COGNITIVE SCIENCE
A Multidisciplinary Journal

Cognitive Science 49 (2025) e70127

ISSN: 1551-6709 online
DOI: 10.1111/cogs.70127

Does Body-Specificity Stand on Solid Ground? *Z*-Curving the Association Between Emotional Valence and Lateral Space

Pablo Dapica, Julio Santiago, Pablo Solana

Mind, Brain and Behavior Research Center, University of Granada

Received 15 March 2025; received in revised form 11 September 2025; accepted 17 September 2025

Abstract

The body-specificity hypothesis proposes that people with different bodies should also have different conceptual systems. The test case of this hypothesis has been the association of emotional valence (good vs. bad) with lateral space (left vs. right) in people of different handedness. As expected, right-handers tend to associate the good with the right space, whereas left-handers show the opposite association. This body-specific effect has been very influential and followed up by an important number of studies. Here, we undertake a systematic examination of the quality of this literature by means of *z*-curve analysis. The results show that the expected replicability rate (statistical power) of this literature is reasonably high (71−76%), especially for those studies using binomial tasks and those that entail the severest tests for the hypothesis, whereas it is lower in reaction time studies. Moreover, the presence of publication bias cannot be statistically asserted. All in all, the literature on space-valence body-specificity appears solid, although there is still room for improvement.

Keywords: Body specificity; Conceptual metaphor; Embodied cognition; Emotional valence; Horizontal space; Handedness; *Z*-Curve analysis

1. Introduction

Embodied theories claim that mental representations, including abstract concepts, are closely tied to perceptuo−motor interactions between the body and the environment

Correspondence should be sent to Pablo Solana, Mind, Brain and Behavior Research Center, University of Granada, Campus of Cartuja s/n, 18071 Granada, Spain. E-mail: solana@ugr.es

(Barsalou, 2008). In this context, Casasanto (2009) proposed the *body-specificity hypothesis* (BSH hereafter): different ways of interacting with the environment should result in different conceptual systems. He supported this hypothesis by showing that right- and left-handers, who interact with the left and right sides of space with different levels of fluency, also have different associations between lateral space and emotional evaluation: right-handers associated "the good" with the right side of space and "the bad" with the left, whereas left-handers showed the opposite association. Casasanto's (2009) seminal paper has been highly influential, with over 500 citations in Scopus and Web of Science at the time of writing these lines.

The original findings were soon followed up by a large number of studies using a variety of methods. The two most frequently used methods are binomial tasks (such as the original Bob task; Casasanto, 2009) and reaction time tasks. In the Bob task, the participant is shown a diagram of a person (Bob) who is going to the zoo and likes zebras but hates pandas.[1] The participant is then asked to place a zebra in the box where good things should go, and a panda in the box where bad things should go. In a typical reaction time task, participants are asked to evaluate the valence of positive and negative stimuli, such as words (e.g., de la Vega, de Filippis, Lachmair, Dudschig, & Kaup, 2012) or faces (e.g., Kong, 2013), by pressing either a right or left key. Follow-up studies have revealed that the body-specificity effect can influence spontaneous gestures (Casasanto & Jasmin, 2010; but see Çatak, Açık, & Göksun, 2018), that it can be already observed from the age of 5 (Casasanto & Henetz, 2012), that experience of asymmetric perceptuo-motor fluency plays a causal role in space-valence associations (Casasanto & Chrysikou, 2011), that just the observation (de la Fuente, Casasanto, & Santiago, 2015a) or even the imagination (de la Fuente, Casasanto, Martínez-Cascales, & Santiago, 2017) of asymmetric fluency is enough to generate the effect, and that linguistic and cultural patterns appear to play no role in the effect (de la Fuente, Casasanto, Román, & Santiago, 2015b; Li & Cao, 2019; but see Mansoory & Nassiri, 2022). But how reliable are these findings?

Cognitive sciences are increasingly aware that the literature is plagued by false positives and unreplicable results, which compromise the credibility of the field's findings (Camerer et al., 2018; Open Science Collaboration, 2015). The issue ultimately stems from a set of questionable research practices that have been part of standard practice for decades: low statistical power (i.e., low probability of observing real effects; Button et al., 2013), *p*-hacking (i.e., flexibly collecting and analyzing data until statistically significant findings emerge; Simmons et al., 2011), and publication bias (i.e., the selective publication of significant findings; Franco et al., 2014). The field is reacting by revising research and publication practices (e.g., Munafò et al., 2017), as well as by carrying out replications and meta-analyses to assess the quality of previous studies. Notably, some key studies and research areas within embodied cognition have recently been shown to be difficult to replicate and to exhibit high levels of publication bias (e.g., Colling et al., 2020; Solana & Santiago, 2022).

The literature on the body-specificity of emotional valence looks healthy, with many successful conceptual replications. Still, there has only been one recent attempt at a highly powered, direct replication of the original finding by Casasanto (2009). Yamada et al. (2024) collected data from 2222 participants across 12 countries and observed that 60% of right-handers placed "the good" to the right, whereas 60% of left-handers placed "the good" to the left (both proportions significantly different from chance). However, these results suggest that the usual sample sizes used in binomial tasks (typically about 100 right-handers and 10–20 left-handers) are insufficient to achieve 80% power: simulations employing one-tailed binomial tests estimate the required sample size in 160 participants. Even more recently, Pohl and Miklashevsky (2025) performed a meta-analysis on the association between valence and space in both the vertical and lateral axes. Focusing on lateral space, they found a significant and medium-sized effect size of $r = .31$, which was crucially mediated by participants' handedness: right-handers tend to associate positive valence with the right side of space ($r = .39$), while left-handers show the opposite pattern ($r = -.38$). Moreover, no evidence of publication bias was found.

The present study adds to these efforts by assessing the quality of the published literature on the body-specific association between valence and lateral space using *z*-curve analysis. Unlike traditional meta-analyses that focus on the average size of an effect, *z*-curve is an innovative meta-analytic tool that allows estimating (1) the average statistical power and (2) the presence and extent of publication bias in a set of studies (Bartoš & Schimmack, 2022; Brunner & Schimmack, 2020). These indices are highly relevant as they provide information on two key aspects of scientific integrity: (1) the probability of replicating the findings of a literature, and (2) whether there are unpublished results that could bias the conclusions we can extract from the available reports. Importantly, this method is highly robust in the presence of heterogeneity, enabling the examination of studies with significant methodological differences, as in the present set, which includes studies employing various types of tests (e.g., binomial tasks, reaction times), populations (e.g., individuals with and without brain damage), and stimuli (e.g., words, faces), among other sources of variation.

Showing that the body-specificity of the association between positive-negative valence and left-right space stands on solid empirical ground has important theoretical implications. Taking a widening lens, it will increase our confidence in the body specificity hypothesis and, more generally, in the idea of body-relativity: people with different bodies think in different ways (Casasanto, 2009). In turn, body-relativity joins cultural and linguistic relativity to support the idea that experiential correlations of different kinds can play a constitutive role of even our more abstract concepts (Casasanto, 2016; Pitt & Casasanto, 2020). Finally, this is consistent with an experiential-realist view of knowledge, which is in direct contrast with Cartesian dualist views and, in general, traditions that see concepts as disembodied, amodal entities (see discussion in Lakoff & Johnson, 1999; see also Barsalou, 1999). All in all, the body-specificity of the space-valence association is a key empirical test of a view that reunites context (physical, social, linguistic, cultural) and mind.

2. Method

2.1. Transparency and open science

The process of article selection, the selected contrasts, the datasets used for the analysis, and the analysis scripts are publicly available at OSF: https://osf.io/wxrmc/

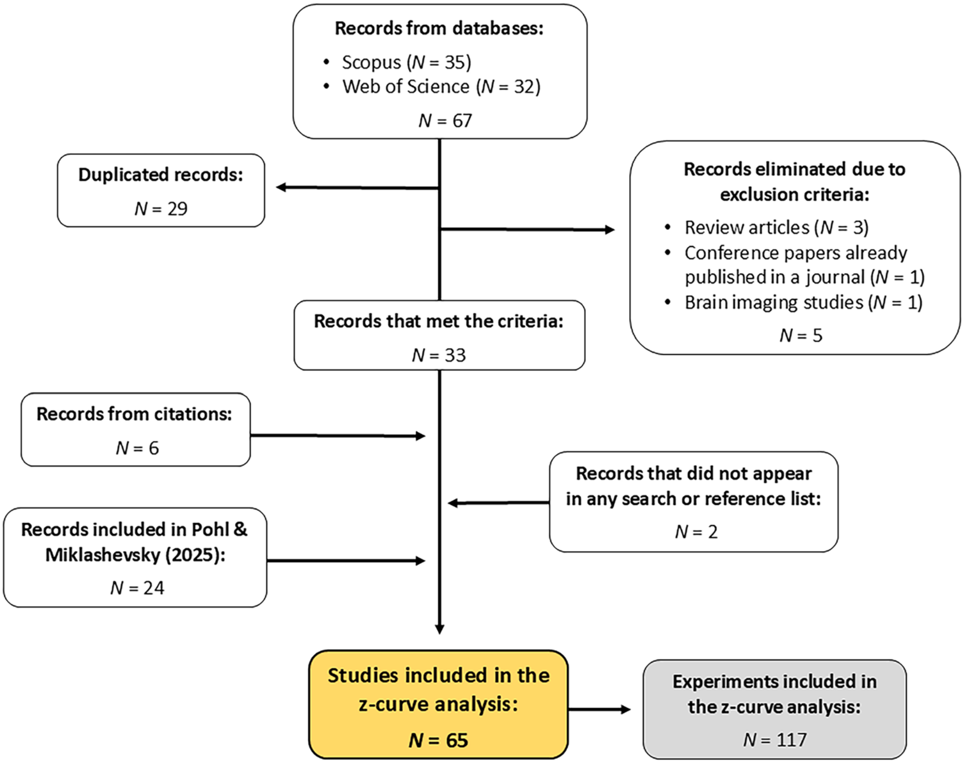

Fig. 1. Diagram illustrating the process of article selection.

2.2. *Literature search and study selection*

We aimed to include all relevant studies for the BSH. The Scopus and Web of Science databases were last searched on October 3rd, 2024, using the following query: ("body specificity") AND ("affective valence" OR "emotional valence" OR "emotion" OR "valence") AND ("space" OR "left" OR "right" OR "laterality" OR "handedness"). This search yielded 67 potential papers, 29 of which were duplicates. Five of the remaining 38 were eliminated as they were either conference articles that had already been published as journal articles ($n = 1$), literature reviews ($n = 3$), or brain imaging studies ($n = 1$). We also included eight additional articles: six were extracted from the reference lists of previous studies, and two were recent articles that had not yet appeared in any search or reference list. Finally, we added 24 extra studies included in Pohl and Miklashevsky's (2025) meta-analysis.[2] In total, 65 papers (117 studies) were included in the *z*-curve analysis. The full list of articles, as well as detailed tables describing their main characteristics (e.g., sample size or design) and the literature search process, can be found in the OSF repository. Fig. 1 illustrates the article selection process.

2.3. Contrast selection

First, we identified the relevant hypothesis of each study. The selected hypotheses should be relevant to the body-specificity effect on emotional valence in the horizontal axis. Then, the *p*-value that best tested the selected hypothesis was extracted. Sometimes, there was more than one *p*-value that could be used for the analysis (e.g., because the central hypothesis could be tested in several ways, or because the study had more than one relevant hypothesis). In those cases, we followed the guidelines of Simonsohn, Nelson, and Simmons (2014) and carried out two complementary analyses: a main analysis and a robustness analysis, which differed in the values they included. That way, the degree of convergence between both analyses indicates whether the results were biased or not by the decisions taken during the selection process. Specifically, following Solana and Santiago (2023) and others, we established the following rules: (1) If there were two relevant *p*-values for the same hypothesis (e.g., one in reaction time and another in accuracy), the first one reported in the paper was included in the main analysis and the next one in the robustness analysis; (2) If there were two relevant hypotheses, the *p*-value associated with the first-mentioned hypothesis was entered into the main analysis and the *p*-value associated with the second hypothesis was entered into the robustness analysis; (3) If there was only one hypothesis and one key result, its corresponding *p*-value was included in both analyses.

By following these rules, we included 109 values in the main analysis and 108 in the robustness analysis. To ensure the reproducibility of the current study, a disclosure table detailing the contrast selection process is available in the OSF repository. All extracted *p*-values were recalculated to ensure there were no reporting errors in the original papers. In some cases, we found discrepancies or *p*-values that could not be recalculated from the statistical information reported in the papers. For brevity, the detailed list of these issues and how we dealt with them can be found as Supplementary Material in the OSF repository.

2.4. Data analysis: Z*-curve*

Z-curve analysis examines the distribution of a set of *p*-values after converting them into *z*-values. The basic idea underlying the analysis of *p*-value distributions (through both the *p*-curve and the *z*-curve techniques, see Brunner & Schimmack, 2020; Simonsohn et al., 2014) is the following: if the selected literature evaluates nonexistent effects, all *p*-values will be equally probable and their distribution will be flat; however, if the studies test real effects, the distribution will become asymmetric, with *p*-values piling up close to zero. Moreover, the asymmetry of this distribution will become more extreme as the power of the studies to detect such effects increases. When transformed into *z*-values, *p*-values follow a normal distribution. If no effects underlie the set of values, then the distribution will be centered on zero, and the number of significant results (false positives) will be 5% (for an alpha of 0.05). Conversely, if real effects exist in the analyzed studies, the mean of the distribution shifts, and the number of significant results increases with power: the higher the power, the greater the number of significant results.

Based on these principles, *z*-curve analysis is able to extract several parameters. First, the *Expected Replicability Rate* (ERR): the likelihood of obtaining the same proportion of significant results upon repeating the studies identically, which corresponds to the statistical power of the studies reporting significant results. Second, the *Expected Discovery Rate* (EDR), which estimates the percentage of significant results that should be obtained from the selected studies, regardless of whether they are statistically significant or not. Third, the EDR can be compared to the *Observed Discovery Rate* (ODR), the proportion of significant results observed in the set of studies, to test for publication bias: if the ODR estimate is outside the 95% confidence interval of the EDR, then there is evidence of publication bias. Finally, the *z*-curve also provides a *File Drawer Rate*, which estimates the number of unpublished *p*-values for each published one.

A total of four analyses were carried out. First, an overall analysis of the whole sample of studies. Second, a comparative analysis of the two most extended types of tasks in this literature: binomial tasks ($n = 17$) versus reaction time (RT) tasks ($n = 24$; the remaining studies used tasks such as memory tests, gesture analysis, line bisection, or word rating). Third, a comparative analysis of the two most frequently used types of stimuli: words ($n = 18$) versus faces ($n = 16$; the remaining studies used sentences, brief texts, or images other than faces). Finally, an analysis with only those experiments that allow a strong inference toward the BSH. Most studies evaluating the association between valence and lateral space have used samples of right-handers. Yet, finding that right-handers associate positive valence with the right can be accounted for by explanations other than the BSH, such as cultural and linguistic influences (e.g., Mansoory & Nassiri, 2022), or the existence of a hemispheric specialization for positive and negative emotions (e.g., Root, Wong, & Kinsbourne, 2006). Strong evidence for BSH requires showing a dissociation between right-handers and either natural left-handers (e.g., Casasanto, 2009) or right-handers with an induced or acquired disfluency toward their dominant hand (e.g., Casasanto & Chrysikou, 2011). Only studies that meet this criterion were included in this analysis ($n = 20$). Each of these three analyses comprised two analyses in turn: a main analysis and a robustness analysis. All the analyses were performed in R, using the *zcurve* package (Bartoš & Schimmack, 2022). The R script and dataset are available at the OSF repository.

3. Results

3.1. Overall analysis

The results are shown in Fig. 2. The ERR was 71% (*95% CI* = 54–86%) in the main analysis and 76% (*95% CI* = 59–91%) in the robustness analysis. The expected proportion of significant findings (EDR) was 20% (*95% CI* = 7–79%; main analysis) and 24% (*95% CI* = 9–89%; robustness analysis). As the observed proportion of significant findings (ODR; main: 75%, robustness: 72%) fell within the EDR's confidence interval, we cannot assert that there is publication bias in the reviewed literature. Yet, the File Drawer Rate suggests 3.93 unpublished results for each published one in the main analysis (*95% CI* = 0.26−13.77) and 3.12 in the robustness analysis (*95% CI* = 0.13−10.05).

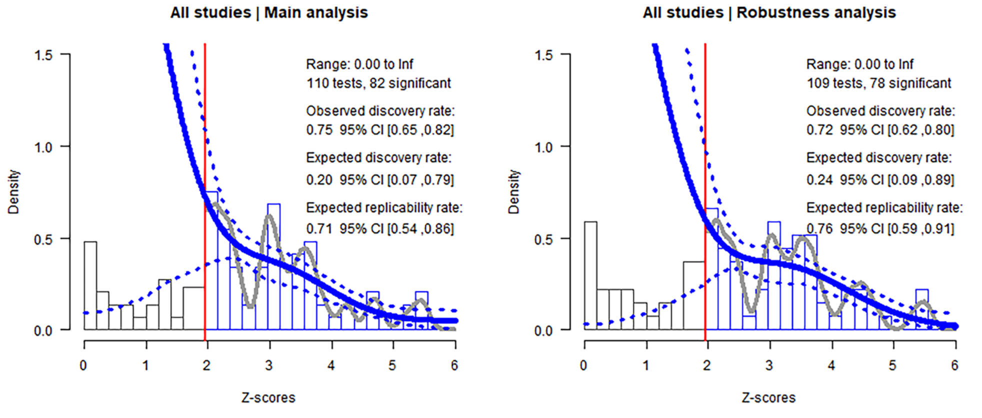

Fig. 2. Results of the z-curve analysis using the whole sample of studies. The bars indicate the observed distribution of z-scores. The solid and dashed blue lines represent the expected distributions of z-scores and their 95% CI, respectively. The vertical red lines represent the significance threshold of $z = 1.96$ ($p = .05$), with significant values to the right.

3.2. Binomial versus RT tasks

For Binomial tasks, the ERRs were 82% (*95% CI* = 57–92%; main) and 87% (*95% CI* = 69–94%; robustness). For RT tasks, however, this index was lower, especially in the main analysis (main: 54%, *95% CI* = 29–87%; robustness: 70%, *95% CI* = 44–96%). As with the overall analyses, we cannot assert the presence of publication bias in either of these subsets of studies, as ODRs fell within their corresponding EDRs (for the exact values, see Fig. S1 at the OSF repository). File drawer estimations were much lower for Binomial tasks (main: 0.8, *95% CI* = 0.07−9.86; robustness: 0.16, *95% CI* = 0.05−3.77) than for RT tasks (main: 2.65, *95% CI* = 0.31−19.00; robustness: 4.50, *95% CI* = 0.05−18.88).

3.3. Words versus faces as stimuli

ERRs were similar between studies using Words (main: 63%, *95% CI* = 35–85%; robustness: 0.74%, *95% CI* = 47–96%) and studies using Faces (main: 75%, *95% CI* = 40–99%; robustness: 0.74%, *95% CI* = 39–97%). Once again, ODRs fell within EDRs (for the exact values, see Fig. S2 at the OSF repository), so we cannot assert the presence of publication bias in these subsets of studies. File drawer estimations were slightly lower for Word studies (main: 1.37, *95% CI* = 0.24−10.32; robustness: 1.49, *95% CI* = 0.07−12.52) than for Face studies (main: 3.88, *95% CI* = 0.01−19.00; robustness: 4.22, *95% CI* = 0.01−19.00).

3.4. Strongest tests for the BSH

The results are shown in Fig. 3. This set of studies exhibited a greater estimated replicability than the literature as a whole: ERR was 81% (*95% CI* = 54–94%) for the main analysis and 90% (*95% CI* = 71–98%) for the robustness analysis. No evidence for publication bias was found: all ODRs fell within EDRs (see the plots). The File Drawer Rate was estimated at 1.76

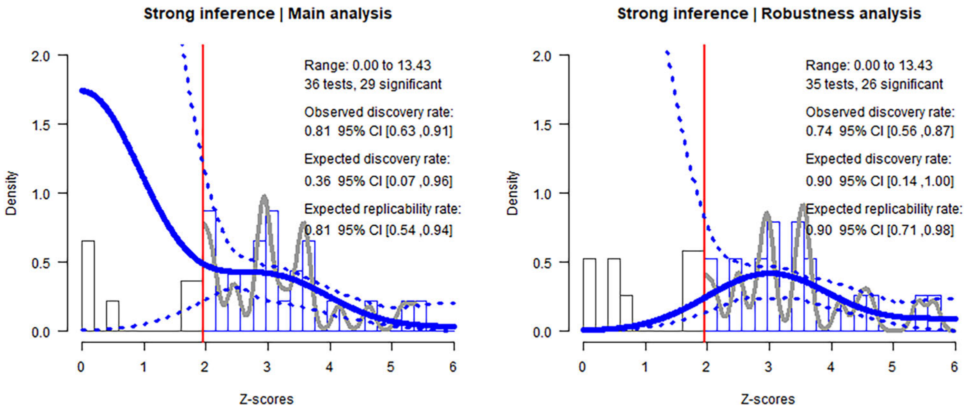

Fig. 3. Results of the z-curve analysis using only those studies that report the strongest tests for the body-specificity hypothesis (i.e., those including a left-handedness or induced left-handedness condition). The bars indicate the observed distribution of z-scores. The solid and dashed blue lines represent the expected distributions of z-scores and their 95% CI, respectively. The vertical red lines represent the significance threshold of $z = 1.96$ ($p = .05$), with significant values to the right.

for the main analysis (*95% CI* = 0.04−13.03) and just 0.11 in the robustness analysis (*95% CI* = 0.00−5.93).

4. Discussion

The present z-curve analysis of studies relevant to the body-specificity hypothesis suggests, first, that the overall ERR of this literature is reasonably high: 71–76% of these studies will replicate if repeated identically. This conclusion holds for studies using either words or faces as stimuli. In contrast, studies using binomial tests showed better estimations of replicability (even surpassing the conventionally accepted level of 80%; Button et al., 2013) than reaction time studies (around 54% in the main analysis). Most importantly, the studies providing the strongest tests for the body-specificity hypothesis also showed estimations higher than 80%.

Second, the observed proportion of significant results (ODR) always fell within the confidence interval of the expected proportion (EDR), which precludes claiming the presence of a significant publication bias in this literature. This may seem surprising given the large difference in absolute values between these indices, with the observed proportion (range 69–82%) often exceeding the expected proportion (range 18–90%). However, in all analyses, there is a wide confidence interval around the EDR index, which allows it to encompass the ODR value. This large confidence interval is probably due to the sample size of the studies being still small with respect to the degree of variability that they show. While this is a potential limitation of the current study, it is important to note that we have arguably included the totality of published studies in the target literature. Moreover, the large confidence interval may also be due to so far unidentified moderating factors of the effect that increase variability in p-values across studies.

Still, the present findings do suggest the presence of some publication bias (range of 0.11−4.50 unpublished *p*-values for each published one), albeit not a particularly strong one when compared to other literatures.[3] For example, Solana and Santiago (2023) reported ODRs that were clearly outside the EDR confidence intervals, and File Drawer indexes of 6.38 and 10.13 (main and robustness, respectively) in their analysis of brain stimulation studies on the involvement of motor cortex in language comprehension (for other examples, see Frese, 2024; Sotola & Credé, 2022). Furthermore, if such a publication bias exists, it appears to have had a negligible impact on studies using binomial tests (main: 0.79; robustness: 0.16) and to be quite low in those studies providing severe tests of the BSH (1.76 in the main analysis, but just 0.11 in the robustness analysis). Further evidence is welcome, but at present, we can conclude that, if there is a publication bias in this literature, it is a small one.

Another interesting result is the huge discrepancy between the EDR and the ERR observed in almost all the analyses. According to Schimmack (2021), this may reflect a high degree of heterogeneity in the statistical power of the reviewed studies: while some may have greater power than average, others may have very low power. This idea is partially supported by the fact that, despite the average sample size in these studies seeming large ($N = 12.5$), its standard deviation is also quite large ($SD = 244.0$), even if the multilab replication by Yamada et al. (2024) is not taken into account ($N = 106.4$, $SD = 219.0$). In short, while this literature has good average power, the reliability of its individual studies should be treated more cautiously.

It is worth commenting that, although the main objective of Pohl and Miklashevsky (2025) was to carry out a meta-analysis on the spatialization of valence, they also performed a complementary *p*-curve analysis. This analysis estimated the replicability of studies on the horizontal axis at 94% (*95% CI*: 90–97%); a value greater than ours. We believe that this difference is due to some limitations that are not present in the current analysis. First, while the *z*-curve includes both significant and nonsignificant results, the *p*-curve omits nonsignificant results. Second, the *z*-curve performs better than the *p*-curve in the presence of high heterogeneity (Brunner & Schimmack, 2020), as is arguably the case. Third, Pohl and Miklashevsky sometimes included more than one *p*-value per experiment, meaning the values are not fully statistically independent, which goes against one of the assumptions of *p*-curve: the inputted values must be independent (Simonsohn et al., 2014). Moreover, our analysis included 21 studies more than theirs, and also included robustness analyses to reduce contrast selection bias. Yonemitsu, Sasaki, and Yamada (2023) also carried out a complementary *p*-curve analysis of their experiments. However, their analysis suffers from the same limitations as Pohl and Miklashevsky (2025), and they only included 20 studies on the lateral axis.

Besides the methodological aspects that make our study more robust than the previous *p*-curve analyses, the present work also provides novel insights into the BSH. First, the *z*-curve analysis provides some new indices, such as the ability to quantify the degree of publication bias, rather than treating it as a dichotomy. While previous *p*-curves suggested high power and no publication bias, our analysis revealed some important nuances: (1) power does not distribute uniformly across studies; and (2) there are some missing studies in the literature, despite publication bias not being statistically significant. Future studies should carry out a priori power analyses (Brysbaert, 2019) and use preregistration, particularly in the form of registered reports (Chambers & Tzavella, 2022), to counter these issues.

Second, while previous *p*-curves have only analyzed the BSH literature as a whole, here we also compared subsets of studies. This revealed that reaction time studies have lower statistical power and contain more publication bias than studies using binomial tasks, which may indicate that binomial tasks more easily capture the space-valence association. Consequently, RT studies may fall short of power more frequently, yielding fewer significant results and, ultimately, a lower probability of publication. It is thus tempting to conclude that binomial tests should be preferred over RT tasks for studying space-valence associations, but what actually follows from the present findings is the importance of carrying out a priori power analyses. In other words, a better planning of the sample size required for RT studies could elevate their credibility to the same level as that of binomial studies.

Finally, another key strength of the present study is that we also conducted specific analyses of those reports providing the strongest evidence for the BSH; that is, those comparing individuals with different handedness or sensorimotor experiences (e.g., Casasanto & Chrysikou, 2011). Crucially, we observed excellent estimates for this subset of studies. This has significant theoretical implications, as it supports that the association between valence and lateral space originates from body-specific patterns of experience, rather than cultural and linguistic influences (e.g., Mansoory & Nassiri, 2022) or innate tendencies related to brain functioning (e.g., Root et al., 2006). Ultimately, this lends weight to the embodiment thesis that sensorimotor experience shapes the way we think and act (Barsalou, 2008; Casasanto, 2009). Avenues for future research include determining which other concepts besides valence are represented in body-specific terms (Willems, Hagoort, & Casasanto, 2010), establishing the extent to which space-valence associations are automatic (Santiago et al., 2011), and identifying the neural correlates of these associations (Brookshire & Casasanto, 2012).

A recent and large multisite study successfully replicated the seminal Casasanto's (2009) findings (Yamada et al., 2024), and a recent meta-analysis showed evidence for body-specificity across multiple studies (Pohl & Miklashevsky, 2025). The present study extends these results by showing that, overall, the body-specificity literature appears to be replicable and not very much affected by publication bias. Taken together, these results provide a solid foundation for confidence in this body of research, supporting the conclusion that studies on the body-specific association between emotional valence and lateral space rest on robust empirical grounds.

Funding statement

The present research was funded by the project PID2022-142583NB-I00 to JS and PS, the FPU grant FPU20/01946 to PS, and the FPI grant PREP2023-002041 to PD.

Conflict of interest statement

The authors declare no conflict of interest.

Data availability statement

Data, analysis scripts, and other resources associated with this research are available at the OSF repository: https://osf.io/wxrmc/

Notes

1 The order of mention and the animal-to-valence mapping are counterbalanced.
2 The majority of these papers test hypotheses other than the BSH (e.g., hemispheric specialization for different valences; e.g., Root, Wong, & Kinsbourne, 2006), but they are relevant to it. This likely explains why our initial search failed to locate them.
3 There are no guidelines to interpret the magnitude of the File-Drawer Rate. According to Bartos and Schimmack (2022), publication bias is statistically asserted when the ODR falls outside the EDR confidence interval.

References

Barsalou, L. W. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, (4), 577–660. https://doi.org/10.1017/S0140525×99002149

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology*, *59*, 617–645. https://doi.org/10.1146/annurev.psych.59.103006.093639

Bartoš, F., & Schimmack, U. (2022). *Z*-curve 2.0: Estimating replication rates and discovery rates. *Meta-Psychology*, *6*, MP.2021.2720. https://doi.org/10.15626/MP.2021.2720

Brookshire, G., & Casasanto, D. (2012). Motivation and motor control: Hemispheric specialization for approach motivation reverses with handedness. *PLoS One*, *7*(4), e36036. https://doi.org/10.1371/journal.pone.0036036

Brunner, J., & Schimmack, U. (2020). Estimating population mean power under conditions of heterogeneity and selection for significance. *Meta-Psychology*, *4*, MP.2018.874. https://doi.org/10.15626/MP.2018.874

Brysbaert, M. (2019). How many participants do we have to include in properly powered experiments? A tutorial of power analysis with reference tables. *Journal of Cognition*, *2*, (1), 16. https://doi.org/10.5334/joc.72

Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., & Munafò, M. R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience*, *14*, (5), 365–376. https://doi.org/10.1038/nrn3475

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johanesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., Isaksson, S., Manfredi, D., Rose, J., Wagenmakers, E. J., & Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature*, *2*, 637–644. http://doi.org/10.1038/s41562-018-0399-z

Chambers, C. D., & Tzavella, L. (2022). The past, present and future of registered reports. *Nature Human Behaviour*, *6*, (1), Article 1. https://doi.org/10.1038/s41562-021-01193-7

Casasanto, D. (2009). Embodiment of abstract concepts: Good and bad in right- and left-handers. *Journal of Experimental Psychology: General*, *138*, (3), 351–367. https://doi.org/10.1037/a0015854

Casasanto, D. (2016). A shared mechanism of linguistic, cultural, and bodily relativity. *Language Learning*, *66*, (3), 714–730. https://doi.org/10.1111/lang.12192

Casasanto, D., & Chrysikou, E. G. (2011). When left is «right»: Motor fluency shapes abstract concepts. *Psychological Science*, *22*, (4), 419–422. https://doi.org/10.1177/0956797611401755

Casasanto, D., & Henetz, T. (2012). Handedness shapes children's abstract concepts. *Cognitive Science*, *36*, (2), 359–372. https://doi.org/10.1111/j.1551-6709.2011.01199.x

Casasanto, D., & Jasmin, K. (2010). Good and bad in the hands of politicians: Spontaneous gestures during positive and negative speech. *PLoS One*, *5*, (7), e11805. https://doi.org/10.1371/journal.pone.0011805

Çatak, E. N., Açık, A., & Göksun, T. (2018). The relationship between handedness and valence: A gesture study. *Quarterly Journal of Experimental Psychology*, *71*, (12), 2615–2626. https://doi.org/10.1177/1747021817750110

Colling, L. J., Szűcs, D., De Marco, D., Cipora, K., Ulrich, R., Nuerk, H. C., Soltanlou, M., Bryce, D., Chen, S-C., Schroeder, P. A., Henare, D. T., Chrystall, C. K., Corballis, P. M., Ansari, D., Goffin, C., Sokolowski, H. M., Hancock, P. J. B., Millen, A. E., Langton, S. R. H., Holmes, K. J., Saviano, M. S., Tummino, T. A., Lindemann, O., Zwaan, R. A., Lukavský, J., Becková, A., Vranka, M. A., Cutini, S., Mammarella, I. C., Mulatti, C., Bell, R., Buchner, A., Mieth, L., Röer, J. P., Klein, E., Huber, S., Moeller, K., Ocampo, B., Lupiáñez, J., Ortiz-Tudela, J., De la Fuente, J., Santiago, J., Ouellet, M., Hubbard, E. M., Toomarian, E. Y., Job, R., Treccani, B., & McShane, B. B. (2020). A multilab registered replication of the attentional SNARC effect. *Advances in Methods and Practices in Psychological Science*, *3*, (2), 143–162. https://doi.org/10.1177/2515245920903079

de la Fuente, J., Casasanto, D., & Santiago, J. (2015a). Observed actions affect body-specific associations between space and valence. *Acta Psychologica*, *156*, 32–36. https://doi.org/10.1016/j.actpsy.2015.01.004

de la Fuente, J., Casasanto, D., Román, A., & Santiago, J. (2015b). Can culture influence body-specific associations between space and valence? *Cognitive Science*, *39*, 821–832. https://doi.org/10.1111/cogs.12177

de la Fuente, J., Casasanto, D., Martínez-Cascales, J. I., & Santiago, J. (2017). Motor imagery shapes abstract concepts. *Cognitive Science*, *41*, (5), 1350–1360. https://doi.org/10.1111/cogs.12406

de la Vega, I., de Filippis, M., Lachma, M., Dudsch, C., & Kaup, B. (2012). Emotional valence and physical space: Limits of interaction. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, (2), 375–385. https://doi.org/10.1037/a0024979

Franco, A., Malhotra, N., & Simonovits, G. (2014). Publication bias in the social sciences: Unlocking the file drawer. *Science*, *345*, (6203), 1502–1505. https://doi.org/10.1126/science.1255484

Frese, J. (2024). Fitting *z*-curves to estimate the size of the UESD file drawer and the replicability of published findings. *Research & Politics*, *11*, (3), 1-8. https://doi.org/10.1177/20531680241277634

Kong, F. (2013). Space-valence associations depend on handedness: Evidence from a bimanual output task. *Psychological Research*, *77*, (6), 773–779. https://doi.org/10.1007/s00426-012-0471-7

Lakoff, G. & Johnson, M. (1999). *Philosophy in the flesh*. New York: Basic Books.

Li, H., & Cao, Y. (2019). The body in religion: The spatial mapping of valence in Tibetan practitioners of Bön. *Cognitive Science*, *43*, (4), e12728. https://doi.org/10.1111/cogs.12728

Mansoory, B., & Nassiri, V. (2022). Testing the body specificity hypothesis: A comparative study of the Persian language and the Persian Sign language. *Acta Psychologica*, *223*, 103496. https://doi.org/10.1016/j.actpsy.2022.103496

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., Ioannidis, John P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour*, *1*, (0021). https://doi.org/10.1038/s41562-016-0021

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*, (6251), aac4716–1–8. https://doi.org/10.1126/science.aac4716

Pitt, B., & Casasanto, D. (2020). The correlations in experience principle: How culture shapes concepts of time and number. *Journal of Experimental Psychology: General*, *149*, (6), 1048–1070. https://doi.org/10.1037/xge0000696

Pohl, J., & Miklashevsky, A. (2025). Vertical and horizontal space-valence associations: A meta-analysis. *Neuroscience & Biobehavioral Reviews*, *170*, 106054. https://doi.org/10.1016/j.neubiorev.2025.106054

Root, J. C., Wong, P. S., & Kinsbourne, M. (2006). Left hemisphere specialization for response to positive emotional expressions: A divided output methodology. *Emotion*, *6*, (3), 473–483. https://doi.org/10.1037/1528-3542.6.3.473

Santiago, J., Román, A., & Ouellet, M. (2011). Flexible foundations of abstract thought: A review and a theory. In T. W. Schubert & A. Maas (Eds.), *Spatial dimensions of social thought* (pp. 41–110). Mounton de Gruyter.

Schimmack, U. (2021). *Z*-Curve: An even better *p*-curve. *Replicability Index*, Retrieved from https://replicationindex.com/2021/04/25/z-curve-an-even-better-p-curve/

Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2014). *P*-curve: A key to the file-drawer. *Journal of Experimental Psychology: General*, *143*, (2), 534–547. https://doi.org/10.1037/a0033242

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology. *Psychological Science*, *22*, (11), 1359–1366. https://doi.org/10.1177/0956797611417632

Solana, P., & Santiago, J. (2022). Does the involvement of motor cortex in embodied language comprehension stand on solid ground? A *p*-curve analysis and test for excess significance of the TMS and tDCS evidence. *Neuroscience & Biobehavioral Reviews*, *141*, 104834. https://doi.org/10.1016/j.neubiorev.2022.104834

Solana, P., & Santiago, J. (2023). Worse than expected: A *z*-curve reanalysis of motor cortex stimulation studies of embodied language comprehension. *Psicologica*, *44*, (2), e15661. https://doi.org/10.20350/digitalCSIC/15661

Sotola, L. K., & Credé, M. (2022). On the predicted replicability of two decades of experimental research on system justification: A *Z*-curve analysis. *European Journal of Social Psychology*, *52*, (5–6), 895–909. https://doi.org/10.1002/ejsp.2858

Willems, R. M., Hagoort, P., & Casasanto, D. (2010). Body-specific representations of action verbs: Neural evidence from right- and left-handers. *Psychological Science*, *21*, (1), 67–74. https://doi.org/10.1177/0956797609354072

Yamada, Y., Xue, J., Li, P., Ruiz-Fernández, S., Özdoğru, A. A., Sarı, Ş., Torres, S. C., Hinojosa, J. A., Montoro, P. R., AlShebli, B., Bolatov, A. K., McGeechan, G. J., Zloteanu, M., Razpurker-Apfeld, I., Samekin, A., Tal-Or, N., Tejada, J., Freitag, R., Khatin-Zadeh, O., Banaruee, H., Robin, N., Briseño-Sanchez, G., Barrera-Causil, C. J., & Marmolejo-Ramos, F. (2024). Where the ‘bad’ and the ‘good’ go: A multi-lab direct replication report of Casasanto (2009, Experiment 1). *Memory & Cognition*, *53*, 1140–1146. https://doi.org/10.3758/s13421-024-01637-1

Yonemitsu, F., Sasaki, K., & Yamada, Y. (2023). The superiority of up/down over left/right in metaphorical association with emotion. *Japanese Journal of Psychonomic Science*, *42*, (1), 11–18. https://doi.org/10.14947/psychono.psychono.42.2

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.