Supplementary Information Computational Modelling of Embodied Visual Perspective-taking

Tobias Fischer, *Member, IEEE* and Yiannis Demiris, *Senior Member, IEEE*

Additional Results

A. Zero degree disparity

The model suggests that there are significant differences for the zero degree angular disparity case between straight and incongruent movements ($t(24) = 3.32$; $P < 0.01$; $n = 24$; two-tailed Welch's t-test), which is opposed to human data [18]. This difference is not examined further for the following reasons. Zero degree angular disparity refers to the case that the self-perspective matches the perspective of the other, and thus no perspective-taking is required. Indeed, it has been argued that even in the case of small angular disparities, a visual matching strategy is employed [62], rather than the embodied matching process proposed in our paper.

There is another notable finding for the zero degree disparity case. As shown in Fig. 3 of the main article, our model data suggests that the response time for zero degrees should be faster compared to 45 degrees angular disparity ($t(33) = -13.6$; $P < 0.001$; $n = 24$; two-tailed Welch's t-test). Our model data is supported by some human data [1], where the same effect was observed. However, our model data is opposed to other human data [18], where the response times for 0 and 40 degrees did not significantly differ. Future behavioral studies with humans should investigate this inconsistency further.

B. Posture congruence

The impact of the target agent's body posture has been empirically studied [18]. While movement congruence indicated whether the self-agent's body posture was congruent with the movement direction towards the target agent, posture congruence here indicates whether the self-agent's body posture and the target agent's body posture match (Suppl. Fig. 3). In other words, if the self-agent's torso is rotated clockwise and the target agent's torso is also rotated clockwise, then the posture is congruent (and similarly for counterclockwise rotations).

In the computational model, differences in the response times in posture congruency experiments are expected due to the alignment strategy where individual joints are matched to some degree. More specifically, the mixing parameter ω controls the impact of the target agent's posture on response times. The smaller ω, the higher the impact as each joint state is individually matched. On the other hand, a large ω leads to a small impact as the combination of all states is matched. This is further detailed in the Methods section of the main article.

In human data, the impact of posture congruence on response times is smaller compared to that of the movement congruence, and there is only a small effect for posture congruence at 120 and 160 degrees angular disparity (but not for 40 and 80 degrees). In model data, there are significant differences between congruent and incongruent postures for all angles ($P < 0.01$ for all angles; $n = 24$; all two-tailed Welch t-tests).

Suppl. Fig. 8 compares the model and human response times for the body posture congruence qualitatively. There is a good qualitative match for the response time profiles of model and human data, which demonstrates that the model can also explain posture congruence behavior.

C. Residual effects on the motor memory also occur with rotated torso postures

This section investigates whether residual effects on the motor memory (see main article Section V-B) also occur in trials where the self-agent's body posture is not straight. This results in two types of congruencies: movement congruency (the self-agent's body posture is compatible with the required movement direction as in Suppl. Fig. 2) and trial congruency (the previous trial's movement direction is compatible with the current trial's movement direction, following the main article and Suppl. Fig. 5). Importantly, however, the self-agent's posture is not changed across two trials, as the physical movement would change the currently active action primitive that was selected by the selection process.

Therefore, two tests can be performed. The first test compares two trials, one with compatible trial congruency and another incompatible one. For the first test, both trials have congruent movements (see Suppl. Fig. 6 for a schematic). A Mann-Whitney U test with a one-tailed null hypothesis is performed. The null hypothesis that both trials have equal response times can be

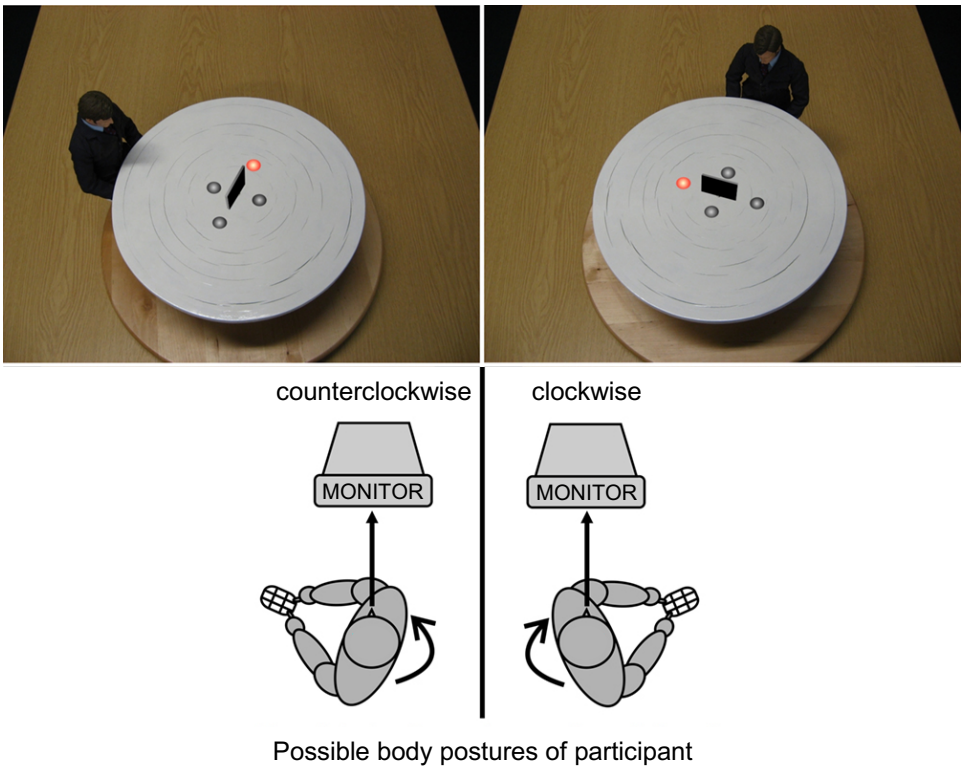
The authors are with the Personal Robotics Laboratory, Department of Electrical and Electronic Engineering, Imperial College London, SW7 2BT, U.K. (e-mail: {t.fischer, y.demiris}@imperial.ac.uk)

rejected in case of congruent trials ($U = 3688.5$; $P = 0.001$; $n = 100$; one-tailed Mann-Whitney U-test). This is not the case for incongruent trials ($U = 4533.0$; $P = 0.127$; $n = 100$; one-tailed Mann-Whitney U-test). The second test is very similar to the first one, with the difference that both trials have incongruent movements (rather than congruent movements as above; see Suppl. Fig. 7 for a schematic). In contrast with the first test, the null hypothesis of equal response times cannot be rejected, neither for congruent trials ($U = 4786.5$; $P = 0.30$; one-tailed Mann-Whitney U-test) nor for incongruent trials ($U = 4896.5$; $P = 0.40$; one-tailed Mann-Whitney U-test).

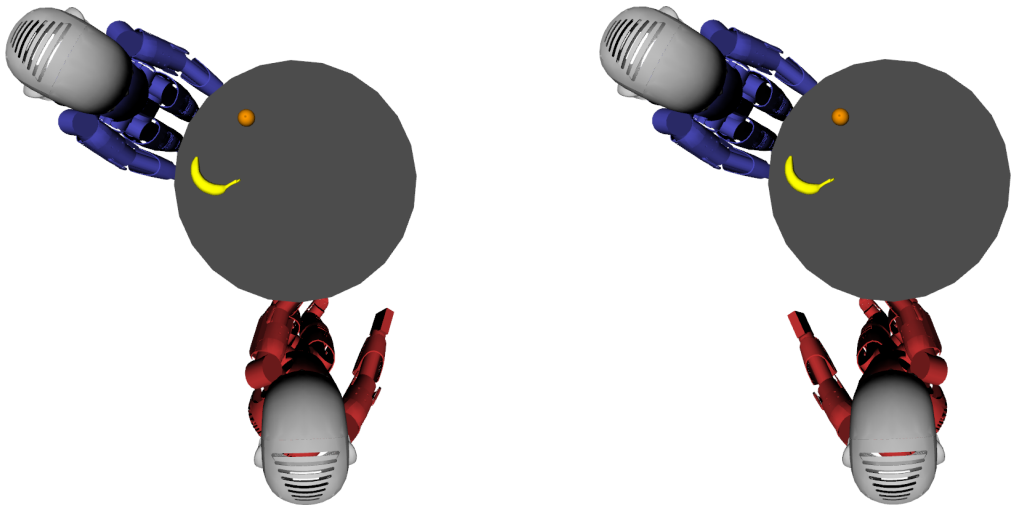
In conclusion, the selection process leads to residual effects on the motor memory even if the self-agent's torso pose is varied, but only if the current trial is congruent to the previous trial, and additionally the movement direction is congruent in these trials.

REFERENCES

[1] P. Michelon and J. M. Zacks, "Two kinds of visual perspective taking," *Perception & Psychophysics*, vol. 68, no. 2, pp. 327–337, 2006.
[18] K. Kessler and L. A. Thomson, "The embodied nature of spatial perspective taking: Embodied transformation versus sensorimotor interference," *Cognition*, vol. 114, no. 1, pp. 72–88, 2010.
[42] K. Kessler and H. Rutherford, "The two forms of visuo-spatial perspective taking are differently embodied and subserve different spatial prepositions," *Frontiers in Psychology*, vol. 1, no. 213, pp. 1–12, 2010.
[62] M. Janczyk, "Level 2 perspective taking entails two processes: Evidence from PRP experiments," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 39, no. 6, pp. 1878–1887, 2013.

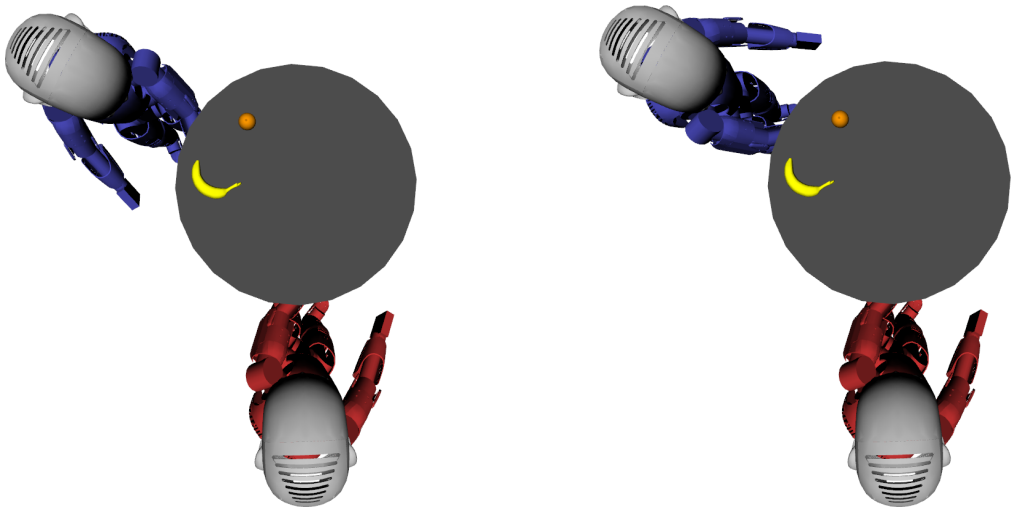


Suppl. Fig. 1. This figure depicts a typical experimental setup with human subjects. The subject sits in front of a computer screen. The subject's torso might be rotated clockwise or counterclockwise. An avatar is displayed on the screen whose perspective the subject has to take. The subject's task is to decide whether the target (in this example, a light) is to the left or right as seen from the avatar's perspective. This figure has been adapted from [42]. The copyright remains with Kessler and Rutherford.



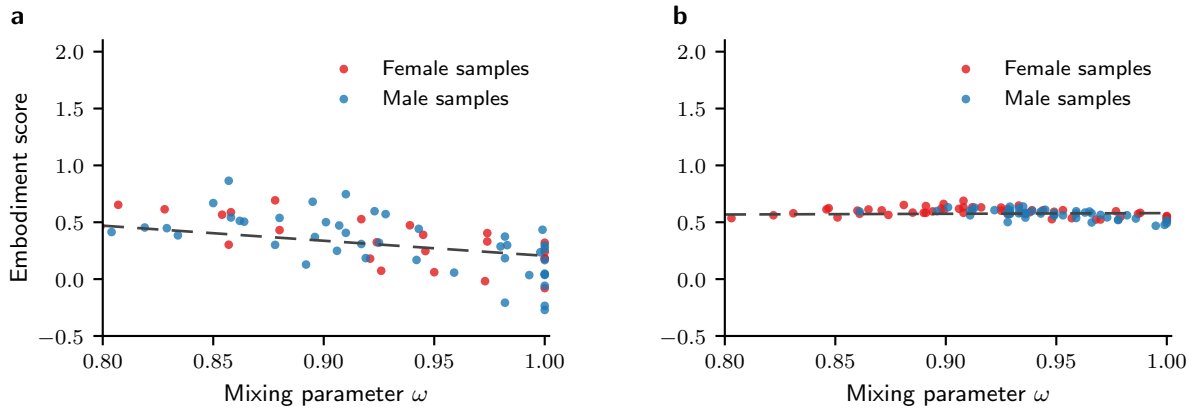
(a) Congruent movement (b) Incongruent movement

Suppl. Fig. 2. Movement congruency schematic. The required mental rotation is clockwise for both examples. In (a), the self-agent's (red robot) torso is also rotated clockwise, which is considered as congruent movement. This is opposed to (b), where the self-agent's torso is rotated counterclockwise and hence the movement direction and torso rotation are incongruent.

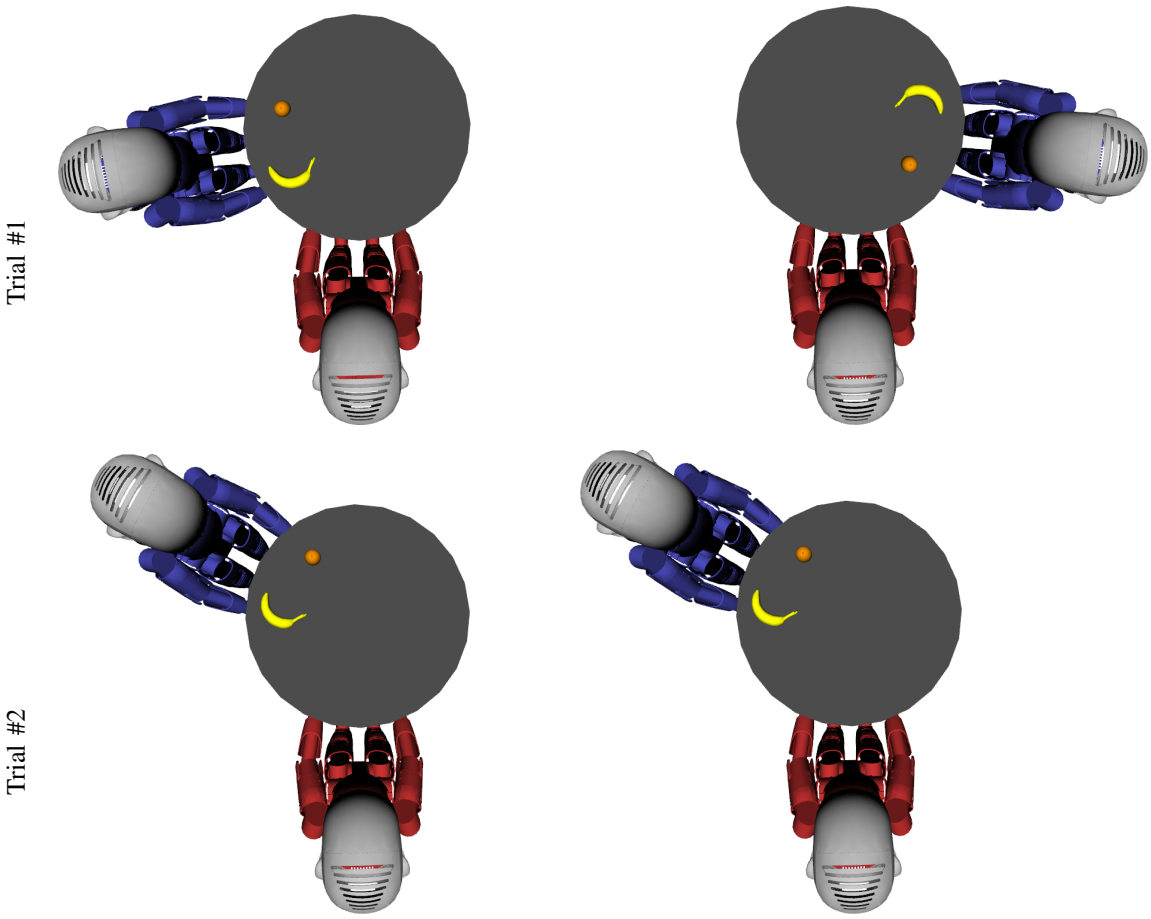


(a) Congruent posture (b) Incongruent posture

Suppl. Fig. 3. Posture congruency schematic. In (a), both the self-agent's (red robot) and target agent's (blue robot) torso are rotated clockwise, which is considered as congruent posture. This is opposed to (b), where the self-agent's torso is rotated clockwise while the target agent's torso is rotated counterclockwise (incongruent posture).

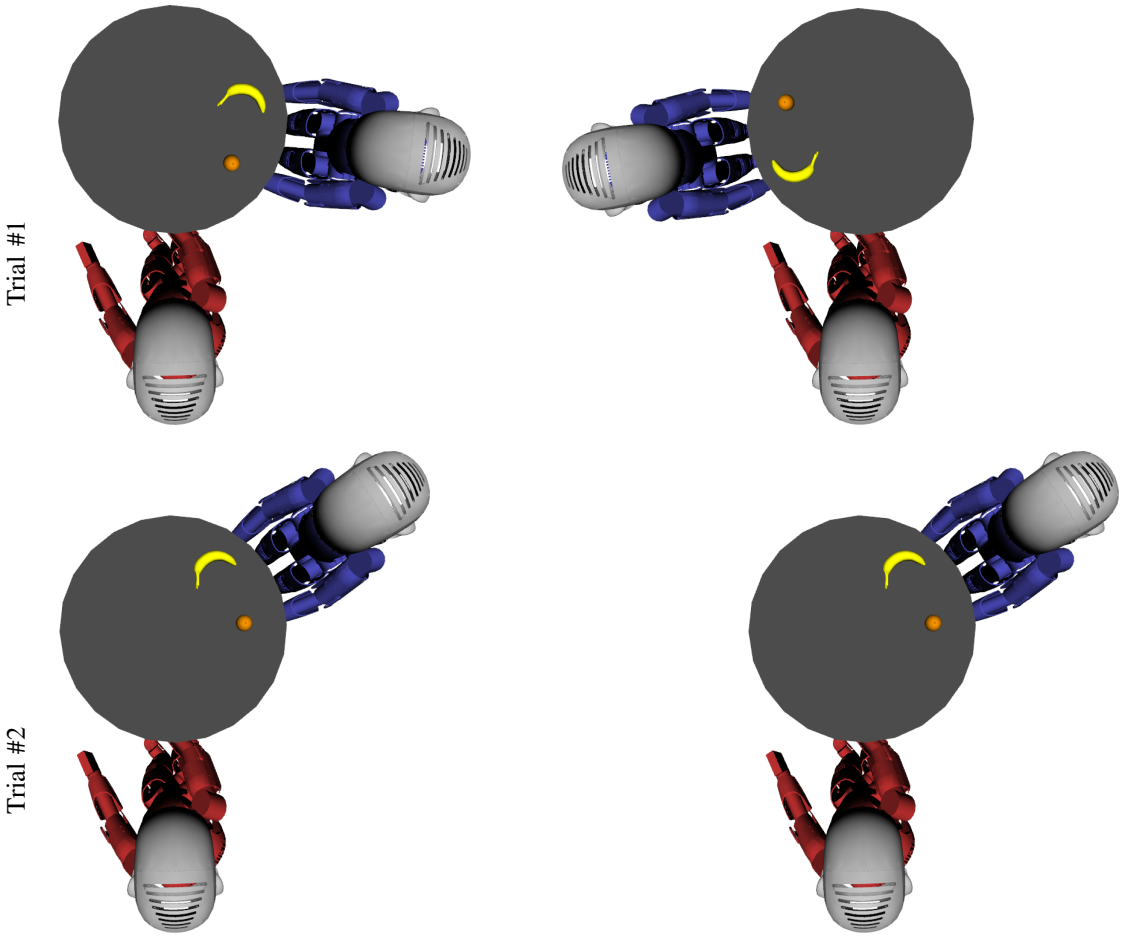


Suppl. Fig. 4. Correlation mixing parameter and embodiment measure. (a) There is a negative correlation between the mixing parameter ω and the corresponding embodiment measure in the computational model ($t(94) = -7.09$; Pearson's $r = -0.584$; $P < 0.001$; $n = 96$; two-tailed Wald's t-test). As detailed in the main text, the "female" and "male" samples have different underlying distributions ($\omega_{\text{females}} \sim \mathcal{N}(0.83, 0.15)$ for $n_{\text{females}} = 51$ instances and $\omega_{\text{males}} \sim \mathcal{N}(0.91, 0.1)$ for $n_{\text{males}} = 45$ instances respectively). (b) In the case that the selection process is not used, we observe no significant correlation between the mixing parameter ω and the embodiment measure ($t(94) = 0.52$; Pearson's $r = 0.053$; $P = 0.601$; $n = 96$; two-tailed Wald's t-test).



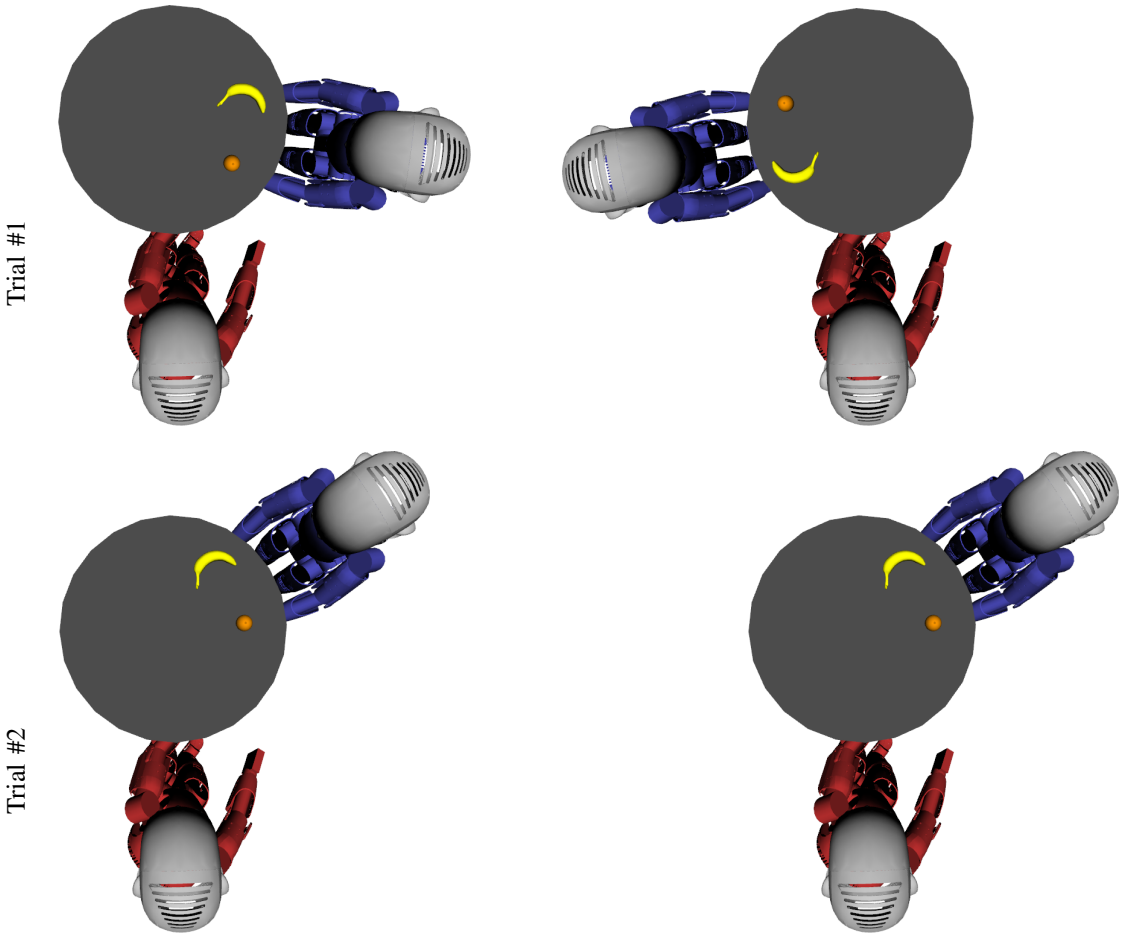
(a) Congruent trials (b) Incongruent trials

Suppl. Fig. 5. Trial congruency schematic (straight body postures of both agents). In (a), the required direction of the mental rotation for trials #1 and #2 remains the same (both clockwise, congruent trials), while the direction changes from counterclockwise to clockwise in (b) (incongruent trials).



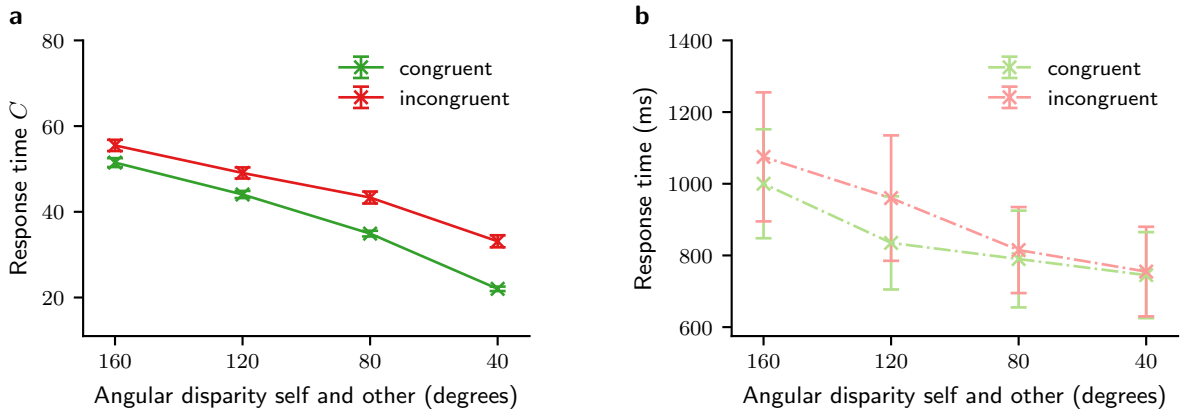
(a) Congruent trials (b) Incongruent trials

Suppl. Fig. 6. Trial congruency schematic (congruent movement direction). While the required mental rotation and the torso rotation are counterclockwise *for trial #2* in both (a) and (b) (congruent movement direction), the required movement direction in trial #1 is congruent with the subsequent trial #2 in (a) (congruent trials) but not in (b) (incongruent trials).

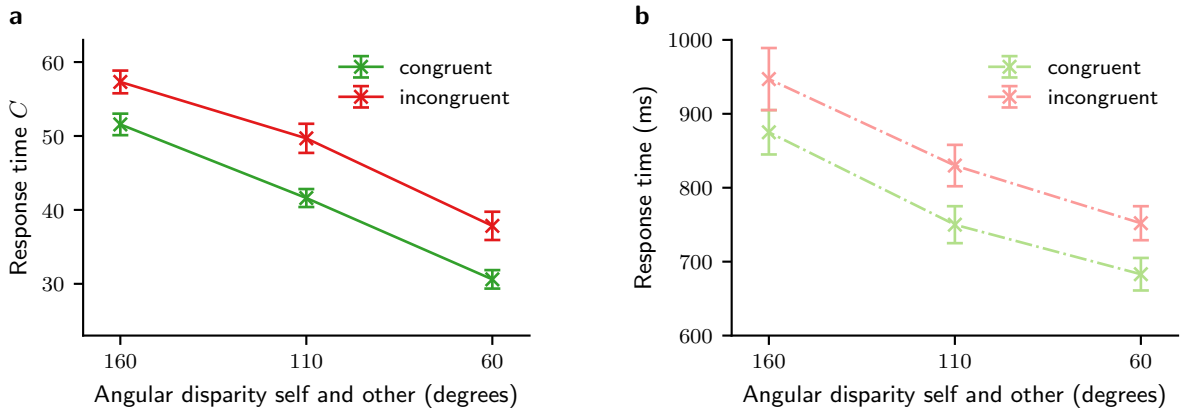


(a) Congruent trials (b) Incongruent trials

Suppl. Fig. 7. Trial congruency schematic (incongruent movement direction). In both (a) and (b), the required mental rotation (counterclockwise) and the torso rotation (clockwise) are incongruent *for trial #2*. However, the required movement direction in trial #1 is congruent with the subsequent trial #2 in (a) (congruent trials) but not in (b) (incongruent trials).



Suppl. Fig. 8. The response times with respect to posture congruence of (a) the computational model are compared to (b) experimental data in humans [18]. There is a good qualitative match of the model and human data for both congruent as well as incongruent postures, which demonstrates that our model can also explain human data with posture congruency variations. All error bars indicate the standard error of the mean across subjects ($n = 24$).



Suppl. Fig. 9. The response times with respect to movement congruence of (a) the computational model are compared to (b) experimental data in humans [42]. There are significant differences between congruent and incongruent response times in both human and model data (all $P < 0.05$). These results supplement Fig. 3c-d of the main text, which compares to human data from [18]. All error bars indicate the standard error of the mean across subjects ($n = 24$).