**Reliability of Self-Prioritization Effect as Measured by the Perceptual Matching Task: Evidence from Multiple Datasets**

Zheng Liu 1#, Mengzhen Hu1#, Yuanrui Zheng2, Jie Sui3, Hu Chuan-Peng 1\*

1 School of Psychology, Nanjing Normal University, Nanjing, China

2 School of Education, Kunming City College, Kunming, China

3 School of Psychology, University of Aberdeen, Old Aberdeen, Scotland

# These authors are equally contributed to this study

\* Corresponding authors: Hu Chuan-Peng ([hu.chuan-peng@nnu.edu.cn](mailto:hu.chuan-peng@nnu.edu.cn); [hcp4715@hotmail.com](mailto:hcp4715@hotmail.com))

# **Abstract**

The self-prioritization effect (SPE) refers to the effect that performance on cognitive tasks is better when stimuli are related to the self than when they are not. In the last decade, the self -perceptual matching task (SPMT) has emerged as a mainstream paradigm for studying SPE due to its simplicity and elimination of familiarity effects. As a simple button-pressing task, SPMT yields two outcomes: reaction time and accuracy. Other indices can be derived from reaction times and accuracy, including sensitivity d prime under signal-detection theory, the efficiency index through a direct division between reaction times and accuracy, and drift rate (v) and starting point (z) estimated using drift-diffusion models. All these indices have been used to quantify SPE in the literature. However, the reliability of these SPE indices remains unexplored. To address this research gap, we conducted a pre-registered study wherein we re-analyzed existing data from 18 datasets from 11 papers using the split-half reliability and intraclass correlation coefficient (ICC). The result of split-half reliability revealed that the reliabilities of RT and Efficiency are relatively high but still lower than that required for good measure in psychometrics: RT, rClose = 0.40 (95% CI = [.18, .62]); rStragner = 0.45 (95% CI = [.25, .62]); Efficiency, rClose = 0.37 (95% CI = [.08, .64]); rStragner = 0.45 (95% CI = [.16, .67]). The estimated split-half reliabilities for the other four indices are lower than 0.3. The results of ICC confirmed the pattern in split-half reliability and found that ICC2 for reaction times is the highest among all six SPE indices: ICC2Close = 0.53 (95%CI = [.39, .69]); ICC2Stranger = 0.58 (95%CI = [.45, .73]). However, estimated ICC2K suggested that SPE assessed by reaction times is reliable for group-level analysis: ICC2KClose = 0.87 (95%CI = [.79, .93]); ICC2KStranger = 0.89 (95%CI = [.82, .94]). Together, these findings call for attention to the reliability of SPMT when researchers are interested in individual differences of SPE.

Keywords: Self-Prioritization Effect (SPE), Self-Perceptual Matching Task (SPMT), Reliability, Multiverse

# **1 Introduction**

The Self-Prioritization Effect (SPE) refers to the phenomenon whereby performance in cognitive tasks is better when stimuli are related to the self than when they are not. This effect has been widely documented and confirmed since the 1950s. In the early days of cognitive psychology, researchers found that subjects were able to recognize their own names, even when they were mixed with a noisy auditory background and not the target of the task in dichotic listening tasks (Cherry, 1953; Moray, 1959). SPE effect was then reported in memory research by Craik and Tulving (1975), who found that participants were able to recall more words when they were related to the self compared to when they were processed at other levels (e.g., semantic). This SPE effect in memory was then replicated by many others (Conway & Dewhurst, 1995; Rogers et al., 1977; Symons & Johnson, 1997). In the following decades, the SPE has also been found to occur with different stimuli, such own face (Keenan et al., 2000; Kircher et al., 2000; Turk et al., 2002), own voice (Hughes & Harrison, 2013; Payne et al., 2021), own name (Constable, Rajsic, et al., 2019), and newly owned object (Strachan et al., 2020). SPE was found across a variety of cognitive tasks, such as perceptual task (Cunningham & Turk, 2017; Desebrock et al., 2018), decision-making task (Sui & Humphreys, 2013), attentional task (Shapiro et al., 1997), and ownership task (Cunningham et al., 2008).

Although SPE is often argued to be a self-specific effect, it can be challenging to disassociate it from the familiarity effect since most studies use stimuli owned by participants or by others. Sui et al. (2012) proposed a paradigm where participants first associate geometrical shapes (e.g., triangle, square, and circle) with labels of persons (e.g., "You," "friend," and "stranger") and then perform a perceptual matching task in which they decide if the shape-label pairs presented on the screen match the learned association or not (Sui et al., 2012). Because the task requires participants to learn the social meaning of different geometric shapes, it is called the Self-Perceptual Matching Task (SPMT). In this task, Sui et al. (2012) found that shapes associated with the self are performed better, with faster response times, better accuracy, and/or higher sensitivity scores, compared to shapes associated with friends and strangers. Because the self-relatedness is acquired immediately right before they start the perceptual matching task, this paradigm eliminated the effect of familiarity of the stimuli.

Since then, the SPMT has become the mainstream method for investigating the mechanism underlying the SPE. For instance, researchers have explored the importance of personality traits in identity labels (Golubickis et al., 2020), the self-relevant labels that include the past, present, and future self (Golubickis et al., 2017), as well as "good self" and "bad self" labels (Hu et al., 2020), and the group advantage effect of in-group labels (Constable, Elekes, et al., 2019; Constable & Knoblich, 2020; Enock et al., 2018; Enock et al., 2020). Moreover, the SPMT has been applied to various fields. In neuroscience and physiology, researchers investigate which brain regions are activated during self-prioritization effect (Feng et al., 2018; Humphreys & Sui, 2015), and gender differences in self-prioritization effect due to oxytocin (Feng et al., 2020). In clinical research, SPMT has been used to understand atypical self-processing in populations such as those with autism or depression (Gillespie‐Smith et al., 2018; Nijhof & Bird, 2019; Sui & Humphreys, 2017). Cross-cultural studies have shown that individuals from individualistic cultures demonstrate a stronger self-prioritization effect (Jiang et al., 2019), and that the language of the experimental stimuli can affect the strength of the effect (Ivaz et al., 2016). Finally, the SPMT has also been applied to child development, with studies examining developmental changes in self-positivity effects (Maire et al., 2020; Zhou et al., 2019).

While SPMT has gained widespread adoption as a prominent method for investigating the underlying mechanism of the self-prioritization effect, there has been microscopic examination and report of the psychometric properties of the outcomes, necessitating a careful evaluation. (Parsons et al., 2019; Zorowitz & Niv, 2023). Given the increasing use of SPMT to assess individual differences in fields such as psychiatry (Liu et al., 2022) and social psychology (Enock et al., 2018). it is crucial to ensure a high degree of measurement consistency to accurately assess human perceptual abilities (Parsons et al., 2019). Furthermore, in tasks as simple as the SPMT, there are multiple approaches to quantify the self-prioritization effect. These include two direct measures based on SPMT, namely reaction times (RT) and accuracy (ACC), as well as derived measures such as efficiency (Humphreys & Sui, 2015; Stoeber & Eysenck, 2008), *d* prime of Signal Detection Theory (SDT) (Hu et al., 2020; Sui et al., 2012), and drift rate (*v*) and starting point (*z*) from Drift Diffusion Model (DDM) (Golubickis et al., 2017).

To address the existing research gap, the present study investigated the reliability of self-prioritization effect (SPE) indices in the self-perceptual matching task (SPMT). We examined six SPE indices as mentioned earlier, that are supposed to capture the disparity between self-related and other-related stimuli of the matching trials. This was achieved by reanalyzing data obtained from previous studies that employed SPMT. Given the diverse methods available for evaluating the reliability of cognitive tasks, we employed both the Split-Half Reliability and Intraclass Correlation Coefficient (ICC) to determine the reliability of each SPE index. These findings deepen our understanding the reliability of SPE as measured by SPMT and facilitate the future usage of SPMT.

# **2 Methods**

## 2.1 Ethics information

This research used secondary data, informed consent is not applicable here.

## 2.2 Datasets

Below, we first provided a brief overview of the original experimental design of SPMT, as described in the Experiment 1 by Sui et al. (2012). Then, we gave an overview of 18 datasets used in the current analysis.

The original SPMT used a 2 by 3 within-subject design. The first independent variable, labeled “Matching,” consisted of two levels: “Matching” and “Nonmatching,” indicating whether the shape and label were congruent. The second independent variable, labeled “Identity,” comprised three levels: “Self”, “Friend”, and “Stranger”, representing the corresponding identity associated with the shape.

The procedure of original SPMT consisted of two phases (see Fig. 1). In the first phase (learning phase), participants completed a learning task in which they associated three geometric shapes (circle, triangle and square) with three labels (self, friend, and stranger) for approximately 60 seconds. The shape-label associations were balanced across participants. In the second phase (formal experimental phase), participants completed a perceptual matching task. Each trial started with a fixation cross displayed in the center of the screen for 500 ms, followed by a shape-label pairing and fixation cross for 100 ms. the screen then went blank for 1500 ms, or until a response was made. Participants were required to judge whether the presented shape and label matched the learned associations from the learning phase and respond as quickly and accurately as possible by pressing one of two buttons within the allotted timeframe. Prior to the formal experimental phase, participants completed a training session consisting of 24 practice trials. After the training, participants completed six blocks of 60 trials in the matching task, with two matching types (matching/nonmatching) and three shape associations, for a total of 60 trials per association. Short breaks lasting up to 60 seconds were provided after each block.



**Fig. 1.** Procedure of the original SPMT in the Experiment 1 (Sui et al., 2012).*Note*: The relation between shape-label pairs is counter-balanced between participants.

In this study, we collected a total of 18 datasets from 11 research articles, and one from our laboratory (Hu et al., 2023). and one from our collaborators (Liu et al., 2023), that included raw data from empirical studies utilizing the SPMT. These datasets were included in the analysis based on two criteria: (1) the experimental design did not deviate from the original SPMT of Sui et al. (2012); (2) the trial-level data is available so that we can estimate at least one reliability index. Seven studies opened their raw data publicly (Constable et al., 2021; Constable & Knoblich, 2020; Golubickis & Macrae, 2021; Navon & Makovski, 2021; Qian et al., 2020; Schäfer & Frings, 2019; Svensson et al., 2022) and did not deviate from the original experimental paradigm. Additionally, we identified five articles that did not have publicly available data but mentioned that data could be obtained upon request (Bukowski et al., 2021; Cheng & Tseng, 2019; Kolvoort et al., 2020; Martínez-Pérez et al., 2020; Woźniak et al., 2018; Xu et al., 2021). One of these articles indicated that data were shared on the Open Science Framework (OSF) platform (https://osf.io/pcv3u/), but the repository was found to be empty (Bukowski et al., 2021). We included datasets with raw data that were shared with us. It is worth noting that the nature of the research culture discourages direct replications (Makel et al., 2012); thus, all datasets included in our analysis involved some degree of modification to the original design, such as incorporating additional independent variables or using different experimental materials (see our preregistration for details). Intraclass Correlation Coefficients (ICC) were calculated if there are test-retest data, otherwise only coefficients for split-half reliability were estimated. The details of these included datasets are described in Table 1.

Table 1. Dataset Information

| Paper | Exp. | Independent Variable | | | | Sample  Size | # of Trials per Condition | Self-Prioritization Effect Indices | | | | | | Reliability | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| IV 1 | IV 2 | IV 3 | IV 4 | RT | ACC | d | Eff | v | z | ICC | SHR |
| Hu et al. (2023) | 1 | Matching | Identity | Emotion  Control, Neutral,  Happy, Sad | Session | 34 | 60 | √ | √ | √ | √ | √ | √ | √ | √ |
| Constable and Knoblich (2020) | 1 | Matching | Identity | Switch Identity  Partner, Stranger | Phase | 92 | 40 | √ | √ | √ | √ | √ | √ |  | √ |
| Constable et al. (2021) | 2 | Matching | Identity  Self; Stranger |  |  | 51 | 24 | √ | √ | √ | √ | √ | √ |  | √ |
| Qian et al. (2020) | 1 | Matching | Identity Self; Stranger; Celebrity | Mood (Session) |  | 24 | 24 | √ | √ | √ | √ | √ | √ |  | √ |
| 2 | Matching | Identity Self; Celebrity | Cue  With, Without |  | 25 | 50 | √ | √ | √ | √ | √ | √ |  | √ |
| Schäfer and Frings (2019) | 1 | Matching | Identity Self; Mother; Acquaintance |  |  | 103 | 24 | √ | √ | √ | √ | √ | √ |  | √ |
| Golubickis and Macrae (2021) | 1 | Matching | Identity | Presentation Mixed; Blocked |  | 30 | 30 | √ | √ | √ | √ | √ | √ |  | √ |
| Navon and Makovski (2021) | 1 | Matching | Identity |  |  | 13 | 60 | √ | √ | √ | √ | √ | √ |  | √ |
| 3 | Matching | Identity  Self; Father; Stranger |  |  | 27 | 60 | √ | √ | √ | √ | √ | √ |  | √ |
| 4 | Matching | Identity |  |  | 26 | 60 | √ | √ | √ | √ | √ | √ |  | √ |
| Svensson et al. (2022) | 1 | Matching | Identity Self; Friend |  |  | 20 | 50 | √ | √ | √ | √ | √ | √ |  | √ |
| 2 | Matching | Identity Self; Friend | Frequency  self > friend |  | 24 | 100 | √ | √ | √ | √ | √ | √ |  | √ |
| 3 | Matching | Identity Self; Friend | Frequency  self < friend |  | 25 | 100 | √ | √ | √ | √ | √ | √ |  | √ |
| Cheng and Tseng (2019) | 1 | Matching | Identity | Go/No-go |  | 22 | 75 | √ | √ | √ | √ | √ | √ |  | √ |
| 2 | Matching | Identity | Go/No-go |  | 26 | 75 | √ | √ | √ | √ | √ | √ |  | √ |
| 3 | Matching | Identity | Go/No-go |  | 22 | 75 | √ | √ | √ | √ | √ | √ |  | √ |
| Bukowski et al. (2021) | 1 | Matching | Identity | Imitation |  | 91 | 60 | √ | √ | √ | √ | √ | √ |  | √ |
| 2 | Matching | Identity | Imitation |  | 109 | 60 | √ | √ | √ | √ | √ | √ |  | √ |
| Kolvoort et al. (2020) | 1 | Matching | Identity | Delay  0, 40, 120, 700 |  | 31 | 25 | √ | √ | √ | √ | √ | √ |  | √ |
| Martínez-Pérez et al. (2020) | 1 | Matching | Identity | Stimulation |  | 90 | 40 | √ | √ | √ | √ | √ | √ |  | √ |
| Xu et al. (2021) | 1 | Matching | Identity | Feedback | Sex | 105 | 60 | √ | √ | √ | √ | √ | √ |  | √ |
| Woźniak et al. (2018) | 1 | Matching | Identity | Facial Gender  Mele; Female |  | 18 | 56 | √ | √ | √ | √ | √ | √ |  | √ |
| 2 | Matching | Identity | Facial Gender  Mele; Female |  | 18 | 60 | √ | √ | √ | √ | √ | √ |  | √ |
| Liu et al. (2023) | 1 | Matching | Identity  Self; Stranger |  |  | 298 | 16 | √ | √ | √ | √ | √ | √ |  | √ |

Note. SPE: Self-Prioritization Effect, ICC: Intraclass Correlation Coefficient, SHR: Split-Half Reliabilit

## 2.3 Analysis



**Fig. 2** Roadmap of the current study. Note: SPE: self-prioritization effect; d-prime is the sensitivity index under the Signal Detection Theory; drift rate (v) and starting point (z) are parameters derived from the Drift-diffusion Model; ICC: Intraclass Correlation Coefficient, SHR: Split-half Reliability.

All analyses in this paper are performed using the statistical software R (R Core Team, 2023). The research flow of the current study is visually represented in Fig. 2. After collecting the data, we performed data cleaning and calculated the six indices' SPE separately for different targets. Finally, we calculated the split-half reliabilities of these SPE values. If there are repeated measurements in the dataset, we also computed the test-retest reliability using the intraclass correlation coefficient (ICC).

### 2.3.1 Data pre-processing

In total, we gathered 18 publicly available datasets, as mentioned earlier and presented in Table 1. We used the following criteria for data exclusion:

1. Participant exclusion criteria
2. Participants who had wrong trial numbers because of procedure errors is excluded from the analysis,
3. Participants with an overall accuracy < 0.5 is excluded from the analysis,
4. Participants with any of the conditions with zero accuracy is excluded from the analysis.
5. Behavioural data exclusion criteria
6. Trials with no response or wrong key press is excluded from the analysis,
7. The practice trials is excluded from the formal analysis,
8. Participants with any of the conditions with zero accuracy is excluded from
9. the analysis,
10. The data under conditions other than the “control condition” would not be used in the current study.

### 2.3.2 Calculating the SPE

For each dataset, we calculated six indices for each experimental condition: Mean RT (MRT), accuracy (ACC), d-prime (*d′*), efficiency (*η*), drift rate (*v*), and starting point

(*z*). Mean RT and ACC are obtained directly from the datasets, while *d′* and *η* are calculated based on Mean RT and ACC using a simple formula (see Table 2). Please note that the condition “Other” may vary across studies, we calculated the SPE for each “Other” condition. More specifically, we calculated the differences for “Self vs Close”, “Self vs Stranger”, “Self vs Celebrities” and “Self vs None condition”.

Table 2. Indices in SPMT and corresponding calculation of indices and SPE

| **Indices** | **Indices Calculation** | **SPE Calculation Based on Indices** | **Source** |
| --- | --- | --- | --- |
| Mean  Reaction Times  (RT) |  |  | Sui et al. (2012) |
| Accuracy (ACC) |  |  | Sui et al. (2012) |
| *d* prime |  |  | Sui et al. (2012) |
| Efficiency |  |  | Humphreys and Sui (2015); Stoeber and Eysenck (2008) |
| Drift rate (*v*) | DDM：parameters will be identified through model selection |  | Golubickis et al. (2017) |
| Starting Point (*z*) |  | Golubickis et al. (2017) |

Note. DDM: Drift Diffusion Model.

### 2.3.3 Estimating the Reliability

**Split-half reliability.** We calculated the split-half reliability of the six indices using four types of split-half reliability measures: first-second, odd-even, permutated, and Monte Carlo (Kahveci et al., 2022; Pronk et al., 2022). The first-second split divides the front and back halves of trials, while the odd-even split divides trials into odd and even numbered sequences. And the permutation split shuffles the trial order and randomly assigns each half to a group. The Monte Carlo split-half is similar to the permutated split-half, but it repeats the process thousands of times to calculate the average and 95% confidence interval of the split-half reliability.

First, the data were stratified according to, Matching, Identity and Session (if applicable). If the data is not stratified, directly splitting it in half will result in an uneven distribution of trials for each experimental condition in the two halves, which can lead to an overestimation or underestimation of split-half reliability (Pronk et al., 2022). Therefore, once the data is stratified, we split it into two halves. For example, when using Monte Carlo Split-Half, we randomly split the data into two halves. Then we repeat this process 5000 times. This will result in 5000 pairs of two halves of the data. Next, we use these 5000 pairs of data to calculate 5000 Pearson correlation coefficients, and then obtain the average and 95% confidence interval of the Monte Carlo split reliability. First-second split, odd-even split, and permutated split are similar to Monte Carlo method, but they only perform one split, so only one split-half reliability is obtained without an interval estimate of the split-half reliability.

**Test-Retest Reliability (ICC).** We assessed the test-retest reliability of the six indices in our dataset that involved test-retest sessions by calculating the Intraclass Correlation Coefficient (ICC). To perform this analysis, we utilized the “psych” package as described by (Revelle, 2017). ICC is a well-established measure used in test-retest, intra-rater, and inter-rater studies to assess reliability (Fisher, 1992). Unlike the Pearson correlation coefficient, ICC considers both the correlation and agreement between multiple measurements, making it a more comprehensive measure of test-retest reliability. Within the ICC family, we specifically employed ICC2 and ICC2k. ICC2 focuses on the individual-level reliability of the indices, while ICC2k evaluates the reliability of mean ratings furnished by a group of judges (Koo & Li, 2016). For the calculation of ICC2 estimates, the formula is:

where MSBS is the mean square between subjects, MSE is the mean square error, MSBM is the mean square between measurements, k is the number of measurements, n is number of participants. For the calculation of ICC2k estimates, the formula is:

Although there is no strict criterion for defining the level of reliability, a widely accepted guideline for Cronbach’s alpha is that a value of 0.60 is “acceptable”, and a value greater than 0.8 means excellent reliability (Cicchetti & Sparrow, 1981; Kupper & Hafner, 1989).

# **3 Deviation from preregistration**

In our initial pre-registration plan, we intended to estimate the drift rate (*v*) and starting point (*z*) of the drift-diffusion model (DDM) using the “fit\_ezddm” function from the “hausekeep” package (Lin et al., 2020). This function was a wrapper for the EZ-DDM function (Wagenmakers et al., 2007). However, during parameter recovery (see supplementary Fig. 1), we observed a substantial disparity between the parameter starting point (z) provided by the "hausekeep" package and those obtained using the original HDDM package (Wiecki et al., 2013) employed in prior studies (Golubickis et al., 2017). This discrepancy may be attributed to the assumption made by "hausekeep" that z = a / 2. As a result, we decided to replace the original package with the “RWiener” package (Wabersich & Vandekerckhove, 2014), which we found to yield the most comparable results to the HDDM package (see our parameter recovery in the supplementary materials). To transparently report our results, we reported the results from ezDDM in supplementary materials.

# **4 Results**

In 18 datasets, 14 datasets have the data for Self vs Close other, 12 datasets have the data for Self vs Stranger, 2 datasets have the data for Self vs Celebrities, 1 dataset has the data for Self vs none condition.

## 4.1 Split-Half Reliability

As described in method part, we utilized four different methods to calculate split-half reliability, namely the first-second, odd-even, permuted, and Monte Carlo methods. Compared to the first-second, odd-even, and permuted split-half methods, the Monte Carlo split-half method yields more stable split-half reliabilities (Pronk et al., 2022). The results of the first three split-half reliabilities will be presented in the supplementary materials. Here, we will focus on discussing the results of Monte Carlo split-half reliabilities.

The vertical axis represents six different indices, and the horizontal axis represents split-half reliability. As depicted in Fig. 3, “RT” and “Efficiency” exhibit higher reliabilities. There are several datasets with split-half reliabilities exceeding 0.6, which is considered an acceptable level of reliability. After averaging the Monte Carlo split-half reliabilities across the 18 datasets, the reliability of RT is 0.40 (95% CI = [.18, .62]) when the Target is set as "Close," and it is 0.45 (95% CI = [.25, .62]) when the Target is set as "Stranger." Similarly, the reliability of Efficiency is 0.37 (95% CI = [.08, .64]) when the Target is set as "Close," and it is 0.45 (95% CI = [.16, .67]) when the Target is set as "Stranger." The Monte Carlo split-half reliabilities of "Accuracy" is approximately 0.3, while the reliabilities of "*d’*" and "rwDDM v" are around 0.2. However, the reliabilities of "rwDDM z" is almost 0. Overall, except for RT and Efficiency, the reliabilities of other indices are completely unacceptable.

图片包含 图表

描述已自动生成

**Fig. 3.** Monte Carlo Split-Half Reliability for different SPE indices.

RT: reaction times; ACC: accuracy; *d’*: sensitivity index in signal detection theory; Efficiency: ratio of mean reaction time to average accuracy in matching group, *v*: drift rate in drift diffusion model; *z*: starting point in drift diffusion model. From top to bottom, each color represents one of the 8 indices of SPE. From left to right, each facet in the figure represents a different target for the Self-Prioritization Effect (SPE), namely close other, stranger, celebrity, and none.

## 4.2 Intraclass correlation coefficient (ICC)

It is important to note that we could only calculate ICC for our own data (Hu et al., 2023). Because all other datasets did not include re-test sessions.

As shown in Fig. 4, The ICC2k values for RT and Efficiency are quite high, while other indices are quite low. When the target is "Close other," the ICC2k for RT is 0.87 (95%CI = [.79, .93]), and for Efficiency, it is 0.86 (95%CI = [.78, .93]). When the target is "Stranger," the ICC2k for RT is 0.89 (95%CI = [.82, .94]), and for Efficiency, it is 0.76 (95%CI = [.61, .87]). However, RT and Efficiency exhibit lower ICC2 values. When the target is "Friend," the ICC2 for RT is 0.53 (95%CI = [.39, .69]), and for Efficiency, it is 0.52 (95%CI = [.38, .68]). When the target is "Stranger," the ICC2 for RT is 0.58 (95%CI = [.45, .73]), and for Efficiency, it is 0.34 (95%CI = [.21, .52]). This suggests that, at the individual level, the SPMT paradigm does not demonstrate a robust test-retest reliability. But, at the group level, this paradigm exhibits a high test-retest reliability.

**电脑屏幕的地图

描述已自动生成**

**Fig. 4** Intraclass correlation coefficient.

RT: reaction times; ACC: accuracy; *d’*: sensitivity index in signal detection theory; Efficiency: ratio of mean reaction time to average accuracy in matching group, *v*: drift rate in drift diffusion model; *z*: starting point in drift diffusion model. The vertical axis represents eight different indices, and the horizontal axis represents intraclass correlation coefficients.

# **5 Discussion**

Evaluating the reliability of a behavioral paradigm is essential for researchers planning to use the paradigm to investigate different research questions, such as individual differences and underlying mechanisms. Despite its importance, this practice is not yet widely adopted among researchers who primarily using cognitive tasks (Green et al., 2016; Hedge et al., 2018; Parsons et al., 2019). In this pre-registered study, our objective is to investigate the reliability of the indices related to the self-prioritization effect (SPE) in the self-perceptual matching task (SPMT). We re-analyzed data from 18 datasets across 11 articles by employing the split-half reliability and intraclass correlation coefficient (ICC2, ICC2k) for this purpose. Our analysis of these datasets collectively demonstrate that RT and Efficiency yield better results compared with other indices, and the result varies between different associations.

[group-level reliability & its implication]

[Importance of reliability in accessing individual differences and the implication of the current results]

[More nuanced results: index matters?]

[More nuanced results: applying DDM may not useful?]

[More nuanced results: control conditions matter?]

[More nuanced results: trial number matters?]

For the split-half reliability, the results indicate variations between the targets and indices. Specifically, response time (RT) and Efficiency yield superior results compared to other indices, suggesting they are the two most reliable indicator for accurately assessing and distinguishing between the self and other targets in the SPMT. Furthermore, when examining the split-half reliability for the self-other difference, self-stranger is the highest among other comparisons. Such finding suggests that the measurement of this particular difference remains consistent across participants and studies, indicating the systematic processing difference between these targets. The self-friend differences demonstrate the lowest reliability, indicating that the measurements obtained from the indices are not stable or consistent when split into two halves. This finding suggests that distinguishing between the self and friend in the paradigm is challenging. The aforementioned result aligns with previous studies conducted using SPMT, consistently demonstrating the establishment of a reliable self-advantage. This advantage is observed when the shapes are associated with the self, in comparison to when they are linked to an unfamiliar person or a neutral label (Feng et al., 2020; Sui et al., 2012).

Although RT and efficiency yield the highest split-half reliability among other indices, the result is still below a commonly considered excellent reliability (higher than 0.8, even higher than 0.9). The low split-half reliability may suggest that the indices in SPMT is subject to random error and inconsistent performance. Several plausible reasons can be identified. Firstly, a potential contributor to low split-half reliability could be the insufficient number of trials per condition. A recent study by Kucina et al. (2023) has highlighted the significance of the number of trials in cognitive tasks for determining reliability. The findings of the study revealed that increasing the number of trials resulted in greater magnitudes of conflict effects and individual differences. Consequently, this led to improved reliability when compared to previous archival data. Specifically, in the case of gamified Flanker task, the study identified that achieving satisfactory reliability required 48 or fewer trials, while achieving a higher level of reliability necessitated 72 trials. Therefore, incorporating a higher number of trials in future employment of the SPMT paradigm may enhance the split-half reliability by enhancing the consistency of measurement. Second, it is worth mentioning the influence of serial dependence effects on task reliability. A recent set of studies has examined serial dependence effects in a variety of cognitive tasks (Braun et al., 2018; Zhang & Alais, 2020). Serial dependence refers to the phenomenon in which the outcome of one trial is influenced by preceding trials, resulting in a systematic relationship between consecutive trials (Pascucci et al., 2023). Notably, studies in the field of perceptual decision making have demonstrated strong serial dependence effects in perception, even when the visual stimuli were reliable and varied randomly over time (Fischer & Whitney, 2014; John-Saaltink et al., 2016). In particular, if the split-half design unintentionally separates temporally adjacent trials in the SPMT, the presence of serial dependence may introduce performance differences between the halves, leading to a reduction in the reliability estimate. Thus, to accurately control for the impact of serial dependence in experiments, further research should employ appropriate statistical methods that account for the temporal dependencies between trials. Time series analysis techniques (Huitema, 1986) or modeling approaches that capture the serial correlation (Mei et al., 2023)can be utilized to obtain more accurate results.

After employing the drift diffusion model to analyze reaction time in the SPMT, we observed that the test-retest reliability of the model parameters was notably low. This finding prompt consideration of several potential explanations. Firstly, the hierarchical drift diffusion model relies on specific assumptions regarding the cognitive processes underlying reaction time. If these assumptions fail to accurately capture the true underlying mechanisms or if they are violated within the SPMT, it can result in inconsistent parameter estimates and diminished reliability (Johnson et al., 2017). Secondly, individual differences among participants in their cognitive processing may contribute to inconsistent parameter estimates. Factors such as fluctuations in attention, motivation, or learning effects could vary across test-retest sessions, thereby impacting parameter reliability. These results raise significant questions regarding the efficacy of applying the hierarchical drift diffusion model to the SPMT. Further investigation is warranted to better understand the sources of low test-retest reliability and explore alternative modeling approaches that may yield more robust parameter estimates.

We observed that the response time (RT) and efficiency measures demonstrated high group-level test-retest reliability (ICC2k), indicating good to excellent consistency over time. However, at the individual level, the reliability (ICC2) of these measures was relatively low. On the other hand, the other indices showed low levels of reliability at both the group and individual levels. Specifically, the RT index exhibited an ICC2k ranging from 0.8 to 0.9, suggesting strong reliability at the group level. Similarly, the Efficiency measures of the SPMT task showed an ICC2k ranging from 0.75 to 0.9, indicating good consistency at the group level. However, when examining individual-level reliability, all indices performed poorly, with ICC2 values ranging from 0.3 to 0.5. It is common for behavioral paradigm to have such result pattern, as demonstrated in previous research testing other cognitive paradigms such as Flanker, Simon, or Stroop (Clark et al., 2022; Mollon et al., 2017). The result also aligns with the reliability paradox that has been previously proposed (Logie et al., 1996). The discrepancy between the high ICC2k and low ICC2 suggests that the SPMT is more influenced by between-participant variability than within-participant variability (Hedge et al., 2018; Liljequist et al., 2019). There are various reasons for this pattern. First, one significant factor could be the prevalence of practice effect, particularly if the practice effect is large enough to cause a substantial change in participants’ performance between measurement occasions, it can introduce additional variability in the measurements. This increased variability may lower the ICC2 (Oswald et al., 2015; Siegelman et al., 2017). The presence of a practice effect underscores the need for alternative measures that can consistently capture performance nuances and reveal individual differences more sensitively (Hedge et al., 2018). Second, behavioral paradigms are susceptible to factors such as external conditions, contextual differences etc.., which contribute to greater within-participant variability and lower ICC2 values. However, when averaging performance between different individuals, the task could still exhibit good consistency, resulting in higher ICC2k values. It's important to note that low ICC values should not be solely interpreted as a measure of a test's overall quality but rather as an indication of the types of questions it can effectively address. In practical terms, the results suggest that the SPMT is better suited for distinguishing performance differences between individuals or groups, rather than capturing consistent performance within the same individuals over time. Thus, the SPMT may be particularly useful for studying inter-individual variability or conducting group-level comparisons, rather than tracking individual-level changes or stability. Therefore, we recommend that researchers take these factors into consideration when investigating individual differences in performance using the SPMT.

Our study has a few limitations that should be acknowledged. Firstly, although we made efforts to enhance sample diversity by including open data as much as possible, it is important to note that a majority of our samples still consisted of individuals from what is commonly referred to as “wired” populations (Rad et al., 2018). Therefore, our findings may not be fully representative of the broader population, and a more diverse sample is needed to ensure greater generalizability of the paradigm. Additionally, it is important to highlight that the majority of the studies included in our analysis focused on adults from healthy populations. Hence, further investigation is needed to determine the reliability of the SPMT across different age groups and clinical populations. Secondly, it is important to clarify the aim of our study, which primarily focused on exploratory purposes and providing information regarding the current state of reliability for the assessed indices. Consequently, it is recommended that future research focuses on modifying the paradigm and conducting tests to assess potential improvements. We propose several approaches that could be considered, such as introducing more challenging task variations, which have the potential to increase the reliability of accuracy measurements. Another suggestion is to include a greater number of trials for each condition, as this may contribute to improved reliability. It is strongly encouraged to undertake further investigation and experimentation in order to refine the paradigm and enhance the reliability of the indices, rather than dismissing the paradigm under certain circumstances.

In conclusion, the current study find that RT and Efficiency provide a more robust result than other indices. Moreover, SPMT is more suitable for group-level analysis rather than assessing individual-level variation. The findings of our study offer significant insights into the reliability of SPMT, shedding light on important factors that require careful consideration when interpreting the reliabilities. These findings also have implications for future task design and data collection protocols aimed at improving reliability. Ultimately, our study paves the way for the prospective utilization of these tasks, in various domains including research, clinical applications, and personal performance monitoring. The information obtained from our study contributes valuable knowledge to the field and sets the stage for further investigations and advancements in utilizing SPMT effectively.

# **Acknowledgements**

The present research is support by.

# **Author contributions**

HCP contributed to the conception and supervision of the study. JS contributed to fund raising, HCP contributed to data collection. ZL, ZYR and HMZ will perform the data pre-processing, analysis and visualize the results. In addition, ZL, JS, HMZ and HCP will contribute to discussing the results and the drafting of the final manuscript. All authors will critically revise the manuscript.

# **Data and Material Availability**

The pre-registration plan is available at <https://osf.io/zv628>. The de-identified raw data from our lab is available at <https://doi.org/10.57760/sciencedb.08117>. The simulated data is accessible on GitHub (<https://github.com/Chuan-Peng-Lab/ReliabilitySPE>).

# **Code Availability**

Code used to simulate and analyze the data is made accessible at <https://github.com/Chuan-Peng-Lab/ReliabilitySPE>.

# **Competing interests**

The authors declare no competing interests.

**Supplementary information**

**图片包含 图示

描述已自动生成**

**Supplementary Fig. 1** Other 3Split-Half Reliabilities for different SPE indices

RT: reaction times; ACC: accuracy; *d’*: sensitivity index in signal detection theory; Efficiency: ratio of mean reaction time to average accuracy in matching group, *v*: drift rate in drift diffusion model; *z*: starting point in drift diffusion model. From left to right, the figure represents the split-half reliabilities calculated using three different methods: first-second, odd-even, and permuted. From top to bottom, each facet in the figure represents a different target for the Self-Prioritization Effect (SPE), namely friend, stranger, celebrity, and none.

图示

描述已自动生成

**Supplementary Fig. 2.** Monte Carlo Split-Half Reliabilities for different SPE indices.

RT: reaction times; ACC: accuracy; *d’*: sensitivity index in signal detection theory; Efficiency: ratio of mean reaction time to average accuracy in matching group, *v*: drift rate in drift diffusion model; *z*: starting point in drift diffusion model. From top to bottom, each color represents one of the 8 indices of SPE. From left to right, each facet in the figure represents a different target for the Self-Prioritization Effect (SPE), namely close other, stranger, celebrity, and none.

Contrary to the main text, this figure also includes ezDDM (“hausekeep”), which also estimates drift rate (v) and starting point (z). Due to the assumption of z = a / 2 in hausekeep, its estimation of parameter a is highly inaccurate. Therefore, we did not report the results obtained from this package in the main text. From this graph, we can observe that ezDDM demonstrates higher stability in estimating the drift rate (v). It appears that the estimation method used in "hausekeep" relies on average reaction time and accuracy, while the estimation method in "RWiener" relies on individual trial-level reaction times and correctness. The chosen split-half procedure may have a greater impact on the estimation by "RWiener," leading to lower split-half reliabilities for both of its indices. This also suggests that the reliability of calculating DDM parameters through split-half reliabilities depends on two factors. Firstly, the stability of the parameter itself and secondly, the stability of the package used to compute that parameter. It is also possible that DDM itself may not be suitable for estimating the SPMT paradigm. It is possible that in the future, DDM variants specifically tailored for the SPMT paradigm might be needed.

图表, 折线图

描述已自动生成

**Supplementary Fig. 3** Intraclass correlation coefficient for different SPE indices.

RT: reaction times; ACC: accuracy; *d’*: sensitivity index in signal detection theory; Efficiency: ratio of mean reaction time to average accuracy in matching group, *v*: drift rate in drift diffusion model; *z*: starting point in drift diffusion model. The vertical axis represents eight different indices, and the horizontal axis represents intraclass correlation coefficient.

图形用户界面, 图示, 应用程序

描述已自动生成

**Supplementary Fig. 4 DDM Packages Comparison**

a: threshold parameter; t: non-decision time; v: drift rate; z: starting point. The y-axis represents three different R packages for estimating DDM parameters: “RWiener”, “hausekeep”, and “FastDMinR”. There are a total of five methods for estimating DDM parameters (3 method in FastDMinR). The x-axis represents the values of the estimated parameters. The dashed line represents the true value of the parameter.

We generated 100 datasets using the DDM package HDDM in Python. These datasets were set with parameters a=2, t=0.3, v=1, and z=0.7. Therefore, if the values obtained from the computation using the three DDM packages in R Package are closer to our set values, it indicates that the package provides more accurate parameter estimation.

**References**

Braun, A., et al. (2018). Adaptive history biases result from confidence-weighted accumulation of past choices. *Journal of Neuroscience, 38*(10), 2418-2429.

Bukowski, H., et al. (2021). Socio-cognitive training impacts emotional and perceptual self-salience but not self-other distinction. *Acta psychologica, 216*, 103297. <https://doi.org/10.1016/j.actpsy.2021.103297>

Cheng, M., & Tseng, C.-h. (2019). Saliency at first sight: instant identity referential advantage toward a newly met partner. *Cognitive Research: Principles and Implications, 4*(1), 1-18. <https://doi.org/10.1186/s41235-019-0186-z>

Cherry, E.C. (1953). Some experiments on the recognition of speech, with one and with two ears. *The Journal of the acoustical society of America, 25*(5), 975-979. <https://doi.org/10.1121/1.1907229>

Cicchetti, D.V., & Sparrow, S.A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *Am J Ment Defic, 86*(2), 127-137. <https://psycnet.apa.org/record/1982-00095-001>

Clark, K., et al. (2022). Test-retest reliability for common tasks in vision science. *Journal of Vision, 22*(8), 18-18.

Constable, M.D., et al. (2021). Affective compatibility with the self modulates the self-prioritisation effect. *Cognition and Emotion, 35*(2), 291-304. <https://doi.org/10.1080/02699931.2020.1839383>

Constable, M.D., et al. (2019). Relevant for us? We-prioritization in cognitive processing. *Journal of Experimental Psychology: Human Perception and Performance, 45*(12). <https://doi.org/10.1037/xhp0000691>

Constable, M.D., & Knoblich, G. (2020). Sticking together? Re-binding previous other-associated stimuli interferes with self-verification but not partner-verification. *Acta psychologica, 210*, 103167. <https://doi.org/10.1016/j.actpsy.2020.103167>

Constable, M.D., et al. (2019). It is not in the details: Self-related shapes are rapidly classified but their features are not better remembered. *Memory & Cognition, 47*, 1145-1157. <https://doi.org/10.3758/s13421-019-00924-6>

Conway, M.A., & Dewhurst, S.A. (1995). The self and recollective experience. *Applied Cognitive Psychology, 9*(1), 1-19. <https://doi.org/10.1002/acp.2350090102>

Craik, F.I.M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*(3), 268-294. <https://doi.org/10.1037/0096-3445.104.3.268>

Cunningham, S.J., & Turk, D.J. (2017, Jun). Editorial: A review of self-processing biases in cognition. *Quarterly journal of experimental psychology, 70*(6), 987-995. <https://doi.org/10.1080/17470218.2016.1276609>

Cunningham, S.J., et al. (2008). Yours or mine? Ownership and memory. *Consciousness and cognition, 17*(1), 312-318. <https://doi.org/10.1016/j.concog.2007.04.003>

Desebrock, C., et al. (2018). Self-reference in action: Arm-movement responses are enhanced in perceptual matching. *Acta psychologica, 190*, 258-266. <https://doi.org/10.1016/j.actpsy.2018.08.009>

Enock, F., et al. (2018). Self and team prioritisation effects in perceptual matching: Evidence for a shared representation. *Acta psychologica, 182*, 107-118. <https://doi.org/10.1016/j.actpsy.2017.11.011>

Enock, F.E., et al. (2020). Overlap in processing advantages for minimal ingroups and the self. *Scientific Reports, 10*(1), 18933. <https://doi.org/10.1038/s41598-020-76001-9>

Feng, C., et al. (2018). Neural representations of the multidimensional self in the cortical midline structures. *NeuroImage, 183*, 291-299. <https://doi.org/10.1016/j.neuroimage.2018.08.018>

Feng, C., et al. (2020). Effect of intranasal oxytocin administration on self-other distinction: Modulations by psychological distance and gender. *Psychoneuroendocrinology, 120*, 104804. <https://doi.org/10.1016/j.psyneuen.2020.104804>

Fischer, J., & Whitney, D. (2014). Serial dependence in visual perception. *Nature neuroscience, 17*(5), 738-743.

Fisher, R.A. (1992). Statistical methods for research workers. *Springer New York*. <https://doi.org/10.1007/978-1-4612-4380-9_6>

Gillespie‐Smith, K., et al. (2018). The I in autism: Severity and social functioning in autism are related to self‐processing. *British journal of developmental psychology, 36*(1), 127-141. <https://doi.org/10.1111/bjdp.12219>

Golubickis, M., et al. (2020). Parts of me: Identity-relevance moderates self-prioritization. *Consciousness and cognition, 77*, 102848. <https://doi.org/10.1016/j.concog.2019.102848>

Golubickis, M., et al. (2017). Self-prioritization and perceptual matching: The effects of temporal construal. *Mem Cognit, 45*(7), 1223-1239. <https://doi.org/10.3758/s13421-017-0722-3>

Golubickis, M., & Macrae, C.N. (2021). Judging me and you: Task design modulates self-prioritization. *Acta psychologica, 218*, 103350. <https://doi.org/10.1016/j.actpsy.2021.103350>

Green, S.B., et al. (2016). Use of internal consistency coefficients for estimating reliability of experimental task scores. *Psychonomic Bulletin & Review, 23*, 750-763.

Hedge, C., et al. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*, 1166-1186.

Hu, C.-P., et al. (2020). Good Me Bad Me: Prioritization of the Good-Self During Perceptual Decision-Making. *Collabra. Psychology, 6*(1), 20. <https://doi.org/10.1525/collabra.301>

Hu, C.-P., et al. (2023, 2023-05-06). Data for Training Effect of Self Prioritization[DS/OL]. V1. *Science Data Bank*. <https://doi.org/10.57760/sciencedb.08117>.

Hughes, S.M., & Harrison, M.A. (2013). I like my voice better: Self-enhancement bias in perceptions of voice attractiveness. *Perception, 42*(9), 941-949. <https://doi.org/10.1068/p7526>

Huitema, B.E. (1986). Autocorrelation in behavioral research: Wherefore art thou? In *Research methods in applied behavior analysis: Issues and advances* (pp. 187-208). Springer.

Humphreys, G.W., & Sui, J. (2015). The salient self: Social saliency effects based on self-bias. *Journal of cognitive psychology, 27*(2), 129-140. <https://doi.org/10.1080/20445911.2014.996156>

Ivaz, L., et al. (2016). The emotional impact of being myself: Emotions and foreign-language processing. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 42*(3), 489. <https://doi.org/10.1037/xlm0000179>

Jiang, M., et al. (2019). Cultural Orientation of Self-Bias in Perceptual Matching. *Front Psychol, 10*, 1469. <https://doi.org/10.3389/fpsyg.2019.01469>

John-Saaltink, E.S., et al. (2016). Serial dependence in perceptual decisions is reflected in activity patterns in primary visual cortex. *Journal of Neuroscience, 36*(23), 6186-6192.

Johnson, D.J., et al. (2017). Advancing Research on Cognitive Processes in Social and Personality Psychology:A Hierarchical Drift Diffusion Model Primer. *Social Psychological and Personality Science, 8*(4), 413-423. <https://doi.org/10.1177/1948550617703174>

Kahveci, S., et al. (2022). Reliability of reaction time tasks: how should it be computed? <https://doi.org/10.31234/osf.io/ta59r>

Keenan, J.P., et al. (2000). Self-recognition and the right prefrontal cortex. *Trends in cognitive sciences, 4*(9), 338-344. <https://doi.org/10.1016/S1364-6613> (00)01521-7

Kircher, T.T., et al. (2000). Towards a functional neuroanatomy of self processing: effects of faces and words. *Cognitive Brain Research, 10*(1-2), 133-144. <https://doi.org/10.1016/S0926-6410(00)00036-7>

Kolvoort, I.R., et al. (2020). Temporal integration as “common currency” of brain and self‐scale‐free activity in resting‐state EEG correlates with temporal delay effects on self‐relatedness. *Human brain mapping, 41*(15), 4355-4374. <https://doi.org/10.1002/hbm.25129>

Koo, T.K., & Li, M.Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of chiropractic medicine, 15*(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>

Kucina, T., et al. (2023). Calibration of cognitive tests to address the reliability paradox for decision-conflict tasks. *Nature Communications, 14*(1), 2234.

Kupper, L.L., & Hafner, K.b. (1989). On Assessing Interrater Agreement for Multiple Attribute Responses. *Biometrics, 45*(3), 957-967. <https://doi.org/10.2307/2531695>

Liljequist, D., et al. (2019). Intraclass correlation–A discussion and demonstration of basic features. *PloS one, 14*(7), e0219854.

Lin, H., et al. (2020). Strong effort manipulations reduce response caution: A preregistered reinvention of the ego-depletion paradigm. *Psychological science, 31*(5), 531-547. <https://doi.org/10.1177/0956797620904990>

Liu, T., et al. (2023). To see or not to see: The parallel processing of self-relevance and facial expressions. *Manuscript submitted for publication.* .

Liu, Y.S., et al. (2022). Depression screening using a non-verbal self-association task: A machine-learning based pilot study. *Journal of Affective Disorders, 310*, 87-95. <https://doi.org/10.1016/j.jad.2022.04.122>

Logie, R.H., et al. (1996). Group aggregates and individual reliability: The case of verbal short-term memory. *Memory & Cognition, 24*, 305-321.

Maire, H., et al. (2020). A Developmental Study of the Self‐Prioritization Effect in Children Between 6 and 10 Years of Age. *Child development, 91*(3), 694-704. <https://doi.org/10.1111/cdev.13352>

Makel, M.C., et al. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science, 7*(6), 537-542. <https://doi.org/10.1177/1745691612460688>

Martínez-Pérez, V., et al. (2020). Examining the dorsolateral and ventromedial prefrontal cortex involvement in the self-attention network: A randomized, sham-controlled, parallel group, double-blind, and multichannel HD-tDCS study. *Frontiers in Neuroscience, 14*, 683. <https://doi.org/10.3389/fnins.2020.00683>

Mei, N., et al. (2023). Using serial dependence to predict confidence across observers and cognitive domains. *Psychonomic Bulletin & Review*, 1-13.

Mollon, J.D., et al. (2017). Individual differences in visual science: What can be learned and what is good experimental practice? *Vision research, 141*, 4-15.

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Quarterly journal of experimental psychology, 11*(1), 56-60. <https://doi.org/10.1080/17470215908416289>

Navon, M., & Makovski, T. (2021). Are Self-related Items Unique? the Self-prioritization Effect Revisited. <https://doi.org/10.31234/osf.io/9dzm4>

Nijhof, A.D., & Bird, G. (2019). Self‐processing in individuals with autism spectrum disorder. *Autism research, 12*(11), 1580-1584. <https://doi.org/10.1002/aur.2200>

Oswald, F.L., et al. (2015). The development of a short domain-general measure of working memory capacity. *Behavior Research Methods, 47*, 1343-1355.

Parsons, S., et al. (2019). Psychological science needs a standard practice of reporting the reliability of cognitive-behavioral measurements. *Advances in methods and practices in psychological science, 2*(4), 378-395.

Pascucci, D., et al. (2023). Serial dependence in visual perception: A review. *Journal of Vision, 23*(1), 9-9.

Payne, B., et al. (2021). Perceptual prioritization of self‐associated voices. *British Journal of Psychology, 112*(3), 585-610. <https://doi.org/10.1111/bjop.12479>

Pronk, T., et al. (2022). Methods to split cognitive task data for estimating split-half reliability: A comprehensive review and systematic assessment. *Psychonomic Bulletin & Review, 29*(1), 44-54. <https://doi.org/10.3758/s13423-021-01948-3>

Qian, H., et al. (2020). Prioritised self-referential processing is modulated by emotional arousal. *Quarterly journal of experimental psychology, 73*(5), 688-697. <https://doi.org/10.1177/1747021819892158>

R Core Team. (2023). R: A Language and Environment for Statistical Computing. <https://www.R-project.org/>

Rad, M.S., et al. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences, 115*(45), 11401-11405.

Revelle, W.R. (2017). psych: Procedures for personality and psychological research. <https://CRAN.R-project.org/package=psych>

Rogers, T.B., et al. (1977, Sep). Self-reference and the encoding of personal information. *J Pers Soc Psychol, 35*(9), 677-688. <https://doi.org/10.1037//0022-3514.35.9.677>

Schäfer, S., & Frings, C. (2019). Understanding self-prioritisation: The prioritisation of self-relevant stimuli and its relation to the individual self-esteem. *Journal of cognitive psychology, 31*(8), 813-824. <https://doi.org/10.1080/20445911.2019.1686393>

Shapiro, K.L., et al. (1997). Personal names and the attentional blink: a visual "cocktail party" effect. *J Exp Psychol Hum Percept Perform, 23*(2), 504-514. <https://doi.org/10.1037//0096-1523.23.2.504>

Siegelman, N., et al. (2017). Measuring individual differences in statistical learning: Current pitfalls and possible solutions. *Behavior Research Methods, 49*, 418-432.

Stoeber, J., & Eysenck, M.W. (2008). Perfectionism and efficiency: Accuracy, response bias, and invested time in proof-reading performance. *Journal of research in personality, 42*(6), 1673-1678. <https://doi.org/10.1016/j.jrp.2008.08.001>

Strachan, J.W., et al. (2020). It goes with the territory: Ownership across spatial boundaries. *Journal of Experimental Psychology: Human Perception and Performance, 46*(8), 789. <https://doi.org/10.1037/xhp0000742>

Sui, J., et al. (2012). Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *Journal of experimental psychology. Human perception and performance, 38*(5), 1105-1117. <https://doi.org/10.1037/a0029792>

Sui, J., & Humphreys, G.W. (2013, Nov). Self-referential processing is distinct from semantic elaboration: Evidence from long-term memory effects in a patient with amnesia and semantic impairments. *Neuropsychologia, 51*(13), 2663-2673. <https://doi.org/10.1016/j.neuropsychologia.2013.07.025>

Sui, J., & Humphreys, G.W. (2017). The self survives extinction: Self-association biases attention in patients with visual extinction. *Cortex, 95*, 248-256. <https://doi.org/10.1016/j.cortex.2017.08.006>

Svensson, S.L., et al. (2022). More or less of me and you: self-relevance augments the effects of item probability on stimulus prioritization. *Psychological Research, 86*(4), 1145-1164. <https://doi.org/10.1007/s00426-021-01562-x>

Symons, C.S., & Johnson, B.T. (1997, May). The self-reference effect in memory: a meta-analysis. *Psychological Bulletin, 121*(3), 371-394. <https://doi.org/10.1037/0033-2909.121.3.371>

Turk, D.J., et al. (2002). Mike or me? Self-recognition in a split-brain patient. *Nature neuroscience, 5*(9), 841-842. <https://doi.org/10.1038/nn907>

Wabersich, D., & Vandekerckhove, J. (2014). The RWiener Package: an R Package Providing Distribution Functions for the Wiener Diffusion Model. *R Journal, 6*(1).

Wagenmakers, E.-J., et al. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14*(1), 3-22. <https://doi.org/10.3758/BF03194023>

Wiecki, T.V., et al. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in neuroinformatics, 7*, 14. <https://doi.org/10.3389/fninf.2013.00014>

Woźniak, M., et al. (2018). Prioritization of arbitrary faces associated to self: An EEG study. *PloS one, 13*(1), e0190679. <https://doi.org/10.1371/journal.pone.0190679>

Xu, Y., et al. (2021). Romantic feedbacks influence self-relevant processing: the moderating effects of sex difference and facial attractiveness. *Current Psychology*, 1-13. <https://doi.org/10.1007/s12144-021-02114-7>

Zhang, H., & Alais, D. (2020). Individual difference in serial dependence results from opposite influences of perceptual choices and motor responses. *Journal of Vision, 20*(8), 2-2.

Zhou, A., et al. (2019). Self-referential processing can modulate visual spatial attention deficits in children with dyslexia. *Frontiers in Psychology, 10*, 2270. <https://doi.org/10.3389/fpsyg.2019.02270>

Zorowitz, S., & Niv, Y. (2023). Improving the reliability of cognitive task measures: A narrative review. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. <https://doi.org/10.1016/j.bpsc.2023.02.004>