



# Self-reported impulsivity does not predict response caution<sup>☆</sup>

Craig Hedge<sup>a,\*</sup>, Georgina Powell<sup>b</sup>, Aline Bompas<sup>a</sup>, Petroc Sumner<sup>b</sup>

<sup>a</sup> CUBRIC - School of Psychology, Cardiff University, UK

<sup>b</sup> School of Psychology, Cardiff University, UK



## ARTICLE INFO

### Keywords:

Response control  
Inhibition  
Impulsivity  
Response caution  
Diffusion model  
UPPS-P  
Self-control  
Boundary separation

## ABSTRACT

The broad construct of impulsivity is one that spans both personality and cognitive ability. Despite a common overarching construct, previous research has found no relationship between self-report measures of impulsivity and people's ability to inhibit pre-potent responses. Here, we use evidence accumulation models of choice reaction time tasks to extract a measure of “response caution” (boundary separation) and examine whether this correlates with self-reported impulsivity as measured by the UPPS-P questionnaire. Response caution reflects whether an individual makes decisions based on more (favouring accuracy) or less (favouring speed) evidence. We reasoned that this strategic dimension of behaviour is conceptually closer to the tendencies that self-report impulsivity measures probe than what is traditionally measured by inhibition tasks. In a meta-analysis of five datasets (total  $N = 296$ ), encompassing 19 correlations per subscale, we observe no evidence that response caution correlates with self-reported impulsivity. Average correlations between response caution and UPPS-P subscales ranged from  $\rho = -0.02$  to  $-0.04$ . While the construct of response caution has demonstrated value in understanding individual differences in cognition, brain functioning and aging; the factors underlying what has been called “impulsive information processing” appear to be distinct from the concept of impulsivity derived from self-report.

## 1. Introduction

The constructs of impulsivity and self-control play a prominent role in our current understanding of personality and neuropsychological disorders (Chambers, Garavan, & Bellgrove, 2009; Eysenck & Eysenck, 1985; Whiteside & Lynam, 2001). However, research into the definition and measurement of impulsivity continues to raise questions about whether the wide range of tasks and measures that are used do in fact measure some common underlying construct. Though different taxonomies of impulsivity exist, one fundamental distinction is that performance in behavioural tasks often shows little to no relationship with self-report measures (Creswell, Wright, Flory, Skrzynski, & Manuck, 2019; Cyders & Coskunpinar, 2011, 2012; Eisenberg et al., 2019; Sharma, Markon, & Clark, 2014). Here, we try to bridge the gap using a cognitive model to isolate a dimension of cognition characterised as “impulsive information processing” (Metin et al., 2013), “decision urgency” or “response caution” (Evans, Rae, Bushmakin, Rubin, & Brown, 2017; Hedge, Powell, Bompas, Vivian-Griffiths, & Sumner, 2018).

One proposed reason for the discrepancy between questionnaires and behavioural tasks is that they were developed with different goals in mind. Cyders and Coskunpinar (2011) suggest that self-report

measures focus on general tendencies or traits while lab-based tasks focus on “snapshots” of behaviour, which may be more sensitive to fluctuations in states (see also Wennerhold & Friese, 2020). For example, the UPPS-P impulsivity questionnaire consists of five subscales labelled negative urgency, positive urgency, (lack of) premeditation, (lack of) perseverance and sensation-seeking (Lynam, Smith, Whiteside, & Cyders, 2006). Individuals rate the extent of their agreement with statements about their general behaviour, such as “I am a cautious person”, an item in the premeditation subscale.

In contrast, behavioural impulsivity tasks are sometimes broadly categorised as either *impulsive choice* or *impulsive action* (Weafer & de Wit, 2014). Impulsive choice tasks typically consist of delayed gratification or gambling tasks, where individuals decide between uncertain or delayed large rewards and certain or immediate lower value rewards. Impulsive action typically refers to tasks designed to measure an individual's ability to rapidly inhibit a response to a salient or pre-potent stimulus (repeatedly). For example, in the Stroop task (Stroop, 1935) participants must quickly and accurately classify the font colour of a word while ignoring its meaning. Performance is traditionally measured by subtracting reaction times or error rates in a congruent or baseline condition (the word ‘red’ in red font) from an incongruent or

<sup>☆</sup> This work was supported by the ESRC (ES/K002325/1); and by the Wellcome Trust (104943/Z/14/Z).

\* Corresponding author at: School of Psychology, Cardiff University, Tower Building, Park Place, Cardiff CF10 3AT, UK.

E-mail address: [hedgec@cardiff.ac.uk](mailto:hedgec@cardiff.ac.uk) (C. Hedge).

conflict condition (the word ‘red’ in blue font). The subsequent RT cost or error cost is taken as an index of an individual’s ability to overcome conflicting information. Further subcategories of inhibition tasks have also been proposed (Friedman & Miyake, 2004; Sharma et al., 2014), though these are not consistently supported by data (Rey-Mermet, Gade, & Oberauer, 2018). Note that the subtraction of performance in a baseline condition is often done for the explicit purposes of removing individual differences in confounding factors such as strategy or caution (Donders, 1969; though recent reviews have highlighted problems with this assumptions; Draheim, Mashburn, Martin, & Engle, 2019; Hedge, Powell, Bompas, et al., 2018).

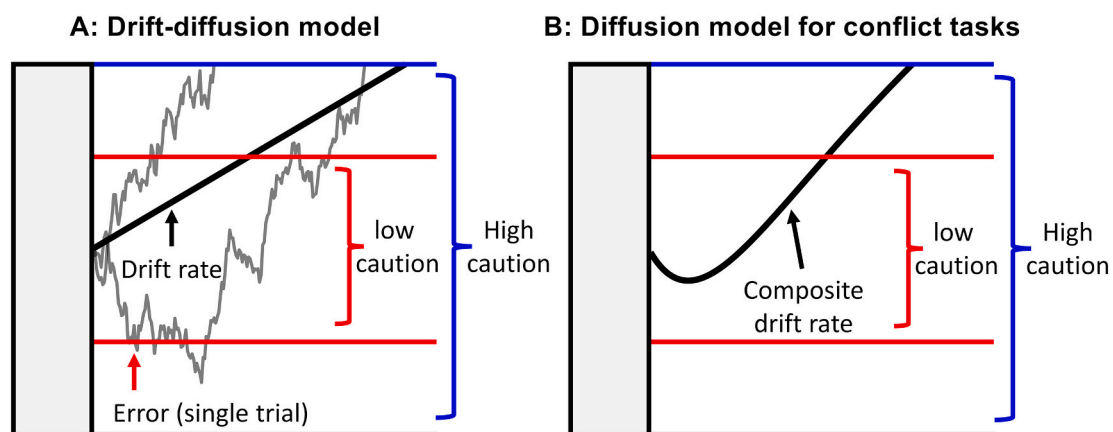
This conceptual distinction between the ability to overcome conflict and the general tendencies for cautious or impulsive behaviours may explain why low correlations are observed between inhibition tasks and self-report measures. We therefore reasoned that correlations might be uncovered if we could extract a measure of ‘caution’ from cognitive tasks. In our recent work, we have used evidence accumulation models to better understand individual differences in choice reaction time tasks, and conflict tasks in particular (Hedge, Powell, Bompas, et al., 2018; Hedge, Powell, & Sumner, 2018a; Hedge, Vivian-Griffiths, Powell, Bompas, & Sumner, 2019). The evidence accumulation framework represents a broad family of models that assume that the decision process can be represented by a process of sampling evidence from the stimulus until a boundary or threshold has been reached (see Fig. 1; for reviews see Donkin, Brown, Heathcote, & Wagenmakers, 2011; Forstmann, Ratcliff, & Wagenmakers, 2016; Ratcliff & Smith, 2004). These models can be used to dissociate and quantify the multiple processes that contribute to behaviour.

For our current purposes, we are particularly interested in the boundary separation parameter, sometimes referred to as “response caution”. In the evidence accumulation model framework response caution represents how much evidence an individual requires before they make a response. An individual with a preference for speed is assumed to set a lower boundary, such that they respond quickly but risk making more errors due to noise in the decision processes. In some studies, such individuals have been referred to as having “impulsive information processing” (Metin et al., 2013). In contrast, a cautious individual waits for more information before responding to ensure accuracy at the expense of reaction time, corresponding to a higher boundary.

The individual’s preference for speed vs. accuracy is mathematically independent of the individual’s ability to extract information and prime the desired response (while inhibiting or controlling unwanted information or response tendencies). This ability is represented by the rate that correct evidence accumulates towards the boundary, rather than the height of the boundary. In the drift-diffusion model (Ratcliff, 1978; Fig. 1A), the underlying rate of evidence accumulation (the drift rate) is assumed to be linear, though subject to moment-to-moment noise. Extensions of the diffusion model, such as the diffusion model for conflict tasks (Ulrich, Schroter, Leuthold, & Birngruber, 2015; Fig. 1B), assume that the information extracted from a stimulus can vary over time as a function of both automatic and controlled processes. This captures the assumption that prepotent response features (e.g., incongruent arrows in a flanker task) initially capture attention but that their influence decreases or is inhibited over time. But in either case, response caution (i.e., the boundary) is a separate parameter from these processes.

The quantitative dissociation of response caution from information processing efficiency (drift rate) has led to valuable insights into cognitive and behavioural changes in fields such as aging (Ratcliff, Thapar, & McKoon, 2006, 2010), as well as neuropsychological conditions such as autistic spectrum disorder and attention-deficit hyperactivity disorder (Karalunas et al., 2018; Pirrone, Dickinson, Gomez, Stafford, & Milne, 2017; Pirrone, Johnson, Stafford, & Milne, 2020; Powell et al., 2019). In most cases (though not all), older adults and individuals with autism show higher levels of response caution relative to young adults and healthy controls respectively. Interestingly, separate studies have shown higher levels of risk aversion in older adults when using self-report measures (the Domain-Specific Risk-Taking scale; Rolison, Hanoch, Wood, & Liu, 2014) and behavioural gambling tasks (Rutledge et al., 2016). Similarly, there is evidence that individuals with autism adopt more risk-averse strategies in gambling tasks (Gosling & Moutier, 2018; South et al., 2014). This co-occurrence of increased response caution in choice reaction time tasks and increased risk aversion in other domains hints at a possible underlying link related to the concept of impulsivity.

We have routinely included the UPPS-P questionnaire in our recent experiments, following completion of the behavioural tasks, with the overarching goal of examining whether using a cognitive model to dissociate caution from other cognitive mechanisms can reveal a



**Fig. 1.** Schematic of two evidence accumulation models. A. In the drift diffusion model (Ratcliff, 1978), the decision on each trial (jagged lines) is represented by the noisy accumulation of evidence to a boundary. The solid black line represents the average rate of evidence accumulation or ‘drift rate’. The upper and lower boundary represent the correct and incorrect response respectively. An individual who sets a low boundary (red lines) waits for less evidence before responding and is more likely to make an error due to noise in the accumulation process. B. In the diffusion model for conflict tasks (Ulrich et al., 2015), the average rate of evidence accumulation is a composite of both controlled processing and automatic activation. The automatic activation function captures the assumption that prepotent response features (e.g., incongruent flankers) are processed via a fast, automatic route (Ridderinkhof, 2002). The solid black line shows the underlying accumulation for an incongruent trial, where automatic activation elicited from (e.g.) flankers in the flanker task contributes to the incorrect response tendency in the early part of the decision phase. See Supplementary material A for more information. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

relationship with impulsivity that was previously hidden. Here, we present a meta-analysis of the evidence we have accumulated over these studies (for a discussion of this approach, see Goh, Hall, & Rosenthal, 2016). We focus on response caution for the theoretical reasons above, and also because we have recently observed that this parameter correlated strongly across conflict tasks (average  $\rho = 0.54$ ), whereas the parameters associated with conflict and inhibition did not ( $\rho = 0.04$ ; Hedge, Powell, Bompas, & Sumner, 2020). We reasoned that if we do not see correlations with a parameter that is both theoretically aligned to impulsivity and that correlates well across tasks, then we are unlikely to see correlations with other sub-factors of cognition (Wennerhold & Friese, 2020). We examine all of the UPPS-P subscales in our analysis, though there is arguably most conceptual overlap between response caution and the items in the *lack of premeditation* and *perseverance* subscales, which refer to planning and attentiveness. Both these dimensions and *negative urgency* have also previously shown weak ( $r = 0.1$ ) but significant correlations with traditional measures of response inhibition (Cyders & Coskunpinar, 2011), though these studies did not dissociate response caution from ability as we do here.

## 2. Method

### 2.1. Datasets

We conducted a new analysis of published data. These studies and their basic details are given in Table 1. We have previously reported the behavioural results and model parameters to examine whether the tasks themselves correlate with each other, but we have not previously reported the correlations between cognitive response caution (or the other model parameters) and the UPPS-P subscales that are the focus of our analysis here.

We adopt a meta-analytic approach to maximise the power of these data to detect a meaningful correlation between caution and the UPPS-P subscales, should one exist. To assess this, we conducted a sensitivity power analysis for a random effects meta-analysis under varying assumptions of heterogeneity in the effect sizes (Griffin, 2020; Pigott, 2012). Based on 19 effect sizes and an average sample size of 68, we have 80% power to detect correlations equal to or greater than  $\rho = 0.09, 0.11,$  and  $0.16$  assuming small, moderate, and large levels of heterogeneity respectively ( $\alpha = 0.05$ , two-tailed).

### 2.2. Behavioural tasks

Fig. 2 shows a schematic of the choice reaction time tasks used. In all tasks, participants were required to decide which of two (four in the Stroop) alternatives to categorise a stimulus. In the flanker task participants must respond to the direction of the central arrow and ignore

the flanking stimuli. In the Stroop task participants must respond to the colour of the font and ignore the written word. In the Simon task participants must respond to the colour of the circle and ignore its location. In the dot motion task participants must respond to the direction of coherent motion in an array of dots. The Simon task in Dataset 3 did not include a neutral condition. All the tasks are commonly used to measure response inhibition except the dot motion task in Dataset 5, which is commonly used in both human and animal studies of decision making (c.f. Ratcliff & McKoon, 2008). Datasets 4 and 5 included a speed-accuracy trade-off manipulation, in which participants were instructed to prioritise speed, accuracy or both equally in separate blocks. In all other datasets, participants were instructed to be both as fast and as accurate as possible.

For our current purposes, the critical similarity between these tasks is that they all allow us to measure the level of caution with which individuals approach speeded decision making. We have previously observed moderate to strong between-task correlations in the response caution parameter (Hedge et al., 2019; Hedge, Powell, et al., 2020). Therefore, we interpret response caution to be a (at least in part) general process or set of processes. Here, where we are interested in the relationship between response caution and impulsivity, we treat the correlations from each task as an estimate of this same relationship.

### 2.3. UPPS-P impulsive behaviour scale

The UPPS-P is a 59-item questionnaire that measures five components of impulsivity: negative urgency, premeditation, perseverance, sensation seeking, and positive urgency (Lynam et al., 2006; Whiteside & Lynam, 2001). Participants rate their agreement with each item on a four-point Likert-type scale, ranging from Agree strongly (1) to Disagree Strongly (4). The subscales have shown high internal consistencies ( $> 0.81$ ; Whiteside & Lynam, 2001), and good to excellent three-week test-retest reliabilities for the subscales in this population (ICCs from 0.70 to 0.91; Hedge, Powell, & Sumner, 2018b). In our test-retest reliability studies (Datasets 2 and 4), participants completed the UPPS-P in both sessions. We average the subscale values across sessions for our correlations here.

### 2.4. Model fitting

For the conflict tasks in Datasets 1–5, we fit the diffusion model for conflict tasks (Ulrich et al., 2015) using Matlab. For the dot motion task in Dataset 5, we fit the standard drift-diffusion model using the DMAT toolbox in Matlab (Vandekerckhove & Tuerlinckx, 2008). We describe both models and their parameters in detail in Supplementary material A. Both models produce a boundary separation parameter which we interpret as response caution (see Fig. 1). The “number of correlations” column in Table 1 corresponds to the number of boundary separation

**Table 1**

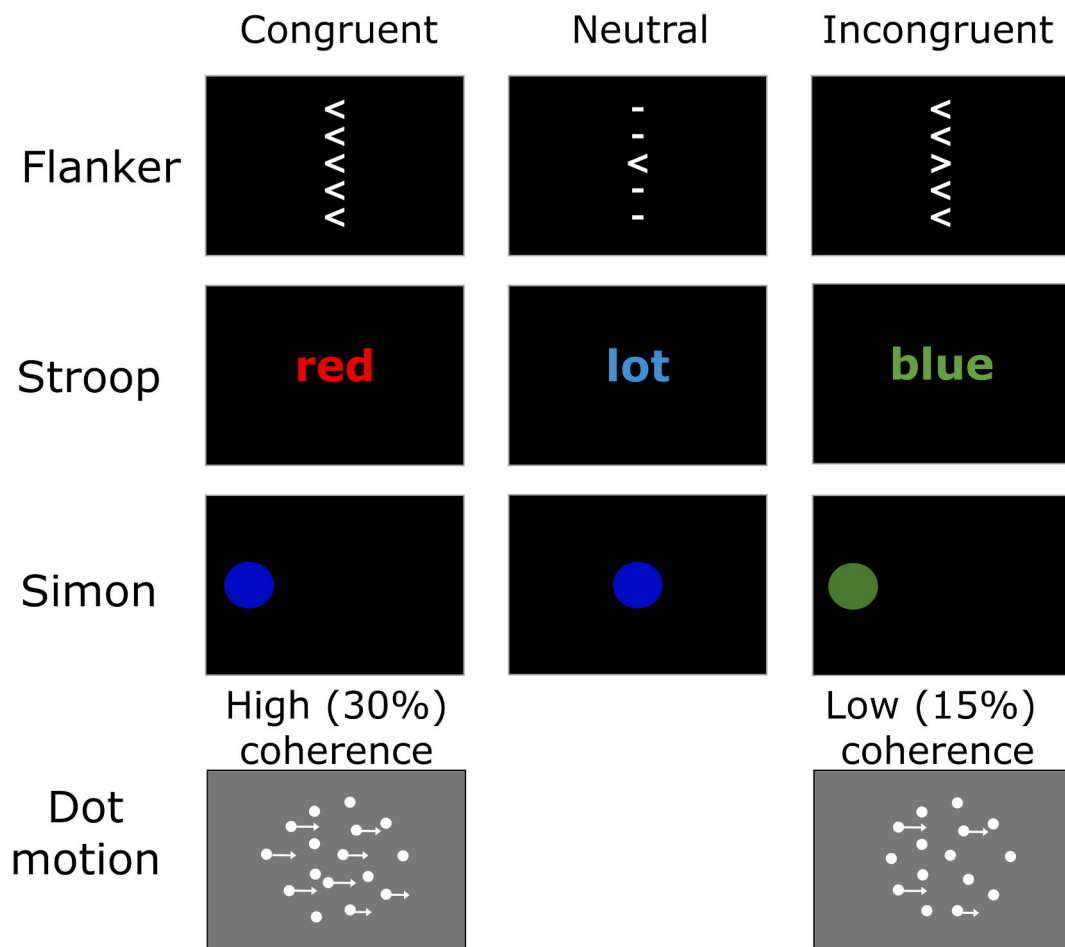
Details of datasets included in meta-analysis. See Source publications for detailed information. The number of correlations is dependent on the number of boundary separation parameters estimated from the dataset (at least one per task). See Fig. 2 for a schematic of the tasks.

| Dataset | Source                               | N   | Tasks                     | Conditions | Trials per condition | Number of correlations |
|---------|--------------------------------------|-----|---------------------------|------------|----------------------|------------------------|
| 1       | Hedge, Powell, et al. (2020)         | 50  | Flanker                   | 3          | 336                  | 1                      |
| 2       | Hedge et al. (2018b)                 | 103 | Simon                     | 3          | 336                  | 1                      |
|         |                                      |     | Flanker                   | 3          | 480                  | 1                      |
| 3       | Hedge, Powell, Bompas, et al. (2018) | 102 | Stroop                    | 3          | 480                  | 1                      |
|         |                                      |     | Simon (blocked trials)    | 2          | 288                  | 2 <sup>a</sup>         |
| 4       | Hedge et al. (2019; Exp. 1)          | 43  | Simon (intermixed trials) | 2          | 288                  | 1                      |
|         |                                      |     | Flanker                   | 9          | 576                  | 3 <sup>b</sup>         |
| 5       | Hedge et al. (2019; Exp. 2)          | 69  | Stroop                    | 9          | 576                  | 3 <sup>b</sup>         |
|         |                                      |     | Flanker                   | 9          | 192                  | 3 <sup>b</sup>         |
|         |                                      |     | Dot motion                | 6          | 240                  | 3 <sup>b</sup>         |

Note. The data were collapsed across two separate testing sessions in Datasets 2 (three weeks apart) and 4 (four weeks apart).

<sup>a</sup> The blocked version of the task includes a separate boundary estimate for congruent and incongruent trials.

<sup>b</sup> Datasets include separate boundary estimates for blocks in which instructions emphasise either speed, accuracy, or both speed and accuracy.



**Fig. 2.** Schematic of the choice reaction time tasks used in all datasets. In the flanker task participants identify the central arrow as pointing to either the left or the right. In the Stroop task participants identified the font colour as red, blue, green or yellow. In the Simon task participants identified the circle as either blue or green. In the dot-motion task, participants identified whether the direction of coherent motion in an array of dots was to the left or right. Trials were separated by an interstimulus interval of 750 ms except in the dot motion task where the interstimulus interval was 500 ms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

parameters estimated from each task. For tasks with no instruction manipulation and where congruent, neutral and incongruent trials are randomly intermixed within blocks (Datasets 1 and 2), we obtained one boundary parameter per task. In the “blocked trials” variant of the Simon task in Dataset 3, we obtained separate estimates for congruent and incongruent trials. For the speed-accuracy trade-off experiments in Datasets 4 and 5, we obtained separate estimates for blocks where speed, accuracy, and both speed and accuracy were emphasised.

The details of our model fitting can be seen in [Hedge, Powell, et al. \(2020\)](#) for Datasets 1 to 4, and [Hedge et al. \(2019\)](#) for Dataset 5. Briefly, parameters are estimated by comparing reaction times for correct and error responses to data simulated from the model. We use optimisation algorithms to find the set of parameters that minimises the discrepancy between the observed and simulated data. Each participant and task are fit independently. We collapsed Datasets 2 and 4 across testing sessions for the purposes of model fitting. For the diffusion model for conflict tasks we adopted a common approach (c.f. [White, Servant, & Logan, 2017](#)) wherein we created six bins based on quantiles of the observed reaction times ([0.1, 0.3, 0.5, 0.7, 0.9]), and counted the number of trials in each bin. We did this separately for correct and error responses, and separately for each condition. When participants made fewer than eleven errors in that condition, error reaction times were instead grouped into three bins ([0.3, 0.5, 0.7]). When fewer than five errors were made, we fit the median reaction time. We initially simulated data from the model using 5000 sets of randomly generated parameters and

compared the deviance ( $-2 \log$ -likelihood) between the observed data and each of our simulated datasets. We then submitted the best 15 parameter sets to a Nelder-Mead simplex ([Nelder & Mead, 1965](#)) based algorithm in Matlab, which we restarted 3 times to avoid local minima. The correlations we entered into our meta-analysis are with the best fitting parameters for each participant and task derived from this process. A similar approach is implemented in DMAT ([Vandekerckhove & Tuerlinckx, 2008](#)), which we used to fit the standard drift diffusion model to the dot-motion task (Dataset 5).

We examined the four-week test-retest reliability for the conflict diffusion model for the speed-accuracy trade-off tasks in Dataset 4 ([Hedge et al., 2019](#)). For boundary separation, these ranged from poor ( $ICC = 0.39$ ) to good ( $ICC = 0.71$ ). These values fall within the range observed elsewhere in the literature with the standard drift-diffusion model across a variety of tasks ([Enkavi et al., 2019](#); [Lerche & Voss, 2017](#); [Schubert, Frischkorn, Hagemann, & Voss, 2016](#)). To our knowledge, there is currently no evidence as to whether there are systematic differences in the reliability of parameters derived from different tasks or administration methods. For our current purposes, these findings suggest that there is some stability in individuals' levels of response caution.

### 2.5. Meta-analysis of caution with impulsivity

For each dataset, we calculated Spearman's rho correlations

between each boundary separation estimate and each UPPS-P subscale. For example, in Dataset 1 we computed correlations between the UPPS-P and both boundary separation in the flanker task and boundary separation in the Simon task. This produced 19 correlations in total across the datasets. These correlations were then meta-analysed using a multilevel random effects meta-analysis, implemented in the metafor package in R (R Core Development Team, 2017; Viechtbauer, 2010). We fit a three-level random effects model, which allows us to account for non-independence of correlations taken from the same dataset. At the first level, we assume that there is variance in the effect size estimates due to sampling error. At the second level, we assume that there is variance in the effect sizes that we are trying to estimate within each dataset. As we take multiple correlations from the same dataset, we refer to these as being nested within datasets. At the third level, we assume variance across datasets. To assess heterogeneity, we examined the  $I^2$  statistic, which represents the percentage of total variance attributable to between- and within-dataset variance relative to sampling variance (Viechtbauer, 2019).

We report the individual correlations, and the results for other parameters, in Supplementary material B. The data and analysis code are available from the Open Science Framework (<https://osf.io/w8va9/>).

### 3. Results

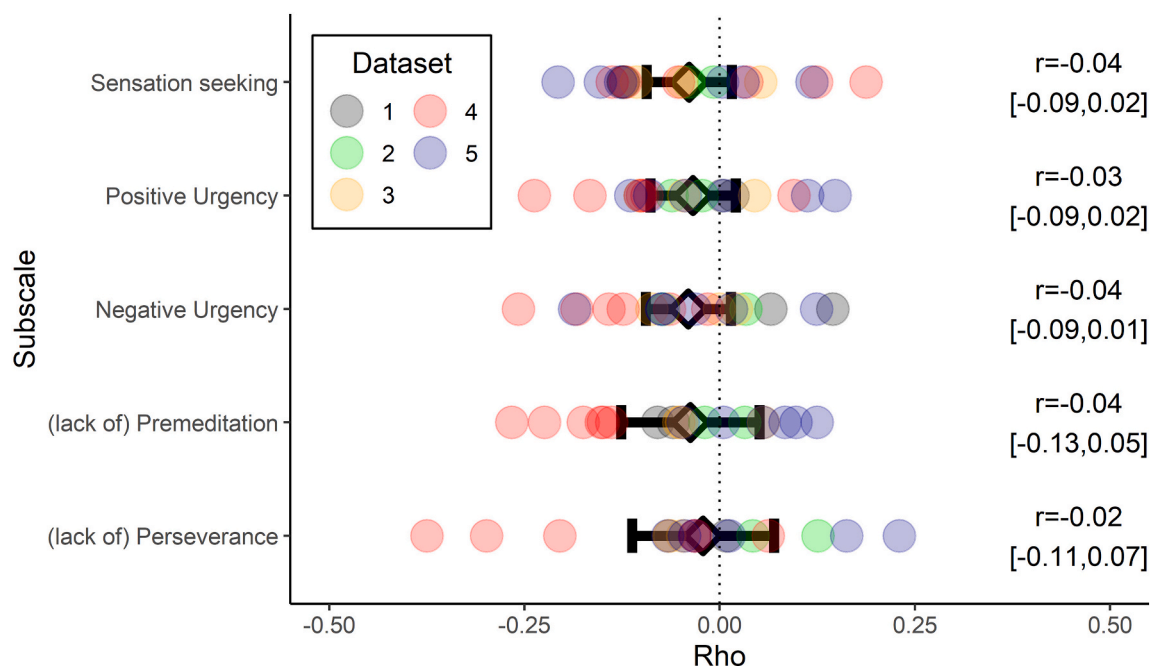
High scores on the UPPS-P subscales are indicative of higher impulsivity, whereas lower values of boundary separation are associated with lower caution. Therefore, if there were a consistent link between impulsivity and caution, we would expect correlations to be consistently negative. Fig. 3 plots the main results from our meta-analysis. For every UPPS-P subscale, the 95% confidence intervals included zero; we therefore observed no evidence that cognitive response caution was associated with impulsivity. Average correlations ranged from  $\rho = -0.04$  to  $\rho = -0.02$  and were consistently low across both subscales and datasets. The 95% confidence intervals for our estimates contain the values of  $\rho$  that we cannot reject based on our data, the

largest (absolute) value of which was observed for the premeditation subscale;  $\rho = -0.04$ , 95% CI  $[-0.13, 0.05]$ . In other words, the largest correlation that our data are consistent with is  $\rho = -0.13$ , and we can reject the hypothesis of a substantial relationship between response caution and any dimension of self-reported impulsivity. The highest total  $I^2$  value for any UPPS-P subscale was 30.7%, which is typically interpreted as a low level of heterogeneity (Higgins, Thompson, Deeks, & Altman, 2003) and was not statistically significant (Cochrane's  $Q = 23.54$ ,  $p = .17$ ). In all cases, this was primarily attributed to variance between datasets rather than within (see Supplementary material B for more details).

### 4. Discussion

Despite conceptual overlap and common terminology, we find no evidence that response caution (or impulsive information processing), as defined by cognitive decision models, is associated with any dimension of impulsivity, as defined by the UPPS-P. It seems that a cautious-impulsive dimension to the way that individuals approach choice reaction time tasks is orthogonal to the propensity for rash action captured by UPPS-P questions.

We extend previous research by attempting to isolate the aspect of cognitive task performance most likely to correlate with self-report impulsivity. Nevertheless, our findings are consistent with several studies and meta-analyses that have shown that self-report measures of impulsivity show weak or no correlation with behavioural tasks (Creswell et al., 2019; Cyders & Coskunpinar, 2011, 2012; Eisenberg et al., 2019; Saunders, Milyavskaya, Etz, Randles, & Inzlicht, 2018; Schluter, Kim, & Hodgins, 2018; Sharma et al., 2014). Theorists have previously noted the “jingle” problem in impulsivity research (Block, 1995; Sharma, Kohl, Morgan, & Clark, 2013); a common terminology has arisen to describe what may be different underlying constructs. Though it is intuitive to refer to individuals on the low end of the response caution dimension as ‘impulsive’ responders, our results indicate it captures something different from the sub-facets of impulsivity that are captured by the UPPS-P.



**Fig. 3.** Meta-analytic (black diamonds) and observed (circles) correlations between boundary separation/response caution and the UPPS-P impulsivity questionnaire subscales. Error bars and brackets show 95% confidence intervals. A multi-level random effects meta-analysis was performed on Spearman's rho correlations calculated for each pair of tasks, allowing for clustering where multiple correlations were taken from the same dataset. Note that all the 95% confidence intervals include zero.

The absence of a correlation with self-report does not mean the construct of response caution is not useful. There is evidence that response caution is sensitive to factors such as aging (e.g., Ratcliff et al., 2006; Ratcliff, Thapar, Gomez, & McKoon, 2004), and that it correlates across conflict tasks (Eisenberg et al., 2019; Hedge et al., 2019; Hedge, Powell, et al., 2020). To our knowledge, only one study has examined the relationship between response caution and real-world behaviours associated with self-control (e.g., self-reported drug use, obesity; Eisenberg et al., 2019). They found that factors derived from behavioural tasks, including response caution, showed poor predictive value relative to questionnaire methods (see also Creswell et al., 2019). Other studies have concluded that both behavioural and self-report measures have independent predictive value for real world outcomes (Sharma et al., 2014). Future efforts to resolve these inconsistencies may benefit from the dissociating response caution from other processes of potential interest in commonly used tasks, such as conflict control and processing speed (Hedge, Powell, et al., 2020). However, while we may gain a better understanding of these tasks through the application of cognitive models, this will not automatically lead us to better real-world prediction (for a discussion, see Hedge, Bompas, & Sumner, 2020).

Though our analysis was not exhaustive given the range of impulsivity measures used in the literature (Sharma et al., 2014), we expect our findings are not specific to the UPPS-P. The UPPS-P was designed to capture factors identified in existing impulsivity questionnaires (Whiteside & Lynam, 2001), and meta-analyses have shown that the UPPS-P and other questionnaires reflect common underlying traits (Sharma et al., 2014). With regard to behavioural tasks, as we have noted, response caution parameters correlate across conflict tasks, both in our data and in a larger battery of tasks (Eisenberg et al., 2019; Hedge, Powell, et al., 2020). We therefore have no reason to expect that response caution would correlate with self-report had we used different tasks, at least when considering spontaneous strategies in a healthy population. Cyders and Coskunpinar (2012) observed small but significant correlations with self-report measures and behavioural measures taken from “prepotent response inhibition” tasks (including the stop-signal, antisaccade, and continuous performance tasks), though variance in these measures may also reflect multiple underlying sources. For example, a recent model-based analysis of the stop-signal task suggests that correlations with self-reported impulsivity may reflect attentional lapses rather than inhibition (Skippen et al., 2019).

If people's disposition to make decisions quickly at the expense of errors is not driven by impulsivity, as commonly understood, then what does it reflect? It is possible that response caution is a behavioural “snap shot” (Cyders & Coskunpinar, 2012), and captures something different to individuals' self-perceived caution in everyday life. However, it appears that individuals are consistent in how they strategically approach similar tasks, given the between-task and reliability correlations we have previously observed (Hedge et al., 2019; Hedge, Powell, et al., 2020). We assume that individuals have some control over their level of caution - if we instruct participants to respond more quickly, they are able to do so, and this is captured in part by lowering their boundary (though other parameters can also change; Rae, Heathcote, Donkin, Averell, & Brown, 2014). We instructed participants to be both fast and accurate in our studies, except for the additional speed- and accuracy-emphasis blocks in Datasets 4 and 5, and we asked whether their default strategy might reflect trait impulsivity. There is evidence that individuals typically favour accuracy over speed by default (Forstmann et al., 2008), though we know little about why they spontaneously adopt the levels of caution that they do. Notably, while evidence accumulation models mathematically dissociate caution from parameters that represent processing ability, they may be correlated in real data (Schmiedek et al., 2007). Individuals who have higher levels of ability (higher drift rates) can achieve the same level of accuracy as individuals with lower ability (lower drift rates) while setting a lower boundary. Advancing our understanding of the mechanisms underlying response caution may benefit from examining why individuals adopt

sub-optimal strategies for their level of ability, or by incentivising participants to favour speed or accuracy (e.g., to optimise a reward; Starns & Ratcliff, 2010), rather than looking at what they do spontaneously.

In conclusion, we show that individuals with a tendency to respond quickly while risking errors do not self-report higher levels of impulsivity. These findings inform the interpretation of individual differences in response caution in domains such as aging (Ratcliff et al., 2006) and neuropsychological conditions (Metin et al., 2013; Powell et al., 2019). They are also a further illustration that the relationship between lab-based and self-report measures of impulsivity is not straightforward (Cyders & Coskunpinar, 2012; Sharma et al., 2014; Wennerhold & Friese, 2020).

#### CRedit authorship contribution statement

**Craig Hedge:** Conceptualization, Methodology, Formal analysis, Writing - original draft, Visualization, Investigation. **Georgina Powell:** Conceptualization, Writing - review & editing. **Aline Bompas:** Conceptualization, Writing - review & editing. **Petroc Sumner:** Conceptualization, Writing - review & editing, Funding acquisition, Supervision.

#### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.paid.2020.110257>.

#### References

- Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, 117(2), 187–215. <https://doi.org/10.1037/0033-2909.117.2.187>.
- Chambers, C. D., Garavan, H., & Bellgrove, M. A. (2009). Insights into the neural basis of response inhibition from cognitive and clinical neuroscience. *Neuroscience and Biobehavioral Reviews*, 33, 631–646. <https://doi.org/10.1016/j.neubiorev.2008.08.016>.
- Creswell, K. G., Wright, A. G. C., Flory, J. D., Skrzynski, C. J., & Manuck, S. B. (2019). Multidimensional assessment of impulsivity-related measures in relation to externalizing behaviors. *Psychological Medicine*. <https://doi.org/10.1017/S0033291718002295>.
- Cyders, M. A., & Coskunpinar, A. (2011). Measurement of constructs using self-report and behavioral lab tasks: Is there overlap in nomothetic span and construct representation for impulsivity? *Clinical Psychology Review*, 31, 965–982. <https://doi.org/10.1016/j.cpr.2011.06.001>.
- Cyders, M. A., & Coskunpinar, A. (2012). The relationship between self-report and lab task conceptualizations of impulsivity. *Journal of Research in Personality*, 46(1), 121–124. <https://doi.org/10.1016/j.jrp.2011.11.005>.
- Donders, F. C. (1969). On the speed of mental processes. *Acta Psychologica*, 30, 412–431. <http://www.ncbi.nlm.nih.gov/pubmed/5811531>.
- Donkin, C., Brown, S., Heathcote, A., & Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin and Review*, 18, 61–69. <https://doi.org/10.3758/s13423-010-0022-4>.
- Draheim, C., Mashburn, C. A., Martin, J. D., & Engle, R. W. (2019). Reaction time in differential and developmental research: A review and commentary on the problems and alternatives. *Psychological Bulletin*, 145(5), 508–535. <https://doi.org/10.1037/bul0000192>.
- Eisenberg, I. W., Bissett, P. G., Zeynep Enkavi, A., Li, J., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, 10(1), 2319. <https://doi.org/10.1038/s41467-019-10301-1>.
- Enkavi, A. Z., Eisenberg, I. W., Bissett, P. G., Mazza, G. L., MacKinnon, D. P., Marsch, L. A., & Poldrack, R. A. (2019). Large-scale analysis of test-retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, 116(12), 5472–5477. <https://doi.org/10.1073/pnas.1818430116>.
- Evans, N. J., Rae, B., Bushmakin, M., Rubin, M., & Brown, S. D. (2017). Need for closure is associated with urgency in perceptual decision-making. *Memory and Cognition*, 45(7), 1193–1205. <https://doi.org/10.3758/s13421-017-0718-z>.
- Eysenck, H. J., & Eysenck, M. W. (1985). *Personality and individual differences: A natural science approach*. Plenum.
- Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmaker, E. J. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 17538–17542. <https://doi.org/10.1073/pnas.0805903105>.
- Forstmann, B. U., Ratcliff, R., & Wagenmakers, E. J. (2016). Sequential sampling models

- in cognitive neuroscience: Advantages, applications, and extensions. *Annual Review of Psychology*, 67(67), 641–666. <https://doi.org/10.1146/annurev-psych-122414-033645>.
- Friedman, N. P., & Miyake, A. (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133(1), 101–135. <https://doi.org/10.1037/0096-3445.133.1.101>.
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, 10(10), 535–549. <https://doi.org/10.1111/spc3.12267>.
- Gosling, C. J., & Moutier, S. (2018). Brief report: Risk-aversion and rationality in autism spectrum disorders. *Journal of Autism and Developmental Disorders*, 48(10), 3623–3628. <https://doi.org/10.1007/s10803-018-3616-8>.
- Griffin, J. W. (2020). metapower: An R package for computing meta-analytic statistical power. R package version 0.1.0. <https://github.com/jasonwgriffin/metapower>.
- Hedge, C., Bompas, A., & Sumner, P. (2020). Task reliability considerations in computational psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. <https://doi.org/10.1016/j.bpsc.2020.05.004>.
- Hedge, C., Powell, G., Bompas, A., & Sumner, P. (2020). Strategy and processing speed eclipse individual differences in control ability in conflict tasks. *PsyArXiv Preprints*. <https://doi.org/10.31234/osf.io/vgpxq>.
- Hedge, C., Powell, G., Bompas, A., Vivian-Griffiths, S., & Sumner, P. (2018). Low and variable correlation between reaction time costs and accuracy costs explained by accumulation models: Meta-analysis and simulations. *Psychological Bulletin*, 144(11), 1200–1227. <https://doi.org/10.1037/bul0000164>.
- Hedge, C., Powell, G., & Sumner, P. (2018a). The mapping between transformed reaction time costs and models of processing in aging and cognition. *Psychology and Aging*, 33(7), 1093–1104. <https://doi.org/10.1037/pag0000298>.
- Hedge, C., Powell, G., & Sumner, P. (2018b). 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>.
- Hedge, C., Vivian-Griffiths, S., Powell, G., Bompas, A., & Sumner, P. (2019). Slow and steady? Strategic adjustments in response caution are moderately reliable and correlate across tasks. *Consciousness and Cognition*, 75. <https://doi.org/10.1016/j.concog.2019.102797>.
- Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327, 557–560. <https://doi.org/10.1136/bmj.327.7414.557>.
- Karalunas, S. L., Hawkey, E., Gustafsson, H., Miller, M., Langhorst, M., Cordova, M., ... Nigg, J. T. (2018). Overlapping and distinct cognitive impairments in attention-deficit/hyperactivity and autism spectrum disorder without intellectual disability. *Journal of Abnormal Child Psychology*, 46, 1705–1716. <https://doi.org/10.1007/s10802-017-0394-2>.
- Lerche, V., & Voss, A. (2017). Rest reliability of the parameters of the Ratcliff diffusion model. *Psychological Research*, 81, 629–652. <https://doi.org/10.1007/s00426-016-0770-5>.
- Lynam, D. R., Smith, G. T., Whiteside, S. P., & Cyders, M. A. (2006). *The UPPS-P: Assessing five personality pathways to impulsive behavior*. (Unpublished technical report).
- Metin, B., Roeyers, H., Wiersema, J. R., van der Meere, J. J., Thompson, M., & Sonuga-Barke, E. (2013). ADHD performance reflects inefficient but not impulsive information processing: A diffusion model analysis. *Neuropsychology*, 27, 193–200. <https://doi.org/10.1037/a0031533>.
- Nelder, J. A., & Mead, R. (1965). A simplex-method for function minimization. *Computer Journal*, 7, 308–313.
- Pigott, T. D. (2012). Advances in meta-analysis. *Advances in meta-analysis* Springer <https://doi.org/10.1007/978-1-4614-2278-5>.
- Pirrone, A., Dickinson, A., Gomez, R., Stafford, T., & Milne, E. (2017). Understanding perceptual judgment in autism spectrum disorder using the drift diffusion model. *Neuropsychology*, 31(2), 173–180. <https://doi.org/10.1037/neu0000320>.
- Pirrone, A., Johnson, I., Stafford, T., & Milne, E. (2020). A diffusion model decomposition of orientation discrimination in children with autism spectrum disorder (ASD). *European Journal of Developmental Psychology*, 17(2), 213–230. <https://doi.org/10.1080/17405629.2018.1561364>.
- Powell, G., Jones, C. R. G., Hedge, C., Charman, T., Happé, F., Simonoff, E., & Sumner, P. (2019). Face processing in autism spectrum disorder re-evaluated through diffusion models. *Neuropsychology*, 33(4), 445–461. <https://doi.org/10.1037/neu0000524>.
- R Core Development Team (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(5), 1226–1243. <https://doi.org/10.1037/a0036801>.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85.
- Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, 20(4), 873–922. <https://doi.org/10.1162/neco.2008.12.06-420>.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111, 333–367. <https://doi.org/10.1037/0033-295X.111.2.333>.
- Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging*, 19, 278–289. <https://doi.org/10.1037/0882-7974.19.2.278>.
- Ratcliff, R., Thapar, A., & McKoon, G. (2006). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin and Review*, 13, 626–635. <https://doi.org/10.3758/BF03193973>.
- Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*. <https://doi.org/10.1016/j.cogpsych.2009.09.001>.
- Rey-Mermet, A., Gade, M., & Oberauer, K. (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(4), 501–526. <https://doi.org/10.1037/xlm0000450>.
- Ridderinkhof, K. R. (2002). Micro- and macro-adjustments of task set: Activation and suppression in conflict tasks. *Psychological Research*, 66, 312–323. <https://doi.org/10.1007/s00426-002-0104-7>.
- Rolison, J. J., Hanoch, Y., Wood, S., & Liu, P. J. (2014). Risk-taking differences across the adult life span: A question of age and domain. *The Journals of Gerontology, Series B, Psychological Sciences and Social Sciences*, 69(6), 870–880. <https://doi.org/10.1093/geronb/gbt081>.
- Rutledge, R. B., Smittenaar, P., Zeidman, P., Brown, H. R., Adams, R. A., Lindenberg, U., ... Dolan, R. J. (2016). Risk taking for potential reward decreases across the lifespan. *Current Biology*, 26(12), 1634–1639. <https://doi.org/10.1016/j.cub.2016.05.017>.
- Saunders, B., Milyavskaya, M., Etz, A., Randles, D., & Inzlicht, M. (2018). Reported self-control is not meaningfully associated with inhibition-related executive function: A Bayesian analysis. *Collabra: Psychology*, 4(1). <https://doi.org/10.1525/collabra.134>.
- Schluter, M. G., Kim, H. S., & Hodgins, D. C. (2018). Obtaining quality data using behavioral measures of impulsivity in gambling research with Amazon's Mechanical Turk. *Journal of Behavioral Addictions*, 7(4), 1122–1131. <https://doi.org/10.1556/2006.7.2018.117>.
- Schmiedek, F., Oberauer, K., Wilhelm, O., Süß, H.-M., Wittmann, W. W., Suss, H. M., & Wittmann, W. W. (2007). Individual differences in components of reaction time distributions and their relations to working memory and intelligence. *Journal of Experimental Psychology: General*, 136(3), 414–429. <https://doi.org/10.1037/0096-3445.136.3.414>.
- Schubert, A.-L., Frischkorn, G., Hagemann, D., & Voss, A. (2016). Trait characteristics of diffusion model parameters. *Journal of Intelligence*, 4(3), 7. <https://doi.org/10.3390/jintelligence4030007>.
- Sharma, L., Kohl, K., Morgan, T. A., & Clark, L. A. (2013). “Impulsivity”: Relations between self-report and behavior. *Journal of Personality and Social Psychology*, 104, 559–575. <https://doi.org/10.1037/A0031181>.
- Sharma, L., Markon, K. E., & Clark, L. A. (2014). Toward a theory of distinct types of “impulsive” behaviors: A meta-analysis of self-report and behavioral measures. *Psychological Bulletin*, 140, 374–408. <https://doi.org/10.1037/a0034418>.
- Skippen, P., Matzke, D., Heathcote, A., Fulham, W. R., Michie, P., & Karayanidis, F. (2019). Reliability of triggering inhibitory process is a better predictor of impulsivity than SSRT. *Acta Psychologica*, 192, 104–117. <https://doi.org/10.1016/j.actpsy.2018.10.016>.
- South, M., Chamberlain, P. D., Wigham, S., Newton, T., Le Couteur, A., McConachie, H., ... Rodgers, J. (2014). Enhanced decision making and risk avoidance in high-functioning autism spectrum disorder. *Neuropsychology*, 28(2), 222–228. <https://doi.org/10.1037/neu0000016>.
- Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed-accuracy compromise: Boundary optimality in the diffusion model. *Psychology and Aging*, 25, 377–390. <https://doi.org/10.1037/a0018022>.
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18, 643–662.
- Ulrich, R., Schroter, H., Leuthold, H., & Birngruber, T. (2015). Automatic and controlled stimulus processing in conflict tasks: Superimposed diffusion processes and delta functions. *Cognitive Psychology*, 78, 148–174. <https://doi.org/10.1016/j.cogpsych.2015.02.005>.
- Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods*, 40, 61–72. <https://doi.org/10.3758/Brm.40.1.61>.
- Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, 36, 1–48.
- Viechtbauer, W. (2019). I<sup>2</sup> for multilevel and multivariate models. Retrieved from: [http://www.metafor-project.org/doku.php/tips:i2\\_multilevel\\_multivariate](http://www.metafor-project.org/doku.php/tips:i2_multilevel_multivariate).
- Weafer, J., & de Wit, H. (2014). Sex differences in impulsive action and impulsive choice. *Addictive Behaviors*, 39(11), 1573–1579. <https://doi.org/10.1016/j.addbeh.2013.10.033>.
- Wennerhold, L., & Friese, M. (2020). Why self-report measures of self-control and inhibition tasks do not substantially correlate. *Collabra: Psychology*, 6(1). <https://doi.org/10.1525/collabra.276>.
- White, C. N., Servant, M., & Logan, G. D. (2017). Testing the validity of conflict drift-diffusion models for use in estimating cognitive processes: A parameter-recovery study. *Psychonomic Bulletin and Review*. <https://doi.org/10.3758/s13423-017-1271-2>.
- Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, 30(4), 669–689.