**Reliability of Self-Prioritization Effect as Measured by the Self-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)

# **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 xxx datasets using the intraclass correlation coefficient (ICC) and split-half reliability. Our results reveal that response time (RT) exhibits high and consistent test-retest reliability across datasets, while accuracy-based measurements yield variable outcomes. These findings establish a benchmark for future investigations utilizing the SPMT and underscore the limitations of accuracy-based measures, which should be considered when employing the SPMT as an assessment tool.

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

# **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). Consequently, two important questions remain unanswered: (1) Do these indices reliably capture the self-prioritization effect across time points? and (2) If so, which index is most suitable for repeated measurements? Addressing these questions is crucial for establishing the reliability and validity of SPMT measurements, allowing for accurate assessment of the self-prioritization effect and its implications in various domains.

To address the existing research gap, the present study aimed to investigate the reliability of self-prioritization effect (SPE) indices in the self-perceptual matching task (SPMT). In order to comprehensively assess the SPE indices derived from SPMT, we examined six indices as mentioned earlier, that capture the disparity between self-related and other-related stimuli of the matching trials. This was achieved by reanalyzing data obtained from multiple sources. 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 aim to provide valuable insights into the reliability and consistency of SPMT and its indices, having the potential to facilitate the future utilization of SPMT in research, clinical settings, and personal performance monitoring.

# **Methods**

## Ethics approval

Since this research involves a secondary analysis of pre-existing data obtained from publicly available datasets or archived data from author’s group, which have used SPMT in recent years, informed consent and confidentiality are not applicable.

## Datasets

In order to assess the reliability of SPMT, we provided a brief overview of its experimental design, as outlined in Experiment 1 by Sui et al. (2012). The original SPMT used a 2 by 3 within-subject design. The first independent variable, "Matching", has two levels: "Matching" and "Nonmatching" and indicates whether the shape and label match. The second independent variable, "Identity", has three levels: "Self", "Friend", and "Stranger" and represents the identity that the shape corresponds to. The original SPMT consisted of two phases (see figure 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.



**Figure 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 the current study, we examined a total of 15 datasets containing raw data from empirical studies utilizing the SPMT. The selection of these datasets was based on two criteria: (1) the experimental design did not deviate from the original SPMT (Sui et al., 2012); (2) the trial-level data is available so that we can estimate at least one reliability index. All these studies shared raw data publicly (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; 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 accessible to 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 pre-reregistration for details). Nonetheless, not all studies incorporated repeated measures. If a publicly available dataset did not include repeated SPMT measurements within a specified time interval, we excluded it from calculating the Intraclass Correlation Coefficient (ICC) and only considered split-half reliability. The details of the datasets used are described below.

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. ICC: Intraclass Correlation Coefficient, SHR: split-half reliability

## Analysis

We used the Python toolkit HDDM of Bayesian Hierarchical Model (Wiecki et al., 2013) to fit the behavioral data into the drift diffusion model (DDM). All the other analyses are performed using the statistical software R (R Development Core Team, 2010). In total, we gathered 15 publicly available datasets, including one from our laboratory and one from our collaborators (Liu et al., 2023), as mentioned earlier and presented in Table 1. The research flow of the current study is visually represented in Figure 3.



**Figure 3.** 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.

### Data Pre-processing

We pre-processed the secondary data using the following criteria:

1. Participant exclusion criteria
2. Participants who had wrong trial numbers because of procedure errors will be excluded from the analysis.
3. Participants with an overall accuracy < 0.5 will be excluded from the analysis.
4. Participants with any of the conditions with zero accuracy will be excluded from the analysis.
5. Behavioural data exclusion criteria
6. Trials with no response or wrong key press will be excluded from the analysis.
7. Responses with reaction times less than 200ms or greater than 1500ms will be excluded from the analysis.
8. The practice trials will be excluded from the formal analysis.
9. The data under conditions other than the “control condition” will not be used in the current study.

### Calculation of SPE

For each dataset, we calculated six indices for each experimental condition: reaction times, accuracy, *d* prime, efficiency, drift rate (*v*), and starting point (*z*). Reaction time and accuracy can be obtained directly from the datasets, while *d* prime and efficiency will be calculated based on reaction time and accuracy using a simple formula (see Table 2). The drift rate (*v*) and starting point (*z*) of the drift-diffusion model is estimated by the "fit\_ezddm" function in the "hausekeep" package (Lin et al., 2020), which wrapped the function from EZ-DDM (Wagenmakers et al., 2007).

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.

### Estimating the Reliability

**Split-half reliability.** We calculated the split-half reliability of the six indices using four types of split-half reliability measures: odd-even, front-back, permutation, and Monte Carlo (Kahveci et al., 2022; Pronk et al., 2022). The odd-even split divides trials into odd and even numbered sequences, while the front-back split divides the first and second halves of trials. The permutation split shuffles the trial order and randomly assigns each half to a group. The Monte Carlo split-half is similar to the permutation split-half, but it repeats the process thousands of times to calculate the average and 95% confidence interval of the split-half reliability. This study will primarily use Monte Carlo split-half to determine the split-half reliability of SPMT for its robustness (Pronk et al., 2022). The results of the other three split-half methods will be presented in the supplementary materials.

First, the data is stratified according to Session (if applicable), Matching, and Identity. 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. 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 1000 times. This will result in 1000 pairs of two halves of the data. Next, we use these 1000 pairs of data to calculate 1000 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 multiple experiment 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 takes into account 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; Liljequist et al., 2019). For the calculation of ICC2 estimates, the formula is:

,

where is the mean square between subjects, is the mean square error, 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:

.

A value less than 0.6 means poor reliability, a value between 0.6 and 0.8 indicates substantial reliability, and a value greater than 0.8 means excellent reliability (Cicchetti & Sparrow, 1981; Kupper & Hafner, 1989).

# **Results**

## Split-Half Reliability

First, we stratified the data based on three variables: Session (if applicable), Matching, and Identity, and then split the stratified data into two halves using four methods. Next, we calculated the SPE for each of the six indices for each half of the data. Finally, we calculated the split-half reliability for each of the six SPEs. As shown in Figure 4, when using the Monte Carlo split-half, the split-half reliability of the six indices obtained is very high, with the highest value of Efficiency, which means that it is the most stable of the six SPE indexing calculations for split-half reliability. The results from the other three split-half methods were like the Monte Carlo method, which is presented in the supplementary material.

**Figure 4. Split-Half Reliability.**

*Note:* the results here are from simulation, this figure will be replaced with plots from real data.

## Intraclass correlation coefficient (ICC)

Intraclass correlation coefficient (ICC) is a measure of the consistency or reliability of measurements made by different raters (observers) or repeated measurements made by the same rater (observer). We calculated ICC only if the study involves repeated measurements of SPMT. Essentially, it tells us how much of the variation in the data can be attributed to differences between raters or repeated measurements, and how much can be attributed to differences within the subjects being measured. In simple terms, it provides an idea of the proportion of total variation in the data that is due to the true differences between subjects, versus due to measurement error or random fluctuations.

The present study aimed to investigate the stability of six indices, including reaction time (RT), accuracy (ACC), *d* prime, Efficiency, drift rate (*v*) and starting point (*z*) in the diffusion decision model (DDM), across time sessions. We utilized the Intraclass Correlation Coefficients (ICC) to evaluate the proportion of variation in SPMT that could be attributed to within-subject repeatability over time and between-subject differences. Specifically, we are most interested in ICC2 and ICC2k, where ICC2 represents the ratio of between-subject variance to total variance, and ICC2k represents the ratio of within-subject variance to total variance. Thus, if SPMT is reliable for measuring individual differences, then the ICC2 is large and the ICC2k is small. As shown in Figure 5, the ICC2 values of the six indices are relatively large and ICC2k values are relatively small, supporting our hypothesis.

**Figure 5. Intraclass correlation coefficient****.**

*Note:* this plot is from simulated data, will be replaced with real data in our formal analyses

# **Discussion**

Assessing the reliability of a behavioral paradigm is crucial when researchers intend to utilize the task for exploring individual differences, although it is not yet a common practice (Hedge et al., 2018). 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) by re-analyzing existing data from 15 datasets. We employ the intraclass correlation coefficient (ICC) and split-half reliability for this purpose. Our analysis includes both our own data and additional data, which collectively demonstrate that response time (RT) consistently exhibits high test-retest reliability across datasets (0.91). However, measurements based on accuracy yield varying outcomes, with some as low as 0 (). The ICC results suggest that, overall, the indices related to the self-prioritization effect (SPE) in the self-perceptual matching task (SPMT) are more suitable for group-level analysis rather than assessing individual-level variation.

We observed that certain indices we tested demonstrated good to excellent group-level test-retest reliability (ICC2k). Specifically, the RT index exhibited an ICC of 0.77, while the Efficiency measures of the SPMT task showed an ICC of 0.74. However, individual-level test-retest reliability results indicated poor performance across all indices, with ICC2 values ranging from 0.2 to 0.4. The discrepancy between the high ICC2k and low ICC2 suggests that the task is more influenced by between-participant variability than within-participant variability (Hedge et al., 2018; Liljequist et al., 2019). It is common for behavioral paradigm to have such result pattern, as demonstrated in previous research (Clark et al., 2022). There are various reasons for this pattern. Behavioral paradigms are susceptible to factors such as momentary fluctuations, external conditions, and contextual differences, 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, these 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. It 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.

For Split-half reliability, the result show that almost all the indices do not perform well. Several plausible reasons could be identified to shed light on these findings. First, the task's inherent complexity, necessitating intricate cognitive processes and challenging actions, may have introduced considerable variability in performance.

The presence of serial dependence in a paradigm can potentially influence split-half reliability. Serial dependence refers to the phenomenon where the outcome of one trial is influenced by the preceding trials, resulting in a systematic relationship between consecutive trials.

In the context of a split-half design, serial dependence can impact reliability estimates if the serial correlation is not adequately accounted for. If the split-half halves are not balanced in terms of trial order or if the temporal structure of the task is not taken into consideration, the systematic dependencies between trials may introduce inconsistencies across the two halves.

Specifically, if the split-half design inadvertently separates temporally adjacent trials, the serial dependence may lead to differences in performance between the halves, reducing the reliability estimate. On the other hand, if the split-half design appropriately considers the temporal structure by ensuring that temporally adjacent trials are kept within the same half, the impact of serial dependence on reliability may be mitigated.

To accurately assess split-half reliability in the presence of serial dependence, researchers should employ appropriate statistical methods that account for the temporal dependencies between trials. Time series analysis techniques or modeling approaches that capture the serial correlation can be utilized to obtain more accurate reliability estimates.

In conclusion, the presence of serial dependence in a paradigm can potentially influence split-half reliability estimates. Careful consideration of the temporal structure and appropriate statistical modeling are crucial to account for these dependencies and obtain reliable estimates of task performance.

Moreover, inadequate participant training or familiarization with the task could have compromised their ability to perform consistently. The task's sensitivity to capture stable individual differences or changes over time might also have played a role. Lastly, the timing of the split-half assessment interval, whether too short or too long, might have influenced participants' performance in unpredictable ways. These potential factors warrant careful consideration when interpreting the reliability of behavioral tasks and provide valuable insights for future task design and data collection protocols to enhance reliability.

The total number of trials to achieve reliabilities (*r*) of 0.8 (lower line, light gray band) and 0.9 (upper line, dark gray band). Gray bands represent the posterior median 95% credible intervals. For every point on the x-axis we are predicting the number of trials required to achieve a reliability of 0.8 (light gray region) or 0.9 (dark gray region); the dotted vertical line descending from the dashed identity line (i.e., a line where x = y), and the values in text, provide the interpolated number of trials required for reliabilities of 0.8 and 0.9. Each panel represents a single task from Hedge et al.: (**a**) Simon, (**b**) Stroop, (**c**) Flanker.

To summarise, the results of our study provide valuable information on SALT for further studies, for example, laying the ground for the prospective uses of SALT in research, clinical usage, and personal performance monitoring.

If accuracy performance is near ceiling for the easier measures, it is unlikely that there is substantial intraparticipant variability to allow for the space to see consistent and reliable differences in performance between individuals. A similar pattern of results is shown by Dai and colleagues (2019) who observed a rising trend in the test-retest reliability coefficients as the memory set size increased: Pearson's rs of 0.50, 0.57, 0.65, and 0.76 were found for set sizes three, four, five, and six, respectively. Soveri, Lehtonen, Karlsson, Lukasik, Antfolk, and Laine (2018), who investigated the test-retest reliability of frequently used executive tasks in healthy adults, also demonstrate a similar trend in results. Among a battery of executive tasks was a visuoverbal N-back working memory task; in this task, numbers one to nine were presented, and participants indicated whether this number matched the number either in the previous trial (1-back condition) or the number three trials back (3-back condition). As difficulty increased (i.e. the load factor increased), RT to respond also increased, as expected. The ICC values for the RT measure increased with increasing difficulty with ICC values of 0.48 and 0.73 for the one-back and three-back conditions, respectively. These findings, together with our results, suggest that task measures which are objectively more difficult may be more reliable. Using a more difficult task measure can help optimize between-participant variation, a core component allowing appropriate exploration of test-retest reliability.

The importance of considering the particular measures and parameters used is also apparent when interpreting our reliability results for our MOT task. We compared performance on the maximum items an individual could track as well as the threshold for the number of items retained in VWM, and the ICCs were quite low (0.41 and 0.36, respectively). At first glance, these results may suggest MOT is a particularly unreliable task and not well suited for the study of individual differences; however, there is an almost infinite range of parameters that can be employed when testing MOT performance (Meyerhoff, Papenmeier, & Huff, 2017; Scholl, 2009). For example, the calculations to assess performance on the task can look quite different, depending on whether the probe-one or mark-all method is used (Hulleman, 2005). Our version of the MOT task used the mark-all method (i.e. participants were asked to correctly identify all target items), but even within the mark-all method, the task itself can vary wildly according to the speed and the trajectory of the items to be tracked. Additionally, the staircase we used varied the number of items presented, which may have limited the variability in terms of what we could explore. Alternatively, the speed of the items can be titrated rather than the number to provide a finer-grained threshold estimation (e.g. Bowers, Anastasio, Sheldon, O'Connor, Hollis, Howe, & Horowitz, 2013; Meyerhoff, Papenmeier, Jahn, & Huff, 2016). Such a measure may yield significantly more intraparticipant variability and thus may be more suitable for evaluating individual differences (e.g. Meyerhoff & Papenmeier, 2020).

However, it must be noted that other work has revealed striking differences between reliability scores for the same task; for instance, van Leeuwen, van den Berg, Hoekstra, and Boomsma (2007) found rather low test-retest reliability for the error cost on the Eriksen flanker task (r = 0.29 and r = 0.14) whereas others have found higher reliability for the same (ICC = 0.65, Wöstmann et al., 2013; ICC = 0.72, Hedge et al., 2018b). These discrepancies could be partly explained by differences in participants’ performance in these studies: for example, participants in van Leeuwen et al.’s (2007) study reached ceiling for both congruent and incongruent conditions, resulting in a very small error cost rate (M = 0.08% and 0.02%) and little between-participant variance. In comparison, participants in Hedge et al.’s (2018b) study showed lower accuracy scores in the incongruent condition relative to the congruent condition, resulting in a much higher error cost rate (M = 8.95%) and more variation between participants.

Standard, well-established cognitive tasks that produce reliable effects in group comparisons also lead to unreliable measurement when assessing individual differences. This reliability paradox has been demonstrated in decision-conflict tasks such as the Simon, Flanker, and Stroop tasks, which measure various aspects of cognitive control.

Over five experiments, we show that a Flanker task and a combined Simon and Stroop task with the additional manipulation produced reliable estimates of individual differences in under 100 trials per task, which improves on the reliability seen in benchmark Flanker, Simon, and Stroop data.

Our work joins a growing effort to improve the reliability of tests of cognitive control. Rather than abandoning conflict-task RT difference measures[28](https://www.nature.com/articles/s41467-023-37777-2" \l "ref-CR28" \o "Draheim, C., Tsukahara, J. S., Martin, J. D., Mashburn, C. A. & Engle, R. W. A toolbox approach to improving the measurement of attention control. J. Exp. Psychol. Gen. 150, 242–275 (2021).), we attempted to improve their reliability.

For Split-half reliability, the result show that

The total number of trials to achieve reliabilities (*r*) of 0.8 (lower line, light gray band) and 0.9 (upper line, dark gray band). Gray bands represent the posterior median 95% credible intervals. For every point on the x-axis we are predicting the number of trials required to achieve a reliability of 0.8 (light gray region) or 0.9 (dark gray region); the dotted vertical line descending from the dashed identity line (i.e., a line where x = y), and the values in text, provide the interpolated number of trials required for reliabilities of 0.8 and 0.9. Each panel represents a single task from Hedge et al.: (**a**) Simon, (**b**) Stroop, (**c**) Flanker.

To summarise, the results of our study provide valuable information on SALT for further studies, for example, laying the ground for the prospective uses of SALT in research, clinical usage, and personal performance monitoring.

# **Acknowledgements**

The present research is support by xxx.

# **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 performed the data pre-processing, analysis and visualize the results. In addition, ZL, JS, HMZ and HCP contributed to the discussion of 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 (Dataset 0) 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**

Table S1 Split-Half Reliability of Other Split Methods based on simulated data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Indices of SPE | Method of SH | SHR |  | Indices of SPE | Method of SH | SHR |
| RT | First-Second | .01 |  | Efficiency | First-Second | .07 |
| RT | Odd-Even | -.05 |  | Efficiency | Odd-Even | -.04 |
| RT | Permuted | .01 |  | Efficiency | Permuted | .05 |
| ACC | First-Second | .02 |  | DDM: v | First-Second | .04 |
| ACC | Odd-Even | -.05 |  | DDM: v | Odd-Even | -.05 |
| ACC | Permuted | .07 |  | DDM: v | Permuted | .10 |
| Dprime | First-Second | .01 |  | DDM: z | First-Second | .07 |
| Dprime | Odd-Even | -.08 |  | DDM: z | Odd-Even | .01 |
| Dprime | Permuted | -.02 |  | DDM: z | Permuted | .13 |

Note. Target: Friend, SH: split-half, SHR: split-half reliability, SPE: self-prioritization effect

Table S2 Split-Half Reliability of Other Split Methods based on simulated data

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Indices of SPE | Method of SH | SHR |  | Indices of SPE | Method of SH | SHR |
| RT | First-Second | .01 |  | Efficiency | First-Second | .07 |
| RT | Odd-Even | -.01 |  | Efficiency | Odd-Even | -.03 |
| RT | Permuted | .08 |  | Efficiency | Permuted | -.01 |
| ACC | First-Second | .02 |  | DDM: v | First-Second | .03 |
| ACC | Odd-Even | -.03 |  | DDM: v | Odd-Even | -.01 |
| ACC | Permuted | -.10 |  | DDM: v | Permuted | -.11 |
| Dprime | First-Second | .03 |  | DDM: z | First-Second | .01 |
| Dprime | Odd-Even | -.07 |  | DDM: z | Odd-Even | .03 |
| Dprime | Permuted | -.07 |  | DDM: z | Permuted | -.14 |

Note. Target: Friend, SH: split-half, SHR: split-half reliability, SPE: self-prioritization effect

# **References**

Bukowski, H., Todorova, B., Boch, M., Silani, G., & Lamm, C. (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., Birch-Hurst, K., Pennington, C. R., Petrie, A. C. P., Lee, J. T., & Hedge, C. (2022). Test-retest reliability for common tasks in vision science. *Journal of vision (Charlottesville, Va.), 22*(8), 18-18. <https://doi.org/10.1167/jov.22.8.18>

Constable, M. D., Becker, M. L., Oh, Y.-I., & Knoblich, G. (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., Elekes, F., Sebanz, N., & Knoblich, G. (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., Rajsic, J., Welsh, T. N., & Pratt, J. (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., Turk, D. J., Macdonald, L. M., & Macrae, C. N. (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., Sui, J., & Spence, C. (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., Sui, J., Hewstone, M., & Humphreys, G. W. (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., Hewstone, M. R., Lockwood, P. L., & Sui, J. (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., Yan, X., Huang, W., Han, S., & Ma, Y. (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., Zhou, X., Zhu, X., Zhu, R., Han, S., & Luo, Y.-J. (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>

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., Ballantyne, C., Branigan, H. P., Turk, D. J., & Cunningham, S. J. (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., Falbén, J. K., Ho, N. S., Sui, J., Cunningham, W. A., & Macrae, C. N. (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., Falben, J. K., Sahraie, A., Visokomogilski, A., Cunningham, W. A., Sui, J., & Macrae, C. N. (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>

Hedge, C., Powell, G., & Sumner, P. (2018, 2018/06/01). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods, 50*(3), 1166-1186. <https://doi.org/10.3758/s13428-017-0935-1>

Hu, C.-P., Lan, Y., Macrae, C. N., & Sui, J. (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., Peng, K., & Sui, J. (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>

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., Costa, A., & Duñabeitia, J. A. (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., Wong, S. K. M., Chung, H. K. S., Sun, Y., Hsiao, J. H., Sui, J., & Humphreys, G. W. (2019). Cultural Orientation of Self-Bias in Perceptual Matching. *Front Psychol, 10*, 1469. <https://doi.org/10.3389/fpsyg.2019.01469>

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

Keenan, J. P., Wheeler, M. A., Gallup, G. G., & Pascual-Leone, A. (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., Senior, C., Phillips, M. L., Benson, P. J., Bullmore, E. T., Brammer, M., ..., & David, A. S. (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., Wainio‐Theberge, S., Wolff, A., & Northoff, G. (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>

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., Elfving, B., & Skavberg Roaldsen, K. (2019). Intraclass correlation - A discussion and demonstration of basic features. *PloS one, 14*(7), e0219854. <https://doi.org/10.1371/journal.pone.0219854>

Lin, H., Saunders, B., Friese, M., Evans, N. J., & Inzlicht, M. (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., Sui, J., & Hildebrandt, A. (2023). To see or not to see: The parallel processing of self-relevance and facial expressions. *Manuscript submitted for publication.* .

Liu, Y. S., Song, Y., Lee, N. A., Bennett, D. M., Button, K. S., Greenshaw, A., Cao, B., & Sui, J. (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>

Maire, H., Brochard, R., & Zagar, D. (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., Plucker, J. A., & Hegarty, B. (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., Campoy, G., Palmero, L. B., & Fuentes, L. J. (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>

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>

Parsons, S., Kruijt, A.-W., & Fox, E. (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. <https://doi.org/10.1177/2515245919879695>

Payne, B., Lavan, N., Knight, S., & McGettigan, C. (2021). Perceptual prioritization of self‐associated voices. *British Journal of Psychology, 112*(3), 585-610. <https://doi.org/10.1111/bjop.12479>

Pronk, T., Molenaar, D., Wiers, R. W., & Murre, J. (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., Wang, Z., Li, C., & Gao, X. (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 Development Core Team. (2010). *R: A language and enviornment for statistical computing.* In R Foundation for Statisticial Computing

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

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (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., Caldwell, J., & Sorensen, R. E. (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>

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., Constable, M. D., & Knoblich, G. (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., He, X., & Humphreys, G. W. (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., Golubickis, M., Maclean, H., Falbén, J. K., Persson, L. M., Tsamadi, D., Caughey, S., Sahraie, A., & Macrae, C. N. (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., Heatherton, T. F., Kelley, W. M., Funnell, M. G., Gazzaniga, M. S., & Macrae, C. N. (2002). Mike or me? Self-recognition in a split-brain patient. *Nature neuroscience, 5*(9), 841-842. <https://doi.org/10.1038/nn907>

Wagenmakers, E.-J., Van Der Maas, H. L., & Grasman, R. P. (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., Sofer, I., & Frank, M. J. (2013). HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python. *Frontiers in neuroinformatics, 7*, 14-14. <https://doi.org/10.3389/fninf.2013.00014>

Woźniak, M., Kourtis, D., & Knoblich, G. (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., Yuan, Y., Xie, X., Tan, H., & Guan, L. (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>

Zhou, A., Duan, B., Wen, M., Wu, W., Li, M., Ma, X., & Tan, Y. (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>