Exploring Reference Frame Integration Using Response Demands in a Tactile Temporal-Order Judgement Task

Kaian Unwalla [1,*], **Daniel Goldreich** [1,**] **and David I. Shore** [1,2,***]

[1] Department of Psychology, Neuroscience & Behaviour, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4K1, Canada

[2] Multisensory Perception Laboratory, a Division of the Multisensory Mind Inc., Hamilton, ON, Canada

Abstract

Exploring the world through touch requires the integration of internal (e.g., anatomical) and external (e.g., spatial) reference frames — you only know what you touch when you know where your hands are in space. The deficit observed in tactile temporal-order judgements when the hands are crossed over the midline provides one tool to explore this integration. We used foot pedals and required participants to focus on either the hand that was stimulated first (an anatomical bias condition) or the location of the hand that was stimulated first (a spatiotopic bias condition). Spatiotopic-based responses produce a larger crossed-hands deficit, presumably by focusing observers on the external reference frame. In contrast, anatomical-based responses focus the observer on the internal reference frame and produce a smaller deficit. This manipulation thus provides evidence that observers can change the relative weight given to each reference frame. We quantify this effect using a probabilistic model that produces a population estimate of the relative weight given to each reference frame. We show that a spatiotopic bias can result in either a larger external weight (Experiment 1) or a smaller internal weight (Experiment 2) and provide an explanation of when each one would occur.

---

* To whom correspondence should be addressed. E-mail: unwallk@mcmaster.ca
** ORCID: 0000-0002-5104-2185
*** ORCID: 0000-0003-2562-1017

1. Introduction

Locating tactile sensations requires knowing where our hands are in space. Two reference frames are used to locate a touch: the internal, anatomical, reference frame and the external, spatial, reference frame. Responding to touch involves integrating these sources of information. The relative contribution of each reference frame can be measured using a tactile temporal-order judgement (TOJ) task (Azañón and Soto-Faraco, 2007; Azañón *et al.*, 2015; Azañón *et al.*, 2016; Badde *et al.*, 2016; Cadieux and Shore, 2013; Cadieux *et al.*, 2010; Craig and Belser, 2006; Crollen *et al.*, 2017; Crollen *et al.*, 2019; Kóbor *et al.*, 2006; Pagel *et al.*, 2009; Roberts and Humphreys, 2008; Röder *et al.*, 2004; Schicke and Röder, 2006; Shore *et al.*, 2002; Unwalla *et al.*, 2020; Wada *et al.*, 2014; Yamamoto and Kitazawa, 2001).

This unspeeded task requires participants to report which of two vibrations applied to each of their hands occurred first, with their hands uncrossed and crossed. Consistently, the crossed-hands condition produces poorer TOJ performance than the uncrossed condition. All accounts of the deficit highlight the automatic transfer of information from the internal to the external reference frame (i.e., spatial remapping; see Badde and Heed, 2016 for review). Models differ with respect to how the two reference frames are treated to determine the final stimulus location. Non-integration models suggest that the touch is located based solely on the external reference frame (Yamamoto and Kitazawa, 2001). The conflict model highlights confusion caused by opposing response requirements for the two locations (i.e., internal versus external; Shore *et al.*, 2002). The integration model builds upon the conflict model by defining the conflict as differential weights placed on the two reference frames in determining the location of the final percept (Badde *et al.*, 2016). Both the conflict and integration models predict that emphasizing the external reference frame will increase the size of the deficit whereas the non-integration model makes the opposite prediction.

One way to bias perception to one reference frame or the other is to emphasize the coordinate system within which the observer must respond (Cadieux and Shore, 2013; Crollen *et al.*, 2019; Shore *et al.*, 2006). For example, by using foot pedals instead of hand buttons to respond, it is possible to place emphasis on either the internal or the external reference frame. The anatomical response demand (i.e., lift the toe corresponding to the hand that was stimulated first) preserves the left–right coding of the vibration and response, making it more internally based. In contrast, the spatiotopic response demand (i.e., lift the toe directly underneath the stimulated hand) requires remapping the response from the hand surface to the corresponding hemispace in the external reference frame. All studies using the response demand manipulation implicitly assume that a tactile stimulus must first be localized to the hand

before a response can be made. Critically, response demands are presumed to affect the ability to localize the tactile stimulus on the hand. The anatomical response demand ties the response to the body (right hand, right foot) biasing the localization towards the internal reference frame which results in a smaller deficit (Cadieux and Shore, 2013; Crollen *et al.*, 2019; Shore *et al.*, 2006). In contrast, the spatiotopic response demand ties the response to the space around the body (left side of space, left foot) biasing the localization to the external reference frame, therefore increasing the deficit.

Quantifying the size of the deficit has typically used behavioural measures (such as the slope of a psychometric curve, or proportion of correct responses), and inferred the weight given to each reference frame by a change in these measures. The recent development of a probabilistic model (Badde *et al.*, 2016) affords us the potential to quantify the response demand effect using estimated reference frame weights. Accordingly, we sought to replicate the response demand effect (larger crossed-hands deficit with a spatiotopic response demand) and apply the probabilistic model (Badde *et al.*, 2016) that maps behaviour onto weights for the internal and external reference frames.

1.1. Measuring the Crossed-Hands Deficit

Multiple measures of the crossed-hands deficit exist. Early work examined the difference in the slope of the psychometric curves for crossed and uncrossed postures (e.g., Shore *et al.*, 2002). The probit analysis converts the proportion of right-first responses into a *z*-score and then fits a straight line to the *z*-score across stimulus onset asynchrony (SOA). This method allows for separate analysis of the uncrossed and crossed performance, and indexes the crossed-hands deficit as the difference in the slope measures. The same analysis can be used to derive a just noticeable difference (JND). The JND is an indicator of the time difference required between the two tactile stimuli to accurately assess their temporal order. These analysis techniques have two shortcomings. First, at longer SOAs, performance reaches a ceiling making the measure less sensitive at detecting performance differences. Second, in the crossed posture, the slope can be negative or approach zero, which can produce unreasonable extrapolation of the data. In terms of the response demand manipulation (Crollen *et al.*, 2019), the uncrossed posture produced similar slopes with both demands. In the crossed posture, the spatiotopic response demand produced significantly shallower slopes (i.e., worse performance) than the anatomical response demand. This larger crossed-hands deficit in the spatiotopic response condition was attributed to a greater reliance on the external reference frame when localizing the tactile stimulus, and supported the conflict model of the deficit.

Another measure, used more recently, is the proportion correct difference (PCD) score (Cadieux *et al.*, 2010). To calculate the PCD score, the difference

in the proportion of correct responses between crossed and uncrossed performance is computed at each SOA, and then summed across SOA. There are several advantages of the PCD score over other measures (i.e., slope or JND). For instance, both the uncrossed and crossed-hands performance are combined into a single score that indexes the magnitude of the deficit. Additionally, the measure is model-free — no assumptions are made about the underlying distribution of responses nor the shape of the psychometric curves. With this measure, a larger deficit was found when participants used a spatiotopic response compared to an anatomical response (Cadieux and Shore, 2013).

These measurements (e.g., slope, JND and PCD) are mostly atheoretical. They provide a description of the data, but not how the underlying theoretically construed reference frame weights change under different task demands. Recently, a probabilistic model was developed to estimate the relative weight placed on the internal and external reference frame during a crossed-hands tactile TOJ task (Badde *et al.*, 2016). The researchers tested two models for the crossed-hands deficit: the integration model and the non-integration model. The integration model explained the crossed-hands deficit as a difficulty integrating the internal and external reference frame in the crossed posture. In contrast, the non-integration model explained the deficit as the result of a difficulty remapping from the internal to the external reference frame. The integration model better accounted for their data. The model estimates a pair of internal and external weights that most likely created both the uncrossed and crossed psychometric curves.

The model assumes the weights are stable within an individual across time; therefore, changing hand position from uncrossed to crossed should not change the weights. Based on these two weights — an internal and an external — the model produces psychometric curves for the uncrossed and crossed postures. In the uncrossed condition, the reference frames provide congruent information, so the model takes the sum of the weights to compute the slope of the curve. In the crossed-hands condition, the two reference frames conflict, with the external reference frame providing incorrect information. As a result, the model takes the difference between the weights to compute the slope of the curve; the more an individual relies on the external reference frame when resolving the conflict, the shallower the slope (i.e., the worse the performance will be). Using this model, performance in the two conditions can be fit simultaneously to the weights placed on the two reference frames. Critically, this measure is theoretical as it is based on the integration/conflict model of the deficit (Badde *et al.*, 2016; Shore *et al.*, 2002).

Although the weights are assumed to remain constant across postural changes, task demands, including instruction, can lead to a change in the weights (Badde *et al.*, 2016). In addition to the typical TOJ task, Badde *et al.* (2016) used two other tasks. The first touch localization (FTL) task asks

participants to indicate, in a speeded fashion, the hand that received the first of two vibrations and ignore the second vibration. The only difference between the TOJ and FTL task is the instruction to ignore the second vibration and respond as quickly as possible. The third task was a single-touch localization (STL) task, where only one tactile stimulus was administered and participants had to indicate which hand was vibrated. Each task was completed in a crossed and an uncrossed posture. In comparison to the tactile TOJ task, both the FTL and STL showed an increased internal weight and a decreased external weight. The weights remained stable within an individual, but simply by changing the instructions provided during the tasks, the emphasis on each reference frame was altered.

1.2. Scope of the Present Study

The present study had two main goals. First, we wanted to confirm whether manipulating response demands would influence the magnitude of the crossed-hands deficit. Second, we applied a probabilistic model to these data to gain insight into the response demands manipulation. Each participant completed a crossed-hands tactile TOJ task under both the anatomical and spatiotopic response demand. To compare the size of the crossed-hands deficit across the two conditions an analysis was conducted on PCD scores. Based on previous studies, we predicted that the use of an anatomical response demand would result in a smaller crossed-hands deficit, whereas a spatiotopic response demand would show a larger crossed-hands deficit. This larger deficit would be revealed by a shallower slope in the crossed posture and a larger PCD score in the spatiotopic response demand, compared to the anatomical demand.

   We employed probabilistic models to estimate how response demands influenced the weights placed on the internal and external reference frame. First, we used a participant-specific model. This provided an estimated internal weight and external weight for each participant in each response demand condition individually. Next, we implemented a hierarchical model that assumed participants were affected equivalently by the response demand manipulation; we used this model to estimate the internal and external weights for the population as well as weights for each participant. We predicted that the larger deficit in the spatiotopic response demand would be explained by a decrease in the internal weight, by an increase in the external weight, or by a combination of the two. All options would result in decreased accuracy in the crossed-hands posture, with minimal changes occurring to the uncrossed posture.

2. Experiment 1

2.1. Methods

2.1.1. Participants

Twenty right-handed participants (10 males; average age: 19.3 years), were recruited from the McMaster University subject pool. All had normal or corrected-to-normal vision, were naïve to the purpose of the experiment, and provided written informed consent prior to participation. All procedures were approved by the McMaster Research Ethics Board and complied with the tri-council statement on ethics (Canada).

2.1.2. Apparatus and Stimuli

Throughout the experiment, participants were seated at a table (height of 73.7 cm) and placed their hands 18 cm apart. Placed in each hand was a small wooden cube with a Plexiglas top; there was a 2 cm hole in the top for participants to place their thumbs on the vibrators, which were mounted under the Plexiglas. Vibrations were delivered with an Oticon-A (100 Ohm; Oticon, Copenhagen, Denmark) bone conduction vibrator (width: 1.6 cm, length: 2.4 cm), that was driven by an amplified 250 Hz sine wave, set by the experimenter to be comfortable and clearly suprathreshold. Mounted beneath the vibrators were response buttons to be pressed by the thumbs on time-out trials. Two foot pedals were positioned beneath the toes of each foot to collect responses. All stimulation was controlled by a set of reed-relays connected to the parallel port of a DELL Dimension 8250 running Windows XP software. Matlab (MathWorks, Natick, MA, USA) was used to administer the stimulation and collect responses. Participants wore over-the-ear headphones, connected to an iPod Touch, playing white noise during the experiment to mask the sounds produced by the tactile vibrators.

2.2. Procedure

Participants held one wooden cube in each hand, with their thumbs in contact with the vibrators. Each trial began 800 ms after the participant's previous response. Two 20-ms vibrations, one to each thumb, were delivered separated by one out of a fixed set of SOAs: ±400, ±200, ±50 ms, where negative SOAs indicate the vibration was to the left hand (anatomical instructions) or hemispace (spatiotopic instructions) first. The task required participants to determine which of two vibrations occurred first under two different response demands. Participants responded by lifting the foot associated with the appropriate response demand. In the anatomical response demand condition, participants were instructed to 'lift the foot pedal corresponding with the hand that was vibrated first.' If the left hand received the first vibration they should lift the left toe (same for right hand and right toe). In the spatiotopic response

demand condition, participants were instructed to "lift the foot pedal directly underneath the hand that was vibrated first." If the left hemispace received the first vibration they should lift their left toe (same for right hemispace with right toe). If no response was made within three and a half seconds of the second vibration, the trial timed out. In this situation, both vibrators vibrated three times, and participants pressed down on both vibrators to activate the buttons mounted underneath. These trials and trials where participants responded in less than 100 ms were removed before analysis. This resulted in the removal of 23 trials across all participants. The next trial began as soon as the participant pressed down on both foot pedals.

Participants initially completed two practice blocks of 18 trials each. During the first practice block, their hands were uncrossed; during the second practice block, their hands were crossed over the midline. The experimenter remained in the room for the practice trials in order to provide feedback and answer any questions. Participants subsequently completed 12 experimental blocks of 60 trials. For one half of the experiment, participants used the anatomical response demand and for the other half the spatiotopic response demand. Hand position was altered every three blocks between crossed and uncrossed positions. The starting response demand and hand position were counterbalanced across participants.

2.3. Analysis

The crossed-hands deficit was assessed using PCD scores (the sum of the difference between the proportion of correct responses in the crossed and uncrossed postures at each SOA; see Cadieux *et al.*, 2010; Heed and Azañón, 2014). The PCD scores were submitted to a 2×2 ANOVA with response demand (anatomical vs spatiotopic) as a within-subject factor and sex (male vs female) as a between-subject factor. We tested whether the crossed-hands deficit was significantly different from zero in each response demand using one-sample t-tests.

The above analysis provides an index of overt behaviour. In contrast, reference frame weight represents a theoretical construct that must be inferred from the data. We first used the equations outlined by Badde *et al.* (2016) to derive psychometric curves from the weights. We took a participant-specific approach by using a maximum likelihood estimation to determine the combination of internal and external weights that best accounted for each individual participant's data. An internal and external weight combination forms a hypothesis, which can be used to generate psychometric curves, $p(t)$, the probability of a right-first response as a function of SOA (t) for the crossed and uncrossed postures (equation 1). Each curve was a logistic function with slope parameter, θ, calculated from a linear combination of the internal and external weights (ω). With the hands uncrossed, the external response was congruent

with the internal response. When the hands were crossed, the external response was incongruent with the internal response. Thus, for the uncrossed posture, θ is the sum of the internal and external weights, whereas for the crossed posture, θ is the difference between the weights. To compute the likelihood of the hypothesis ($H = \omega_{\text{int}\,i}, \omega_{\text{ext}\,i}$), the probability of the participant's responses at each SOA was calculated from a binomial distribution with expected value $p(t)$. Each participant's internal and external weights were fit to the participant's uncrossed and crossed data simultaneously, reflecting the assumption that the weights do not change across these postures (see Badde *et al.*, 2016). Using a brute force algorithm, we discretized each participant's internal and external weights into bins of 0.5 spanning the range 0 to 40, calculated the log-likelihood (equation 2) at each combination of internal and external weights for each participant, and read out the maximum likelihood estimate. Estimates for the weight parameters were determined separately for the spatiotopic and anatomical response demands.

$$p(t) = \frac{1}{1 + e^{-\theta t}} \tag{1}$$

where

$$\theta_{\text{uncrossed}} = \omega_{\text{int}} + \omega_{\text{ext}};$$
$$\theta_{\text{crossed}} = \omega_{\text{int}} - \omega_{\text{ext}};$$

t is contained in the set $\{\pm 400, \pm 200, \pm 100, \pm 50\}$

$$\log\big[p(d_{i\,\text{rd}} \mid \omega_{i\,\text{nt}\,i\,\text{rd}}, \omega_{\text{ext}\,i\,\text{rd}})\big]$$
$$= \sum_{t} r_{ti\,\text{rd}} \cdot \log\big[p_{i\,\text{rd}}(t)\big] + l_{ti\,\text{rd}} \cdot \log\big[1 - p_{i\,\text{rd}}(t)\big] \tag{2}$$

where $d_{i\,\text{rd}}$ are the data from participant i for the given response demand; $\omega_{\text{int}\,i\,\text{rd}}, \omega_{\text{ext}\,i\,\text{rd}}$ are the hypothesized internal and external weight values for each response demand; $r_{t\,i\,\text{rd}}$ is the number of right-first responses; $l_{t\,i\,\text{rd}}$ is the number of left-first responses (i.e., the number of trials minus the number of right first responses); and t is contained in the set $\{\pm 400, \pm 200, \pm 100, \pm 50\}$.

The participant-specific model assumed that participants' data were statistically independent of one another. We next implemented a hierarchical model that encoded the arguably more plausible assumption that participants had similar weights and were affected equivalently by the response demand manipulation. For this purpose, we modified the hierarchical model proposed by Badde *et al.* (2016). The hierarchical model encodes the assumption that the response demand manipulation will have the same effect on all participants. Each participant's weights in one condition (we used the anatomical response demand) were multiplied by a population task parameter to obtain the weights in the other condition (the spatiotopic response demand).

A Markov Chain Monte Carlo (MCMC) sampler using a Metropolis–Hastings algorithm, implemented in R, was used to estimate the task parameters and the population means and standard deviations for the internal and external weights (see Badde *et al.*, 2016). We assumed that population distributions of internal and external weights were approximated by truncated Gaussian distributions (limits $= 0, \infty$), with unknown means and standard deviations. These population distributions served to generate priors for individual participant weights.

The MCMC procedure provides an approximation for the posterior distribution of the model parameters. This is accomplished by comparing the posterior probability of the current location in parameter space (which we refer to as a hypothesis) with the posterior probability of a proposed hypothesis. The proposed hypothesis is selected by randomly choosing a value from a Gaussian distribution with a mean of the current value and a proposal standard deviation (specified before starting the simulation). If the proposed hypothesis has a higher posterior probability, then the simulation accepts the proposed hypothesis. A new proposed hypothesis is then generated from this location. If the current hypothesis has a greater probability than the proposed hypothesis, the probability of accepting the new hypothesis is computed as the ratio of the probability of the proposed hypothesis to that of the current hypothesis.

The parameter set for the hierarchical model consisted of 46 parameters, six population-level parameters and 40 participant-level parameters. The six population-level parameters were: the population mean internal and external weights (μ_{internal} and μ_{external}) and standard deviations (σ_{internal}, σ_{external}) in the anatomical response demand, and an internal and external weight task context parameter (δ_{internal} and δ_{external}). All population parameters had strictly positive, uniform hyperpriors. The standard deviation parameters represent the model's estimation of the spread of the individual weights. The task context parameters were multiplied to the respective anatomical response demand mean weights to obtain the spatiotopic response demand weights. Each of the 20 participants' data were fitted with two parameters: an internal and external weight (ω_{internal} and ω_{external}) for the anatomical response demand. Given that the response demand manipulation was assumed to affect all participants equivalently, the weights for the spatiotopic response demand were calculated by multiplying each participants' anatomical response demand weights by the population task context parameters. For instance, if the population external task context parameter was 2, then the external weight in the spatiotopic condition for all participants would be twice as large as their external weight in the anatomical response demand. One hypothesis generated four psychometric curves for each participant — for each of the two response demands there

were two curves, one for each hand posture.

$$p(H \mid D) \propto \prod_i p(d_i|\omega_{\text{int}\,i}, \omega_{\text{ext}\,i})_i$$
$$\times \prod_i p(\omega_{\text{int}\,i}, \omega_{\text{ext}\,i} \mid \mu_{\text{int}}, \mu_{\text{ext}}, \delta_{\text{int}}, \delta_{\text{ext}}, \sigma_{\text{int}}, \sigma_{\text{ext}}) \qquad (3)$$
$$\times p(\mu_{\text{int}}, \mu_{\text{ext}}, \delta_{\text{int}}, \delta_{\text{ext}}, \sigma_{\text{int}}, \sigma_{\text{ext}})$$

The posterior probability of a hypothesis, H, given the dataset (D) was calculated by Bayes' formula (equation 3). The probability of each participant's data (d_i) given the participant's hypothesized weights was multiplied by the prior probability of those weights. The probability of each participant's data given the weights was calculated using the binomial formula as described in equation (2). To avoid underflow errors, equation (3) was evaluated by calculating the logarithms of all likelihoods and priors, and then summed across participants. The resulting log-likelihood was exponentiated prior to the probability comparison.

Five Markov chains with 250 000 samples each were run, with the first 50 000 samples removed as the burn-in period. The convergence metric, $\hat{R}$, was close to 1, indicating that the chains had converged (Brooks and Gelman, 1998). Each chain was initialized with random values for each of the 46 parameters. Future parameter values were chosen from a Gaussian distribution with a mean centred on the previous parameter value and proposal standard deviations of 0.14 for the weights, 0.06 for the population standard deviations, and 0.01 for the task context parameter. All runs had acceptance rates between 25% and 26%, and $\hat{R}$ between 0.98 and 1.05 (see Supplementary Material).

A posterior predictive model check (PPMC) was conducted to evaluate the goodness of fit of the model (Gelman *et al.*, 2014, Ch. 6, pp. 141–164). During one Markov chain, simulated data were created for the chosen hypothesis on each MCMC trial using the hypothesized participant weights and population task parameters. Using the simulated data, we looked at two measures of goodness of fit: the average PCD score (in the anatomical response demand, spatiotopic response demand, and the difference between the spatiotopic and anatomical response demands), and the correlation between the PCD scores in the anatomical and spatiotopic response demands. The average PCD score would determine whether the model provided a good fit for the overall data; the correlation between conditions would be an indicator of the model's fit for individual participants. The response demands manipulation is expected to bias all participants' weights to either the internal or the external reference frame. For this reason, we might expect a systematic difference in participants' anatomical and spatiotopic response demand PCD scores. This should result in a correlation between the scores from the two response demands. Indeed,
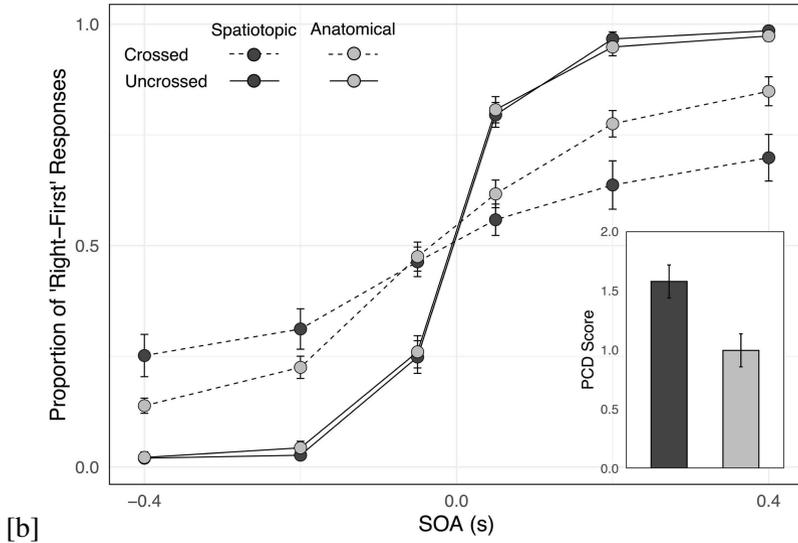
Figure 1. Proportion of right (hand for the anatomical condition, hemispace for the spatiotopic condition) first responses across stimulus onset asynchrony (SOA) from twenty participants (10 males) for both the crossed and uncrossed hand postures under both response demand conditions. Inset bar graph represents the average proportion correct difference (PCD) score for each response demand. Error bars represent standard error corrected for a within-subject design (Cousineau, 2005; Morey, 2008).

this is built into the model by having the weights for every participant in the anatomical response demand altered by the same magnitude when calculating the weights for the spatiotopic response demand. The posterior predictive model check on the correlation would therefore test whether the model and the observed data agree on this relation between participants.

2.4. Results

2.4.1. PCD Scores

PCD scores were calculated separately for each participant in both the anatomical and spatiotopic response demand trials (see Fig. 1). PCD scores were significantly smaller (i.e., a smaller crossed-hands deficit) using the anatomical response demand than the spatiotopic response demand (anatomical: $M = 1.00$, SD $= 0.51$; spatiotopic: $M = 1.58$, SD $= 0.84$; $F_{1,18} = 8.57$, $p = 0.009$, $\eta_g^2 = 0.16$). There was no significant difference in performance between males and females (male: $M = 1.15$, SD $= 0.69$; female: $M = 1.42$, SD $= 0.79$; $F_{1,18} = 1.24$, $p = 0.281$, $\eta_g^2 = 0.04$), and no significant interaction between response demand and sex ($F_{1,18} = 0.70$, $p = 0.42$, $\eta_g^2 = 0.02$). One-sample *t*-tests confirmed the presence of a crossed-hands deficit for both the anatomical ($t_{19} = 8.67$, $p < 0.001$, $d = 1.94$) and spatiotopic ($t_{19} = 8.44$,
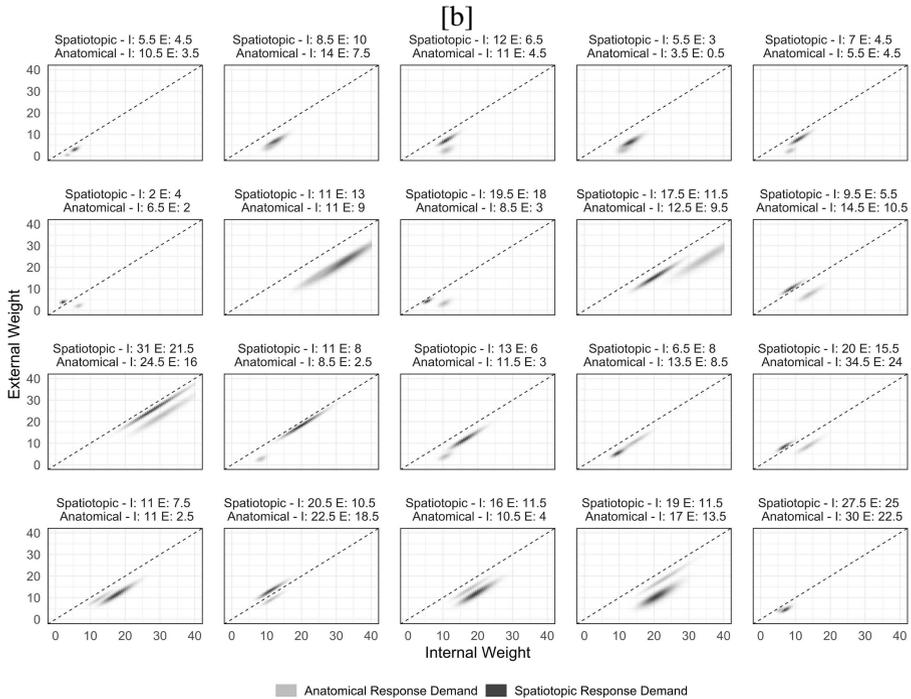
Figure 2. Joint likelihood distributions for the internal and external weights for each participant. The darker the point the greater the likelihood of the weight pair.

$p < 0.001$, $d = 1.89$) response demand. Given the lack of a sex difference in this dataset, all further analyses will not include sex as a factor.

2.4.2. Participant-Specific Model

We computed the maximum likelihood weight pair for each participant by calculating the probability of the data given each hypothesized weight pair (see Fig. 2 for the joint likelihoods of all tested weight pairs, see Fig. 3A for the maximum likelihood weight pair). Based on these weights, we calculated each participant's expected data (Fig. 4). The expected data fit well with the participants' data ($R^2 = 0.93$, p < 0.001), suggesting the internal and external weight combination successfully captures each participant's crossed and uncrossed performance. Based on the most likely weights, we computed an expected PCD score for each individual, which was highly correlated with their actual PCD score (Fig. 3B; $r = 0.97$, $p < 0.001$). Finally, we computed an average internal and external weight for the different response demand conditions (Fig. 3C). Overall, there was a higher weight placed on the internal reference frame ($F_{1,18} = 71.3$, $p < 0.001$, $\eta_g^2 = 0.1$), and no significant difference in the overall weight value between the response demands ($F_{1,18} = 0.39$, $p = 0.54$, $\eta_g^2 = 0.003$). There was an interaction between the
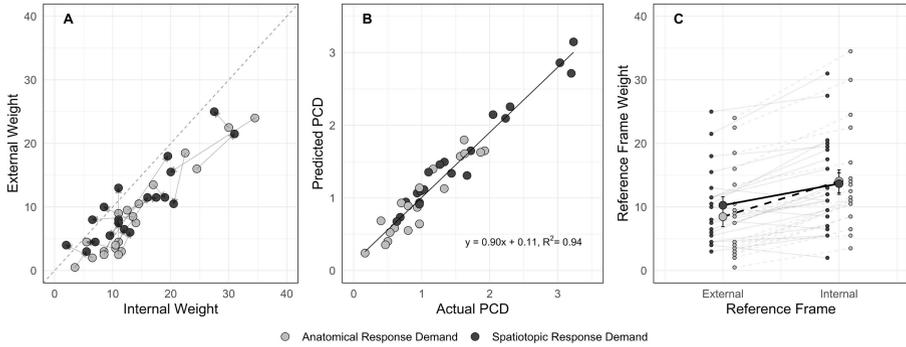
Figure 3. (A) The combination of internal and external weights that best fit the data, for each of the response demands, for each participant (connected by a line). (B) Comparison of the proportion correct difference (PCD) score obtained from the participants' raw data to the PCD score calculated from their most likely weights. (C) Overall internal and external weight in each response demand. This was found by taking the average of each participants' weights. Error bars represent standard error of the mean corrected for within-subject comparisons. The smaller circles represent the weights from individual participants.
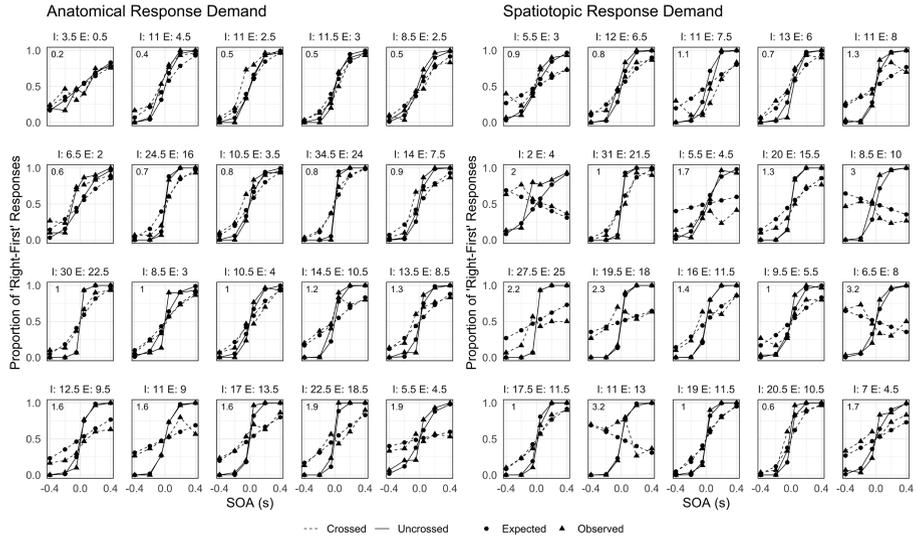
Figure 4. Individual participant performance. The circles represent the participants' raw data, and the triangles represent the expected performance calculated from their most likely weights. The numbers on top of each figure represent the participant's maximum likelihood estimated internal and external weights. The number in the top left of each figure is the participant's actual proportion correct difference (PCD) score. The graphs are sorted by PCD score in the anatomical condition.
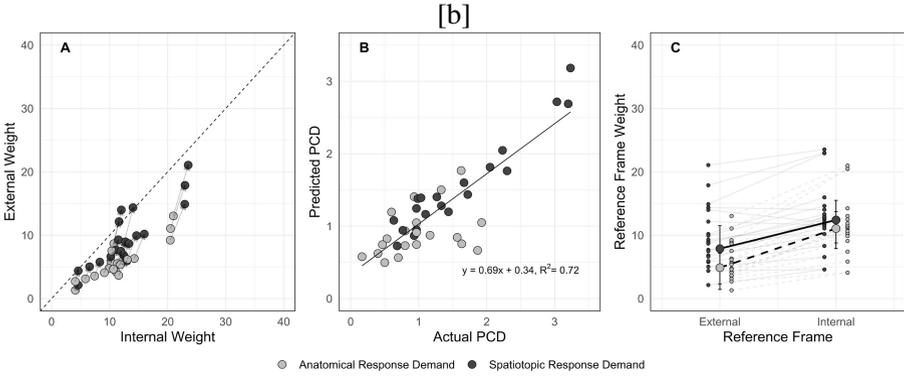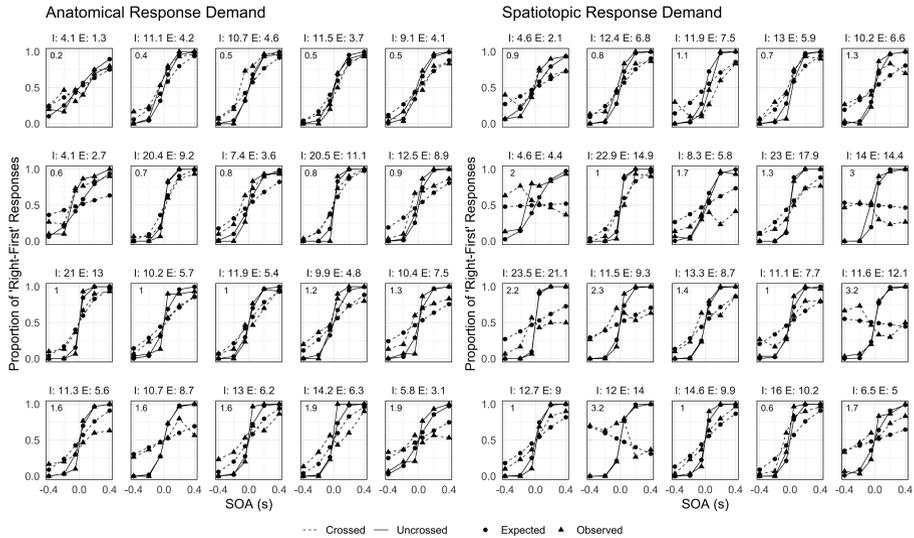
Figure 5. (A) Most likely weight for each participant in each response demand. The arrow connects the weights for each participant. (B) Comparison of the proportion correct difference (PCD) score obtained from the participants' raw data to the PCD score calculated from the most likely weights. (C) Overall population internal and external weight in each response demand. Error bars represent 95% credible intervals. The small circles represent individual participant weights.

reference frame and response demand, ($F_{1,18} = 6.23$, $p = 0.02$, $\eta_g^2 = 0.006$) such that the anatomical response demand appeared to place less weight on the external reference frame ($t_{19} = 1.53$, $p = 0.14$, $d = -0.05$) than did the spatiotopic response demand, while the internal weights between the two response demands were not significantly different ($t_{19} = -0.30$, $p = 0.77$, $d = 0.28$).

2.4.3. Hierarchical Model

The population internal and external weights were calculated for the anatomical response demand by taking the mean values across all five Markov chains. The weights in the spatiotopic response demand were calculated by taking the weights in the anatomical response demand on each trial, and multiplying it by that trial's task parameter, then averaging across the five Markov chains (Fig. 5C). This resulted in an internal population weight of 11.05 [95% credible interval (CI): 7.89, 13.73] with a population standard deviation of 5.85. The external population weight was 4.87 (95% CI: 1.43, 7.16) with a population standard deviation of 4.16 for the anatomical condition. The population internal task parameter was 1.12 (95% CI: 1.03, 1.23) and the external task parameter was 1.62 (95% CI: 1.42, 1.87), resulting in an internal weight of 12.40 and external weight of 7.87 for the spatiotopic response demand. The credible intervals for the task parameters did not include 1, indicating an increase in both reference frame weights in the spatiotopic condition.

For each participant, we estimated the internal and external weights for each condition by taking posterior means — i.e., the average value of each weight parameter across the Markov chains (Fig. 5A). Using the posterior mean weight, we computed expected data for each participant (Fig. 6). The expected

Figure 6. Individual participant performance. The circles represent the participants' raw data, and the triangles represent the performance calculated from their most likely weights in the hierarchical model. The numbers on top of each figure represent the participant's maximum likelihood estimated internal and external weights. The number in the top left of each figure is the participant's actual proportion correct difference (PCD) score. The graphs are organized by PCD score in the anatomical condition.

data provided a good fit for the participants' data ($R^2 = 0.9$, $p < 0.001$). The expected PCD scores were correlated with the participants' actual PCD scores ($r = 0.85$, $p < 0.001$; Fig. 5B). When compared to the weights obtained from the participant-specific approach (Fig. 7) these weights showed a strong correlation for the internal weight ($r = 0.89$, $p < 0.001$) and external weight ($r = 0.77$, $p < 0.001$).

The PPMC (Fig. 8) revealed that the mean from the posterior PPMC distribution matched closely with the average PCD score for each condition, meaning the model successfully captures the average PCD scores of the participants for each condition. However, the model does not reproduce the correlation between PCD scores from the two response conditions. The posterior mean of the PPMC distribution suggests a strong positive correlation between the anatomical and spatiotopic conditions ($r = 0.76$); however, this correlation is not observed in the raw data ($r = 0.215$, $p = 0.36$).

2.5. Discussion

Overall, we observed a larger crossed-hands deficit when using a spatiotopic response demand compared to an anatomical response demand. This was evident from the proportion of right-first responses, where the crossed-hands condition showed closer to chance performance under the spatiotopic response
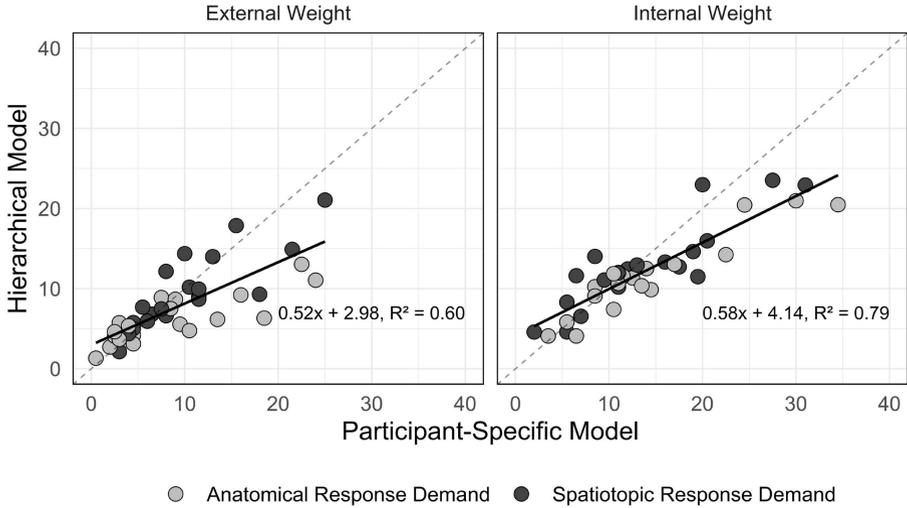
Figure 7. Comparison of the weights obtained from the maximum likelihood estimation and the hierarchical model. There was a strong correlation obtained for both the internal weight and external weight.
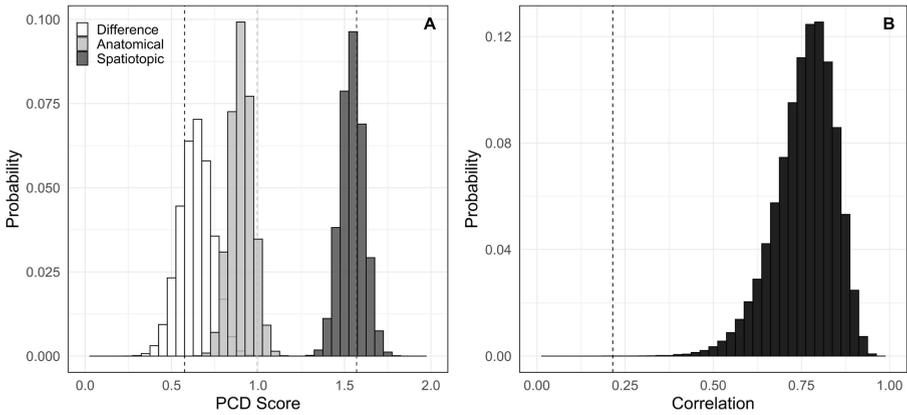
Figure 8. (A) Posterior predictive distributions of proportion correct difference (PCD) scores for the anatomical and spatiotopic conditions, and the difference between the spatiotopic and anatomical conditions. (B) Posterior predictive distribution of the correlation between the anatomical and spatiotopic PCD scores. The vertical dotted lines represent the observed values from the raw data from Experiment 1.

demand. Larger PCD scores were also observed in the spatiotopic response condition. Both behavioural analyses support our initial hypothesis of a larger deficit when the external reference frame was emphasized.

A probabilistic model was used to estimate the weights placed on the internal and external reference frame in each response demand. Using a participant-specific approach, we determined the internal and external weights for individual participants. Overall, a spatiotopic response demand resulted in a greater external weight and a slightly lower internal weight than the anatomical response demand. A hierarchical model showed that, at the population level, the internal weight increased slightly with the response demand manipulation, while the external weight was 1.5 times greater in the spatiotopic response demand. This was similar to the changes observed in the participant-specific approach. The hierarchical model more accurately reflects the true relation between the conditions, as the population parameter estimates are based on more information than just an average of the participant values.

The participant-specific model used a maximum likelihood technique to determine the probability of the data given all combinations of internal and external weight pairs. By looking at the weight pairs with higher likelihoods, it is evident that the difference between the internal and external weights remains constant. Given that the crossed-hands curve is estimated as the difference between the weights, the crossed posture seems to constrain the weights that are plausible. In contrast, the uncrossed posture is fitted based on the sum of the weights. This posture typically results in steeper slopes, resulting in many sums that can give rise to similar psychometric functions.

The hierarchical model provided a good fit for the participants' data, but not as good a fit as the participant-specific model. This is expected of a hierarchical model, because in such a model the population parameters relate the participants to one another, refining the inference about each participant based on the data from all the others. Consequently, the inference regarding the true parameter values of each participant depends on more information than merely the data from that one participant. The hierarchical model assumed that the response demands manipulation would affect each participant to the same degree. Therefore, the participants' weights in the anatomical response demand were multiplied by the corresponding population task parameter to obtain their weights for the spatiotopic response demand. Provided that this model's structure realistically reflects the effect of the response demand manipulation, parameter estimates from the hierarchical model will be more robust than the participant-specific estimation against noise in the individual participant's data.

Posterior predictive model checks revealed the average PCD score of the model was similar to the average PCD score from the observed data. This would suggest that the hierarchical model structure realistically captures that aspect of the observed data. In contrast, PPMC applied to the correlation between the anatomical and spatiotopic response demand PCD scores reveals a poor fit with the observed data. In the observed data there is a small positive

(non-significant) correlation between the two response conditions, while the hierarchical model consistently predicts a moderate to strong positive correlation. The strong correlation in the hierarchical model is likely a by-product of the population task parameter. Given that the weights for every participant in the spatiotopic response demand are multiplied by the same values, this is perhaps predicting a cleaner relation between the two response conditions than actually exists. Because the external task parameter is greater than the internal task parameter, the model requires there to be a larger crossing effect in the spatiotopic response demand than in the anatomical response demand. While this is the case for the majority of participants, five participants showed the opposite effect. It is likely that these few participants are driving the low correlation in the observed data. To test this, we checked the magnitude of the correlation with these few participants removed ($r = 0.84$, $p < 0.001$), and it is similar to the correlation predicted by the PPMC. It is possible that these participants are performing the task using a different strategy, and these participants might be better fitted with a different model. Future studies could explore additional model variants that can accommodate individual differences in the relation between response conditions.

3. Experiment 2

Next, we wanted to replicate the model results using a different dataset. This dataset was chosen because of its larger size and more homogeneous sample (only right-handed females). The identical task to that used in Experiment 1 was completed as part of a larger experiment investigating the relation between body image, the rubber hand illusion, and the crossed-hands deficit. For the purpose of this paper, we will only be focusing on the crossed-hands tactile TOJ portion.

3.1. Methods

3.1.1. Participants
Forty-seven right-handed, female participants (average age: 18.4 years), were recruited from the McMaster University subject pool. Four participants were removed for not following the task instructions. All had normal or corrected-to-normal vision, were naïve to the experiment, and gave written informed consent before participation. All procedures were approved by the McMaster Research Ethics Board and complied with the tri-council statement on ethics (Canada).

3.1.2. Apparatus and Stimuli
Apparatus and stimuli were identical to those in Experiment 1.
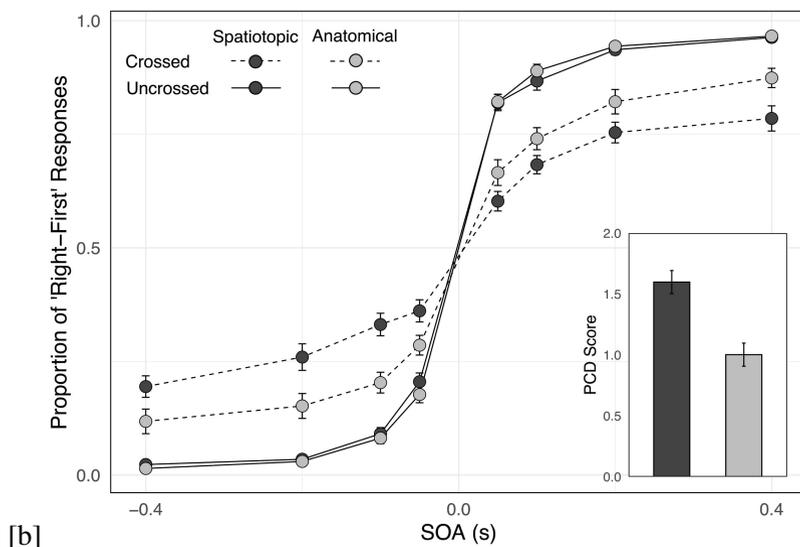
Figure 9. Proportion of right-first (hand for the anatomical condition, hemispace for the spatiotopic condition) responses across stimulus onset asynchrony (SOA) from forty-three female participants for both the crossed and uncrossed hand postures under both response demand conditions. Inset bar graph represents the average proportion correct difference (PCD) score for each response demand. Error bars represent standard error corrected for a within-subject design (Cousineau, 2005; Morey, 2008).

3.1.3. Procedure
The procedure was identical to that in Experiment 1, except two additional SOAs were tested (±100 ms). A total of 222 time-out and premature trials were removed.

3.1.4. Analysis
The analyses were identical to those in Experiment 1, except the ANOVA on PCD score did not include the between-subject factor of sex.

3.2. Results

3.2.1. PCD Scores
PCD scores were calculated separately for each participant in both the anatomical and spatiotopic response demand (see Fig. 9). PCD scores were significantly smaller in the anatomical response demand than the spatiotopic response demand (anatomical: $M = 1.00$, SD $= 0.89$; spatiotopic: $M = 1.60$, SD $= 0.86$; $t_{43} = 4.45$, $p < 0.001$, $d = 0.68$). The crossed-hands deficit was reduced by the use of an anatomical response demand. One-sample t-tests revealed a crossed-hands deficit in both the anatomical ($t_{43} = 7.34$, $p < 0.001$, $d = 1.12$) and spatiotopic ($t_{43} = 12.23$, $p < 0.001d = 1.87$) response demand conditions.

3.2.2. Participant-Specific Model

The maximum likelihood estimate for each participant was the weight pair with the highest likelihood (see Fig. 10 for the joint posterior probability of all tested weight pair, see Fig. 11A for the maximum likelihood weight pair). Using these weights we calculated each participant's expected responses (Fig. 12). The expected responses fit well with the participant's data ($R^2 = 0.93$, $p < 0.001$), suggesting the internal and external weight combination successfully captures participant performance. Based on the most likely weights, we computed an expected PCD score (Fig. 11B). The expected PCD scores were highly correlated with the participants' actual PCD scores ($r = 0.96$, $p < 0.001$). We calculated an average internal and external weight for each response demand (Fig. 11C). Overall, a higher weight was placed on the internal reference frame ($F_{1,42} = 129.82$, $p < 0.001$, $\eta_g^2 = 0.29$). There was no significant difference in the weights between the different response demands ($F_{1,42} = 3.82$, $p = 0.06$, $\eta_g^2 = 0.01$). An interaction between reference frame and response demand was observed ($F_{1,42} = 26.54$, $p < 0.001$, $\eta_g^2 = 0.04$), such that when switching from an anatomical to a spatiotopic response, less weight was placed on the internal reference frame ($t_{19} = -4.62$, $p < 0.001$, $d = -0.51$) while the weights placed on the external reference frame between two response demands were not significantly different ($t_{19} = 1.37$, $p = 0.18$, $d = 0.23$).

3.2.3. Hierarchical Model

Each hypothesis was initialized with random values for each of its 92 parameters. The additional parameters are due to the increased number of participants. The same parameters from Experiment 1 were used for each participant and the population. The population internal weight was 16.00 (95% CI: 13.58, 18.17) with a population standard deviation of 6.80. The external population weight was 7.01 (95% CI: 5.47, 8.30) with a population standard deviation of 3.55 for the anatomical condition (Fig. 13C). The population internal task parameter was 0.77 (95% CI: 0.73, 0.81), and the external task parameter was 1.00 (95% CI: 0.91, 1.10), leading to an internal weight of 12.27 and an external weight of 7.00 for the spatiotopic condition. The credible interval on the internal task parameter was below 1, indicating a decreased internal weight for the spatiotopic response demand, while the credible interval of the external task parameter included 1, implying no change in this weight.

The highest probability internal and external weights for each participant were estimated by the posterior means (Fig. 13A). The expected data computed from these weights were a good fit for the participants data ($R^2 = 0.92$, $p < 0.001$; Fig. 14). The participant's observed PCD score was correlated with the expected PCD score ($r = 0.86$, $p < 0.001$; Fig. 13B). The internal
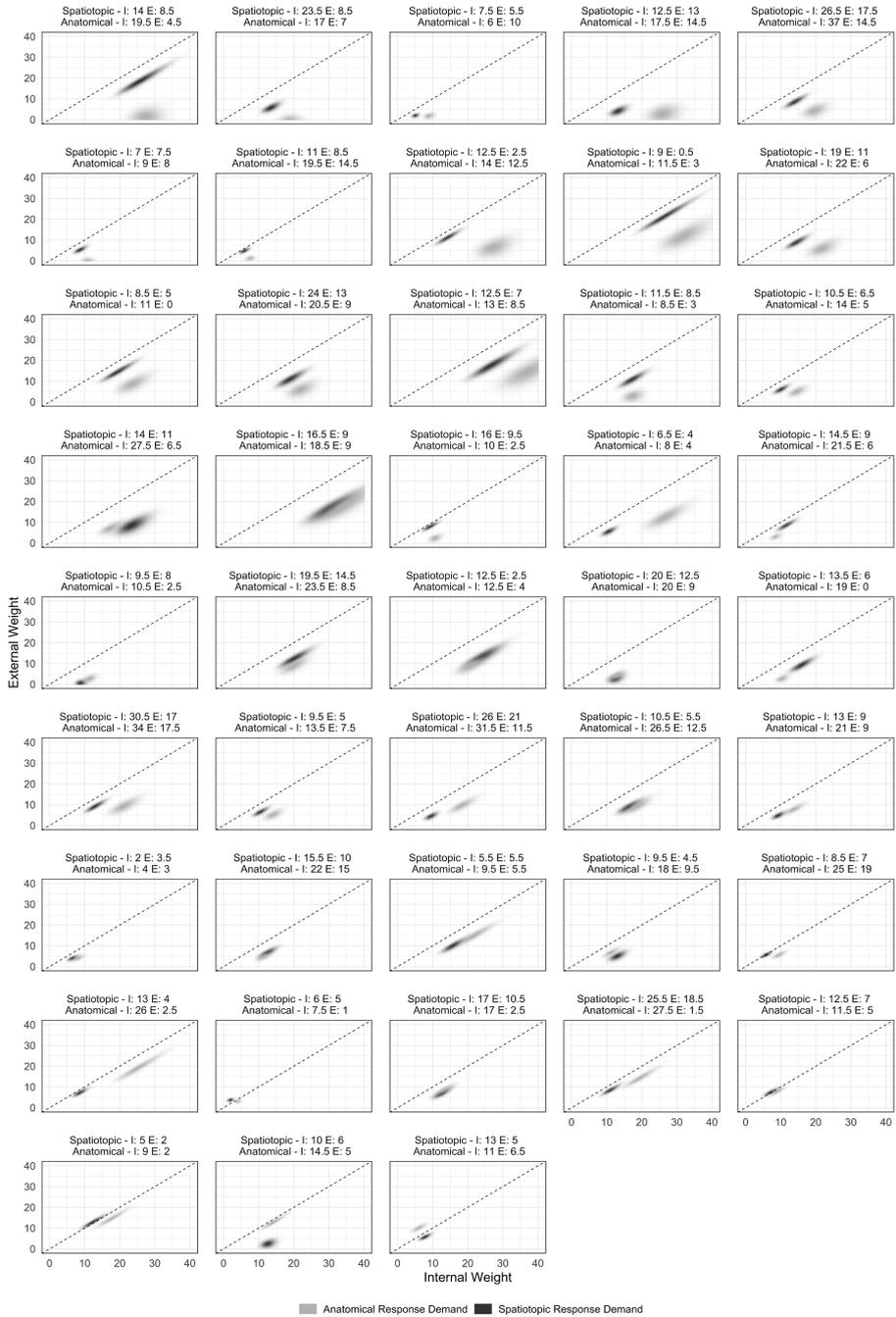
Figure 10. Joint likelihood distributions for the internal and external weights for each participant. The darker the point the greater the likelihood of the weight pair.
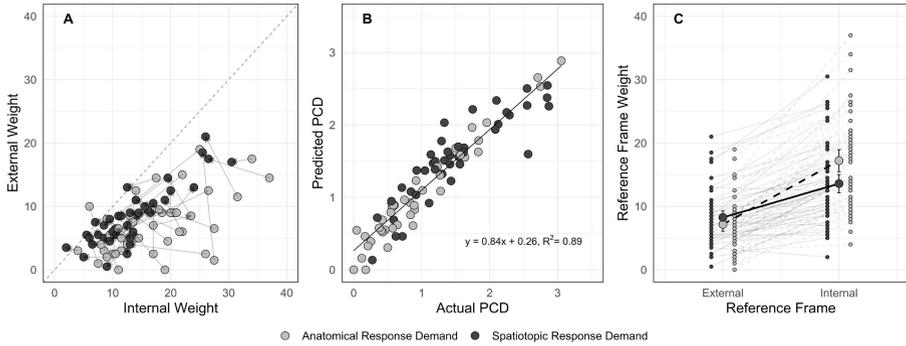
Figure 11. (A) The combination of internal and external weights that best fit the data, for each of the response demands, for each participant (connected by a line). (B) Comparison of the proportion correct difference (PCD) score obtained from the participants' raw data to the PCD score calculated from their most likely weights. (C) Overall internal and external weight in each response demand. This was found by taking the average of each participants' weights. Error bars represent standard error of the mean corrected for within-subject comparisons. The smaller circles represent the weights from individual participants.

weight ($r = 0.93$, $p < 0.001$) and external weight ($r = 0.77$, $p < 0.001$) were strongly correlated with the participant-specific weights (Fig. 15).

The PPMC successfully captured the average PCD score, but not the correlation between the two response demands (Fig. 16). The observed data show a moderately positive correlation between the anatomical and spatiotopic response conditions ($r = 0.50$, $p < 0.001$), while the model predicts a stronger correlation between conditions ($r = 0.79$).
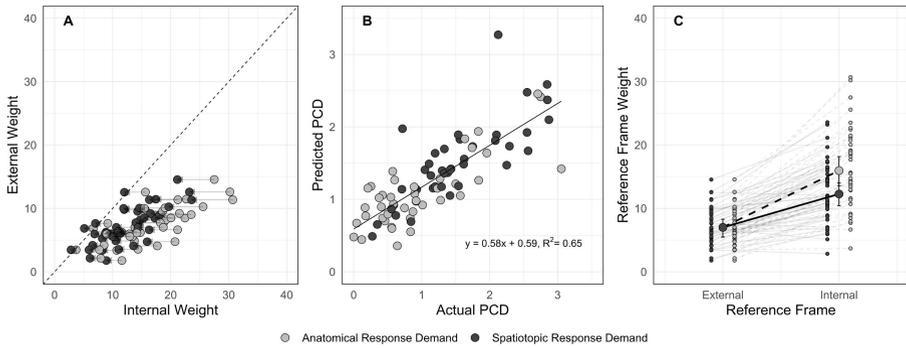
3.3. Discussion

In Experiment 2, we replicated Experiment 1: a larger crossed-hands deficit was observed using a spatiotopic response demand compared to an anatomical response demand. This was measured through a larger PCD score in the spatiotopic condition. This supports the theory that the spatiotopic response demand emphasized the external reference frame.

When using a participant-specific model to determine the internal and external weight for each response demand using a maximum likelihood estimation, both the spatiotopic and anatomical conditions had similar external weights; the internal weight was lower in the spatiotopic condition compared to the anatomical response demand. The hierarchical model replicated the results from the participant-specific model, whereby the external weight remained the same in both conditions and the internal weight was 1.3 times lower for the spatiotopic response demand.
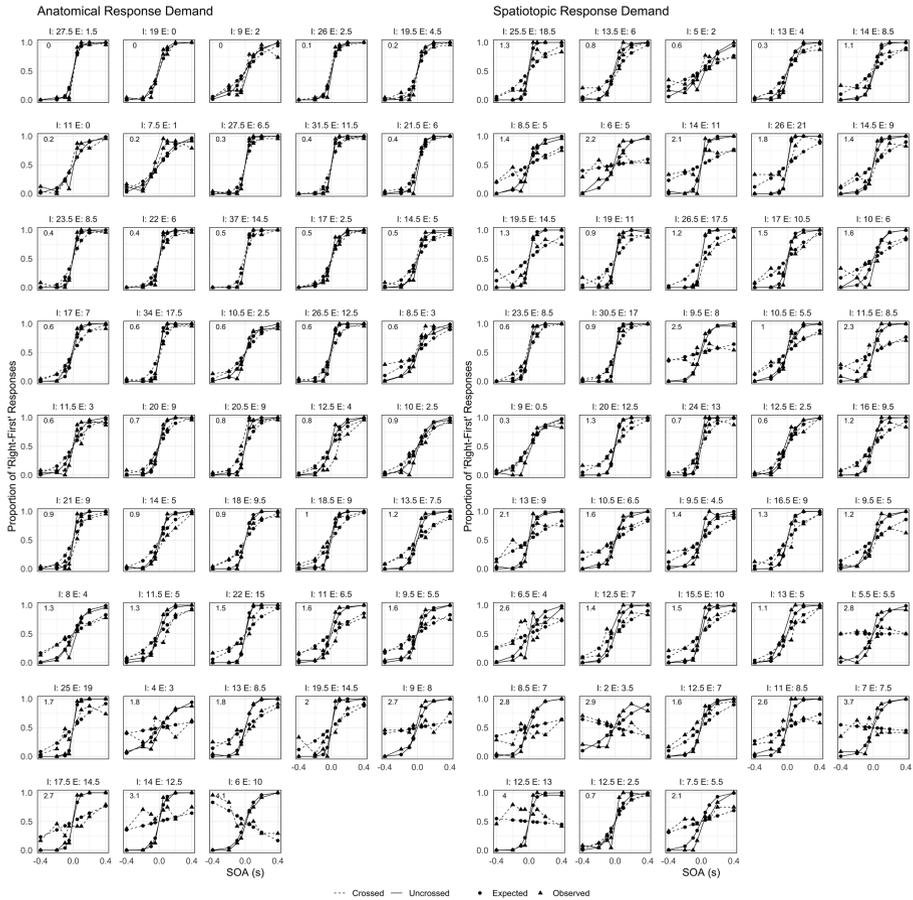
Figure 12. Individual participant performance from Experiment 2. The circles represent the participants' raw data, and the triangles represent the expected performance calculated from their most likely weights. The numbers on top of each figure represent the participant's maximum likelihood estimated internal and external weights. The number in the top left of each figure is the participant's actual proportion correct difference (PCD) score. The graphs are sorted by PCD score in the anatomical condition.

Posterior predictive checks on the average PCD score revealed that the model provided a good estimate of the participant's average PCD score. However, the model again predicted a stronger correlation between the participants' anatomical and spatiotopic response demand PCD scores than was observed in the data. If the eight participants who revealed a smaller crossed-hands deficit in the spatiotopic condition were removed, the observed correlation ($r = 0.79$, $p < 0.001$) more closely matched the predicted correlation from the PPMC.

Figure 13. (A) Most likely weights for each participant in each response demand. The arrow connects the weights for each participant. (B) Comparison of the proportion correct difference (PCD) score obtained from the participants' raw data to the PCD score calculated from the most likely weights. (C) Overall population internal and external weight in each response demand. Error bars represent 95% credible intervals. The small circles represent individual participant weights.

4. General Discussion

Across two experiments we observed a larger crossed-hands deficit under a spatiotopic response demand compared to an anatomical response demand. This effect was measured using the PCD score (a measure of performance difference between the uncrossed and crossed postures), which was larger for the spatiotopic response demand compared to the anatomical response demand. These results replicated previous studies using this manipulation (Cadieux and Shore, 2013; Crollen *et al.*, 2019). The spatiotopic condition requires responses to be made in external spatial coordinates. This has led to the hypothesis that the spatiotopic response demand should place more emphasis on the external reference frame. The fact that these behavioural measures showed worse performance in the spatiotopic response demand supported this hypothesis.

To measure the weight placed on the internal and external reference frame, we employed a modified version of the model designed by Badde *et al.* (2016). We first fitted each participant's data based on their equations for creating the psychometric functions, using a participant-specific model. Next, we employed a modified version of the hierarchical model. The participant-specific approach fits each participant independently. The only estimate of the population weights from this technique is through the mean of the participant weights. While this results in a slightly better fit for the individual participant data, it also increases the chance of the parameter fits being influenced by noise in the individual data. This technique is appropriate if the participants' weights are in fact statistically unrelated, such that the parameter estimates
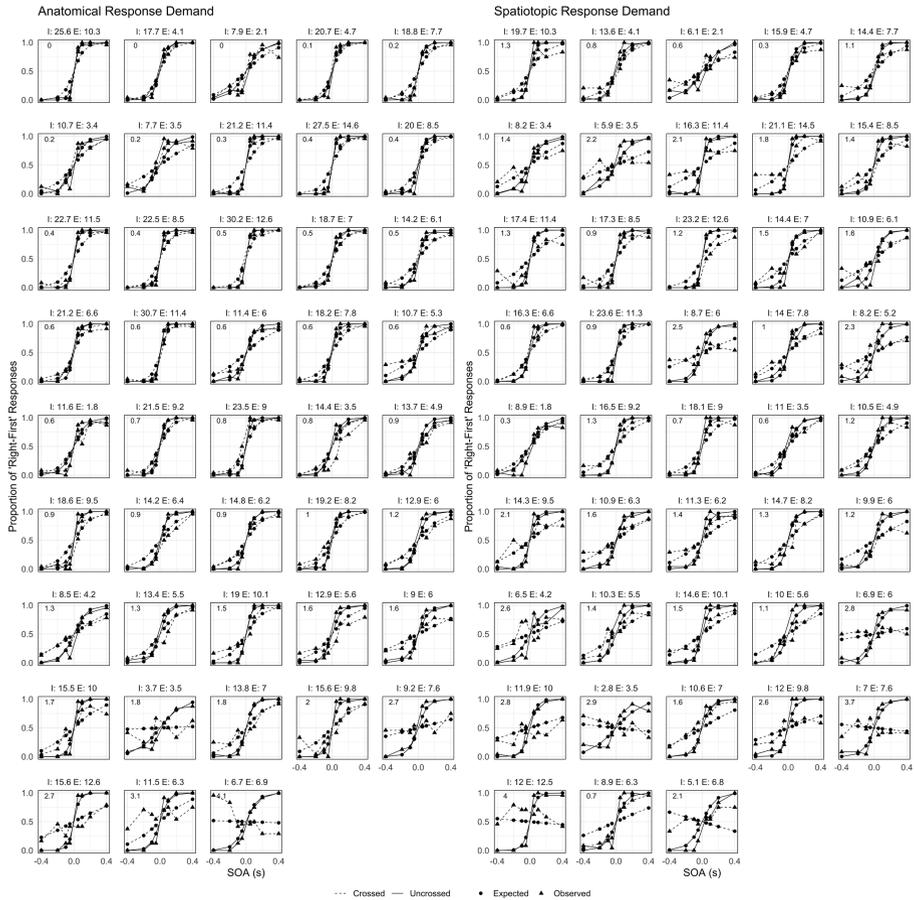
Figure 14. Individual participant performance from Experiment 2. The circles represent the participants' raw data, and the triangles represent the performance calculated from their most likely weights in the hierarchical model. The numbers on top of each figure represent the participant's maximum likelihood estimated internal and external weights. The number in the top left of each figure is the participants actual proportion correct difference (PCD) score. The graphs are organized by PCD score in the anatomical condition.

for each participant can be based on that participant's data alone. The hierarchical model, in contrast, assumes that participants' weights come from a Gaussian population distribution, with unknown mean and standard deviation parameters. Therefore, each participant's data contribute by influencing the population parameter fits, to the parameter estimates of the other participants. This results in the individual parameter estimates being less swayed by noise in the individual participant data, and also allows for a more sophisticated estimation of the population weights.
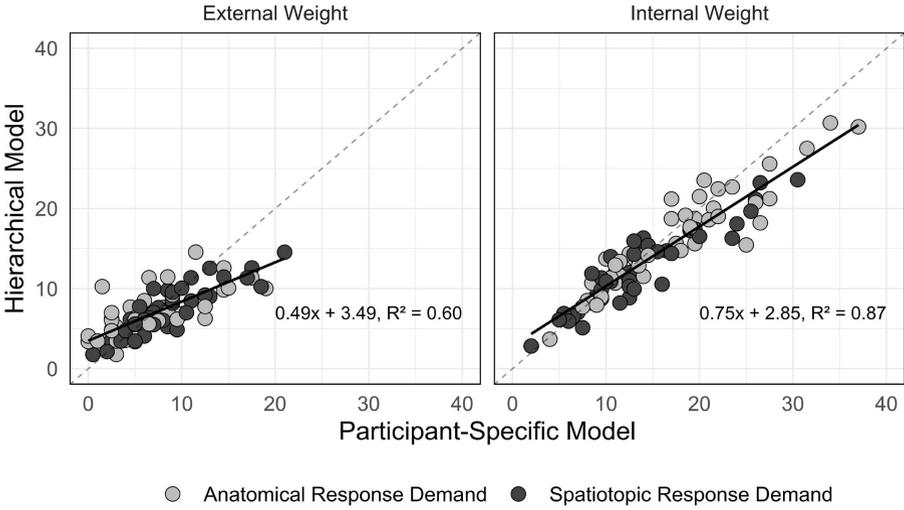
Figure 15. Comparison of the weights obtained from the maximum likelihood estimation and the hierarchical model for Experiment 2. There was a moderate correlation obtained for both the internal weight and external weight.
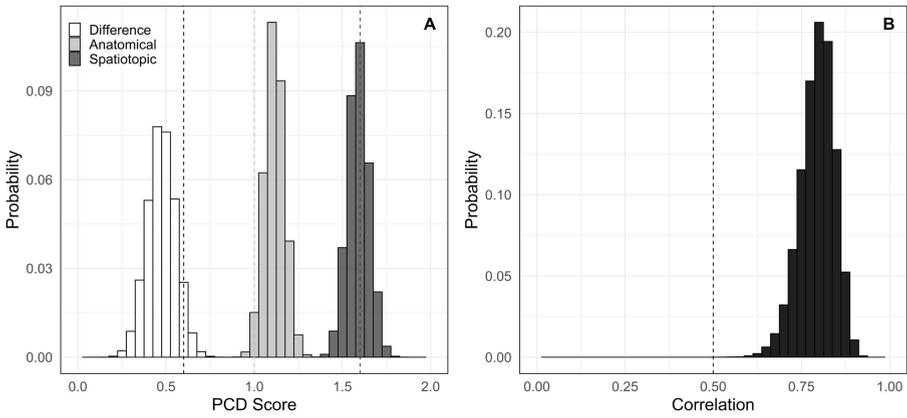
Figure 16. (A) Posterior predictive distributions of proportion correct difference (PCD) scores for the anatomical and spatiotopic conditions, and the difference between the spatiotopic and anatomical conditions. (B) Posterior predictive distribution of the correlation between the anatomical and spatiotopic PCD scores. The vertical dotted lines represent the observed values from the raw data from Experiment 2.

Both the participant-specific model and the hierarchical model provided similar results for each experiment. For Experiment 1, both methods revealed a larger external weight for the spatiotopic response demand compared to the anatomical response demand, while the internal weight remained unchanged. In Experiment 2, the spatiotopic condition resulted in a smaller internal weight

compared to the anatomical response demand, while the external weights remained the same. The slopes of the crossed-hands conditions are computed as the difference between the internal and external weight; as such, an increase in the external weight or an equivalent decrease in the internal weight will result in the same slope. Given that each experiment showed worse crossed-hands performance in the spatiotopic condition than the anatomical condition, both options are able to fit the crossed-hands data. The difference between these two options is only evident in the slopes of the uncrossed psychometric functions. Since the uncrossed posture is created from the sum of the internal and external weights, an increase in the external weight results in a slightly steeper uncrossed slope for the spatiotopic response demand, whereas a decreased internal weight would result in a shallower uncrossed slope. While the uncrossed performance in the two experiments was similar, in Experiment 1 the spatiotopic condition had slightly better uncrossed performance compared to the anatomical condition, which the model attributes to an increased external weight. In contrast, the spatiotopic condition had slightly worse uncrossed performance compared to the anatomical condition in Experiment 2, which the model attributes to a decreased internal weight.

The participant parameters from the participant-specific model and the hierarchical model were slightly different (Figs 6 and 13), as a result of the additional population parameters, specifically the population standard deviation. As the participant-specific approach fits all participants independently, this technique assumes there is no relation between participants, or that the standard deviation is extremely large. In the hierarchical model, as the estimated standard deviation approaches infinity, the weights would become equivalent to those in the participant-specific approach. Here, for both participant parameters, higher weights in the participant-specific approach were slightly smaller in the hierarchical model, and lower weights were slightly larger in the hierarchical model compared to the participant-specific approach. The smaller range of participant weights estimated by the hierarchical model suggests that the participants' weights are indeed related via a population distribution with finite standard deviation.

At the individual level, not all participants showed the same trend of worse performance in the spatiotopic response demand. A small subset of participants in each experiment showed the opposite, a smaller deficit in the spatiotopic response demand. There were no commonalities regarding these participants' performance or the order in which they completed the response demands. One possibility is that these participants more successfully ignored the external reference frame in the spatiotopic condition than the anatomical condition. Alternatively, these participants could be performing the task differently than the rest. The spatiotopic response demand requires participants to locate the tactile stimulus in external coordinates (hemispace instead

of hand). In this condition, instead of a direct mapping from the location of the hand to the response, there is a direct mapping from hemispace to response. Given that the response matches the location of the stimulus in external space, the more a person can focus and utilize the external reference frame, the better their performance should be. The conceptualization for the anatomical response demand task remains the same. The critical difference between the anatomical and spatiotopic response demand, based on this conceptualization, is which reference frame the participant must ignore when the hands are crossed in order to respond correctly. In the anatomical condition the response is mapped internally, therefore ignoring the external reference frame results in better performance. In the spatiotopic condition where the response is mapped externally, ignoring the internal reference frame, and focusing on external information, would improve performance. It is possible that some participants are conceptualizing the spatiotopic response demand in this way. This hypothesis for how individuals are performing the task would require slight modifications to the equations for constructing the spatiotopic psychometric curves.

This may also be causing the low correlation between the anatomical and spatiotopic response demand PCD score in the observed data. When the hierarchical model fits all participants using the same strategy, a large correlation is observed as performance in the spatiotopic condition always has a larger PCD score than the anatomical condition. When the participants showing the opposite effect are removed, the observed correlation is much closer to the correlation estimated by the PPMC.

Given that participants might use different strategies, the use of a single performance model for all participants might be a limitation. The ability to assign different participants to different performance models might help differentiate which participants are using similar strategies. This could be implemented as another level in the hierarchy. Future studies with more explicit instructions are needed to better understand the different strategies that may be used on this task. One such instruction could be to ask participants to locate the stimulus based on the hemispace, rather than the hand of the vibration in the spatiotopic response demand. This instruction explicitly ties the response to external coordinates; therefore, if participants show a smaller deficit under these new instructions it would suggest some participants in the original study were adopting the strategy of responding based on the hemispace.

The results from this study support previous research showing that task instructions influence the weights placed on the internal and external reference frames. In one study, participants received one low and one high frequency vibration, one on each hand (Badde et al., 2015). The participant had to make two responses to the vibrations. First, the participants indicated the

hand that received the first stimulus. After making the temporal response, participants were asked to determine the location of the stimulus of a certain frequency (either high or low depending on the participants). This secondary response used either internal instructions (location tied to the hand) or external instructions (location tied to a side of space). A smaller deficit was observed under internal compared to external instructions. Even though the task instructions only affected the second response, performance on the primary temporal response was altered by the task instructions, showing that task instructions result in a reweighting of internal and external information.

Task instructions have also been shown to affect performance during a tactile congruency task (Gallace *et al.*, 2008; Schubert *et al.*, 2017). In this task, participants had to locate a tactile target on the hand while ignoring a tactile distractor presented on the opposite hand. Under internal instructions participants located the target based on where on the hand it occurred; external instructions had participants indicate the target location relative to gravity. Accuracy was higher when the distractor occurred at a congruent compared to incongruent location, however what was considered congruent changed based on the instructions. Under internal instructions congruency was judged anatomically (e.g., target and distractor on palm), while under external instructions congruency was based on gravity (e.g., target and distractor at upper location). The weights applied to each reference frame are affected by the task instructions.

Both the behavioural data, as well as the model results, support integration accounts for the deficit (Badde *et al.*, 2016; Shore *et al.*, 2002), as opposed to non-integration models (Yamamoto and Kitazawa, 2001). Integration models posit that both reference frames are used when localizing the tactile stimulus, and the different weights placed on each reference frame determine the final perceived location. In the crossed posture when the external reference frame is more heavily relied on, this can sometimes lead to erroneous localization. According to integration accounts, a larger crossed-hands deficit will occur when an individual places greater weight on the external reference frame, and will decrease as weight is transferred to the internal reference frame. This was supported by the results of the present study, where the anatomical and spatiotopic response demand manipulations biased participants towards the internal and external reference frames respectively. As a result, a larger crossing effect was observed under a spatiotopic response demand. Modelling revealed that the spatiotopic condition caused either a greater external weight, or a smaller internal weight. Both options placed a relatively larger emphasis on external information. Other manipulations to the crossed-hands tactile TOJ task have shown support for an integration of the reference frames. Altering visual information through blindfolding (Cadieux and Shore, 2013), placing the hands behind the back (Kóbor *et al.*, 2006), or viewing uncrossed hands

(Azañón and Soto-Faraco, 2007), results in a smaller crossed-hands deficit, presumably by removing conflicting external information. Furthermore, congenitally blind individuals do not show a crossed-hands deficit (Crollen *et al.*, 2019; Röder *et al.*, 2004) unless the response modality emphasizes the external reference frame (Crollen *et al.*, 2019) suggesting they do not automatically integrate internal and external reference frames.

The use of this probabilistic model allowed direct exploration about how various manipulations affect the use of each reference frame. This provides a deeper understanding of how information is weighted when locating a touch. Without this model, the theoretically construed weights could only be inferred based on the size of the deficit. Future studies could apply this model to other manipulations assumed to influence reference frame weights (i.e., visual information) in order to test these assumptions quantitatively.

Acknowledgements

Supplementary Material

Supplementary material is available online at:
https://doi.org/10.6084/m9.figshare.14872695

References

Azañón, E. and Soto-Faraco, S. (2007). Alleviating the 'crossed-hands' deficit by seeing uncrossed rubber hands, *Exp. Brain Res.* **182**, 537–548. DOI:10.1007/s00221-007-1011-3.

Azañón, E., Stenner, M.-P., Cardini, F. and Haggard, P. (2015). Dynamic tuning of tactile localization to the body posture, *Curr. Biol.* **25**, 512–517. DOI:10.1016/j.cub.2014.12.038.

Azañón, E., Mihaljevic, K. and Longo, M. R. (2016). A three-dimensional spatial characterization of the crossed-hands deficit, *Cognition* **157**, 289–295. DOI:10.1016/j.cognition.2016.09.007.

Badde, S. and Heed, T. (2016). Towards explaining spatial touch perception: weighted integration of multiple location codes, *Cogn. Neuropsychol.* **33**, 26–47. DOI:10.1080/02643294.2016.1168791.

Badde, S., Röder, B. and Heed, T. (2015). Flexibly weighted integration of tactile reference frames, *Neuropsychologia* **70**, 367–374. DOI:10.1016/j.neuropsychologia.2014.10.001.

Badde, S., Heed, T. and Röder, B. (2016). Integration of anatomical and external response mappings explains crossing effects in tactile localization: a probabilistic modeling approach, *Psychon. Bull. Rev.* **23**, 387–404. DOI:10.3758/s13423-015-0918-0.

Brooks, S. P. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations, *J. Comput. Graph. Stat.* **7**, 434–455. DOI:10.1080/10618600.1998.10474787.

Cadieux, M. L. and Shore, D. I. (2013). Response demands and blindfolding in the crossed-hands deficit: an exploration of reference frame conflict, *Multisens. Res.* **26**, 465–482. DOI:10.1163/22134808-00002423.

Cadieux, M. L., Barnett-Cowan, M. and Shore, D. I. (2010). Crossing the hands is more confusing for females than males, *Exp. Brain Res.* **204**, 431–446. DOI:10.1007/s00221-010-2268-5.

Cousineau, D. (2005). Confidence intervals in within-subject designs: a simpler solution to Loftus and Masson's method, *Tutor. Quant. Methods Psychol.* **1**, 42–45. DOI:10.20982/tqmp.01.1.p042.

Craig, J. C. and Belser, A. N. (2006). The crossed-hands deficit in tactile temporal-order judgments: the effect of training, *Perception* **35**, 1561–1572. DOI:10.1068/p5481.

Crollen, V., Albouy, G., Lepore, F. and Collignon, O. (2017). How visual experience impacts the internal and external spatial mapping of sensorimotor functions, *Sci. Rep.* **7**, 1022. DOI:10.1038/s41598-017-01158-9.

Crollen, V., Spruyt, T., Mahau, P., Bottini, R. and Collignon, O. (2019). How visual experience and task context modulate the use of internal and external spatial coordinate for perception and action, *J. Exp. Psychol. Hum. Percept. Perform.* **45**, 354–362. DOI:10.1037/xhp0000598.

Gallace, A., Soto-Faraco, S., Dalton, P., Kreukniet, B. and Spence, C. (2008). Response requirements modulate tactile spatial congruency effects, *Exp. Brain Res.* **191**, 171–186. DOI:10.1007/s00221-008-1510-x.

Gelman, A., Carlin, J. B., Stern, H. S., Dunsen, D. B., Rubin, D. B. and Vehtari, A. (2014). *Bayesian Data Analysis*, 3rd edn. Chapman & Hall/CRC, Boca Raton, FL, USA.

Heed, T. and Azañón, E. (2014). Using time to investigate space: a review of tactile temporal order judgments as a window onto spatial processing in touch, *Front. Psychol.* **5**, 76. DOI:10.3389/fpsyg.2014.00076.

Kóbor, I., Füredi, L., Kovács, G., Spence, C. and Vidnyánszky, Z. (2006). Back-to-front: improved tactile discrimination performance in the space you cannot see, *Neurosci. Lett.* **400**, 163–167. DOI:10.1016/j.neulet.2006.02.037.

Morey, R. D. (2008). Confidence intervals from normalized data: a correction to Cousineau (2005), *Tutor. Quant. Methods Psychol.* **4**, 61–64.

Pagel, B., Heed, T. and Röder, B. (2009). Change of reference frame for tactile localization during child development, *Dev. Sci.* **12**, 929–937. DOI:10.1111/j.1467-7687.2009.00845.x.

Roberts, R. D. and Humphreys, G. W. (2008). Task effects on tactile temporal order judgments: when space does and does not matter, *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 592–604. DOI:10.1037/0096-1523.34.3.592.

Röder, B., Rösler, F. and Spence, C. (2004). Early vision impairs tactile perception in the blind, *Curr. Biol.* **14**, 121–124. DOI:10.1016/j.cub.2003.12.054.

Schicke, T. and Röder, B. (2006). Spatial remapping of touch: confusion of perceived stimulus order across hand and foot, *Proc. Natl Acad. Sci. U. S. A.* **103**, 11808–11813. DOI:10.1073/pnas.0601486103.

Schubert, J. T. W., Badde, S., Röder, B. and Heed, T. (2017). Task demands affect spatial ref-
  erence frame weighting during tactile localization in sighted and congenitally blind adults,
  PLoS ONE **12**, e0189067. DOI:10.1371/journal.pone.0189067.

Shore, D. I., Spry, E. and Spence, C. (2002). Confusing the mind by crossing the hands, *Cogn.
  Brain Res.* **14**, 153–163. DOI:10.1016/s0926-6410(02)00070-8.

Shore, D. I., Gallace, A., Mimnagh, K. and Spence, C. (2006). *Assessing the Frames of Refer-
  ence Involved in the Crossed Hands Temporal Order Judgments Deficit: the Role of Response
  Demands.* Presented at 7th International Multisensory Research Forum, Dublin, Ireland.
  http://imrf.mcmaster.ca/2006/viewabstract.php%3fid=186&symposium=0.html.

Unwalla, K., Kearney, H. and Shore, D. I. (2020). Reliability of the crossed-hands deficit in
  tactile temporal order judgements, *Multisens. Res.* **34**, 387–421. DOI:10.1163/22134808-
  bja10039.

Wada, M., Suzuki, M., Takaki, A., Miyao, M., Spence, C. and Kansaku, K. (2014). Spatio-
  temporal processing of tactile stimuli in autistic children, *Sci. Rep.* **4**, 5985. DOI:10.1038/
  srep05985.

Yamamoto, S. and Kitazawa, S. (2001). Reversal of subjective temporal order due to arm cross-
  ing, *Nat. Neurosci.* **4**, 759–765. DOI:10.1038/89559.