rotation/mirroring manipulations as applied to the body stimuli were then applied to these control stimuli. This gave us 12 different stimuli (2 * rotation (equivalent to “perspective” in the body stimuli condition), 2 * vertical symmetry (equivalent to “hand” in the body stimuli condition), 3 * letters (equivalent to “finger” in the body stimuli condition). Like the body stimuli, the positions of the letters f and h could be spatially congruent (i.e. equivalent right hand, first-person perspective and left hand, third-person perspective) or incongruent (i.e. equivalent left hand, first-person perspective and right hand, third-person perspective). The central position of the letter g meant that it was always spatially congruent, regardless of the orientation of the stimulus. As in the body stimuli condition, participants completed two blocks, each block comprising 120 trials with 10 trials for each stimulus, for a total of 240 trials with the same pseudorandomisation.

Protocol

Participants always began with the familiarisation condition, after which the order of conditions was counterbalanced across participants; half of the participants completed the body stimuli condition first, followed by the control stimuli condition, and vice-versa. The median completion time was 24 minutes (including breaks between blocks).

Data analysis

Analyses were performed in R version 4.3.2 (R Core Team 2024).

We measured two variables of interest: reaction time (defined as the time between the appearance of the stimulus and the response given by the participants) and error rate (percentage of incorrect responses out of the total number of responses).

Exclusion criteria. Of the 13920 total trials conducted, 370 trials (representing 2.66% of the total) were excluded from the analyses for having no response.

Reaction time. As we performed a separate analysis for the error rate, trials with incorrect responses were not included in the reaction time analyses (n=1421), leaving a total of n=12129 trials for analysis.

Global Analysis: Data were first analysed using a "Global" model that considered the data in its most granular form. A complex random intercept model (the choice of model was made following the guidelines in (Scandola & Tidoni, 2024)) examined effects and possible interactions between several factors, labelled here in relation to the body stimuli. This included the factors of "Stimulus Type" (2 levels; body or control), "Perspective" (2 levels; first-person or third-person), "Hand" (2 levels; left or right), and "Finger" (3 levels; index, middle, or ring). The model included between-participant variability as a random effect (i.e. random intercept) and used the formula: reaction time ~ stimulus type * perspective * hand * finger + (1|participant) + (1|participant: stimulus type) + (1|participant: perspective) + (1|participant: finger) + (1|participant: perspective: hand) + (1|participant: stimulus type: finger) + (1|participant: stimulus type: perspective: hand) + (1|participant: perspective: hand: finger) + (1|participant: stimulus type: perspective: hand: finger).

Spatial Congruence Analysis: We anticipated that spatial congruence would have important effects on participant performance, and therefore conducted an analysis to specifically examine this effect. Stimuli were grouped according to the spatial congruence between the observed stimulus and required response. Spatial congruence could therefore be either "neutral" (i.e. all middle finger stimuli regardless of perspective, as the middle finger was always in the same central position on screen), "congruent" (i.e. index and ring finger

stimuli for the first-person right hand stimuli, and third-person left hand stimuli), or "incongruent" (i.e. index and ring finger stimuli for the first-person left hand stimuli and third-person right hand stimuli). We therefore used a complex random intercept model to see if reaction time was affected by the factors of "Stimulus Type" (2 levels; body or control) and "Spatial Congruence" (3 levels; neutral, congruent, incongruent), using the formula: reaction time ~ stimulus type * congruence + (1|participant) + (1|participant: stimulus type) + (1|participant: congruence) + (1|participant: stimulus type: congruence).

Error Rate. The analysis of error rates was carried out in a similar way to the analysis of reaction times, but using binomial generalized linear mixed models to take account of the binary nature of the outcome (0=incorrect, 1=correct). The same general analyses were conducted as for the reaction time. A "Global Analysis" used the formula: error rate ~ stim_type * perspective * hand * finger + (1 | participant) + (1 | participant: stimulus_type) + (1|participant: perspective: hand)+(1|participant: stimulus type: finger)+(1|participant: stimulus type: perspective: hand: finger), while a "Spatial Congruence Analysis" examined the data using the formula: error rate ~ stimulus type * congruence + (1|participant) + (1|participant: stimulus type) + (1|participant: congruence) + (1|participant: stimulus type: congruence).

Response type. We looked at what kind of response participants made when they saw spatially incongruent stimuli. To do this, we first filtered to keep only the spatially incongruent trials. Then the trials were labelled as follows: correct (for trials with a correct response), spatially induced error (for responses given with the index finger when the correct response was the ring finger or vice versa) and spatially neutral error (for incorrect responses given with the middle finger). A generalized linear mixed model was then calculated using the following

formula: number of responses ~ stimulus type * response type + (1|participant) + (1|participant: stimulus type) + (1|participant: response type) + (1|participant: stimulus type: response type).

We were also interested in the reaction time for each type of response. We calculated the following linear mixed model: reaction time ~ stimulus type * response type + (1|participant) + (1|participant: stimulus type) + (1|participant: response type) + (1|participant: stimulus type: response type).

Post hoc comparisons. For all models, post hoc comparisons were conducted using z-tests. All p-values obtained were adjusted using Tukey's method for multiple comparison, and alpha was set to 0.05 for statistical significance.

302 Experiment 1 Results

303 *Reaction Time*

304 *Global Analysis (fig 2A,B).* Results of the global model are presented in Table 1.

Term	Reaction Time		Accuracy	
	χ2	*P*	χ2	*P*
Stimulus Type	*14.9*	*< 0.001**	10.8	*< 0.001**
Perspective	*29.2*	*< 0.001**	0.0	0.970
Hand	*4.1*	*0.043**	3.9	*0.048**
Finger	145.6	< 0.001*	97.4	*< 0.001**
Stimulus Type * Perspective	5.4	0.020*	2.1	0.144
Stimulus Type * Hand	1.0	0.315	6.1	0.014*
Perspective * Hand	181.6	*< 0.001**	119.7	*< 0.001**
Stimulus Type * Finger	26.1	*< 0.001**	35.8	*< 0.001**
Perspective * Finger	13.1	*0.001**	1.9	0.390
Hand * Finger	4.0	0.136	1.5	0.474
Stimulus Type * Perspective * Hand	28.3	*< 0.001**	18.7	*< 0.001**
Stimulus Type * Perspective * Finger	0.8	0.660	0.9	0.627
Stimulus Type * Hand * Finger	4.1	0.126	5.9	0.052
Perspective * Hand * Finger	136.8	*< 0.001**	48.0	*< 0.001**
Stimulus Type * Perspective * Hand * Finger	5.8	0.056	1.4	0.509

305 Table 1: Results from the Global analyses of Reaction Times and Accuracy. * Indicates
306 statistically significant results.

307 The global model identified several notable factors. First, main effects identified that
308 participants were significantly faster when responding to control stimuli (562 ± 57ms)
309 compared to body stimuli (586 ± 53ms, z = 4.0, p< 0.001)(fig S2), when responding to first-
310 person stimuli (566 ± 53ms) compared to third-person stimuli (582 ± 51ms, z = 5.2, p<
311 0.001)(fig S4), and when responding to right hand (571 ± 51ms) stimuli compared to left hand

stimuli (576 ± 51ms, $z = 2.1$, $p = 0.033$). This analysis also identified finger-related effects; post-hoc comparisons identified that participants were generally faster when presented with stimuli requiring responses with the middle finger (543 ± 50ms), compared to the index (584 ± 55ms, $z = 8.9$, $p<0.001$) or ring fingers (597 ± 54ms, $z = 11.6$, $p<0.001$), while responses with the index finger were generally faster than those with the ring finger ($z = 2.7$, $p=0.020$). Secondly, important interactions between the perspective and observed hand were identified; generally, in the first-person perspective, participants were faster when responding to a right hand (549 ± 55ms) than a left hand (584 ± 52ms) ($z = 11.4$, $p< 0.001$), while in the third-person perspective, participants were faster when responding to a left hand (569 ± 51ms) than a right hand (596 ± 51ms) ($z = 8.4$, $p< 0.001$). This overall pattern of results was highly consistent with effects that depended on the spatial congruence between the observed stimulus and the required response. As such, further analysis concentrated on examining this effect directly.

Spatial Congruence Analysis (fig 2C). The model for spatial congruence showed a significant interaction between stimulus type and spatial congruence ($\chi^2 = 39.1$, $p < 0.001$) as well as a main effects of stimulus type ($\chi^2 = 15.3$, $p < 0.001$) and spatial congruence ($\chi^2 = 266.8$, $p < 0.001$). Post-hoc comparisons focusing on body stimuli showed that reaction time was shortest for neutral (547 ± 57ms) then congruent (580 ± 57ms) then incongruent stimuli (647 ± 55ms) (all $z > 5.9$, all $p < 0.001$). Results for the control stimuli showed shorter reaction times for neutral (540 ± 53ms) and congruent stimuli (556 ± 64ms) compared to incongruent stimuli (592 ± 60ms) (both $z > 5.9$, all $p < 0.001$). Interestingly, while there was no significant difference in reaction times between neutral body and control stimuli ($z= 0.8$, $p = 0.97$) and between congruent body and control participant ($z = 2.9$, $p = 0.05$), participants were significantly faster when responding to control stimuli compared to body stimuli spatially incongruent stimuli ($z = 6.4$, $p< 0.001$).

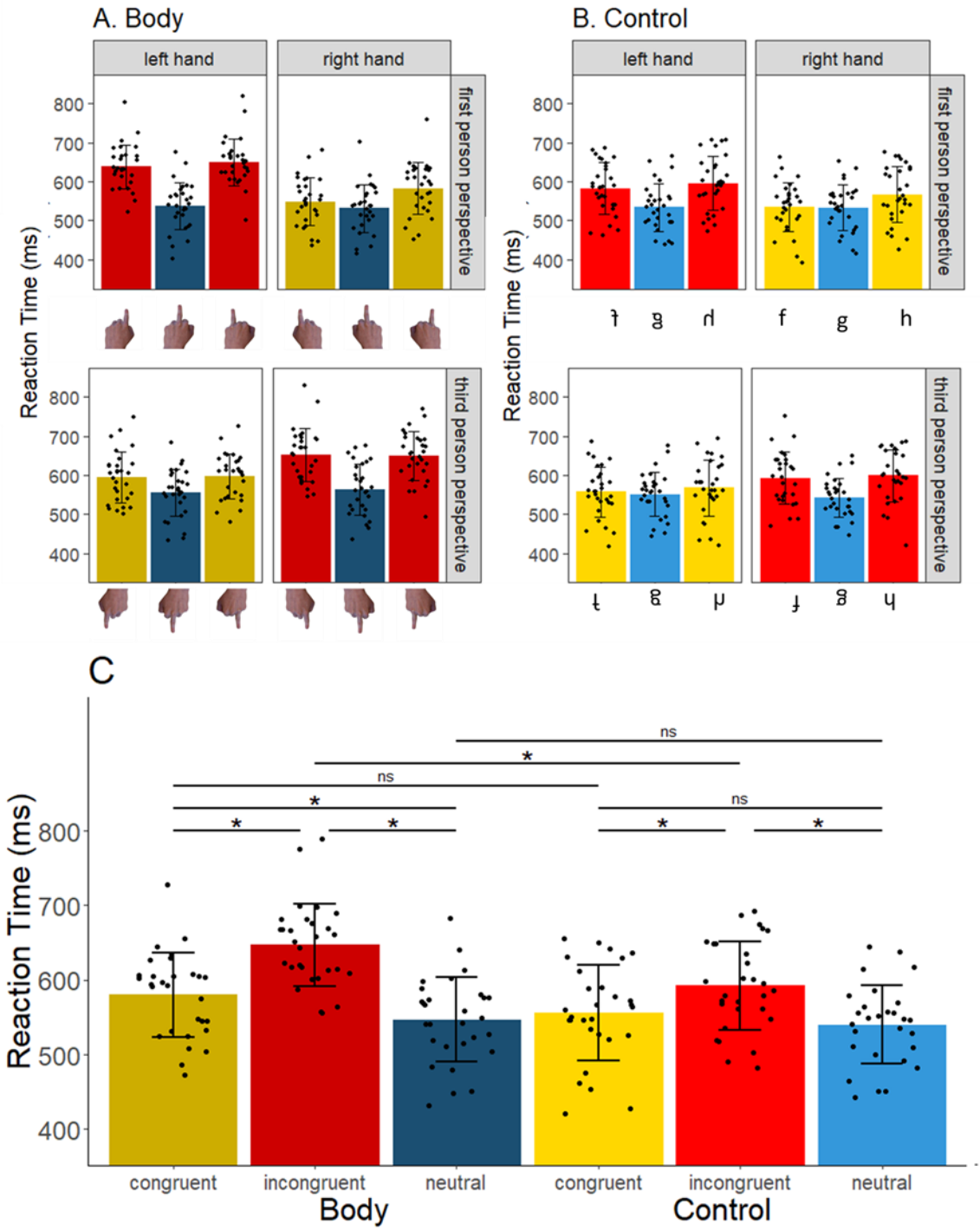

Figure 2. Reaction time analysis. For all plots, yellow bars indicate spatially congruent stimuli, red bars indicate spatially incongruent stimuli, and blue bars indicate spatially neutral stimuli. A darker shade is used for body stimuli and a lighter shade for control stimuli. Error bars represent ±1SD, each point represents a single participant. A) Mean reaction time for all body stimuli. Panels are separated according to hands (columns) and perspective (lines). B) Mean

reaction time for all control stimuli. The panels are located in the same relative position as the body stimuli whose characteristics they mimic (see A). C) Mean reaction time for congruent (index or ring finger for the right hand in first-person perspective or the left hand in third-person perspective), incongruent (index or ring finger for the left hand in first-person perspective or the right hand in third-person perspective) and neutral stimuli (all middle finger stimuli).

Error rate

Global Analysis (fig 3A,B). Results of the global analysis for error rates is presented in Table 1. Post-hoc comparison showed no significant difference in the error rate between body (10.81 ± 6.11%) and control stimuli (7.30 ± 4.54%)($z = 1.5$, $p = 0.125$). We also identified a finger effect, whereby there were less errors when answering with the middle finger (3.98 ± 3.13%) than with the index (10.06 ± 5.40%, $z = 6.1$, $p < 0.001$) and ring fingers (13.16 ± 7.58%, $z = 10.4$, $p < 0.001$), whereas the index finger had less errors than the ring finger ($z = 4.4$, $p < 0.001$). The analysis also identified several interactions (see Table 1), and the overall pattern of results were again highly consistent with effects depending on spatial congruence. We therefore conducted a specific analysis to examine this effect in greater detail.

Spatial Congruence Analysis (fig 3C). The model for spatial congruence showed a significant interaction between stimulus type and spatial congruence ($\chi^2 = 40.4$, $p < 0.001$) as well as main effects of stimulus type ($\chi^2 = 8.4$, $p < 0.001$) and spatial congruence ($\chi^2 = 149.9$, $p < 0.001$). Post-hoc comparisons for the body stimuli showed that the percentage of error was lowest for the neutral (2.85 ± 3.03%), then congruent (7.26 ± 6.53%), then incongruent (22.58 ± 12.63%) stimuli (all $z > 4.8$, all $p < 0.001$). Post hoc-analysis for the control stimuli revealed the percentage error did not differ significantly for the neutral (5.13 ± 4.31%) and congruent stimuli (5.84 ± 5.26%) ($z = 0.6$, $p = 0.988$), and both were lower than the percentage error for the incongruent stimuli (10.93 ± 7.63%) (both $z > 4.3$, both $p < 0.001$).

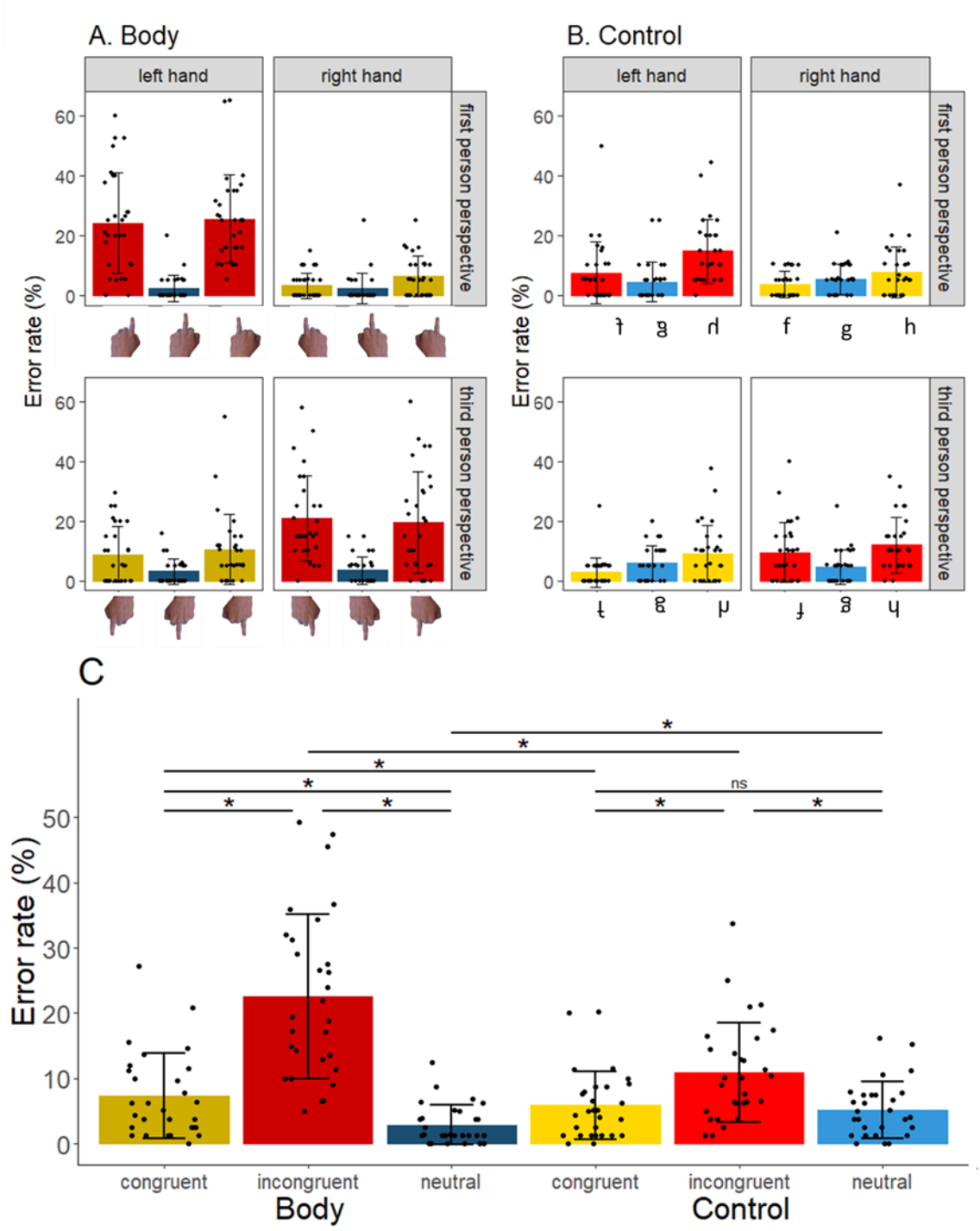

Figure 3. Error rate analysis. Bars are color coded as followed; yellow is spatially congruent stimuli, red is spatially incongruent stimuli, and blue is neutral stimuli. A darker shade is used for body stimuli and a lighter shade for control stimuli. Error bars represent ±1SD, each point represents a single participant. A) Mean error rate for all body stimuli. Panels are separated according to hands (columns) and perspective (lines). B) Mean error for all control stimuli. The

panels are located in the same relative position as the body stimuli whose characteristics they mimic (see A). C) Mean error for congruent (index or ring finger for the right hand in first-person perspective or the left hand in third-person perspective), incongruent (index or ring finger for the left hand in first-person perspective or the right hand in third-person perspective) and neutral stimuli (all middle finger stimuli). Body stimuli are on the left and control stimuli on the right.

Response type Analysis

Number of responses (fig 4A). For this model, the interaction between response type and stimulus type was significant ($\chi^2 = 141.7$, $p < 0.001$). The main effect of response type was also significant ($\chi^2 = 807.6$, $p < 0.001$) unlike the main effect of stimulus type ($\chi^2 = 0.7$, $p = 0.400$). For body stimuli, the number of correct responses (60 ± 10) was higher than the number of spatial errors (15 ± 8, $z = 14.1$, $p<0.001$) and the number of 'spatially neutral' errors (3 ± 3, $z = 21.5$, $p< 0.001$). The number of spatial errors was also higher than the number of 'spatially neutral' errors ($z = 11.3$, $p< 0.001$). For control stimuli the number of correct responses (70 ± 6) was higher than the number of spatial errors (4 ± 5, $z = 22.1$, $p< 0.001$) and the number of 'spatially neutral' errors (4 ± 3, $z = 22.2$, $p< 0.001$) but there was no difference between the number of spatial and 'spatially neutral' errors ($z = 0.2$, $p = 1$). All comparisons between corresponding body and control conditions showed less errors for control stimuli (all $z > 3.2$, all $p< 0.018$).

Reaction time (fig 4B). For the reaction times of these different types of response, the interaction between response type and stimulus type was not significant ($\chi^2 = 0.2$, $p = 0.891$). However, the main effect of response type was significant ($\chi^2 = 77.4$, $p < 0.001$), as was the main effect of stimulus type ($\chi^2 = 34.1$, $p < 0.001$). For body stimuli, the reaction time for correct responses (647 ± 55ms) did not differ to that for 'spatially neutral' errors (602 ± 111ms) ($z = 2.0$, $p = 0.337$), but was greater than the reaction time for spatial errors (546 ± 62ms) ($z =$

6.8, $p < 0.001$). For control stimuli the reaction time for correct responses was the highest (592 ± 60ms) compared to both 'spatially neutral' errors (535 ± 122ms) and spatial errors (484 ± 158ms) (both significant $z > 3.1$, $p < 0.022$).

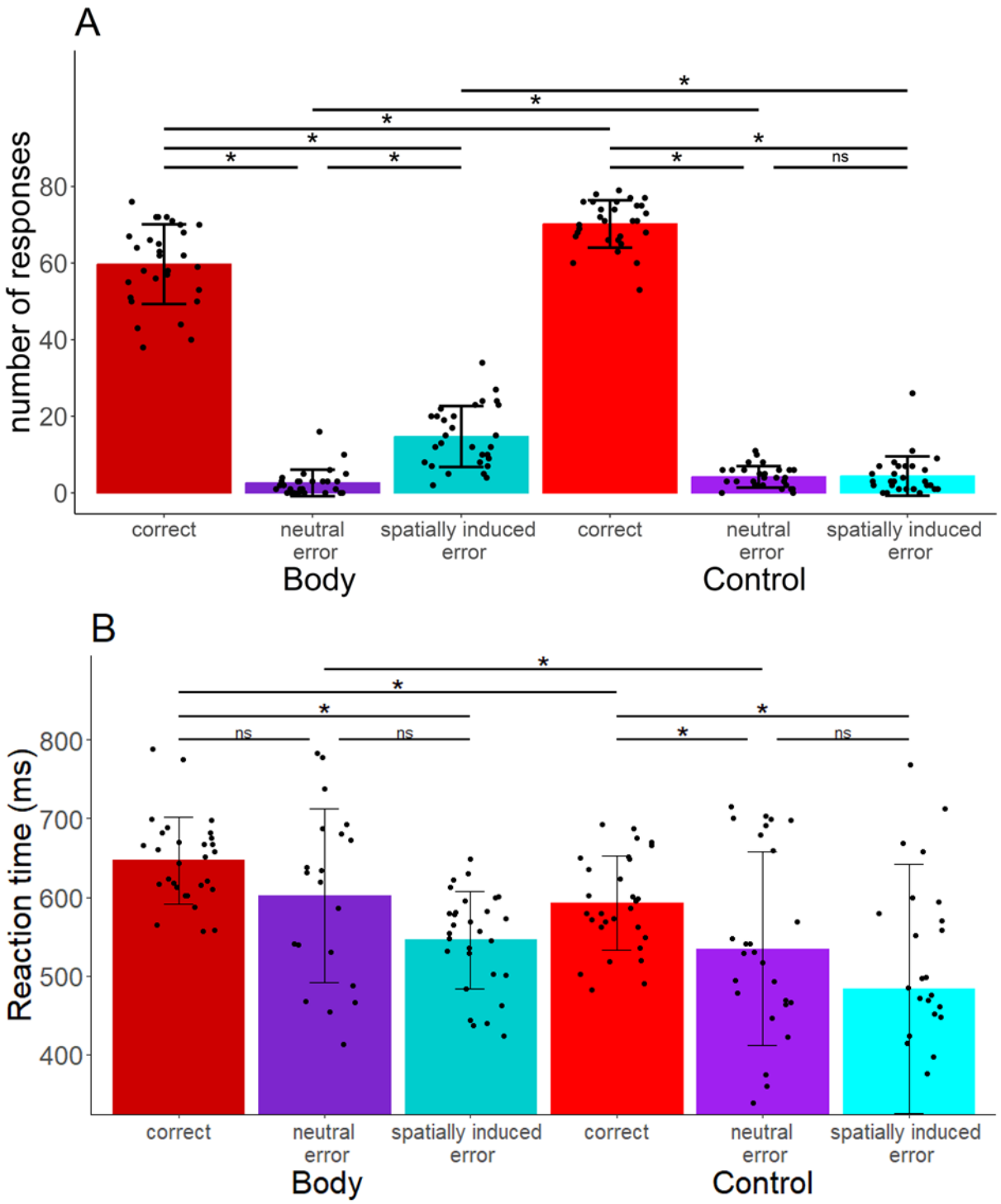

Figure 4. Error type analysis. For all plots, error bars represent SD, each point represents a participant and the color code is as followed; red for correct response, purple for spatially

neutral error and cyan for spatially induced error. A darker shade is used for body stimuli and a lighter shade for control stimuli. Error bars represent ±1SD, each point represents a single participant. A) number of response for each type of response for spatially incongruent body and control stimuli. B) Mean reaction time for each type of response for spatially incongruent body and control stimuli.

Experiment 1 Discussion

Several noteworthy results emerged from this experiment. Firstly, spatial congruence had a major impact on performance, with longer reaction times and poorer accuracy for spatially incongruent stimuli than for spatially congruent stimuli. Responses to spatially neutral stimuli were fastest with low levels of error, probably because these stimuli were always centred and therefore not subject to spatial congruence effects. Secondly, participants were generally faster and more accurate when responding to control stimuli compared to body stimuli. However, spatial congruence again modulated this effect, as there were no differences in reaction time when participants responded to control and body stimuli presented in a neutral position. Furthermore, our analysis of the errors made indicates that body stimuli were more likely to induce spatial errors, whereas there was no difference between the likelihood of 'spatial' and 'spatially neutral' errors for control stimuli. These data suggest that the spatial information from body stimuli may be more compelling than for control stimuli. Thirdly, our results indicate that perspective may be less important than would be assumed. Although an effect of perspective was present for reaction time, with faster reaction times for first-person perspective than for third-person perspective, this effect was not present for accuracy. Furthermore, when we break down the results, we see that in first-person perspective the responses are faster and more accurate for stimuli from the (spatially congruent) right hand compared with the left hand (spatially incongruent), whereas for third-person perspective, participants provided faster and more accurate responses to the (spatially

431 congruent) left hand compared to the right hand (spatially incongruent). This again suggests
432 the importance of spatial congruence.

433 The results of Experiment 1 highlight notable effects of spatial congruence during
434 action observation. They also show a conflict between spatial and anatomical information,
435 suggesting a difference in their processing. This fits well with the hypothesis of two streams
436 of vision: the dorsal stream for processing spatial information and the ventral stream for
437 anatomical recognition (Goodale & Milner, 1992). It should be noted, however, that having
438 two variables of interest (i.e. reaction time and accuracy), can make interpretation of the
439 results complex (e.g the main effect of perspective on reaction time, but not error rates). In
440 addition, the reaction time task gives us information about the 'end point' of stimulus
441 processing, but does not give us information about its time-course. Our second experiment
442 addressed these issues using a forced-response paradigm.

443 **Experiment 2**

444 Experiment 2 Method

445 The second experiment was inspired by previous studies that used a 'forced response'
446 paradigm to effectively control the time participants have available to prepare a response
447 (Haith et al., 2016; Hardwick et al., 2019; Vleugels et al., 2020) (fig 5A). This allows the
448 measurement of how accuracy changes as a function of the time that participants had to
449 prepare their response. Using this procedure in the present experiment allowed us to assess
450 the amount of time required to process different features of the presented stimuli.

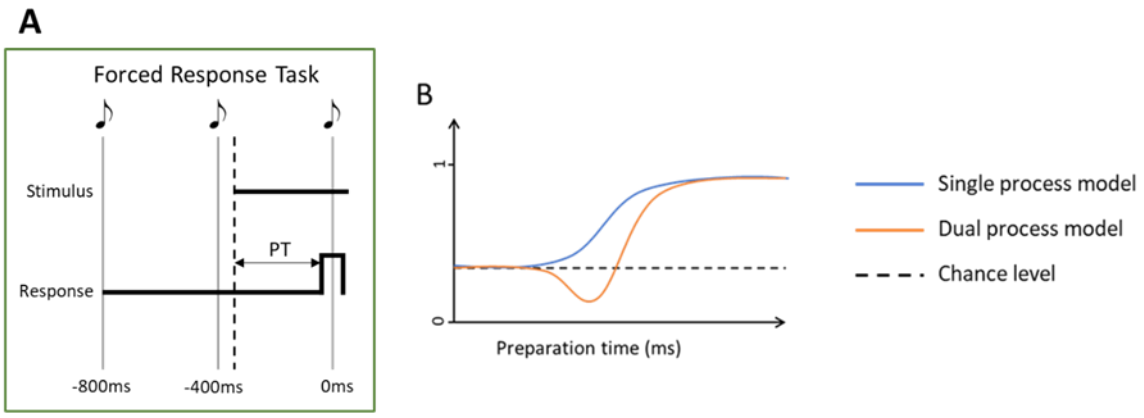

Figure 5. A. Forced Response paradigm. Participants heard a sequence of three equally spaced tones, but here the stimulus appeared at a variable time between the first and third tones. Participants were instructed to respond synchronously with the third tone. This effectively controlled the amount of time that the participant had to process the stimulus and prepare the corresponding response. The preparation time (PT) was defined as the time between the appearance of the stimulus and the participant's response. B. Models used for analysis of the forced response task. The single process model (blue) follows a sigmoidal response profile, with an initial plateau where the probability of responses begins around chance levels, before increasing to a final plateau where the probability for that response is close to 100%. In orange, the dual process model allows for situations where different responses could be produced as a function of the time available; in the situation shown here an initial decrease below chance level of the probability to produce that type of response before increasing to reach the final plateau.

Participants

A group of 57 healthy young participants were recruited via the Prolific online platform (https://www.prolific.com/). Seven participants were excluded from the analyses for failing to follow task instructions (see the exclusion criteria section for more details). This left a total of 50 participants aged between 18 and 35 years (mean = 25.5 years, SD = 4.5 years). All the participants were right handed, 12 were women, 38 were men. Two sample size calculations were performed with Gpower 3.1 based on the results of Hardwick and al., 2019 who used a similar task. The first was based on a period where the two conditions of interest differed by

0.25, based a paired sample, two tailed t-test with alpha = 0.05, power = 0.8 and an effect size (dz) of 2.38. The second followed the same principle but was based on a period with a smaller difference between the two conditions (0.07), with an effect size (dz) of 0.67. The results showed that sample sizes of more than 4 and 20 respectively would be sufficient. To ensure sufficient power in case of additional variability and to ensure that the model fitting would be of high quality, the number of participants was increased to 50. They were all naive about the purpose of the experiment. All participants gave their consent before the start of the experiment and were financially compensated (£5.50). The study was approved by the IPSY ethics committee of UCLouvain.

General Task Procedures

The task was created in PsychoPy2 version 2023.2.2 (Peirce et al., 2019) and run online via PsychoPy2's associated webpage Pavlovia (https://pavlovia.org/) (Bridges et al., 2020).

The general procedure for Experiment 2 was similar to that of Experiment 1. Participants responded to visual stimuli presented in three different conditions; familiarisation, body stimuli, and control stimuli. The stimuli and associated responses were the same as in Experiment 1.

As in Experiment 1, participants heard three short 50ms tones spaced 400ms apart on each trial. Here, however, in each trial a stimulus appeared at a variable time between the first and last tones, and participants were instructed to provide their response synchronously with the third tone (fig 5A). This allowed us to systematically vary the amount of allowed "preparation time" (i.e time between the presentation of the stimulus and the intended time of the response at the third tone; note that all analyses used the 'actual' preparation time, i.e. the time between the presentation of the stimulus and the participant's response) participants

had to process the stimulus and prepare a response; presenting the stimulus very early during the sequence of tones allowed a long preparation time, whereas presenting the stimulus very late in the trial provided a very limited preparation time. Note that the timings used included trials in which the stimulus was presented so late that the participant would not realistically have time to process it, forcing them to guess as to which response would be correct. As such, participants were expected to perform at around chance levels when allowed only very limited preparation times, whereas longer preparation times should reflect stimulus processing, allow examination of how response accuracy varied with time. To encourage participants to respond synchronously with the final tone, they were instructed to respond both at the right time and with the correct finger, but that their highest priority was to respond at the right time, even if this meant that they had to guess the correct response. Visual feedback was provided in the same manner as in Experiment 1, and additional feedback was provided to aid participants to respond with good timing. If participants responded 100ms before the third tone, the message "too early" appeared in red on the screen, and if they responded 100ms after the third tone, the message "too late" appeared in red. If they responded between -100ms and +100ms of the third tone, the message "on time" appeared in green. After the third tone there was a period of 800ms during which feedback remained on screen. If the participant provided a response during the trial, there was an interval of 200ms before the next trial. If the participant did not provide a response, the message "no key pressed" appeared on the screen for 2 seconds. As in Experiment 1, between each block, the participants had to rest for at least 10 seconds, during which the instructions were displayed before they could restart. They then had to press the three keys to check that their fingers were correctly in place.

Familiarisation Condition. Participants completed 1 block of 144 trials. Each of the three stimuli was presented at the start of the trial (i.e. 0ms, leaving 800ms of preparation

time) and then every 17ms thereafter (modelled on the most common 60Hz computer monitor refresh rate) until 799ms (i.e. leaving 1ms of preparation time) for a total of 48 different timings. The order of the trials was pseudo-randomised to ensure that the different timings were distributed throughout the block, and to limit the number of consecutive responses with the same finger to 2. The first two trials were specifically chosen to have longer preparation times to let the participants get used to the task.

Body stimuli Condition. Participants completed 4 blocks of 144 trials for a total of 576 trials. The 12 different stimuli each had one trial for each of the 48 timings ranging from 0ms to 799ms before the last tone in steps of 17ms. The trials were distributed over the 4 blocks to ensure an equivalent distribution of each stimulus and each timing. In addition, the order of trials in each block was pseudo-randomised to ensure that the different timings were distributed throughout the block, to limit up to 3 consecutive responses requiring the same button to be pressed, and up to 2 consecutive repeats of the same stimulus.

Control stimuli blocks. As with the body stimuli condition, participants completed 4 blocks of 144 trials. The distribution of trials in the different blocks and the pseudo-randomisation were the same as for the body stimuli blocks.

Protocol

Participants always began by completing the familiarisation condition, after which condition order was counterbalanced across subjects as in Experiment 1, with half the participants completing the body stimulus condition before the control stimulus condition, and the other half performing the conditions in the opposite order. The median time of completion was 53 minutes, including breaks.

Data analysis

Analyses were performed in R version 4.3.2 (R Core Team 2024).

Exclusion criteria. In the forced response task, to verify that the participants complied with the instruction to respond at the same time as the third tone, the percentage of responses given within ±100ms of the third tone was calculated for the hand and control blocks. Participants were excluded from our analyses if less than 50% of their responses across the whole experiment were outside the instructed time window. In addition, to ensure that participants were not just concentrating on the timing of their responses without ever trying to answer correctly, participants with a proportion of correct answers of less than chance (33%) throughout the whole experiment, and participants with more than 50% of their answers made using the same finger/key were excluded. As noted above, a total of 7 participants were excluded based on these criteria, leaving a final sample of 50 participants for analysis.

Data Visualization. Data were visualized using a sliding window (Haith et al., 2016; Hardwick et al., 2019; Vleugels et al., 2020), considering the proportion of correct/incorrect responses within ±50ms from each time point. We calculated 95% confidence intervals on this data using bootstrapping (1000 iterations) with replacement.

Computational Models (fig 5B). Previous work employing forced response tasks has used computational models to identify different forms of stimulus processing (Haith et al., 2016; Hardwick et al., 2019). The present study compared the ability of two possible models to explain behaviour in the forced response task. These models assume that if participants are forced to respond with very short preparation times, their response will be at random. However, once a participant has sufficient time (t – a random value taken from a gaussian distribution) they will be able to produce a specific response with a high level of accuracy.

The first “single process” model considers only a single process; as such, this model has a sigmoidal form, beginning with responses being given at random, then, as the likelihood that time t has elapsed increases, the proportion of correct responses begins to increase until it reaches a plateau at close to 100% correct responses. This model uses 4 parameters: (1) the initial plateau around responding by chance, (2) the mean and (3) standard deviation of the time t, and (4) the proportion of correct responses at the final plateau. This single process model assumes that there is no conflict between different types of information in the stimulus, allowing the stimulus to be processed as a single global piece of information.

The second “dual process” model assumes that two separate processes take place. This model assumes that after the initial plateau at chance level, an initial, prepotent response becomes available from time t_1. However, a second response, which requires more processing time to prepare, becomes available at time t_2, replacing the initially prepared response as the action that will be performed. This would capture phenomenon in which participants would initially prepare a more ‘intuitive’ response (e.g. based on previous experience or salient features of the stimulus), but later replace this with a more reasoned answer. This more complex model includes 7 parameters: (1) the initial plateau around chance levels, (2) the mean and (3) standard deviation of the time t_1, and (4) the point at which responses based on t1 reach a plateau, (5) the mean and (6) standard deviation of the time t_2, and (7) a final plateau. This model allows the possibility that there are conflicts between different types of information in the stimulus which are processed in parallel, but require differing amounts of processing time to become available to act upon.

In the present study the data from all the participants was pooled across the relevant condition, and each model was fit to this pooled data. The Akaike information criterion (AIC)

589 was calculated for each model, with comparisons of these scores being used to identify the
590 most favourable model. The code used in the present manuscript was adapted from that used
591 in previous papers using the same model (Haith et al., 2016; Hardwick et al., 2019).

592 *SMART method*. The SMART (Smoothing Method for Analysis of Response Time-
593 course) method allows investigation of the continuous time-course of the behaviour using
594 cluster-based permutation analysis, providing greater temporal precision when compared to
595 traditional binning methods (for a detailed overview and examples of the code used see (van
596 Leeuwen et al., 2019). First, the time-series of responses was calculated for each participant
597 using a gaussian kernel of width 15ms. The different time-series of the participants were then
598 average weighted across the participants. The 95% confidence intervals were then calculated
599 via the estimated standard error of the weighted mean. Finally, a cluster-based permutation
600 test was performed to determine which cluster was statistically different from a baseline
601 (obtained via the initial plateau parameter of the corresponding model fit described above) or
602 another condition. The code used was that supplied in the original publication describing the
603 procedure (van Leeuwen et al., 2019).

604 *Analyses*. Results from Experiment 1 indicate that the factor of spatial congruence had
605 important effects on participant performance. Analyses presented in Experiment 2 therefore
606 focus primarily on this factor (though note further analysis of corresponding main effects and
607 interactions are presented in the appendices).

608 *Stimulus Type and Spatial Congruence*: To examine this interaction, data were
609 categorized into one of 6 conditions based on a 2x3 design according to their stimulus type
610 (i.e. body or control) and spatial congruency (i.e. spatially congruent, incongruent and
611 neutral). Computational models were fit to each condition to establish the complexity of

612 stimulus processing, after which the SMART method was used to identify regions where
613 participant performance differed from their baseline level. Notably, Experiment 1 established
614 that participants had a preference for responding with the middle finger, presumably due to
615 its spatially neutral nature. Therefore, for each condition we used the model fit to establish
616 the baseline likelihood of a given response (i.e. the initial plateau parameter of the fastest
617 process of the model) as the baseline for comparison using the SMART method. Finally, the
618 SMART method was used to compare data for body and control stimuli across corresponding
619 spatial conditions.

620 *Response Type Analysis*: Following on from the results obtained in the first experiment,
621 we were interested in the type of responses, and more specifically the types of errors, that
622 participants produced for spatially incongruent stimuli. Specifically, if errors were induced by
623 the spatially incongruent nature of the stimuli, we would expect to see more 'spatially
624 induced' errors than spatially neutral types of errors. This analysis provided a further in-depth
625 examination of trials in which spatially incongruent stimuli were presented. Again, data were
626 split into 6 conditions based on a 2x3 design based on their stimulus type (body or control)
627 and the response provided (correct response, spatial error, spatially neutral error). Further
628 analysis steps follow the same procedure as that for analysis of stimulus type and spatial
629 congruence above.

630 *Perspective analyses*: Results from Experiment 1 indicated that effects of visual
631 perspective were secondary to those resulting from spatial congruence. We therefore
632 conducted a series of analyses to assess the presence or absence of effects of perspective
633 without separating the data in relation to spatial congruence. An "interaction" analysis split

634 the data into 4 conditions based on a 2x2 design, considering possible interactions between
635 perspective (first or third-person) and stimulus type (body, control).

636 All data and code used for the task and data analysis can be found here.

637 Experiment 2 Results

638 *Stimulus type and spatial congruence*

639 Visual inspection of the results (fig 6A) shows that at very short (less than 200ms) preparation
640 times, participants do not have time to process the stimulus and respond with approximately
641 chance levels of accuracy. However, this initial plateau varies somewhat according to the type
642 of stimulus; spatially neutral stimuli have a slightly higher initial proportion of correct
643 responses. After this initial period as the allowed preparation time increases, there is an
644 increase in the proportion of correct responses up to a plateau close to 100%. There is,
645 however, a notable exception to this pattern of results; for spatially incongruent body stimuli,
646 as the allowed preparation time increases, we initially observe a decrease in the proportion
647 of correct responses, before the increase.

648 This difference is confirmed by the model fitting results, where the simpler, single process
649 model was able to better explain the data for the majority of conditions (see Table 2);
650 however, data from body, spatially incongruent stimuli were best explained by the dual
651 process model, which could account for a decrease in accuracy below chance levels of
652 performance.

Condition	**AIC Values**		
	Single Process Model	Dual Process Model	Difference (Single - Dual)

Body	Neutral	**10154**	10174	-20
	Congruent	**11280**	11290	-10
	Incongruent	11681	**11672**	9
Control	Neutral	**10338**	10354	-16
	Congruent	**10297**	10320	-23
	Incongruent	**10543**	10563	-20

Table 2: AIC (Akaike Information Criterion) (Akaike, 1974) values for models fit to data from each condition. Bold text indicates the lower AIC score, indicating the better fit model. For difference scores, negative values indicate a better fit for the single process model, and positive values indicate a better fit for the dual process model.

We then used the SMART method to determine detailed comparisons of the time courses for each condition. In the majority of conditions, this analysis demonstrated a relationship whereby accuracy increased with the amount of time participants had to prepare their responses (for full results see Table S2). However, consistent with the computational modelling results, the SMART analysis confirmed that accuracy for body, spatially incongruent data first fell significantly below initial baseline levels from 341-458ms, later rising above the baseline from 500ms onwards.

We then made comparisons between corresponding conditions to identify whether there were any clusters during which the proportion of correct responses was different (fig 6B; for full results see Table S3). For spatially congruent stimuli, a cluster from 394ms indicated a lower final plateau in accuracy for the body stimuli. For spatially incongruent stimuli, a cluster from 334ms indicated that body stimuli showed a lower final plateau and a decrease below the baseline. Finally, we found no significant differences for spatially neutral stimuli presenting body or control images.

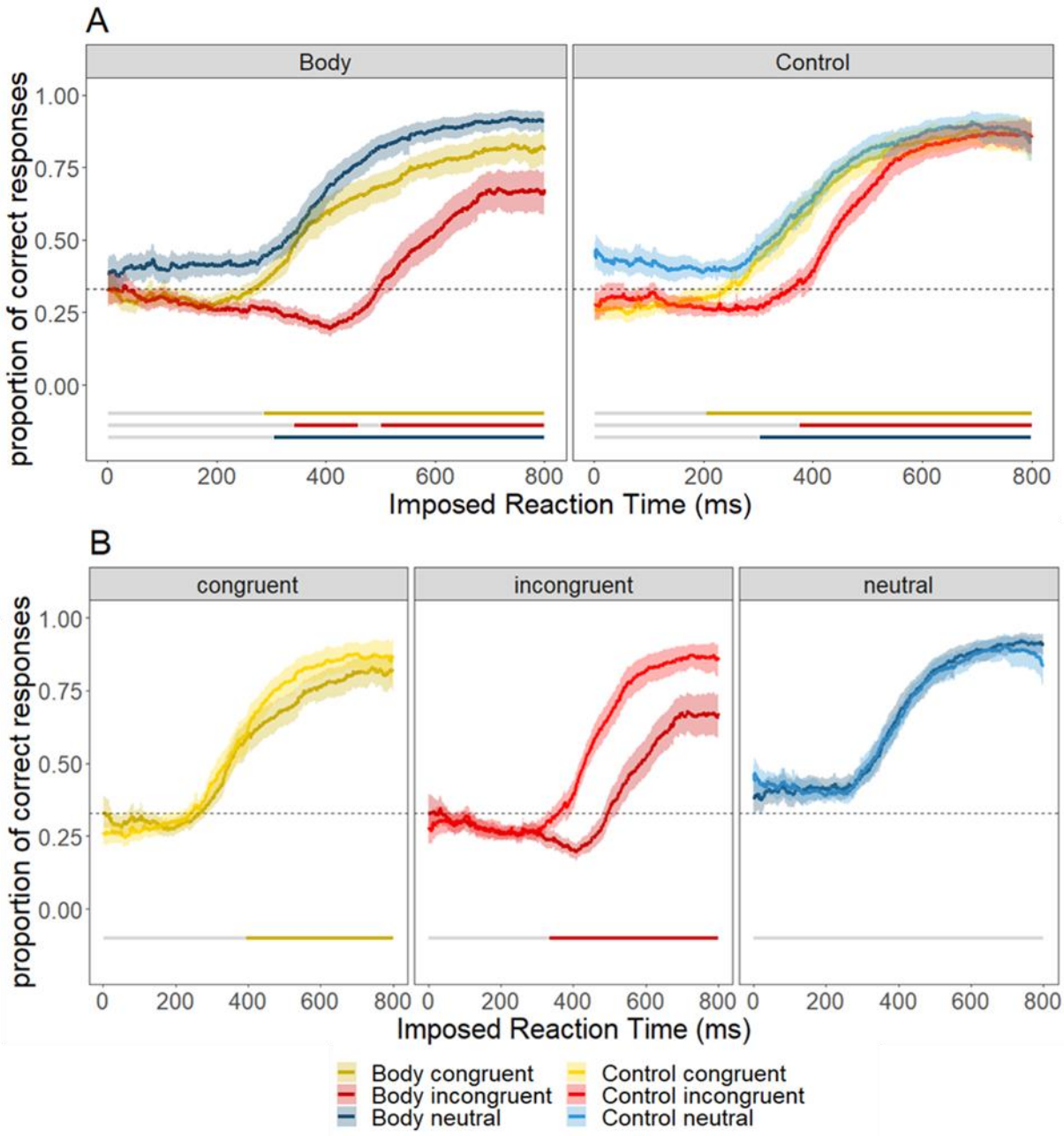

Figure 6. Sliding windows presenting the interaction between stimulus type and spatial congruence. The color code is as followed; yellow is spatially congruent stimuli, red is spatially incongruent stimuli, and blue is neutral stimuli. A darker shade of these colours is used for body stimuli and a lighter shade for control stimuli. Shaded areas present 95% confidence intervals. Lines at the bottoms of the plot indicate significant differences identified using the SMART method, either against baseline (A) or against another condition (B). A. Comparison against baseline. Evolution of the proportion of correct responses as a function of preparation time. The left panel is for body stimuli and the right is for control stimuli. B. Comparison between body and control conditions.

Response Type Analysis

When visually inspecting the data (Figure 7), we see that the line presenting the likelihood of producing a correct response is the same as the data for the spatially incongruent condition in the “stimulus type and spatial congruence analysis” (i.e. red lines in figures 6 and 7), with a decrease in the proportion of correct responses below chance level only for body stimuli. Visual inspection of the body spatial error shows us that at the same time as the decrease in proportion of correct responses, there is an increase in the proportion of spatial errors (fig 7). Although also present for the control stimuli, this effect is much smaller. The ‘spatially neutral errors’ show the same pattern for the body and control stimuli, with a period at chance level before decreasing to a plateau close to 0%.

These observations are confirmed by the model fitting results (Table 3). For body stimuli, both correct responses and spatial errors are best explained by the dual process model, whereas spatially neutral responses are best explained by the single process model. For control stimuli, only spatial errors were better explained by the dual process model.

Response Type		AIC Values		
		Single Process Model	Dual Process Model	Difference (Single - Dual)
Body	correct	11681	**11672**	9
	spatial	12309	**11991**	318
	Spatially neutral	**9585**	9609	-24
Control	correct	**10543**	10563	-20
	spatial	8623	**8609**	14
	Spatially neutral	**9864**	9892	-28

Table 3: AIC values for models fit to data from each condition. Bold text indicates the lower AIC score, indicating the better fit model. For difference scores, negative values indicate a better fit for the single process model, and positive values indicate a better fit for the dual process model.

When applying the SMART method (table S2), starting with body stimuli, correct responses first dipped significantly below baseline from 341 to 458ms, before rising significantly above baseline from 500ms onward. By contrast, spatial errors rose significantly above baseline from 307ms to 615ms, before falling significantly below baseline from 730ms onwards. In comparison, spatially neutral errors followed a simpler trajectory, dropping significantly below baseline from 310ms onwards, and remaining low thereafter. For the control stimuli, in contrast to the body stimuli, correct responses never fell below baseline levels, and instead increasingly only significantly above baseline from 374ms. By contrast, spatial errors rose significantly above baseline from 282ms to 383ms, before falling significantly below baseline levels from 429ms. Finally, spatially neutral errors fell significantly below baseline from 303ms.

We then made comparisons between corresponding conditions (fig 7B)(table S3). For correct responses, while the curves were similar, body responses fell significantly below control

stimuli from 334ms onwards, with a slower inflection and a lower final plateau. We found a corresponding effect on spatial errors; again, the curves had similar shapes, but differed from 355ms onwards, with a greater likelihood of errors for body compared to control stimuli. Finally, comparison of the curves for spatially neutral errors found no differences between the body and control conditions.

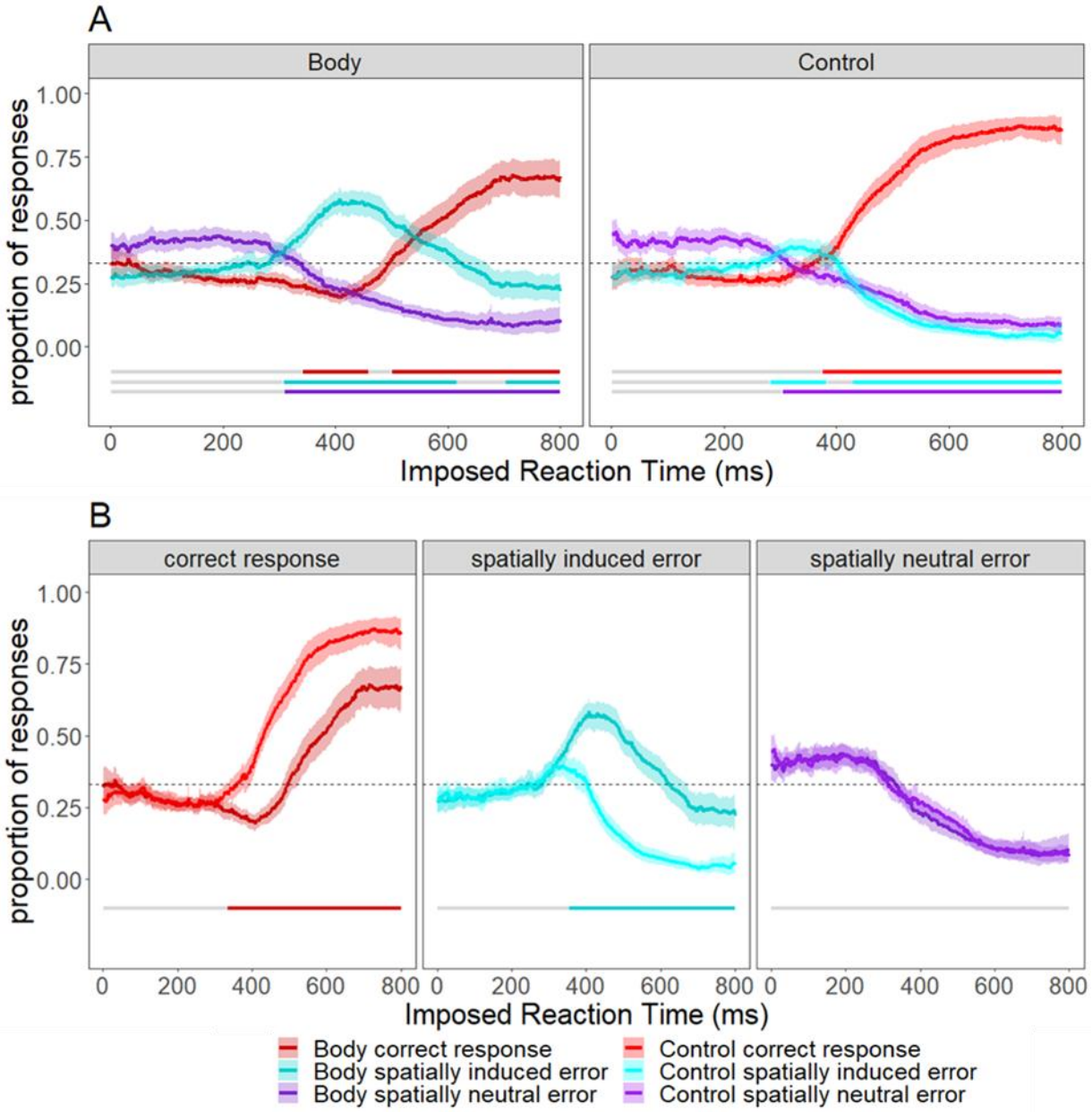

Figure 7. Sliding windows presenting the interaction between stimulus type and type of responses to spatially incongruent stimuli. The color code is as followed; red for correct response, purple for spatially neutral error, and cyan for spatial error. A darker shade of these

colours is used for body stimuli and a lighter shade for control stimuli. Shaded areas present 95% confidence intervals. Lines at the bottoms of the plot indicate significant differences identified using the SMART method, either against baseline (A) or against another condition (B). A. Comparison against baseline. Evolution of the probability of a given type of responses as a function of preparation time. B. Comparison between body and control for each type of response.

Perspective analysis

Visual inspection of the data indicated that all data followed a generally sigmoidal shape, with accuracy generally rising more slowly for body than control stimuli. Model fitting identified that for all conditions, the single process model was the most favourable (all AIC diff < -17)(table S1-2). Using the SMART method, we found that the proportion of correct responses increased significantly above baseline for all conditions (see Table S3), though the rise occurred more slowly for body compared to control stimuli. Notably, when using the SMART method to directly examine possible effects of perspective, we found no significant differences between first- and third-person stimuli for the body condition, nor were there any differences when comparing the corresponding stimuli in the control condition.

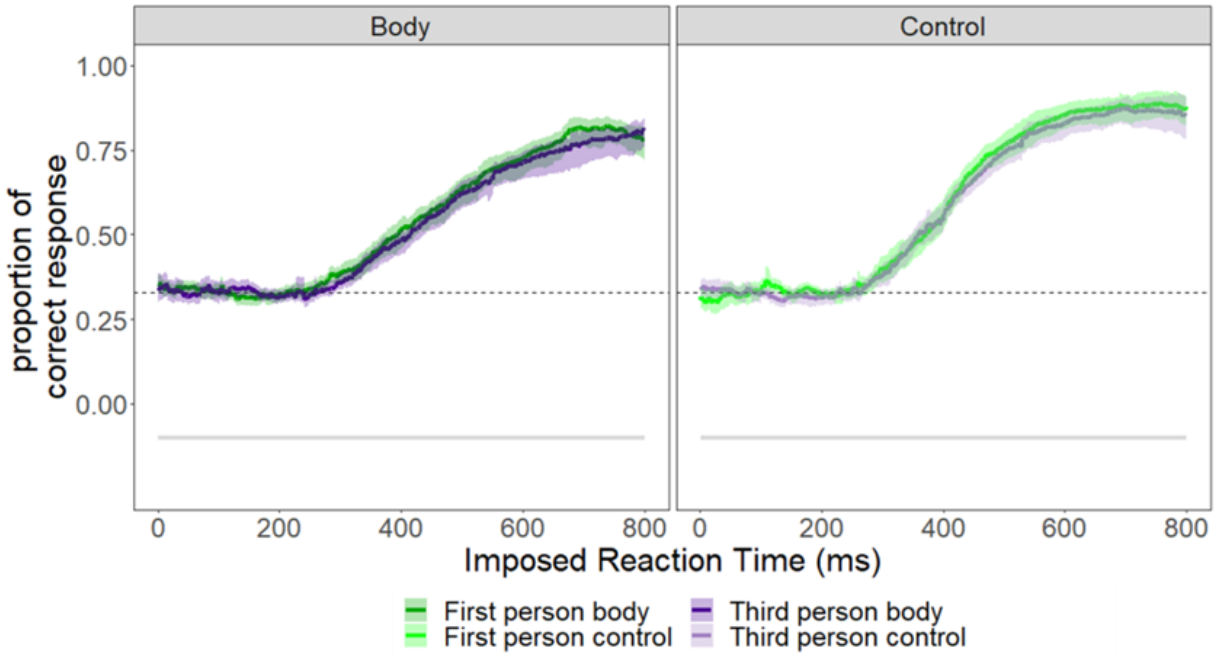

Figure 8. Sliding window visualizations of the data in relation to visual perspective for body and control stimuli. In beige is the evolution for first-person perspective stimuli. In green for third-person perspective stimuli. The lines at the bottom of the plot are coloured when there is a statistical difference between the 2 conditions; note here there were no significant differences across the whole timescale measured for either type of stimuli.

Experiment 2 Discussion

Several noteworthy results can be highlighted in this experiment. Firstly, as in Experiment 1, we can see a major effect of spatial congruence with, in particular, a difference in behaviour for spatially incongruent body stimuli. Indeed, these stimuli are better explained by a dual-process model, which can account for an initial decrease in the proportion of correct responses before an increase. These results are consistent with participants making incorrect responses based on *where* the stimulus is presented, rather than *what* it presents. This is supported by the type of response participants made to spatially incongruent stimuli; as the proportion of correct responses decreases, there is a corresponding increase in spatial errors. Interestingly, this pattern of results was markedly stronger for body stimuli. Indeed, for control

753 stimuli, there was no decrease in the proportion of correct responses, and the increase in
754 spatially induced responses, although significant, was much smaller than for the body stimuli.
755 These results, which are in line with those of Experiment 1, therefore show a difference
756 between the processing of body and control stimuli, and more particularly of spatial
757 information. Finally, while Experiment 1 identified a weak effect of perspective, in Experiment
758 2 there were no significant differences between first- and third-person perspective for either
759 body or control stimuli. This result again emphasizes that the importance of perspective during
760 action observation may be overestimated when compared to spatial congruence.

761 **General Discussion**

762 The present study characterized the role of the factors of stimulus type (body or control),
763 spatial congruence, and perspective (first- and third-person) during action observation.
764 Experiment 1 used a classical reaction time paradigm to examine how participants responded
765 to these factors. Experiment 2 used a forced response paradigm to examine the time-course
766 of stimulus processing, and how the effects of these factors varied depending on the allowed
767 processing time.

768 Results across the two experiments indicate that spatial congruence has a major impact on
769 the observation of actions, as introducing conflicts between the required response and the
770 spatial location of the stimulus led to a significant increase in incorrect responses.
771 Interestingly, this effect was stronger in body compared to control images, consistent with a
772 difference in the processing of these two types of stimuli. Furthermore, in contrast with
773 previous reports, we found little difference for first-person perspective over third-person
774 perspective, and our results suggest such effects could be mostly driven by spatial congruence.

775 Stronger spatial congruence effects for body stimuli

The present experiments indicate that the spatial congruence of the stimulus has a major impact on participants' performance. Interestingly, this effect was more pronounced for body-related stimuli than for control stimuli. In particular, results of Experiment 2 indicate that participants initially respond on the basis of the stimulus' spatial information, even if this is irrelevant for this task, before this initial automatic response is replaced by a response based on anatomical recognition. It therefore seems that there are different timescales of information processing for spatial and anatomical information. Interestingly, these effects were smaller or absent for control stimuli, consistent with previous work proposing differences in their processing (for review see (Gowen & Poliakoff, 2012)).

These results are consistent with the ‘dual visual streams hypothesis’ (Goodale & Milner, 1992), which proposes that spatial processing ('where’) and object recognition (‘what’) are respectively processed by separate dorsal and ventral visual steams. In particular, the dorsal stream is known for its role in visually guided actions, which require rapid visual processing to allow us to quickly adapt movements according to visual feedback (Króliczak et al., 2008). By contrast, the ventral stream processes information related to visual form, and does not have the same level of immediacy. Thus, in the context of Experiment 2, which examined the time-course of information processing, spatial information processed via the dorsal stream would be available to influence participant responses earlier than anatomical information processed via the ventral stream. The differences observed between the body and control stimuli could be explained by three non-exclusive proposals. The first is that the dorsal stream is particularly sensitive to body-related stimuli. Motor simulation theory proposes that action observation activates the same neural circuitry as action execution (Jeannerod, 2001), and since the dorsal stream is known for its role in visuomotor control (Goodale & Milner, 1992), it is possible that it is more strongly attuned to body stimuli. In addition, the dorsal stream contains areas such

800 as the inferior parietal lobule (Freud et al., 2016), known to contain mirror neurons that are
801 activated during observation of action, and thus in the present experiment would be recruited
802 only by our body stimuli. Increased activation of these regions by observation of body stimuli
803 could in turn make the information processed by the dorsal stream more compelling. The
804 second is that the dorsal stream passes information to premotor areas that are activated
805 during action observation (Goodale & Milner, 1992), and this activation could be necessary to
806 induce this response based on spatial information. The third is that the differences could be
807 explained by the choice of control stimuli (see strengths and limitations section for more
808 details on the choice of stimuli). In fact, for control stimuli, the letter is represented on its own
809 without being linked to any other reference element, which is not the case for body stimuli
810 where other information is present (i.e. the hand). These differences between the stimuli
811 could also partly explain the results obtained.

812 Effect of perspective

813 The results suggest that perspective has a relatively minor effect on action priming, and that
814 effects of perspective may in fact mostly be driven by spatial congruence. Effects of
815 perspective in Experiment 1 were limited only to reaction times, which were faster for first-
816 than third-person perspective, with differences that were smaller in magnitude than those
817 identified for spatial congruence. Similarly, no effects of perspective were identified in
818 Experiment 2. These results suggest the effects of perspective described in the literature so
819 far could be partially or even mostly driven by spatial congruence. Indeed, several studies on
820 perspective have compared right hands in first-person perspective with right hands in third-
821 person perspective (Angelini et al., 2021; Ge et al., 2018; Oosterhof et al., 2012);our present
822 results suggest that the differences observed in such experiments may be driven more by a

823 change in spatial congruence than by a change in perspective. This interpretation is in
824 agreement with other results (Avikainen et al., 2003; Bekkering et al., 2000; Chiavarino et al.,
825 2007; Shmuelof & Zohary, 2008; Wohlschläger & Bekkering, 2002) that show an interaction
826 between perspective and the identity of the hand observed. It therefore seems important that
827 future studies on the effects of perspective during action observation choose their stimuli
828 carefully. An optimal solution would be to compare both right and left hands in first- and third-
829 person perspective, but in cases where the overall number of trials needs to be limited,
830 comparing a right hand in a first-person perspective with a left hand in a third-person
831 perspective would allow 'true' comparisons between perspectives that are not potentially
832 confounded by differences in spatial congruence.

833 Effect of stimulus type

834 Results of the present study indicate that participants were generally faster and more accurate
835 when responding to control stimuli compared to body stimuli. We note, however, that the
836 main effect of stimulus type was in fact likely driven by the interaction between stimulus type
837 and spatial congruence, as we identified minimal differences in responses to spatially neutral
838 body and control stimuli in Experiment 1, and no differences in Experiment 2. These results go
839 against the current literature, which shows a tendency towards a preference for body stimuli,
840 or a status quo between the two (Gowen & Poliakoff, 2012). One possible explanation is that,
841 as proposed by Gowen and Poliakoff, the priming effect induced by different types of stimuli
842 depends mainly on their characteristics (e.g. saliency, size) and not on whether they are "body-
843 related" in nature. It is therefore possible that the low-level properties of the stimuli were one
844 of the causes of the preference for control stimuli. Specifically, even though the position of
845 the relevant part of the stimulus to respond to (i.e. the letter vs the finger) was presented in

a consistent manner across our body and control stimuli, the body stimuli presented more overall information (i.e. not only the relevant finger, but also the rest of the hand) than the control stimuli (which showed only the relevant letter on an otherwise blank screen). We note that the selection of control stimuli for action observation studies is a complex issue, (see for example (Loporto et al., 2011)) and provide further discussion of this point in the 'strengths and limitations' section below. (Holahan et al., 1978)

Strengths and limitations

One of the main strengths of this paper is the information provided by the forced response task. Without it, we would not have been able to observe the differences in information processing, particularly between spatially incongruent body and control stimuli. The use of this paradigm therefore enabled us to obtain important new temporal information. In addition, analysis using the SMART method meant that we did not lose the precision of the temporal information that would have been the case with binning analysis (van Leeuwen et al., 2019).

The choice of suitable control stimuli is a difficult one as there are an infinite number of possibilities. The approach used in the present study allowed us to maintain the position of the most relevant information of interest across equivalent stimuli. This approach appears to have been somewhat successful, as responses to spatially neutral stimuli did not differ for any of our comparisons between body-related and control stimuli. However, it is not possible to generalise our results to all possible control stimuli. Certain parameters, such as the salience of the stimuli, may not have been perfectly matched up with their body-related counterparts. A limitation of the stimuli we have chosen is the fact that for body stimuli, the fingers are represented in relation to the hand, which is not the case with the control stimuli, where the

letters stand alone. It is possible that the difference in the effect of spatial congruence between body and control stimuli is partly influenced by this. Similarly, low-level perceptual properties of the stimuli (e.g. the number of white/coloured pixels) differed between our stimuli. However, we again note that there were no significant differences between spatially neutral stimuli in the body-related and control conditions, suggesting that reaction times were mainly influenced by other factors. In addition, the body stimuli represented raised fingers, while the participants had to lower theirs to respond. This could have the effect of globally increasing the reaction time for the body stimuli, as the stimuli are imitatively incompatible with the movement the participants have to perform. Critically, the limitations outlined above do not undermine our main findings, which highlight an effect of spatial congruence on body-related stimuli, which will still be correct even if the control results are not generalizable. Nonetheless, further studies with different types of control stimuli are needed to test this generalisation.

For the body stimuli, we used static stimuli. There were two main reasons for this choice. Firstly, the aim of the forced response task was to have precise control over the time participants had to prepare a response and process the stimuli. The use of static stimuli allowed more precise timing control than the use of video. Secondly, previous studies have already used static stimuli in the context of action observation. For these reasons, static stimuli seemed the most appropriate. However, future studies are needed to confirm that these results can be generalized to video viewing.

Finally, the results of Experiment 2 suggest that a ‘finger effect’ was present in the forced response task. At very short reaction times (approximately <200ms), if responses were given truly randomly, we would expect the proportion of responses with the index, middle and ring

892 fingers should be the same (i.e. 33%). However, we found that around 40% of responses were
893 with the middle finger and 30% with the index and ring fingers. This preference could be
894 explained by the fact that these stimuli never presented any conflict between spatial and
895 anatomical information for responses with the middle finger. Future studies using similar
896 methodology could ask whether this represents a conscious strategy by interviewing
897 participants during debriefing. This preference has the effect of modifying the baseline when
898 the stimuli requiring a response with the middle finger are separated from the others (thus in
899 the analysis of spatial congruence) with a higher baseline for neutral stimuli than for
900 congruent or incongruent stimuli. The possible impact that this effect could have is minimised
901 (or even eliminated) thanks to our analysis method, which allows us to calculate the baseline
902 specific to each condition, and then compare how the proportion corresponding responses
903 varies in relation to this baseline (note that with SMART method, we only made pairwise
904 comparisons between conditions with similar baselines).

905 It is important to note that the results of the present study are exclusively at the behavioural
906 level; proposals of the neural mechanisms behind the effects on reaction time and accuracy,
907 while based on results from the previous literature, are therefore speculative at this stage.
908 Future experiments could explore the neural correlates of these effects in further detail, for
909 example, by using repetitive Transcranial Magnetic Stimulation (TMS) to disrupt areas of the
910 dorsal and ventral visual streams, and see how this affects performance in the task. In
911 addition, given the time-dependent nature of the present results, using techniques such as
912 TMS to test how corticospinal excitability is modulated by different stimuli at different timings,
913 or electroencephalography to examine changes in evoked potentials at different stages of
914 stimulus processing, present interesting avenues for future research

Conclusion

These experiments show that spatial congruence has a major impact on action observation, with longer reaction times and more errors when there is a conflict between the response requested and the spatial position of the stimulus. This phenomenon can be explained by the fact that spatial information is processed more quickly than anatomical information, thus inducing a first automatic response based on spatial information that may be incorrect. Interestingly, this phenomenon is more pronounced for body-related than control stimuli, which is consistent with a difference in processing between these two types of stimuli. Furthermore, we found little difference between first- and third-person perspective. Our results also suggest that the perspective effect previously identified in the literature may be driven mainly by spatial congruence. We believe that a better understanding of how stimuli are processed during action observation is important for future applications. For example, future research into the use of action observation for rehabilitation after stroke may consider optimizing the spatial compatibility of the stimulus and the observer, as this seems to be optimal for participant performance.

Data availability statement

All data and code used for the task and data analysis can be found here:

https://osf.io/q2rd8/?view_only=e492221e800a4d78b44040910e10985b

Author Contributions (CRediT)

Author	Conceptualization	Methodology	Software	Validation	Formal Analysis	Investigation	Resources	Data Curation	Writing – original draft	Writing – reviewing & editing	Visualization	Supervision	Project Administration	Funding Acquisition
Baptiste M. Waltzing	x	x	x	x	x	x	x	x	x	x	x			
Siobhan McAteer			x		x					x				
Marcos Moreno-Verdú			x		x					x				
Elise E. Van Caenegem										x				
Yue Du			x											
Robert M. Hardwick	x	x	x	x	x		x		x	x		x	x	x

Funding

BMW, SMcA, MMV, and RMH are supported by an FNRS MIS Grant (FNRS F.4523.23). RMH is supported by an FNRS CDR award (FNRS J.0084.21). MMV is supported by an FNRS CR Fellowship (FNRS 1.B359.25). EVC is supported by an FNRS ASP fellowship (FNRS 1.AB19.24). YD is supported by an NSF grant (2218406).

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance.

Acknowledgments

None

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. IEEE Transactions on Automatic Control. https://doi.org/10.1109/TAC.1974.1100705

Angelini, M., Del Vecchio, M., Lopomo, N. F., Gobbo, M., & Avanzini, P. (2021). Perspective-dependent activation of frontoparietal circuits during the observation of a static body effector. *Brain Research*, *1769*, 147604. https://doi.org/10.1016/j.brainres.2021.147604

Avikainen, S., Wohlschläger, A., Liuhanen, S., Hänninen, R., & Hari, R. (2003). Impaired Mirror-Image Imitation in Asperger and High-Functioning Autistic Subjects. *Current Biology*, *13*(4), 339–341. https://doi.org/10.1016/S0960-9822(03)00087-3

Bekkering, H., Wohlschläger, A., & Gattis, M. (2000). Imitation of gestures in children is goal-directed. *Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, *53*(1), 153–164. Scopus. https://doi.org/10.1080/713755872

Biermann-Ruben, K., Jonas, M., Kessler, K., Siebner, H. R., Bäumer, T., Schnitzler, A., & Münchau, A. (2008). Observing repetitive finger movements modulates response times of auditorily cued finger movements. *Brain and Cognition*, *68*(1), 107–113. https://doi.org/10.1016/j.bandc.2008.03.005

Blakemore, S.-J., & Frith, C. (2005). The role of motor contagion in the prediction of action. *Neuropsychologia*, *43*(2), 260–267. https://doi.org/10.1016/j.neuropsychologia.2004.11.012

Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica*, *106*(1), 3–22. https://doi.org/10.1016/S0001-6918(00)00024-X

Bridges, D., Pitiot, A., MacAskill, M. R., & Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ*, *8*, e9414. https://doi.org/10.7717/peerj.9414

Caligiore, D., Mustile, M., Spalletta, G., & Baldassarre, G. (2017). Action observation and motor imagery for rehabilitation in Parkinson's disease: A systematic review and an integrative hypothesis. *Neuroscience and Biobehavioral Reviews*, *72*, 210–222. https://doi.org/10.1016/j.neubiorev.2016.11.005

Catmur, C., & Heyes, C. (2011). Time course analyses confirm independence of imitative and spatial compatibility. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(2), 409–421. https://doi.org/10.1037/a0019325

Chiavarino, C., Apperly, I. A., & Humphreys, G. W. (2007). Exploring the functional and anatomical bases of mirror-image and anatomical imitation: The role of the frontal lobes. *Neuropsychologia*, *45*(4), 784–795. https://doi.org/10.1016/j.neuropsychologia.2006.08.007

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences, Rev. Ed* (pp. xv, 474). Lawrence Erlbaum Associates, Inc.

Curioni, A., Knoblich, G. K., Sebanz, N., & Sacheli, L. M. (2020). The engaging nature of interactive gestures. *PloS One*, *15*(4), e0232128. https://doi.org/10.1371/journal.pone.0232128

di Pellegrino, G., Fadiga, L., Fogassi, L., Gallese, V., & Rizzolatti, G. (1992). Understanding motor events: A neurophysiological study. *Experimental Brain Research*, *91*(1), 176–180. https://doi.org/10.1007/BF00230027

Edwards, M. G., Humphreys, G. W., & Castiello, U. (2003). Motor facilitation following action observation: A behavioural study in prehensile action. *Brain and Cognition*, *53*(3), 495–502. https://doi.org/10.1016/S0278-2626(03)00210-0

Faber, T. W., van Elk, M., & Jonas, K. J. (2016). Complementary Hand Responses Occur in Both Peri- and Extrapersonal Space. *PloS One*, *11*(4), e0154457. https://doi.org/10.1371/journal.pone.0154457

Fadiga, L., Fogassi, L., Pavesi, G., & Rizzolatti, G. (1995). Motor facilitation during action observation: A magnetic stimulation study. *Journal of Neurophysiology*, *73*(6), 2608–2611. https://doi.org/10.1152/jn.1995.73.6.2608

Freud, E., Plaut, D. C., & Behrmann, M. (2016). 'What' Is Happening in the Dorsal Visual Pathway. *Trends in Cognitive Sciences*, *20*(10), 773–784. https://doi.org/10.1016/j.tics.2016.08.003

Gatti, R., Tettamanti, A., Gough, P. M., Riboldi, E., Marinoni, L., & Buccino, G. (2013). Action observation versus motor imagery in learning a complex motor task: A short review of literature and a kinematics study. *Neuroscience Letters*, *540*, 37–42. https://doi.org/10.1016/j.neulet.2012.11.039

Ge, S., Liu, H., Lin, P., Gao, J., Xiao, C., & Li, Z. (2018). Neural Basis of Action Observation and Understanding From First- and Third-Person Perspectives: An fMRI Study. *Frontiers in Behavioral Neuroscience*, *12*, 283. https://doi.org/10.3389/fnbeh.2018.00283

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*(1), 20–25. https://doi.org/10.1016/0166-2236(92)90344-8

Gowen, E., Bradshaw, C., Galpin, A., Lawrence, A., & Poliakoff, E. (2010). Exploring visuomotor priming following biological and non-biological stimuli. *Brain and Cognition*, *74*(3), 288–297. https://doi.org/10.1016/j.bandc.2010.08.010

Gowen, E., & Poliakoff, E. (2012). How does visuomotor priming differ for biological and non-biological stimuli? A review of the evidence. *Psychological Research*, *76*(4), 407–420. https://doi.org/10.1007/s00426-011-0389-5

Haith, A. M., Pakpoor, J., & Krakauer, J. W. (2016). Independence of Movement Preparation and Movement Initiation. *Journal of Neuroscience*, *36*(10), 3007–3015. https://doi.org/10.1523/JNEUROSCI.3245-15.2016

Hardwick, R. M., Caspers, S., Eickhoff, S. B., & Swinnen, S. P. (2018). Neural correlates of action: Comparing meta-analyses of imagery, observation, and execution. *Neuroscience and Biobehavioral Reviews*, *94*, 31–44. https://doi.org/10.1016/j.neubiorev.2018.08.003

Hardwick, R. M., & Edwards, M. G. (2011). Observed reach trajectory influences executed reach kinematics in prehension. *Quarterly Journal of Experimental Psychology*, *64*(6), 1082–1093. https://doi.org/10.1080/17470218.2010.538068

Hardwick, R. M., & Edwards, M. G. (2012). Motor interference and facilitation arising from observed movement kinematics. *Quarterly Journal of Experimental Psychology*, *65*(5), 840–847. https://doi.org/10.1080/17470218.2012.672995

Hardwick, R. M., Forrence, A. D., Krakauer, J. W., & Haith, A. M. (2019). Time-dependent competition between goal-directed and habitual response preparation. *Nature Human Behaviour*, *3*(12), Article 12. https://doi.org/10.1038/s41562-019-0725-0

Holahan, C. J., Culler, R. E., & Wilcox, B. L. (1978). Effects of Visual Distraction on Reaction Time in a Simulated Traffic Environment. *Human Factors*, *20*(4), 409–413. https://doi.org/10.1177/001872087802000404

Jeannerod, M. (2001). Neural simulation of action: A unifying mechanism for motor cognition. *NeuroImage*, *14*(1 Pt 2), S103-109. https://doi.org/10.1006/nimg.2001.0832

Kemmerer, D. (2021). What modulates the Mirror Neuron System during action observation?: Multiple factors involving the action, the actor, the observer, the relationship between actor and observer, and the context. *Progress in Neurobiology*, *205*, 102128. https://doi.org/10.1016/j.pneurobio.2021.102128

Króliczak, G., McAdam, T. D., Quinlan, D. J., & Culham, J. C. (2008). The Human Dorsal Stream Adapts to Real Actions and 3D Shape Processing: A Functional Magnetic Resonance Imaging Study. *Journal of Neurophysiology*, *100*(5), 2627–2639. https://doi.org/10.1152/jn.01376.2007

López, N. D., Monge Pereira, E., Centeno, E. J., & Miangolarra Page, J. C. (2019). Motor imagery as a complementary technique for functional recovery after stroke: A systematic review. *Topics in Stroke Rehabilitation*, *26*(8), 576–587. https://doi.org/10.1080/10749357.2019.1640000

Loporto, M., McAllister, C., Williams, J., Hardwick, R., & Holmes, P. (2011). Investigating central mechanisms underlying the effects of action observation and imagery

through transcranial magnetic stimulation. *Journal of Motor Behavior*, *43*(5), 361–373. https://doi.org/10.1080/00222895.2011.604655

Machado, T. C., Carregosa, A. A., Santos, M. S., Ribeiro, N. M. da S., & Melo, A. (2019). Efficacy of motor imagery additional to motor-based therapy in the recovery of motor function of the upper limb in post-stroke individuals: A systematic review. *Topics in Stroke Rehabilitation*, *26*(7), 548–553. https://doi.org/10.1080/10749357.2019.1627716

Mibu, A., Kan, S., Nishigami, T., Fujino, Y., & Shibata, M. (2020). Performing the hand laterality judgement task does not necessarily require motor imagery. *Scientific Reports*, *10*(1), 5155. https://doi.org/10.1038/s41598-020-61937-9

Moreno-Verdú, M., Hamoline, G., Van Caenegem, E. E., Waltzing, B. M., Forest, S., Valappil, A. C., Khan, A. H., Chye, S., Esselaar, M., Campbell, M. J., McAllister, C. J., Kraeutner, S. N., Poliakoff, E., Frank, C., Eaves, D. L., Wakefield, C., Boe, S. G., Holmes, P. S., Bruton, A. M., … Hardwick, R. M. (2024). Guidelines for reporting action simulation studies (GRASS): Proposals to improve reporting of research in motor imagery and action observation. *Neuropsychologia*, *192*, 108733. https://doi.org/10.1016/j.neuropsychologia.2023.108733

Nicholson, V., Watts, N., Chani, Y., & Keogh, J. W. (2019). Motor imagery training improves balance and mobility outcomes in older adults: A systematic review. *Journal of Physiotherapy*, *65*(4), 200–207. https://doi.org/10.1016/j.jphys.2019.08.007

Oosterhof, N. N., Tipper, S. P., & Downing, P. E. (2012). Viewpoint (In)dependence of Action Representations: An MVPA Study. *Journal of Cognitive Neuroscience*, *24*(4), 975–989. https://doi.org/10.1162/jocn_a_00195

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Press, C., Cook, J., Blakemore, S.-J., & Kilner, J. (2011). Dynamic modulation of human motor activity when observing actions. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(8), 2792–2800. https://doi.org/10.1523/JNEUROSCI.1595-10.2011

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Brain Research. Cognitive Brain Research*, *3*(2), 131–141. https://doi.org/10.1016/0926-6410(95)00038-0

Scandola, M., & Tidoni, E. (2024). Reliability and Feasibility of Linear Mixed Models in Fully Crossed Experimental Designs. *Advances in Methods and Practices in Psychological Science*, *7*(1), 25152459231214454. https://doi.org/10.1177/25152459231214454

Schouten, J. F., & Bekker, J. A. M. (1967). Reaction time and accuracy. *Acta Psychologica*, *27*, 143–153. https://doi.org/10.1016/0001-6918(67)90054-6

Shmuelof, L., & Zohary, E. (2008). Mirror-image representation of action in the anterior parietal cortex. *Nature Neuroscience*, *11*(11), 1267–1269. https://doi.org/10.1038/nn.2196

Simon, J. R., & Wolf, J. D. (1963). Choice Reaction Time as a Function of Angular Stimulus-Response Correspondence and Age. *Ergonomics*, *6*(1), 99–105. https://doi.org/10.1080/00140136308930679

van Leeuwen, J., Smeets, J. B. J., & Belopolsky, A. V. (2019). Forget binning and get SMART: Getting more out of the time-course of response data. *Attention, Perception, & Psychophysics*, *81*(8), 2956–2967. https://doi.org/10.3758/s13414-019-01788-3

Vleugels, L. W. E., Swinnen, S. P., & Hardwick, R. M. (2020). Skill acquisition is enhanced by reducing trial-to-trial repetition. *Journal of Neurophysiology*, *123*(4), 1460–1471. https://doi.org/10.1152/jn.00741.2019

Wohlschläger, A., & Bekkering, H. (2002). Is human imitation based on a mirror-neurone system? Some behavioural evidence. *Experimental Brain Research*, *143*(3), 335–341. https://doi.org/10.1007/s00221-001-0993-5

Appendices

1. Results Experiment 1

1.1 Speed -accuracy trade-off

A speed-accuracy trade-off was plotted primarily to spot a potential outlier. After visual inspection, one participant was identified as an outlier due to a very high percentage of error (close to 50%) and very short reaction time. This participant was excluded for all analysis. A Pearson correlation coefficient was calculated. It was significant ($r(28) = -0.42$, $p = 0.02$) before the exclusion of the participant but didn't hold after the exclusion ($r(28) = -0.17$, $p = 0.37$).

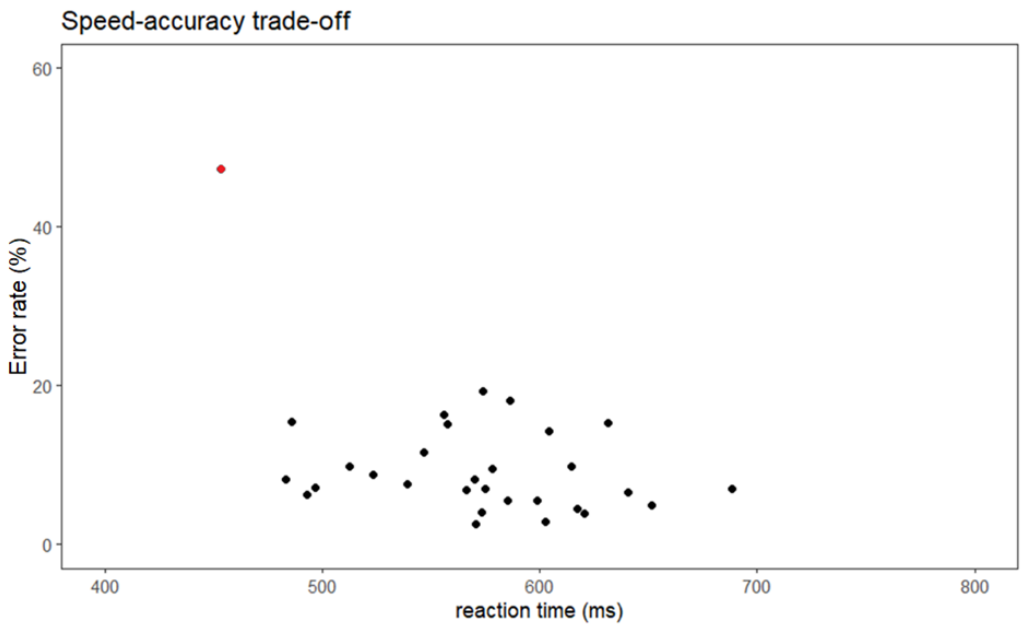

Figure S1. Speed-accuracy trade off. Each point represents a participant. The red point is the excluded participant.

1.2 Stimulus type

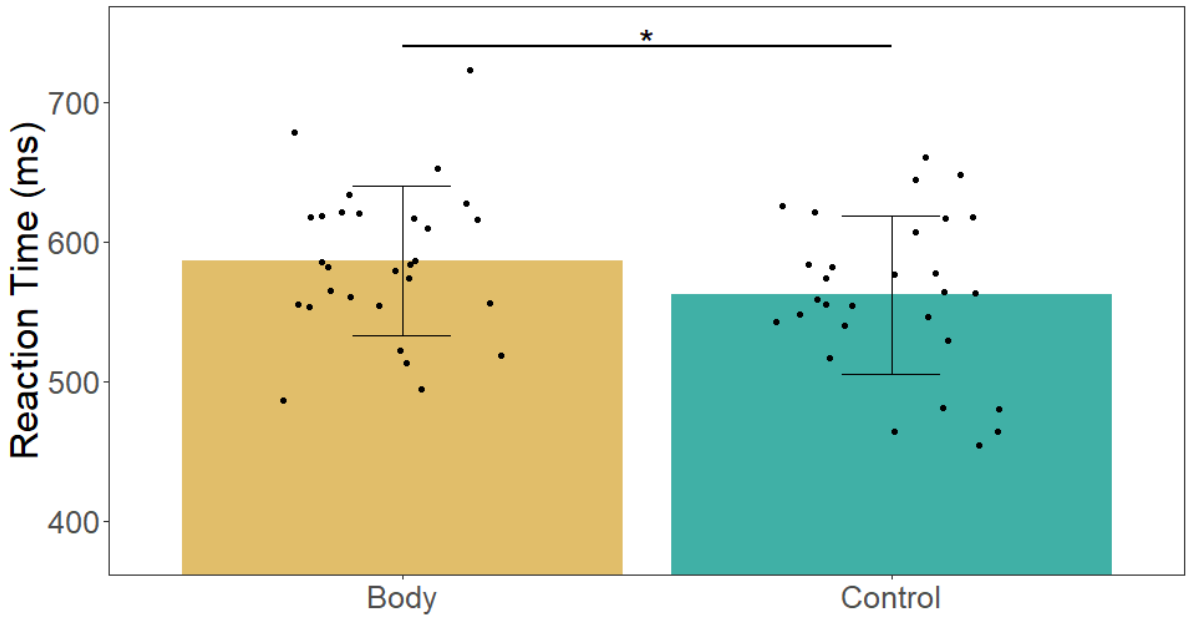

Figure S2. Reaction time for body and control stimuli. Error bars represent SD, each point represents a participant.

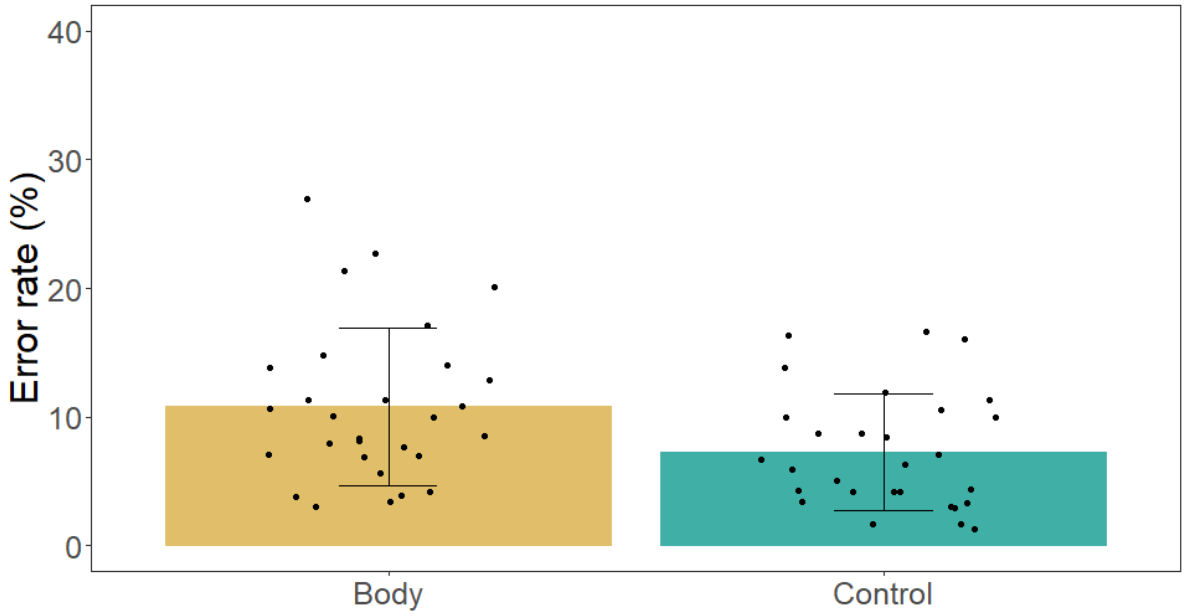

Figure S3. Error rate for body and control stimuli. Error bars represent SD, each point represents a participant.

1.3 Perspective

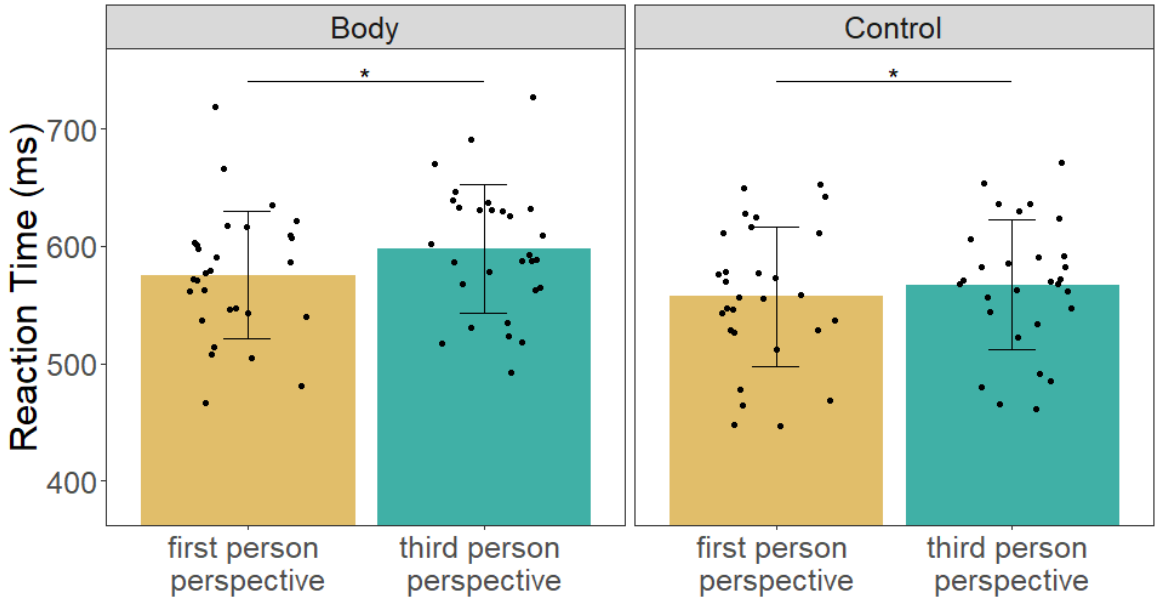

Figure S4. Mean reaction time for stimuli in first-person perspective and third-person perspective. The left panel is for body stimuli and the right is for control stimuli. Error bars represent SD, each point represents a participant.

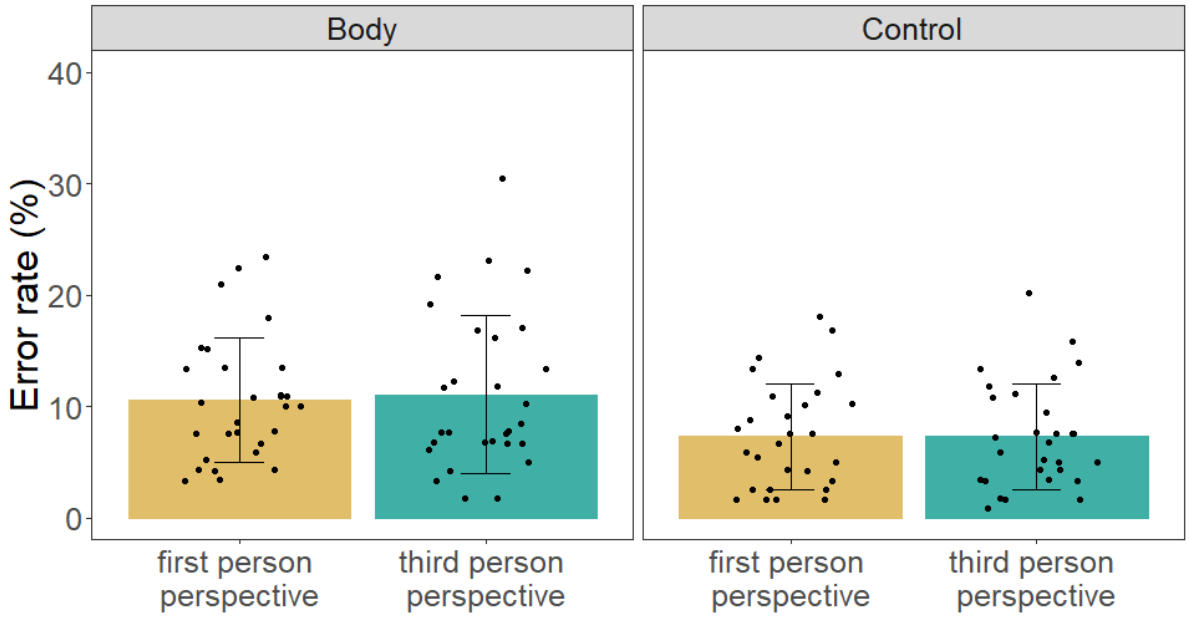

Figure S5. Mean reaction time for stimuli in first-person perspective and third-person

perspective. The left panel is for body stimuli and the right is for control stimuli. Error bars represent SD, each point represents a participant.

2. Results Experiment 2

Term	initial plateau (%)	mean t1	sd t1	plateau t1	mean t2	sd t2	final plateau (%)	AIC	AIC difference
single process model body, congruent	28,71	na	na	na	389,4	107,2	80,06	11280	
dual process model body, congruent	31,13	170,9	76,8	0,304 41	324,3	149,1	80,87	11290	-10
single process model body, incongruent	26,18	na	na	na	574,1	72,6	71,07	11681	
dual process model body, incongruent	28,41	370	60	0,394 24	525	110,5	72,28	11672	9
single process model body, neutral	40,58	na	na	na	396,6	80,4	89,73	10154	
dual process model body, neutral	40,66	255,6	63,2	0,446 76	329,4	115,7	90,2	10174	-20
single process model control, congruent	27,33	na	na	na	368,4	107,9	86,85	10297	
dual process model control, congruent	28,67	198,4	98,5	0,372 11	320,9	133,7	87,29	10320	-23

single process model control, incongruent	27,68	na	na	na	454,9	76,5	86,05	10543	
dual process model control, incongruent	29,22	310,5	86,5	1	418,5	99,9	86,73	10563	-20
single process model control, neutral	39,87	na	na	na	403	90,3	89,56	10338	
dual process model control, neutral	42,63	230,7	95,6	0,367 58	331,2	130,3	90,26	10354	-16
single process model body, first person	32,66	na	na	na	454	129,6	81,7	17605	
dual process model body, first person	34,12	240,6	121,8	188,3	374,1	168,8	99,95	17629	-24
single process model body, third person	32,17	na	na	na	455,5	123,2	79,47	17751	
dual process model body, third person	32,41	220,4	86,2	204,4	368	176,9	99,84	17768	-17
single process model control, first person	32,13	na	na	na	409	98	88,34	15639	
dual process model control, first person	32,18	242,3	70,7	0,222 55	350,8	131,8	88,95	15659	-20
single process model control, third person	31,99	na	na	na	416,9	99,2	87,25	15958	

dual process model									
control, third person	34,31	236	104,5	115,1	360,4	132,3	98,68	15977	-19
single process model									
body, correct responses	26,18	na	na	na	574,1	72,6	71,07	11681	
dual process model				0,394					
body, correct responses	28,41	370	60	24	525	110,5	72,28	11672	9
single process model									
body, spatial errors	39,43	na	na	na	635,5	33,3	21,46	12309	
dual process model				0,410					
body, spatial errors	29,35	375,5	62,3	87	476,3	133,1	20,37	11991	318
single process model									
body, other errors	43,14	na	na	na	397,6	116,2	8,21	9585,3	
dual process model									
body, other errors	42,97	235	110,5	86,2	264,9	160,4	0,01	9608,7	-23,4
single process model									
control, correct responses	27,68	na	na	na	454,9	76,5	86,05	10543	
dual process model									
control, correct responses	29,22	310,5	86,5	1	418,5	99,9	86,73	10563	-20
single process model									
control, spatial errors	32,51	na	na	na	467,4	57,6	4,97	8623	
dual process model				0,129					
control, spatial errors	30,19	323,5	59,1	24	353,3	108,2	4,53	8609	14

single process model									
control, other errors	42,05	na	na	na	414,7	119,7	8,81	9864,6	
dual process model									
control, other errors	41,47	301,1	134,5	94,3	301,1	149,8	0,01	9892,4	-27,8

Table S1. Tables with the values of the various parameters for all the models. Column 1 corresponds to the inflection point for the decrease in the proportion of correct responses and is only present for the complex model. Column 2 corresponds to the variance between participants in column 1. Column 3 corresponds to the inflection point for the increase in the proportion of correct responses. Column 4 corresponds to the variance between participants in column 3. column 5 corresponds to the final plateau calculated by the model. column 6 corresponds to the initial plateau calculated by the model. column 7 corresponds to xxx and is only present for the complex model. Column 8 is the AIC calculated for each model. Column 9 is the difference between the AIC of the simple model and the complex model. A negative value indicates a favour for the simple model, while a positive value indicates a favour for the complex model.

Term	cluster significantly above baseline (ms)	cluster significantly below baseline (ms)	threshold for t value	t value for cluster above baseline	t value for cluster below baseline
Body	294-800		142,89	7729	
Control	288-800		144,76	10660,29	
Body, congruent	285-800		155,58	7357,53	
Body, incongruent	500-800	341-458	141,31	2617,44	455,91
Body, neutral	304-800		145,08	9890,24	
Control, congruent	205-800		147,26	10450,31	
Control, incongruent	374-800		145,26	7800,17	
Control, neutral	302-800		142,87	9680,27	
Body, first-person perspective	299-800		149,67	739525	
Body, third-person perspective	300-800		141,34	6755,82	
Control, first-person perspective	288-800		136,57	10384,24	
Control, third-person perspective	294-800		147,43	9825,76	
Body, correct responses	500-800	341-458	145,45	2617,44	455,91
Body, spatially induced errors	307-615	703-800	157,88	2100,2	249,8
Body, random errors		310-800	151,47		7537,74
Control, correct responses	374-800		140,09	7800,17	
Control, spatially induced errors	282-383	429-800	142,36	320,09	6695,73
Control, random errors		303-800	143,89		7257,68

1167

Table S2. Clusters significantly different from baselines for the different conditions. In column 1 and 2 are the clusters respectively significantly above and below the baseline. Column 3 is the t-value that a cluster must reach to be significant. And columns 4 and 5 are the t-values of the different clusters identified in column 1 and 2 respectively.

Term	Significantly different cluster	threshold	T value for significantly different cluster
Body vs control, congruent	394-764	152,11	999,74
Body vs control, incongruent	334-800	154,68	3560,79
Body vs control, neutral		151,59	
Body, first- vs third-person perspective		156,43	
Control, first- vs third-person perspective		155,19	
Body vs control, correct responses	334-800	154,68	3560,79
Body vs control, other errors		151,96	
Body vs control, spatial errors	355-800	153,25	4143,23

Table S3. Cluster identified as significantly different by the SMART method for the comparison of two conditions. Column 1 corresponds to the significantly different cluster. Column 2 is the t-value that a cluster must reach to be significant. And column 3 is the t-values of the different clusters identified in column 1