Crossing effects in the tactile temporal order judgment task: A meta-analysis

Nina Schmelter[1], Benedikt Langenberg[2], Axel Mayer[1], Tobias Heed[3,4]

[1] Department of Psychology, Faculty of Psychology and Sports Science, Bielefeld University, Bielefeld, Germany

[2] Department of Methodology and Statistics, Maastricht University, Netherlands

[3] Department of Psychology, University of Salzburg, Salzburg, Austria

[4] Centre for Cognitive Neuroscience, University of Salzburg, Salzburg, Austria

Correspondence

T.H., tobias.heed@plus.ac.at

Acknowledgments

We thank Tine-Marie Wujciak for helpful discussion during conceptualization, data curation, and formal analysis.

Author Contributions

Contributions according to https://credit.niso.org

NS: conceptualization, data curation, formal analysis, investigation, methodology, validation, visualization, writing – original draft

BL: formal analysis

AM: conceptualization, methodology, formal analysis, project administration, supervision, validation, writing – original draft

TH: conceptualization, project administration, supervision, writing – original draft

Abstract

The tactile temporal order judgment (TOJ) task is widely used in multisensory neuroscience. Participants judge which of two tactile stimuli, one on each hand, came first. A key finding is that TOJ performance declines when the arms are crossed, likely due to interactions between tactile, proprioceptive, and visual information. The TOJ crossing effect has been widely reported, but studies have employed various analysis methods, leaving open whether the choice of method influences the effect's magnitude. Moreover, some studies have reported modulations by the availability of visual information, response modality, or speeded response requirements. Through an exhaustive, systematic literature search, we identified 37 experiments that investigated the TOJ crossing effect. Meta-analysis estimated the effect size at approximately 1.4. The moderators did not significantly affect this estimate, though this is likely due to too few studies being available to obtain sufficient statistical power. The large effect size supports the TOJ task's use in other, e.g. developmental and clinical, research. However, the lack of statistically significant effect modulation by the moderators calls for caution when applying it for research questions beyond the crossing effect itself.

Keywords

meta-analysis; touch; remapping; multisensory; tactile temporal order judgement

Introduction

Crossing the arms over the body midline impairs tactile choice tasks, a phenomenon known as the (tactile) crossing effect[1]. A prominent experimental paradigm for its investigation is the tactile Temporal Order Judgment (TOJ) task, in which participants report which of two consecutive stimuli, applied one to each hand, occurred first (or second), typically by pressing a button with the respective hand[2,3]. Intuitively, it would seem irrelevant for choosing a hand for a button press whether the arms are held uncrossed or crossed. However, when participants cross their arms, their TOJ are strongly impaired and sometimes even systematically reversed. Therefore, the TOJ task has become a prime example of a tactile crossing effect, and it has been used extensively over the last two decades to investigate tactile spatial processing.

The origin of crossing effects such as that observed in the TOJ task has been debated. The most prominent idea has been that crossing effects arise from a conflict between different spatial codes: a skin-based, or anatomical, code that signals where on the skin the touch occurred; and a spatiotopic, or external, code that signals where the touch was in space[1,4]. With crossed arms, these two codes carry opposing information; the right hand, for instance, is then located in left space, and it may be this kind of conflict that impairs tactile judgments. Newer studies have cast doubt on this theory and have suggested that crossing effects, instead, stem from confusion of where the touched limb is usually located – its default posture – and where it is placed at the moment of touch[5–7].

Irrespective of its theoretical underpinnings, the TOJ crossing effect is often large, and it is probably this feature that has contributed the most to the task's popularity. However, there are multiple versions of the TOJ paradigm, and the task has been analyzed in multiple ways. This variability in how the task is run and analyzed has made it difficult to compare studies directly, beyond assessing qualitative similarities. Moreover, no attempt has been made to estimate a statistical effect size based on evidence from multiple studies.

An effect size estimate would aid future research, for instance, for planning sample sizes based on statistical power or for defining priors in Bayesian analysis. The TOJ task has been employed, among others, in developmental and clinical studies, including research

on mental disorders, physical illnesses, and brain processes[8–11]. Such studies typically aim at optimizing experimental efficiency and determining the minimum number of required study participants. One option to reduce experimental time may be to apply yet another version of the TOJ task which applies adaptive testing[12]. Yet, a proper estimate of the expected effect size is necessary to determine an appropriate sample size. In practice, therefore, experiments are often either inefficient or underpowered[13].

The TOJ task has been employed for a wide range of scientific questions, and numerous manipulations have been applied next to limb crossing for this purpose. Some of them have been scrutinized across multiple studies, potentially allowing investigating their effect sizes as well. One key variable that has differed across studies is the response modality. Often, participants have responded with the hands, that is, with the same limbs that received stimulation. Some studies, however, have required button responses with the feet. With this response modality, it is necessary to instruct participants how to match stimulus hand and response foot: under anatomical instructions, responses are given with the foot that belongs to the same body side as the hand that received the first stimulus. In contrast, under external instructions, responses are given with the foot that is on the same side of external space as the stimulated hand. These two instructions require different responses when the hands are crossed: for instance, when the participant chooses the right hand, (s)he must press the right foot under anatomical instructions (stimulus on right body side), but the left foot under external instructions (stimulus on left side of space due to crossed posture). Several studies have reported a larger TOJ crossing effect for external than anatomical instructions[14,15]. In yet other studies, participants responded verbally[11,16], but, to our knowledge, no direct comparison of verbal vs. button press responses has been published so far. However, the apparent differences between these response modalities have not been systematically explored.

A second manipulation that has been investigated in multiple studies is the availability of visual information. The crossing effect was reported to be diminished when the hands were crossed behind the back[17], when the eyes were closed[18], and when uncrossed rubber arms were placed above the hidden, crossed arms[19]. While most studies have reported that the absence of visual input reduces the crossing effect, some findings have been inconsistent[20].

A third aspect which may affect TOJ are task instructions regarding speed and accuracy. The TOJ has sometimes been run as an unspeeded task that emphasized correct responding[2], whereas many studies have asked participants to respond both as fast and as correct as possible[3,21–23]. Emphasizing speed may compel participants to respond before they have fully processed the stimuli, leading to more frequent guessing and errors in the crossed posture, which may, in turn, amplify the crossing effect. We are, however, not aware of any study systematically addressing this detail for the TOJ task.

Finally, the reported size of experimental effects may systematically depend on the analysis method[1]. TOJ performance has often been assessed using a range of stimulus onset asynchronies (SOAs) between the two tactile stimuli. By convention, SOAs are marked as negative when the left hand stimulation occurred first and plotted as "percentage of right-first" responses (y-axis) against the SOA (x-axis). With uncrossed hands, performance follows a typical psychophysical, S-shaped curve. In contrast, with crossed hands, the curve is often N-shaped, with responses for long SOAs correct but responses for short SOAs reversed[3]. This N-shaped result pattern is unusual – in fact, we are not aware that it has been observed in any other psychophysical task to date. Reaction times in the TOJ task are longer for shorter SOAs, and are significantly prolonged in the crossed compared to the uncrossed posture[3,23].

Whether speeded or not, most reports have focused on analyzing TOJ accuracy rather than RT. One way to analyze accuracy has been to probit-transform percentage of right-first responses to linearize the response curve's S-shape, and to derive the slope of the resulting straight line[2]. A second measure has been to derive the just noticeable difference as half the distance between the SOA required for 25% and 75% correct right-first responses after probit transformation[2]. A third measure has been to simply sum up accuracy across all SOAs[21]. With all three methods, the crossing effect was defined as the difference of the respective measure between the uncrossed and crossed postures.

Finally, a fourth analysis method has been to fit the N-shaped response curves of the crossed posture. This can be done by adding up three Gaussians[3]. Two of the three Gaussians fit the N's bulks; their amplitudes, termed A_{right} and A_{left}, are then interpreted as a measure of the crossing effect,[3,24,25]. Note, the two "flip Gaussians" are derived solely from the crossed condition; this distinguishes the flip fitting procedure from all

other crossing effect measures, which all rely on subtracting performance between uncrossed and crossed conditions.

There is no consensus on which analysis method is most appropriate, nor does any method exhibit generally greater sensitivity or consistently yield a larger crossing effect[23]. In addition to these analysis choices, the choice of SOAs used in the experiment, as well as some further analysis options, may result in differences across studies[26].

In sum, both the need for an effect size estimate to aid future study planning and the lack of consensus about the formal analysis for TOJ regarding the crossing effect, and the so far unsystematic exploration of the four discussed potential moderators, all call for a systematic review of the research, which we provide by conducting a meta-analysis for the TOJ crossing effect and its moderation by response mode, visual input, speed and accuracy instructions, and analysis method.

Methods

Search Strategy

We searched the databases BASE, EMBASE, Pubmed, and Web of Science, accessed through Bielefeld University on April 28, 2023. We chose these databases because they have demonstrated high levels of precision, recall, and reproducibility[27]. We employed the search string "(tactile OR hand OR arm) temporal order judg*" for all databases with the respective, appropriate syntax (see Appendix A, Table A1). We included only English articles when this option was available. Search an article screening were performed by NS and discussed with TH and AM.

Selection Criteria

Studies were included in the meta-analysis if they: (1) conducted the tactile TOJ task with a crossed and an uncrossed condition, (2) presented stimuli sequentially on both hands, (3) asked participants to judge the temporal order of both stimuli, (4) collected responses from participants using their hands, feet, or verbally, (5) included participants aged 18 and above, (6) were published peer-reviewed articles, (7) were in English. Studies were excluded if they: (1) reported mental or health impairment of participants and no unaffected control group was available, (2) did not provide sufficient data within the

published articles to calculate a crossing effect. This criterion led us to exclude studies that reported only flip parameters obtained by fitting Gaussians (see Introduction, fourth analysis method). As flip parameters are unique to the crossed condition, they cannot be tested against the uncrossed condition, making them formally incomparable to the remaining measures which all compare crossed and uncrossed conditions. Qualitatively, however, the respective studies support the conclusions we draw from our formal analysis (see Discussion).

Data Extraction and Derived Variables

All studies were collected in Zotero (Corporation for Digital Scholarship, Vienna, Virginia, USA, https://www.zotero.org). We employed automated procedures to eliminate duplicates and removed undetected duplicates manually. The assessment of the remaining studies regarding the fulfillment of the inclusion criteria was performed by one coder (N.S.). The same coder extracted data from each study, including statistical data for calculating effect sizes and standard errors. All results were double-checked by the same coder to ensure accuracy.

Additionally, potential moderator variables were operationalized and coded for each study in the following manner:

(1) outcome measure

a. “accuracy” – the difference in performance between the crossed and uncrossed condition was calculated for each SOA and then pooled over all SOAs, as introduced in *(21)*

b. “slope” – the S-shaped curve of the percentage of right-first responses was probit, logit, or Weibull transformed to linearize the psychometric curve, and the resulting straight line displayed the performance in the TOJ task as a function of the SOA

c. “JND 75%” – the JND was obtained from linearized response data (see b) and calculated as the time window between the two applied stimuli in which 75% of the judgments were correct)

d. “JND 84%” – response data were pooled across participants and fitted with a cumulative four-parametric cumulative Gaussian function, as introduced in *(3)*. Subsequently, the standard deviation

(σ) was calculated, reflecting the smallest time window between the two applied stimuli in which 84% of the judgments were correct)

If multiple outcome measures were available, we entered only one into our analysis, using the following order: accuracy, slope, JND 75%, JND 84%.

(2) eyes
   a. “open” – participants’ eyes were open; we did not differentiate whether fixation was required or not, or whether participants' hands were hidden or not
   b. “closed” – participants were blindfolded
(3) response modality and response demands
   a. “hands” participants responded with the hands or a finger of the hand that was stimulated first [or second]
   b. “feet external” – participants responded with the foot on the same side of space as the hand stimulated first [or second]
   c. “feet anatomical” – participants responded with the foot on the same body side as the hand stimulated first [or second]
   d. “verbal” participants responded verbally
(4) task instruction regarding speed and accuracy
   a. “unspeeded” – participants responded without any time restriction and/or were instructed to prioritize accuracy
   b. “time-restricted” – participants had to respond within a specific time interval or were instructed to respond as accurately and as quickly as possible; the latter were usually combined)

We did not conduct a formal risk of bias assessment, reasoning that the TOJ task is typically performed in controlled laboratory settings.

Statistical Analysis

Effect Measures

All included studies used within-subject designs, and all studies except two aggregated the dichotomous dependent variable (correct or incorrect) into a percentage correct score that can be treated as continuous for the range of values that is not too close to the limits. Therefore, we calculated a version of Cohen’s *d* known as Cohen’s d_z or d_D

(hereafter referred to as Cohen's d_z), which is appropriate for within subject-designs with a continuous outcome measure. Using Cohen's d_z allowed us to pool primary studies with different outcome measures, such as slope versus percentage correct. One possibility to calculate Cohen's d_z is to compute the mean difference between crossed and uncrossed conditions and then use the standard deviation of these difference scores as the denominator[28]:

$$Cohen's\ d_z = \frac{M_{crossed} - M_{uncrossed}}{SD_{difference}} \qquad (1)$$

The mean difference and its standard deviation were only available in studies that used accuracy as an outcome measure. Alternative derivations are available for Cohen's d_z depending on the available information:

a. calculation based on the *t*-value and the sample size[29]:

$$Cohen's\ d_z = \frac{t}{\sqrt{n}} \qquad (2)$$

b. calculation based on the *F*-value (with one numerator degree of freedom) and the sample size[30]:

$$F = t^2, \text{consequently } Cohen's\ d_z = \frac{\sqrt{F}}{\sqrt{n}} \qquad (3)$$

c. from the χ^2- value and the sample size, with the χ^2- value for categorical outcomes treated as *F*-value to obtain comparability with the analysis of continuous variables as described in b:

$$Cohen's\ d_z = \frac{\sqrt{\chi^2}}{\sqrt{n}} \qquad (4)$$

d. from the regression coefficient (β) in a generalized linear mixed model (GLMM) and the sample size, employing the additonal assupmtions that the β-value is unstandardized, and the *z*-value is approximately equivalent to the *t*-value:

$$z = \frac{\beta}{SE}, Cohen's\ d_z = \frac{z}{\sqrt{n}} \qquad (5)$$

e. from the means of the crossed and uncrossed conditions, their standard deviations, and the correlation between the two conditions[28]; we estimated the required correlation from three openly available datasets (see Supplementary Information for details):

$$Cohen's\ d_z = \frac{M_{crossed} - M_{uncrossed}}{\sqrt{SD_{crossed}^2 + SD_{uncrossed}^2 - 2 * r * SD_{crossed} * SD_{uncrossed}}} \quad (6)$$

Equations (4) and (5) are approximations based on the properties of *F* and *t*, which converge against the χ^2 and the normal distribution for large *n*.

If multiple statistical metrics were available to calculate the effect size, formulas were chosen in the order reported above.

Standard Error

The standard error of each effect size was calculated as shown in equation (7). The derivation of this formula is shown in Appendix B.

$$SE(d_z) = \sqrt{\frac{1}{n} * \frac{n-1}{(n-1)-2} + d_z^2 * \left(\frac{n-1}{(n-1)-2} - \frac{1}{\left(1 - \frac{3}{4*(n-1)-1}\right)^2} \right)} \quad (7)$$

Meta-Analysis

We conducted statistical analyses with R[31] and R Studio[32]. We employed the packages meta[33] to conduct the meta-analysis, dmetar[34] for outlier and influence analysis, metasens[35] for Rücker's Limit Meta-Analysis Method, and tidyverse[36] for data handling. We used a random-effects model to pool effect sizes, because we anticipated considerable between-study heterogeneity. We calculated the heterogeneity variance (τ^2) using the restricted maximum likelihood (REML) estimator[37] and the confidence interval around the pooled effect using Knapp-Hartung adjustments[38].

$$Cohen's\ d_z = \frac{\sum_{k=1}^{K} Cohen's\ d_{z\,k} w_k^*}{\sum_{k=1}^{K} w_k^*} \quad (8)$$

where:

w_k^* is the weight assigned to each study k

K is the number of studies included in the meta-analysis

(9)

with:

$$w_k^* = \frac{1}{se_k{}^2 * n_k + \tau^2}$$

where:

n_k is the sample size of each study k

We used Cochran's Q to test if between-study variation exceeded sampling error and estimated the proportion of variability due to true differences using Higgins and Thompson's I^2[39]. Additionally, we calculated prediction intervals to estimate the expected range of effects in future studies[40]. To detect divergent studies, that is, studies that differ markedly from the study pool, we set as outlier criterion that a study's 95% confidence interval was beyond the 95% confidence interval of the pooled effect. Additionally, we identified studies with a high influence on the pooled effect size using the leave-one-out method. To this end, we calculated, for each study, the difference in Cohen's d_z when the study was included in the analysis compared to when it is not included. The resulting differences were plotted in a Baujat plot on the y-axis, with overall heterogeneity, measured by Cochran's Q, on the x-axis[41].

Analysis of moderator effects

To explore the impact of the four potential moderators of the crossing effect's effect size – analysis method, visual input, response mode, and speed and accuracy instructions – we conducted separate moderator analyses, each over those articles that reported the respective moderator characteristics. Ten studies lacked information on one moderator each; therefore, we opted for separate analyses, so that we retained as many studies as possible for each individual analysis[42]. We applied a random-effects (plural) model for these analyses[43]. We used a pooled version of the between-study heterogeneity variance τ^2 across all studies subsets when any subset consisted of fewer than 6 studies[44].

Data availability statement

All variables created for our report, as well as the R code used for analysis, are freely available at strategy https://osf.io/xtd9g/. Moreover, the data contain several variables we extracted during data extraction but did not use for our report, such as the reported *p* and *df* values for *t*, *F*, and $\chi 2$ tests.

Results

Search Results

The database search yielded 973 studies (Fig. 1). After removing duplicates, 346 eligible studies remained. 301 of these were removed during abstract screening. Forty-five studies underwent full-text review; they comprised a total of 85 experiments. Of these, 18 studies/51 experiments were excluded, 18 experiments due to insufficient data and 33 experiments due to not meeting the inclusion criteria (see Supplementary Table B1 for references and reasons for exclusion of fully screened studies).

The remaining 27 studies comprised a total of 37 experiments and comprised 636 participants (see Table 1 for study citations and characteristics). The difference between studies and experiments results from our including between-subject experimental groups, such as separate groups that conducted the experiment with their eyes open vs. closed, as well as different experiments within a single study if they comprised different individuals.

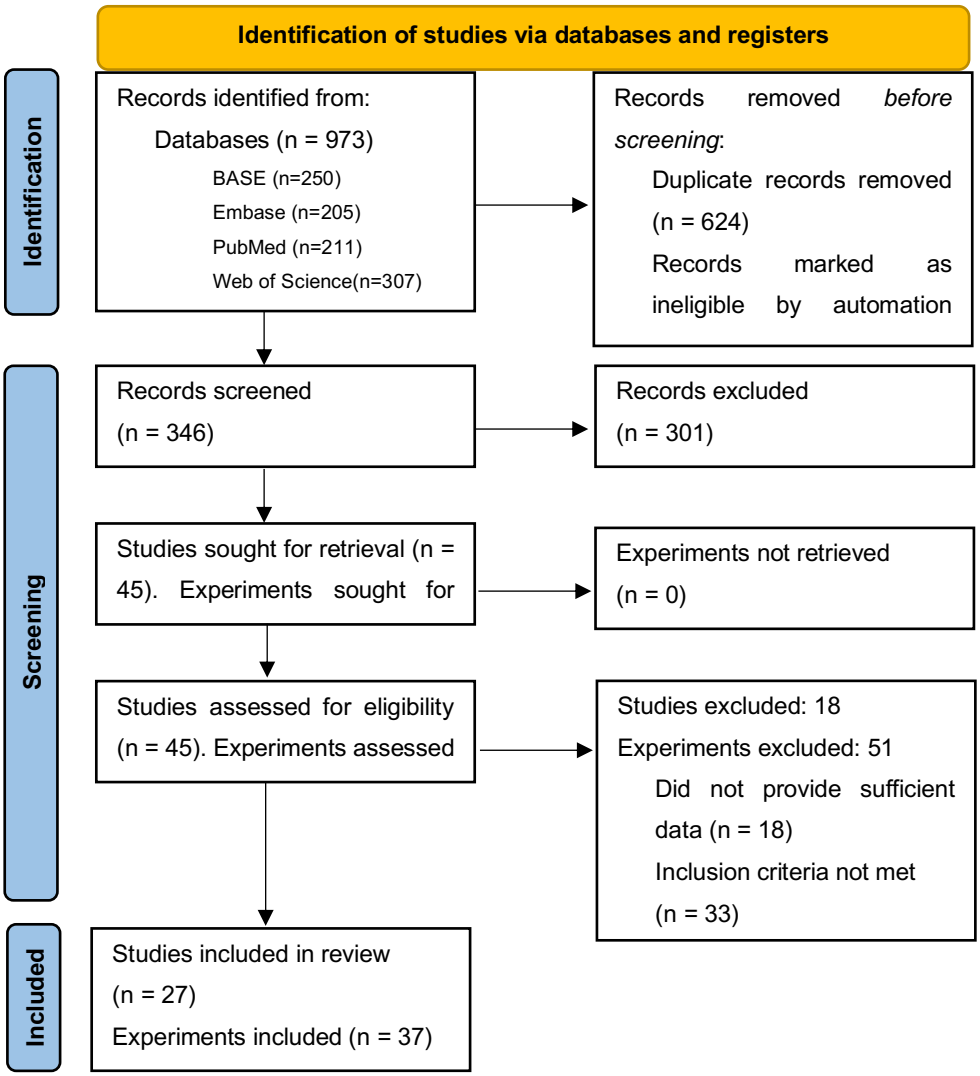

Figure 1. Flowchart of study selection. Design adapted from the PRISMA 2020 statement[45].

Seven studies comprised 2-4 experiments each (see Supplementary Table B2). As the experiments were conducted with different persons, we treat them as independent experiments. For further information and limitations regarding the coding of subgroup analysis variables outcome measure, response modality, and task instructions about speed see Supplementary Tables B3, B4, and B5.

Table 1. Included studies and characteristics used for the meta-analysis.

Author	*n*	Statistical characteristics	Outcome measure	Eyes	Response modality	Speed
Azañón & Soto-Faraco (2007)[19] - exp. 1	19	*M, SE*	JND 84%	open	hands	unspeeded
Azañón & Soto-Faraco (2007)[19] - exp. 2	20	*M, SE*	JND 84%	open	hands	unspeeded
Azañón et al. (2015)[46]	12	*F*	JND 75%	open	hands	unspeeded
Azañón, Mihaljevic et al. (2016)[47]	16	*t*	JND 75%	closed	hands	unspeeded
Azañón, Radulova et al. (2016)[48]	11	*t*	slope	NA	hands	unspeeded
Badde et al. (2014)[49] - exp. 1	17	*F*	Accuracy	open	hands	time restriction
Badde et al. (2014)[49] – exp. 2	17	*F*	Accuracy	open	hands	time restriction
Badde et al. (2015)[50] - exp. 1	13	*F*	Accuracy	open	hands	unspeeded
Badde et al. (2015)[50] - exp. 2	14	*F*	Accuracy	open	hands	unspeeded
Badde et al. (2017)[51]	16	χ^2	Accuracy	NA	hands	time restriction
Cadieux & Shore (2013)[14] - eyes closed, feet external	11	*t*	Accuracy	closed	feet external	time restriction
Cadieux & Shore (2013)[14] - eyes closed, hands	20	*t*	Accuracy	closed	hands	time restriction
Cadieux & Shore (2013)[14] - eyes open, feet external	11	*t*	Accuracy	open	feet external	time restriction
Cadieux & Shore (2013)[14] - eyes open, hands	20	*t*	Accuracy	open	hands	time restriction
Cadieux et al. (2010)[21]	48	*t*	slope	open	feet external	time restriction
Crollen, Albouy, et al. (2017)[52]	11	*t*	slope	closed	hands	time restriction
Crollen, Lazzouni, et al. (2017)[22]	11	*t*	slope	closed	hands	time restriction
Crollen et al. (2019)[53]	15	*t*	slope	closed	feet external	time restriction
Heed et al. (2012)[23] - exp. 1	15	*F*	Accuracy	open	hands	unspeeded

Heed et al. (2012)[23] - exp. 2	14	*t*	Accuracy	open	hands	unspeeded
Heed et al. (2016)[4]	24	*β*	Accuracy	open	hands	NA
Kóbor et al. (2006)[17]	18	*F*	JND 75%	closed	feet anatomical	unspeeded
Landry & Champoux (2018)[54]	20	MD, SD_{within}	Accuracy	NA	feet external	unspeeded
Moseley et al. (2012)[11]	10	*t*	JND 75%	open	verbal	NA
Roberts & Humphreys (2008)[55] – exp. 1	9	*t*	slope	closed	feet external	unspeeded
Röder et al. (2004)[20] - eyes closed	13	*F*	JND 75%	closed	hands	unspeeded
Röder et al. (2004)[20] - eyes open	12	*F*	JND 75%	open	hands	unspeeded
Sambo et al. (2013)[56]	12	*t*	JND 75%	closed	hands	time restriction
Schicke & Röder (2006)[26]	10	*F*	JND 75%	closed	hands	unspeeded
Sharp et al. (2018)[57]	13	MD, SD_{within}	Accuracy	NA	feet external	unspeeded
Shore et al. (2002)[2]	9	*F*	slope	open	hands	unspeeded
Soto-Faraco & Azañón (2013)[10]	32	*F*	JND 84%	open	hands	NA
Unwalla et al. (2020)[58]	20	*t*	Accuracy	NA	hands	time restriction
Unwalla, Goldreich, et al. (2021)[15] – exp. 1	20	*t*	Accuracy	NA	feet anatomical	time restriction
Unwalla, Goldreich, et al. (2021)[15] - exp. 2	43	*t*	Accuracy	NA	feet anatomical	time restriction
Unwalla, Cadieux, et al. (2021)[18] - eyes closed	20	*t*	Accuracy	closed	hands	time restriction
Unwalla, Cadieux, et al. (2021)[18] - eyes open	20	*t*	Accuracy	open	hands	time restriction

Note. Exp. = experiment, *n* = simple size, *M* = mean, *SE* = standard error, *MD* = mean difference, *SD* = standard deviation, NA = not available

Our main analysis involved the data from all identified 27 studies with their 37 experiments. The estimated effect size of the crossing effect, quantified as Cohen's d_z, was 1.42 ($t(36) = 19.44$, $p < .001$, 95% CI [1.27, 1.56]). The effect size is very large, based on Cohen's classification of effects as large when $d_z >= 0.8$[59]. Fig. 2 breaks these results down by study. The estimated between-study heterogeneity variance, τ^2, was 0.06 (95%CI [0.00, 0.15]), with Q(36) = 47.67, p = .09, and an I^2 value of 25% (95%CI [0%, 50%]). The prediction interval ranged from Cohen's d_z = 0.87 to 1.96, indicating that the TOJ crossing effect is large and should be reliably observable in future studies.

Study	Cohens dz	SE	Weight	Cohens dz IV, Random, 95% CI
(1) Azañón & Soto-Faraco (2007) - experiment 1	1.0784	0.3158	3.8%	1.08 [0.46; 1.70]
(2) Azañón & Soto-Faraco (2007) - experiment 2	1.4714	0.3556	3.3%	1.47 [0.77; 2.17]
(3) Azañón et al. (2015)	0.9309	0.3995	2.8%	0.93 [0.15; 1.71]
(4) Azañón, Mihaljevic, et al. (2016)	0.9125	0.3296	3.6%	0.91 [0.27; 1.56]
(5) Azañón, Radulova, et al. (2016)	1.1880	0.4706	2.2%	1.19 [0.27; 2.11]
(6) Badde et al. (2014) - experiment 1	1.5312	0.4023	2.8%	1.53 [0.74; 2.32]
(7) Badde et al. (2014) - experiment 2	2.0772	0.4913	2.0%	2.08 [1.11; 3.04]
(8) Badde et al. (2015)- experiment 1	1.7541	0.5233	1.9%	1.75 [0.73; 2.78]
(9) Badde et al. (2015)- experiment 2	1.9977	0.5437	1.7%	2.00 [0.93; 3.06]
(10) Badde et al. (2017)	0.7224	0.3082	3.9%	0.72 [0.12; 1.33]
(11) Cadieux & Shore (2013) - eyes closed, feet external	1.6101	0.5583	1.7%	1.61 [0.52; 2.70]
(12) Cadieux & Shore (2013) - eyes closed, hands	1.3014	0.3333	3.6%	1.30 [0.65; 1.95]
(13) Cadieux & Shore (2013) - eyes open, feet external	2.3970	0.7433	1.0%	2.40 [0.94; 3.85]
(14) Cadieux & Shore (2013) - eyes open, hands	1.6010	0.3733	3.1%	1.60 [0.87; 2.33]
(15) Cadieux et al. (2010)	2.1362	0.2731	4.5%	2.14 [1.60; 2.67]
(16) Crollen, Albouy, et al. (2017)	1.0885	0.4518	2.3%	1.09 [0.20; 1.97]
(17) Crollen, Lazzouni, et al. (2017)	0.9166	0.4217	2.6%	0.92 [0.09; 1.74]
(18) Crollen et al. (2019)	1.6086	0.4493	2.4%	1.61 [0.73; 2.49]
(19) Heed et al. (2012) - experiment 1	1.4733	0.4265	2.5%	1.47 [0.64; 2.31]
(20) Heed et al. (2012) - experiment 2	1.7238	0.4916	2.0%	1.72 [0.76; 2.69]
(21) Heed et al. (2016)	1.6311	0.3384	3.5%	1.63 [0.97; 2.29]
(22) Kóbor et al. (2006)	1.0878	0.3274	3.6%	1.09 [0.45; 1.73]
(23) Landry & Champoux (2018)	2.4298	0.4982	2.0%	2.43 [1.45; 3.41]
(24) Moseley et al. (2012)	1.8025	0.6474	1.3%	1.80 [0.53; 3.07]
(25) Roberts & Humphreys (2008)	1.1817	0.5462	1.7%	1.18 [0.11; 2.25]
(26) Röder et al. (2004) - eyes closed	1.6129	0.4957	2.0%	1.61 [0.64; 2.58]
(27) Röder et al. (2004) - eyes open	1.0567	0.4198	2.6%	1.06 [0.23; 1.88]
(28) Sambo et al. (2013)	1.5964	0.5211	1.9%	1.60 [0.57; 2.62]
(29) Schicke & Röder (2006)	2.2107	0.7521	1.0%	2.21 [0.74; 3.68]
(30) Sharp et al. (2018)	2.0924	0.5921	1.5%	2.09 [0.93; 3.25]
(31) Shore et al. (2002)	1.9351	0.7422	1.0%	1.94 [0.48; 3.39]
(32) Soto-Faraco & Azañón (2013)	0.7888	0.2117	5.7%	0.79 [0.37; 1.20]
(33) Unwalla et al. (2020)	1.6435	0.3793	3.0%	1.64 [0.90; 2.39]
(34) Unwalla, Goldreich, et al. (2021) - experiment 1	1.9387	0.4223	2.6%	1.94 [1.11; 2.77]
(35) Unwalla, Goldreich, et al. (2021) - experiment 2	1.1193	0.2020	5.9%	1.12 [0.72; 1.52]
(36) Unwalla, Cadieux, et al. (2021) - eyes closed	1.3685	0.3419	3.5%	1.37 [0.70; 2.04]
(37) Unwalla, Cadieux, et al. (2021) - eyes open	1.3774	0.3431	3.4%	1.38 [0.71; 2.05]
Total (95% CI)			**100.0%**	**1.42 [1.27; 1.56]**
Prediction interval				**[0.87; 1.96]**

Heterogeneity: $Tau^2 = 0.0648$; $Chi^2 = 47.67$, df = 36 (P = 0.09); $I^2 = 24\%$

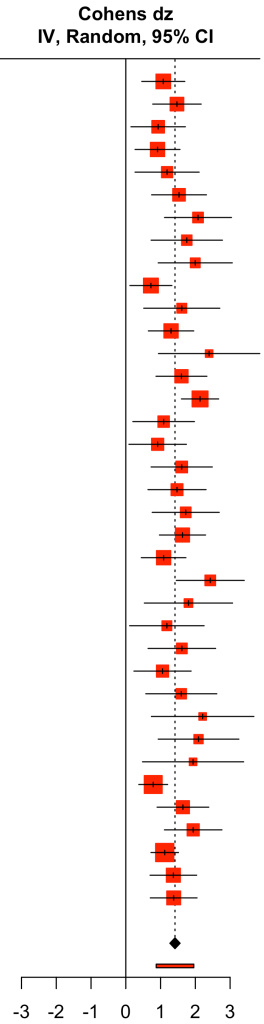

Figure 2. Forest plot of the effect sizes of included studies. *SE* = standard error, CI = confidence interval, *df* = degrees of freedom. Red squares represent the point estimates for each study and black lines are 95% confidence intervals. The size of the squares reflects the study weight with which a study contributes to the pooled effect size. The

black diamond symbolizes the pooled effect estimate, while the red line below denotes the prediction interval.

The outlier analysis identified two studies: Cadieux et al. (2010)[21] and Soto-Faraco and Azanon (2013)[10]. The same two studies also contributed strongly to the overall heterogeneity measured by Cochran's Q and exhibited a strong influence on the pooled result measured by the leave-one-out method (see Fig. 3). Notably, the effect size in Cadieux et al. (2010) was higher than the pooled effect size, while it was lower in Soto-Faraco & Azañón (2013). As a result, excluding these two studies affected the numerical results only slightly (Cohen's d_z, = 1.38, $p < .001$, 95% CI [1.25, 1.51], 95% PI [1.18, 1.59; I^2 value of 0%, 95%CI [0%, 38%]).

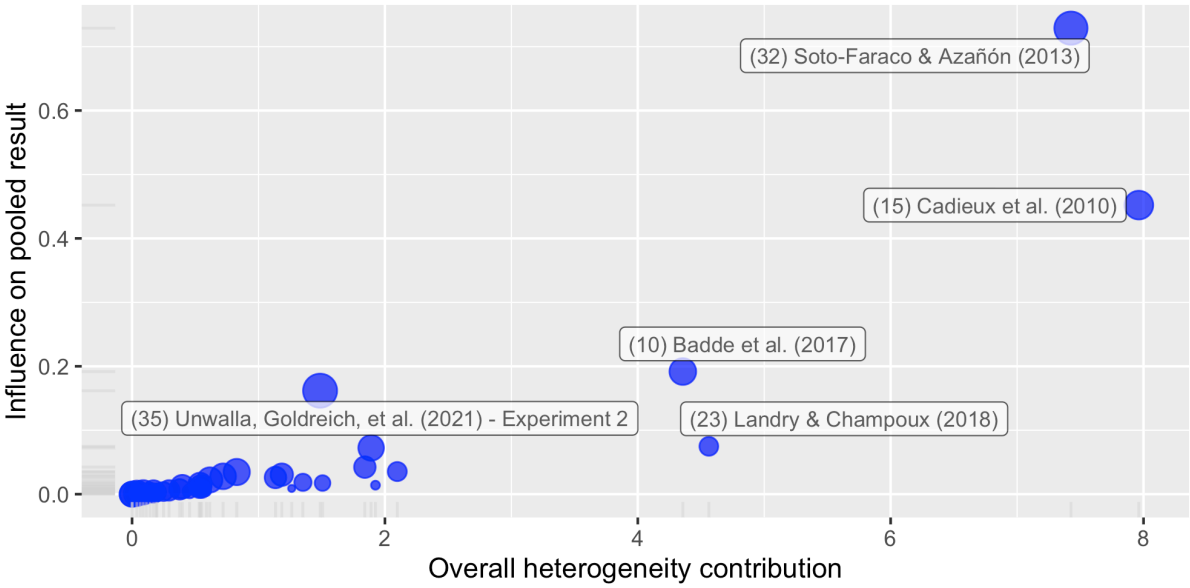

Figure 3: Baujat Plot. Illustration of the heterogeneity with which each study included in the main meta-analysis contributes to the effect size estimate. X-axis: Cochran's Q, indicating the contribution to heterogeneity; Y-axis: Cohen`s d_z, representing the difference in the overall effect when a study is included or not assessed with the leave-one-out method. Note, the Y-axis only displays positive values, meaning that both positive and negative deviations from the pooled effect are represented with positive values. Circle size indicates the study weight with which a study contributes to the pooled effect size. See main text for details.

Influence of Moderators on Effect Size Estimate

Table 2 summarizes influence of the four potential moderators – analysis method, visual input, response mode, and speed and accuracy instructions – in the same order as above.

Regarding the analysis method, the largest descriptive crossing effects was present for slope and accuracy, a slightly smaller crossing effect for JND 75%, and the smallest for JND 84%. Nevertheless, all effect size estimates were still very large, ranging from 1.04-1.53, and these differences did not reach statistical significance.

For the moderator visual input, the crossing effect was slightly larger when participants had their eyes open than when they had them closed. However, this descriptive difference was not statistically significant, and again, the effect size was very large for both cases.

For response modality, the crossing effect was smaller for foot responses under anatomical instructions and for hand responses. The effect was larger when participants responded verbally and when they gave foot responses under external-spatial instructions. Note, however, that only a single study employed verbal responses, and three studies employed anatomical foot responses. Only the difference between the two most extreme effect size values – hand responses and external foot responses – was statistically significant.

The crossing effect was smaller when participants gave unspeeded than speeded responses, but again this descriptive difference did not reach statistical significance.

Table 2. Effect size estimates when different potential moderators were considered.

Moderator	*k*	Cohen's d_z	95% CI	*p*	I^2	$p_{subgroup}$
			Analysis method			
JND 84%	3	1.04	0.2 - 1.88	.03	29%	
JND 75%	8	1.23	0.9 - 1.57	< .001	0%	
Slope	7	1.5	1.03 – 1.98	< .001	35%	
Accuracy	19	1.53	1.33 – 1.73	< .001	13%	
Total	37					.07
			Visual input			

Open	18	1.5	1.28 - 1.72	< .001	35%	
Closed	12	1.27	1.07 – 1.46	< .001	0%	
Total	30					.1
			Response modality			
Hands	26	1.3	1.14 – 1.46	< .001	1%	
Feet external	7	1.95	1.56 – 2.33	< .001	0%	
Feet anatomical	3	1.25	0.29 – 2.22	.03	40%	
Verbal	1	1.8	0.5 – 3.11	.007	-	
Total	37					.002
			Task instruction of speed			
Unspeeded	17	1.41	1.17 – 1.64	< .001	1%	
Time-restricted	17	1.46	1.24 – 1.68	< .001	29%	
Total	34					.72

Note. *k* = number of studies, CI = confidence interval. *p*-values for each moderator characteristic indicate whether the crossing effect is significant when only including studies possessing that characteristic. *p*-value for the total indicates whether the moderator has a significant effect.

Discussion

Our meta-analysis over 27 studies, which comprised 37 experiments, provides robust evidence for a significant crossing effect in tactile TOJ tasks. Specifically, arm crossing leads to a pronounced performance decline, as evidenced by a pooled effect size of Cohen's d_z = 1.42. The prediction interval ranged from 0.87 to 1.96, suggesting that we can be certain that the effect size is large (> 0.8). Moreover, between-study heterogeneity was low. This formal result confirms the colloquial perception of the TOJ crossing effect in the tactile research community, encouraging its further use in research into tactile spatial processing. It is noteworthy that the task has also been attested high reliability[58], further supporting this conclusion.

The effect size was not consistently affected by the analysis method. This finding is in line with a report that analyzed three experiments with several analysis approaches[23]: All approaches yielded qualitatively comparable results, and none of them resulted in consistent biases, such as always identifying larger effects than others. This suggests that methodological variations in analysis should not be a major concern when comparing studies in this field. Notably, the comparison included the use of the flip gaussian fitting procedure, which we excluded in the present meta-analysis because it does not involve a direct comparison of uncrossed and crossed performance. The result pattern of the studies that used this method is well in line with all others. One could consider that the $A_{r/l}$ parameter be converted into an effect size (e.g. by assessing its t-value tested against zero) and including flip fitting studies in the meta-analysis. Our data and code are freely available to implement this strategy (https://osf.io/xtd9g/), and we have listed flip fitting studies in all lists provided in the Supplementary Information.

Our moderator analysis addressing the response mode suggest that the spatial content of the instructions affects the TOJ crossing effect. Specifically, the crossing effect was larger when foot responses had to be given with external instructions (i.e., "on which side of space did the first stimulus occur") than with anatomical instructions (i.e., "respond with the foot of the same body side as the hand that was stimulated first"). This difference was the numerically largest effect in our analysis; we consider it likely that it was not significant because the low number of studies that have implemented foot responses leads to a lack of statistical power. Furthermore, the difference between external foot

responses and hand responses was almost as large, with a significantly greater crossing effect for external foot responses. Note, that hand responses can be considered implicitly anatomically instructed, because participants respond with the limb that was touched first, without reference to the limbs' spatial location. We are not aware that these two response modes have been directly tested against each other. Together, these results suggest that TOJ are impaired by external response instructions, independent of which limbs are used.

Descriptively, the crossing effect was smaller when participants were blindfolded than when they could see. While this effect did not reach statistical significance, its numerical direction aligns with the hypotheses that have been formulated regarding availability of visual information[20]. Notably, some prior studies have demonstrated a strong influence of visual input[14], and the relatively small effect observed here may have been affected by our grouping criteria. We categorized studies only based on whether participants' eyes were open or closed, without accounting for factors such as room illumination and fixation control, which may play a moderating role[4,14]. Future studies could adopt more precise classifications to clarify the impact of visual input on the crossing effect, though currently the number of available studies is too low for such a meta-analysis.

We observed a slight tendency for a more pronounced crossing effect when studies imposed time constraints than when they did not. However, this difference was minimal and did not reach statistical significance. We found this experimental detail surprisingly difficult to extract from the original papers and, therefore, had to apply coarse categorization. More detailed reporting in primary studies would enable more fine-grained analyses in future research and may help resolve whether time pressure amplifies the crossing effect.

Limitations

Our screening process was mainly performed by one author. Therefore, it is possible that we overlooked some publications that should have been included in our analyses. However, our internal policy was to discuss inclusion of papers whenever uncertainty arose. We therefore judge the probability of having misclassified a publication incorrectly as low.

In contrast to the large crossing effect, moderation effects are much smaller. Larger number of studies would be needed to reliably detect them, but these are not currently

available. Therefore, our subset analyses addressing potential moderators of the crossing effect suffered from a lack of power, so that even apparently large effect size differences were not statistically significant. On the one hand, future studies can help such endeavors by diligently reporting on all experimental details, even if they appear irrelevant at the time of writing up the study. This may be even more relevant as the TOJ paradigm is often employed to test for additional moderators, rather than crossing per se. On the other hand, our review may encourage further dedicated research into these moderators. It is our perception that, colloquially, researchers in the field already assume that some of the effects we tested for do exist – for instance roles of vision availability and spatial response mode – but our results show that the evidence is not yet firm.

An important aspect of meta-analyses is whether a publication bias exists. We did not control for this for two reasons. First, given that there is a considerable number of tactile TOJ publications, and they consistently report a large effect size, it is unlikely that even studies with small sample sizes would have non-significant results. In fact, in our own work, we have seen highly significant results with as few as five participants. Second, commonly used small-study effect methods might produce biased results when the standard error depends on the effect size[60], as is the case here, and p-curve analyses only provide accurate results when the specific p-value analyzed is directly related to the hypothesis[61], which is rarely the case in the TOJ literature. Notably, whereas we are confident about the validity of the crossing effect size, the same cannot be said about the moderators we have addressed: there are markedly fewer publications for them, and accordingly we cannot say with confidence that some studies have not landed in the drawer.

Conclusion

In summary, we have confirmed that the tactile TOJ crossing effect is a very large effect, irrespective of the specific criteria one uses to include a particular study or not. The obtained effect size value of approximately 1.4 may be of help in planning future studies, especially if they target the crossing effect per se, as may be the case, for instance, in clinical and developmental studies. The picture is much less clear regarding the

moderators we identified as prevalent in the tactile TOJ literature. If we are truly interested in them, more research and diligent reporting will be necessary.

Competing interests

The author(s) declare no competing interests.

References

1. Heed, T. & Azanon, E. Using time to investigate space: a review of tactile temporal order judgments as a window onto spatial processing in touch. *Frontiers in Psychology* **5**, (2014).
2. Shore, D., Spry, E. & Spence, C. Confusing the mind by crossing the hands. *Cognitive Brain Research* **14**, 153–163 (2002).
3. Yamamoto, S. & Kitazawa, S. Reversal of subjective temporal order due to arm crossing. *Nat. Neurosci.* **4**, 759–765 (2001).
4. Heed, T., Backhaus, J., Röder, B. & Badde, S. Disentangling the external reference frames relevant to tactile localization. *PLoS ONE* **11**, (2016).
5. Badde, S., Röder, B. & Heed, T. Feeling a touch to the hand on the foot. *Current Biology* **29**, 1491-1497.e4 (2019).
6. Heed, T., Burbach, J., Rödenbeck, N., Habets, B. & Fuchs, X. Limb, not touch location, is coded in 3D space. 2024.02.01.578364 Preprint at https://doi.org/10.1101/2024.02.01.578364 (2024).
7. Maij, F., Seegelke, C., Medendorp, W. P. & Heed, T. External location of touch is constructed post-hoc based on limb choice. *eLife* **9**, e57804 (2020).
8. Hense, M., Badde, S., Köhne, S., Dziobek, I. & Röder, B. Visual and proprioceptive influences on tactile spatial processing in adults with autism spectrum disorders. *Autism Res.* **12**, 1745–1757 (2019).
9. Espenhahn, S. *et al.* Atypical Tactile Perception in Early Childhood Autism. *J Autism Dev Disord* (2022) doi:10.1007/s10803-022-05570-7.
10. Soto-Faraco, S., 1970- & Azañón, E. Electrophysiological correlates of tactile remapping. *Neuropsychologia* **51**, 1584–1594 (2013).

11. Moseley, G. L., Gallace, A. & Iannetti, G. D. Spatially defined modulation of skin temperature and hand ownership of both hands in patients with unilateral complex regional pain syndrome. *Brain* **135**, 3676–3686 (2012).
12. Vanderclausen, C., Filbrich, L., De Volder, A. & Legrain, V. Measuring the sensitivity of tactile temporal order judgments in sighted and blind participants using the adaptive psi method. *Atten Percept Psychophys* **83**, 2995–3007 (2021).
13. Lakens, D. Sample Size Justification. *Collabra: Psychology* **8**, 33267 (2022).
14. Cadieux, M. L. & Shore, D. I. Response demands and blindfolding in the crossed-hands deficit: an exploration of reference frame conflict. *Multisens Res* **26**, 465–482 (2013).
15. Unwalla, K., Goldreich, D. & Shore, D. I. Exploring reference frame integration using response demands in a tactile temporal-order judgement task. *Multisens Res* **34**, 1–32 (2021).
16. Ritterband-Rosenbaum, A., Hermosillo, R., Kroliczak, G. & van Donkelaar, P. Hand position-dependent modulation of errors in vibrotactile temporal order judgments:the effects of transcranial magnetic stimulation to the human posterior parietal cortex. *Experimental Brain Research* **232**, 1689–1698 (2014).
17. Kóbor, I., Füredi, L., Kovács, G., Spence, C. & Vidnyánszky, Z. Back-to-front: Improved tactile discrimination performance in the space you cannot see. *Neurosci. Lett.* **400**, 163–167 (2006).
18. Unwalla, K., Cadieux, M. L. & Shore, D. I. Haptic awareness changes when lying down. *Scientific reports* **11**, 13479 (2021).
19. Azañón, E. & Soto-Faraco, S. Alleviating the ‘crossed-hands’ deficit by seeing uncrossed rubber hands. *Exp. Brain Res.* **182**, 537–548 (2007).
20. Röder, B., Rösler, F. & Spence, C. Early vision impairs tactile perception in the blind. *Curr. Biol.* **14**, 121–124 (2004).

21. Cadieux, M. L., Barnett-Cowan, M. & Shore, D. Crossing the hands is more confusing for females than males. *Experimental brain research,* **204**, 431–446 (2010).
22. Crollen, V. *et al.* Visual experience shapes the neural networks remapping touch into external space. *Journal of neuroscience* **37**, 10097–10103 (2017).
23. Heed, T., Backhaus, J. & Roder, B. Integration of hand and finger location in external spatial coordinates for tactile localization. *Journal of Experimental Psychology: Human Perception and Performance* **38**, 386–401 (2012).
24. Nishikawa, N., Shimo, Y., Wada, M., Hattori, N. & Kitazawa, S. Effects of aging and idiopathic Parkinson's disease on tactile temporal order judgment. *PLoS ONE* **10**, (2015).
25. Wada, M., Yamamoto, S. & Kitazawa, S. Effects of handedness on tactile temporal order judgment. *Neuropsychologia* **42**, 1887–1895 (2004).
26. Schicke, T. & Röder, B. Spatial remapping of touch: confusion of perceived stimulus order across hand and foot. *Proc Natl Acad Sci U S A* **103**, 11808–11813 (2006).
27. Gusenbauer, M. & Haddaway, N. R. Which academic search systems are suitable for systematic reviews or meta-analyses? Evaluating retrieval qualities of Google Scholar, PubMed, and 26 other resources. *Research Synthesis Methods* **11**, 181–217 (2020).
28. Goulet-Pelletier, J.-C. & Cousineau, D. A review of effect sizes and their confidence intervals, Part I: The Cohen's d family. *TQMP* **14**, 242–265 (2018).
29. Lakens, D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology* **4**, (2013).
30. Langenberg, B., Janczyk, M., Koob, V., Kliegl, R. & Mayer, A. A tutorial on using the paired t test for power calculations in repeated measures ANOVA with interactions. *Behav Res* **55**, 2467–2484 (2022).
31. R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing (2023).

32. Posit team. RStudio: Integrated Development Environment for R. Posit Software, PBC (2023).
33. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: A practical tutorial. **Evidence-Based Mental Health**, 153–160 (2019).
34. Harrer, M., Cuijpers, P., Furukawa, T. & Ebert, D. D. dmetar: Companion R package for the guide 'Doing meta-analysis in R'. (2019).
35. Schwarzer, G., Carpenter, J. R. & Rücker, G. metasens: Statistical methods for sensitivity analysis in meta-analysis. (2023).
36. Wickham, H. *et al.* Welcome to the {tidyverse}. *Journal of Open Source Software* **4**, 1686 (2019).
37. Veroniki, A. A. *et al.* Methods to estimate the between-study variance and its uncertainty in meta-analysis. *Research Synthesis Methods* **7**, 55–79 (2016).
38. Knapp, G. & Hartung, J. Improved tests for a random effects meta-regression with a single covariate. *Statistics in Medicine* **22**, 2693–2710 (2003).
39. Higgins, J. P. T. & Thompson, S. G. Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine* **21**, 1539–1558 (2002).
40. IntHout, J., Ioannidis, J. P. A., Rovers, M. M. & Goeman, J. J. Plea for routinely presenting prediction intervals in meta-analysis. *BMJ Open* **6**, e010247 (2016).
41. Baujat, B., Mahé, C., Pignon, J.-P. & Hill, C. A graphical method for exploring heterogeneity in meta-analyses: application to a meta-analysis of 65 trials. *Stat Med* **21**, 2641–2652 (2002).
42. Thompson, S. G. & Higgins, J. P. T. How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine* **21**, 1559–1573 (2002).
43. Borenstein, M. & Higgins, J. P. T. Meta-analysis and subgroups. *Prev Sci* **14**, 134–143 (2013).

44. Borenstein, M., Hedges, L. V., Higgins, J. & Rothstein, H. R. Subgroup Analyses. in *Introduction to Meta-Analysis* 149–186 (John Wiley & Sons, Ltd, 2009). doi:10.1002/9780470743386.ch19.

45. Page, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* n71 (2021) doi:10.1136/bmj.n71.

46. Azañón, E., Stenner, M., Cardini, F. & Haggard, P. Dynamic tuning of tactile localization to body posture. *Current Biology* **25**, 512–517 (2015).

47. Azañón, E., Mihaljevic, K. & Longo, M. R. A three-dimensional spatial characterization of the crossed-hands deficit. *Cognition* **157**, 289–295 (2016).

48. Azañón, E., Radulova, S., Haggard, P. & Longo, M. R. Does the crossed-limb deficit affect the uncrossed portions of limbs? *Journal of Experimental Psychology: Human Perception and Performance* **42**, 1320–1331 (2016).

49. Badde, S., Heed, T. & Röder, B. Processing load impairs coordinate integration for the localization of touch. *Atten Percept Psychophys* **76**, 1136–1150 (2014).

50. Badde, S., Röder, B. & Heed, T. Flexibly weighted integration of tactile reference frames. *Neuropsychologia* **70**, 367–374 (2015).

51. Badde, S., Röder, B. & Bruns, P. Task-irrelevant sounds influence both temporal order and apparent-motion judgments about tactile stimuli applied to crossed and uncrossed hands. *Attention, Perception, & Psychophysics* **80**, 773–783 (2017).

52. Crollen, V., Albouy, F., Lepore, F. & Collignon, O. How visual experience impacts the internal and external spatial mapping of sensorimotor functions. *Scientific reports* **7**, 1022 (2017).

53. Crollen, V., Spruyt, T., Mahau, P., Bottini, R. & Collignon, O. How visual experience and task context modulate the use of internal and external spatial coordinate for perception and

action. *Journal of Experimental Psychology: Human Perception and Performance* **45**, 354–362 (2019).

54. Landry, S. P. & Champoux, F. Long-term musical training alters tactile temporal-order judgment. *Multisens Res* **31**, 373–389 (2018).
55. Roberts, R. D. & Humphreys, G. W. Task effects on tactile temporal order judgments: when space does and does not matter. *J. Exp. Psychol. Hum. Percept. Perform.* **34**, 592–604 (2008).
56. Sambo, C. F., Gallace, A., Liang, M., Moseley, G. L. & Iannetti, G. D. The temporal order judgement of tactile and nociceptive stimuli is impaired by crossing the hands over the body midline. *Pain* **154**, 242–247 (2013).
57. Sharp, A., Landry, S. P., Maheu, M. & Champoux, F. Deafness alters the spatial mapping of touch. *PLoS ONE* **13**, e0192993 (2018).
58. Unwalla, K., Kearney, H. & Shore, D. I. Reliability of the crossed-hands deficit in tactile temporal order judgements. *Multisens Res* **34**, 387–421 (2020).
59. Cohen, J. *Statistical Power Analysis for the Behavioral Sciences*. (Erlbaum, Hillsdale, NJ u.a., 1988).
60. Zwetsloot, P.-P. *et al.* Standardized mean differences cause funnel plot distortion in publication bias assessments. *eLife* **6**, e24260 (2017).
61. Simonsohn, U., Nelson, L. D. & Simmons, J. P. P-curve: A key to the file-drawer. *Journal of Experimental Psychology: General* **143**, 534–547 (2014).
62. Hogben, D., Pinkham, R. S. & Wilk, M. B. The Moments of the Non-Central t-Distribution. *Biometrika* **48**, 465–468 (1961).
63. Hedges, L. V. Distribution Theory for Glass's Estimator of Effect size and Related Estimators. *Journal of Educational Statistics* **6**, 107–128 (1981).

64. Tothfalusi, L. & Endrenyi, L. An Exact Procedure for the Evaluation of Reference-Scaled Average Bioequivalence. *AAPS J* **18**, 476–489 (2016).

Appendix A

Table A1. Search String for Each Database

Database	Search string
BASE	(tactile OR hand OR arm) temporal order judg* doctype:121; english manually selected
EMBASE	('tactile temporal order' OR ((hand OR tactile OR arm) AND temporal AND order AND judg*)) AND ([article]/lim OR [article in press]/lim) AND [english]/lim
PUBMED	(tactile OR hand OR arm) AND temporal order judg* AND (english[Filter]) NOT(review pt])
Web of Science	ALL=((hand OR tactile OR arm) temporal order judg*); articles manually selected

Appendix B

Transformation of noncentral t variance to variance of Cohen's d_z

To calculate the standard error of Cohen's d_z, typically either the standard deviation of the mean difference or the correlation between the two measurements (crossed and uncrossed conditions) is needed[28,29]. However, these data were not available in all studies. Therefore, to avoid using an estimated correlation for the standard error calculation, the relationship between t and Cohen's d_z was utilized to calculate the standard errors of the effect sizes:

1. Variance of the random variable t which follows a non-central t-distribution[62] (using an approximation[63,64]):

$$Var(t) = \frac{v(1+\mu^2)}{v-2} - \frac{\mu^2 * v}{2} * \left(1 - \frac{3}{4*v-1}\right)^{-2} \quad \text{(B1)}$$

   where v is the degrees of freedom and μ is the noncentrality parameter.

2. Given the relation between t and Cohen's d_z[29], t can be calculated from d_z, and the sample size:

$$Cohen's\ d_z = \frac{t}{\sqrt{n}} \text{ and therefore } t = d_z * \sqrt{n} \quad \text{(B2)}$$

3. Substitution of v and μ for within-subject designs: where v = n-1 and $\mu = t = d_z * \sqrt{n}$:

$$Var(t) = \frac{(n-1)\left(1+(d_z*\sqrt{n})^2\right)}{(n-1)-2} - \frac{(d_z*\sqrt{n})^2 * (n-1)}{2} * \left(1 - \frac{3}{4*(n-1)-1}\right)^{-2} \quad \text{(B3)}$$

4. Given equation (B2) the relationship between $Var(t)$ and $Var(d_z)$ can be expressed as:

$$Var(t) = n * Var(d_z) \quad \text{(B4)}$$

   therefore:

$$Var(d_z) = \frac{Var(t)}{n} \quad \text{(B5)}$$

5. Substitution of $Var(t)$ into the equation for $Var(d_z)$:

$$Var(d_z) = \frac{1}{n} * \left(\frac{(n-1)\left(1 + (d_z * \sqrt{n})^2\right)}{(n-1)-2} - \frac{(d_z * \sqrt{n})^2 * (n-1)}{2} * \left(1 - \frac{3}{4 * (n-1) - 1}\right)^{-2} \right) \quad \text{(B6)}$$

6. Simplified to:

$$Var(d_z) = \frac{1}{n} * \frac{n-1}{(n-1)-2} + d_z^2 * \left(\frac{n-1}{(n-1)-2} - \frac{1}{\left(1 - \frac{3}{4 * (n-1) - 1}\right)^2} \right) \quad \text{(B7)}$$

7. Derive the standard error SE(d_z), which in this case equals to the standard deviation and is therefore defined as the square root of the variance:

$$SE(d_z) = \sqrt{\frac{1}{n} * \frac{n-1}{(n-1)-2} + d_z^2 * \left(\frac{n-1}{(n-1)-2} - \frac{1}{\left(1 - \frac{3}{4 * (n-1) - 1}\right)^2} \right)} \quad \text{(B8)}$$