The Forced-Response Method: A New Chronometric Approach to Measure Conflict Processing

Taraz Lee, Jacob Sellers, John Jonides, and Han Zhang

University of Michigan

Author Note

This research was supported in part by National Institute of Mental Health (NIMH) grant R21MH129909 to the University of Michigan (Principal Investigator: JJ) and in part by a grant from the College of Literature, Science, and the Arts to TL. Correspondence concerning this article should be addressed to Taraz Lee, 3063 East Hall, Ann Arbor, MI, 48109. E-mail: tarazlee@umich.edu.

Abstract

Despite long-standing concerns about the use of free reaction times (RTs) in cognitive psychology, they remain a prevalent measure of conflict resolution. This report presents the forced-response method as a fresh approach to examine speed-accuracy tradeoff functions (SATs) in conflict tasks. The method involves fixing the overall response time, varying the onset of stimuli, and observing response expression. We applied this method to an arrow flanker task. By systematically varying the time between stimulus onset and response, we reveal a comprehensive time course of the flanker interference effect that is rarely observed in previous literature. We further show that influential manipulations observed in free-RT paradigms similarly affect accuracy within the forced-response technique, suggesting that the forced-response method retains the core cognitive processing characteristics of traditional free-RT conflict tasks. As an RT-irrelevant, accuracy-based method, the forced-response method provides a novel and more nuanced look into the dynamics of conflict resolution.

Keywords: conflict resolution; response conflict; conflict processing; speed-accuracy tradeoff; response signal

The Forced-Response Method: A New Chronometric Approach to Measure Conflict Processing

Imagine visiting a country where traffic flows on the opposite side of the road compared to the U.S. (e.g., the U.K.) Most American drivers who visit such a country struggle to adapt to this change. Or imagine that you are on a diet, but you smell the delicious flavor of fried chicken from a restaurant and feel compelled to stop in for a bite. These scenarios exemplify situations in which well-learned and nearly automatic responses conflict with goal-directed responses. The question arises: How do we overcome such habits to produce responses appropriate to a goal at hand?

For decades, cognitive psychologists have used so-called “conflict tasks” to model response conflicts of this sort. One widely used such task is the flanker task (Eriksen & Eriksen, 1974). In one popular version of the flanker task, participants are asked to indicate the direction of a middle arrow when flanking arrows could either be congruent (e.g., → → → → →) or incongruent (e.g., ← ← → ← ←) in direction compared to the middle arrow. Typically, participants can respond to the stimulus whenever they are ready to respond on each trial[1]. The existence and efficacy of conflict resolution is typically indexed by the difference in free responses times (RTs) between congruent and incongruent conditions.

The flanker task is but one of many tasks in which a prepotent response conflicts with a goal-oriented response. Two other popular conflict tasks are the Stroop task (Stroop, 1935) and the Simon task (Simon, 1969). The congruency effect, as indexed by the difference in mean response time between congruent and incongruent trials in these tasks, has been the driving force

[1] In some cases, a response deadline may be imposed, but it is rather lenient in order to time out extremely long responses. We still consider these tasks as “free RT” tasks because participants have substantial freedom to decide when to respond.

behind the development of theories and models of conflict resolution over the years (cite early and recent studies).

However, there have been long-standing concerns about using free RT to index cognitive processes (Cronbach & Furby, 1970; Draheim et al., 2019; Edwards, 2001; Friedman & Miyake, 2004; Hedge et al., 2018, 2020; Miller & Ulrich, 2013). Despite these concerns, the practice of using free RTs to measure conflict resolution persists, with a recent survey finding that 84% of the difference scores in conflict tasks were based on mean RTs differences (von Bastian et al., 2020). The sheer reliance on mean RT differences, coupled with the issues associated with free RTs (discussed more below), calls for the development of alternative approaches to study conflict resolution.

In this paper, we revisit several issues having to do with using free RT to study conflict resolution. Then, we introduce a "forced-response" method as a new approach to studying conflict resolution that is complementary to the traditional free-RT-based approach. We will discuss how the forced-response method overcomes many issues with using free RT as a dependent measure of conflict-resolution. We will also demonstrate that the forced-response method provides novel insights into the dynamics of conflict resolution beyond existing methods.

Free RT-based Approach

Since at least the 1960s, cognitive researchers have appreciated many of the shortcomings of free RT as a dependent measure, and several alternative methods have been adopted which allow for the explicit examination of the time course of cognitive processing. A complete review of these issues is beyond the scope of this paper and can be found elsewhere

(Dosher, 1976; Draheim et al., 2019; Heitz, 2014; Wickelgren, 1977). However, we do point out several issues relevant to the current paper.

First, free RT difference scores (and free RTs in general) do not consider the speed-accuracy trade-off, which differs from person to person and task to task (Heitz, 2014). Participants who are fast but error-prone may appear better than participants who are slower but more careful when only free RT is considered. One might think that this is resolvable by somehow combining RT and accuracy into a single composite measure of performance, but it is not possible to create such a composite measure because there is no theoretically justified basis for adjusting free RTs based on how accurate a person is (Pachella, 1974).

Second, RT difference scores are confounded with general processing speed in that they are proportionally larger for overall slower individuals than for overall faster individuals. This makes it difficult to compare conflict resolution across populations that might differ in general processing speed.

Yet another problem has to do with whether RT itself is a pure measure of the cognitive processing required for a task. One study found that when participants were *forced* to respond earlier than they would normally, they could produce accurate responses with much less response time than suggested by their free RT (Haith et al., 2016). This indicates that RT is not an ideal measure of the time to complete processing of all component stages as there is a substantial delay between this time point and the actual initiation of the response. Another recent study found that response times exhibit a use-dependent bias, meaning that people are biased to respond at times similar to their previous response times. This result again suggests that a person's response time reflects factors above and beyond the processing required to produce an accurate response. Overall, then, whereas RT is traditionally thought of as the duration of various information processing steps, an alternative view is that RT is itself an independent motor-control parameter (Wong et al., 2017). By this view, people decide *when* to respond as well as *what* to respond.

Speed-Accuracy Tradeoff Approaches

Many cognitive psychologists have argued for the superiority of methods that specifically examine speed-accuracy tradeoff functions (SATs) as a tool to better understand cognitive processes of interest (Dosher, 1976; Heitz, 2014; Wickelgren, 1977). Though dwarfed by the volume of research examining mean RT, the explicit examination of SATs has allowed researchers to advance our understanding of the dynamics of conflict resolution since at least the 1990s (Gratton et al., 1992; Heitz & Engle, 2007; Ridderinkhof, 2002).

Conditional Accuracy Functions. One frequently used approach is to simply partition data from free RT tasks into a number of different time bins (e.g., 200-300 milliseconds, 301-400 milliseconds, 401-500 milliseconds, etc.) and then calculate an accuracy rate in each time bin. These are sometimes called Conditional Accuracy Functions (CAFs). With this binning method, researchers can examine the speed accuracy tradeoff inherent in the responses participants make. This approach has been used to attempt to examine the time course of interference resolution (Gratton et al., 1992; Heitz & Engle, 2007; Manohar et al., 2015).

There are several shortcomings of this approach, though. Binning RT leaves the sampling within each bin up to chance, which often leads to certain time bins being less well represented for a particular individual than others. A consequence of this is that some participants may provide a good deal of data for a particular bin whereas others may leave that bin underrepresented. Also, if the true SAT function varies randomly trial-to-trial, accuracy will be overestimated at fast RTs and underestimated at slow RTs. Unfortunately, it is the short reaction times that are often the most important when examining conflict resolution. This is because researchers are interested in investigating the latency at which fast, incorrect responses driven by prepotent or automatic cognitive processing give way to slower, more goal-directed responses. Additionally, as noted above, using free RT here still gives participants strategic control over when to respond and may not reflect the time course of the underlying cognitive processing required to emit a response.

Response deadlines and/or payoffs. Another class of methods that has been used to empirically obtain SATs has been to use response deadlines and/or payoffs (Fitts, 1966; Pachella & Pew, 1968). Participants can be instructed to respond before some time deadline following stimulus presentation. This time deadline can be varied from one block of trials to the next and participants can be given feedback to ensure that they produce enough responses that fall below each deadline. An experimental manipulation that is somewhat similar is to give participants monetary payoffs for speed vs. accuracy. Correct responses can be rewarded, errors can be punished by taking away accumulated rewards, and these values can be proportional to RT. For example, participants can be paid $P - k$*RT for each correct response and punished -k*RT for incorrect responses. P and k can be varied in different blocks of trials to induce participants to respond at different speeds and with different accuracy rates (Swensson & Edwards, 1971). Although these methods allow researchers to plot SATs, they also leave a substantial opportunity for participants to exert strategic control over how they are responding. Especially when payoffs and deadlines are blocked, participants my choose to adopt different strategies that vary systematically over the different conditions.

Response signals. To deal with some of the drawbacks in these SAT methods, some researchers have adopted what is often referred to as “the response-signal paradigm” (Reed, 1973). In a typical response-signal task, a stimulus is presented and is then followed at some short time lag by a signal that tells the research participant to initiate a response. This procedure allows the researcher to use the timing of the response signal as an independent variable and allows an examination of accuracy as the dependent measure. The timing of this response signal is usually varied randomly from trial to trial from very short times at which performance is at chance to long times at which performance reaches an asymptote.

Although this procedure may seem similar to tasks which employ a time deadline with a fixed lag following the stimulus, it is much more difficult for subjects to adjust their strategy as a function of the given processing time in that they are not informed of the timing before the trial begins. It is often assumed that a subject's strategy does not vary in any controlled way. By varying the timing of the response signal relative to the timing of stimulus presentation from trial to trial, researchers can be more certain that a subject is in the same state following stimulus presentation at any given time across all trials up until the timing of the response cue.

The response-signal paradigm is the only SAT method that does not require that a subject be informed of the time condition in advance of presenting the stimulus that requires a response. A major drawback of the response-signal paradigm is that participants, not being able to predict when the response signal will be presented, are working with great uncertainty from trial to trial. This leaves open the development of idiosyncratic strategies to prevent one from being "surprised" by an early signal and anticipating such a signal disproportionately.

A path forward

We believe a path forward in the study of conflict resolution is to move from RT as a dependent variable and instead treat it as an independent variable by adopting a significant variant of the response-signal paradigm which we call the "forced-response method". A version of this method was one of the first response-signal paradigms ever reported (Schouten & Bekker, 1967), but it has never been widely adopted, likely due to its difficulty in administration prior to the use of modern computer-based stimulus presentation. In the balance of this article, we outline the forced-response method and note its specific advantages for studying conflict resolution. To allay fears that adopting a forced-response method might alter the nature of processing in conflict tasks, we additionally present evidence showing that using this method to study conflict preserves several of the classic effects observed when using free-RT methodology.

METHODS

The forced-response method involves fixing overall response time from trial to trial, varying the onset of stimuli prior to a demanded response, and observing response expression. In this way, the forced-response method differs importantly from the response-signal paradigm. As we comment above, with that paradigm, there is constant uncertainty in when a response will be required. With the forced-response method, there is complete certainty as illustrated with the procedure shown in Figure 1. Participants receive *N* equally spaced signals (e.g., an auditory beep or a visual cue presented every 500 milliseconds) and are instructed to respond exactly at the onset of the final “go” signal. So, the timing of the demanded response is always predictable. Processing time (PT) is manipulated by varying the time from stimulus onset to the “go” signal, but response time is fixed on each trial by virtue of participants having to respond at the time of the ”go” signal. Response accuracy at each PT is the dependent measure. Next, we describe in detail an application of the forced-response method to an arrow flanker task.

Forced-response flanker task

Participants. We recruited 137 participants from Prolific.co to participate in this study. Participants were rewarded at a rate of $10 per hour for participating. Participants completed the forced-response flanker task (described in detail below) plus several questionnaires which are not relevant to the presentation of the method that is the topic of this paper. The duration for completing the forced-response task alone was about 35 minutes. Data from 11 participants were discarded due to accuracy in producing a response at the "go" signal (< 30%), leaving a total of 126 participants for analysis (mean age = 36.9, SD age = 12.4, 64.3% Female).

Task and Experimental Timeline. The task consisted of two practice phases and one experimental phase. In the first practice phase, participants trained for 40 trials on a free-response flanker task. On each trial, 5 arrows appeared on the screen and participants were asked to indicate the direction of the center arrow while ignoring the direction of the flanking arrows. They were instructed to press the "W" key (for left) or the "P" key (for right). There were two types of trials that appeared equally often: congruent trials (e.g., <<<<<<) and incongruent trials (e.g., <<><<). Each arrow had a size of (.06, .06) with the unit being percentage of screen height. On each trial, two outlined rectangles that had the same width as the row of arrows were also shown on the screen, one above one and one below the arrows, as illustrated in the example presented in Figure 2. These outlined rectangles served as response signals in later phases of the task. The arrows appeared in between the two outlined rectangles as shown in the figure. The arrows remained on the screen until response expression. Once a response was made, participants received feedback about their response accuracy. The feedback message was displayed for 400 milliseconds.

In the second practice phase, participants were given 30 trials of training to learn to produce a response at the time of a "go" signal. In this phase, there were no arrows shown on the screen. Instead, on each trial the two outlined rectangles were incrementally filled with white ink (see Figure 1A). As a trial began, white ink filled 25% of each rectangle's inner space every 500 milliseconds, resulting in a fully filled rectangle at 2000 milliseconds. The filling began at the center of each rectangle and expanded to the edges, and the rectangles were filled synchronously. The rectangles were removed from the screen after 2100 milliseconds. Participants were instructed to respond exactly when both rectangles were filled with white ink, (i.e., 2000 milliseconds). Participants were told that they could respond with either the W key or the P key, but they were encouraged to practice timing with both keys. After each trial, participants were given feedback for 1000 milliseconds about whether they responded too quickly (RT < 1900 milliseconds), too slowly (RT > 2100 milliseconds), or with perfect timing.

The experimental phase was a combination of the two practice phases (see Figure 1A). Participants performed 10 blocks of 48 trials of the flanker task with forced-response timing and with stimulus presentation time that was uniformly varied prior to the "go" signal. As in the first practice phase, participants were asked to respond to the direction of the center arrow using the "W" and "P" keys. As in the second practice phase, participants were asked to respond exactly when the two rectangles were filled with white ink, (i.e., 2000 milliseconds). The onset time of the arrows was selected from a uniform distribution between 1000 milliseconds and 2000 milliseconds in increments of 20 milliseconds with the value on each trial unpredictable. In other words, the arrows would appear at a random time within one second before the "go" signal. This approach allowed us to measure the accuracy of responses when the exact amount of time allowed for stimulus processing and response preparation was controlled. After each trial, participants received feedback about their timing accuracy, as in the second practice phase. After the feedback, there was an intertrial interval of 1000 milliseconds before the next trial began.

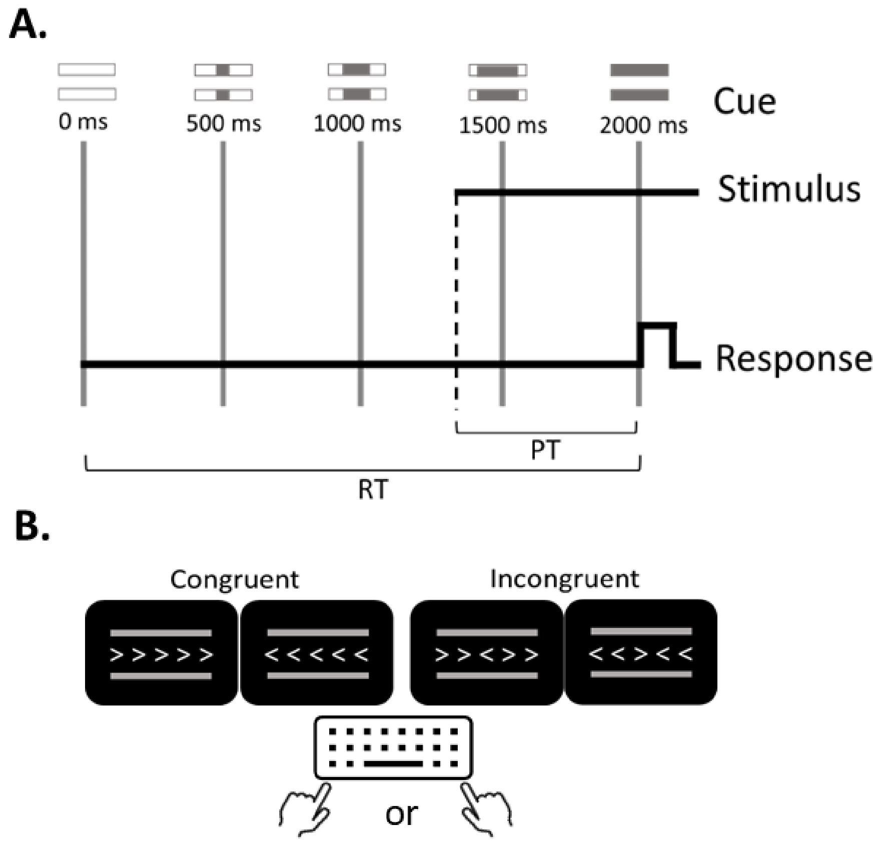

Figure 1. A schematic illustration of the forced-response flanker task. Figure 1A shows the forced-response procedure. As a trial began, white ink filled 25% of each rectangle's inner space every 500 milliseconds, resulting in a fully filled rectangle at 2000 milliseconds. Participants were asked to respond exactly when the two rectangles were completely filled, namely at 2000 milliseconds. The imperative stimulus appears unpredictably during the last 1000 milliseconds of the countdown. Figure 1B shows the task display at 2000 milliseconds, at which participants must press the left or the right key to respond to the direction of the middle arrow.

Analysis. We define processing time (PT) as PT = RT – *t* (target onset). For example, if on a given trial the stimulus appeared at 1500 milliseconds, and the response was made at 2050 milliseconds, PT is then calculated as 2050 – 1500 = 550 milliseconds. This calculation method accounts for the fact that participants are not always accurate at responding exactly at the "go" signal, and it reflects the actual time it took for participants to generate a response. Note that a negative PT occurs when participants responded before the stimulus appeared. We removed trials if PT was smaller than 0 millisecond or greater than 1000 milliseconds (3.99% of trials).

The data produced by the forced-response method consist of response accuracies and their corresponding processing times. These data can be analyzed using SMART, a state-of-the-art technique for reconstructing the time-course from one-sample-per-trial data, which also permits statistical analysis of the accuracies (van Leeuwen et al., 2019). Simply put, the method involves first smoothing the data of each participant and then using a weighted average to construct a group-average time course. Furthermore, a cluster-based permutation test can be used to examine differences between time courses or from a baseline. Python scripts for implementing the SMART method can be found at the project's OSF website: https://osf.io/qa2uc/.

RESULTS

The relationship between response accuracy and PT for congruent and incongruent trials, as analyzed using the SMART method, is shown in Figure 2. Several features in these data are worth highlighting. First, when PT is short (before approximately 200 milliseconds), response curves are no different from chance (50%). This suggests that, on these trials, participants were simply guessing as there was little time for any meaningful processing of the stimulus to take place.

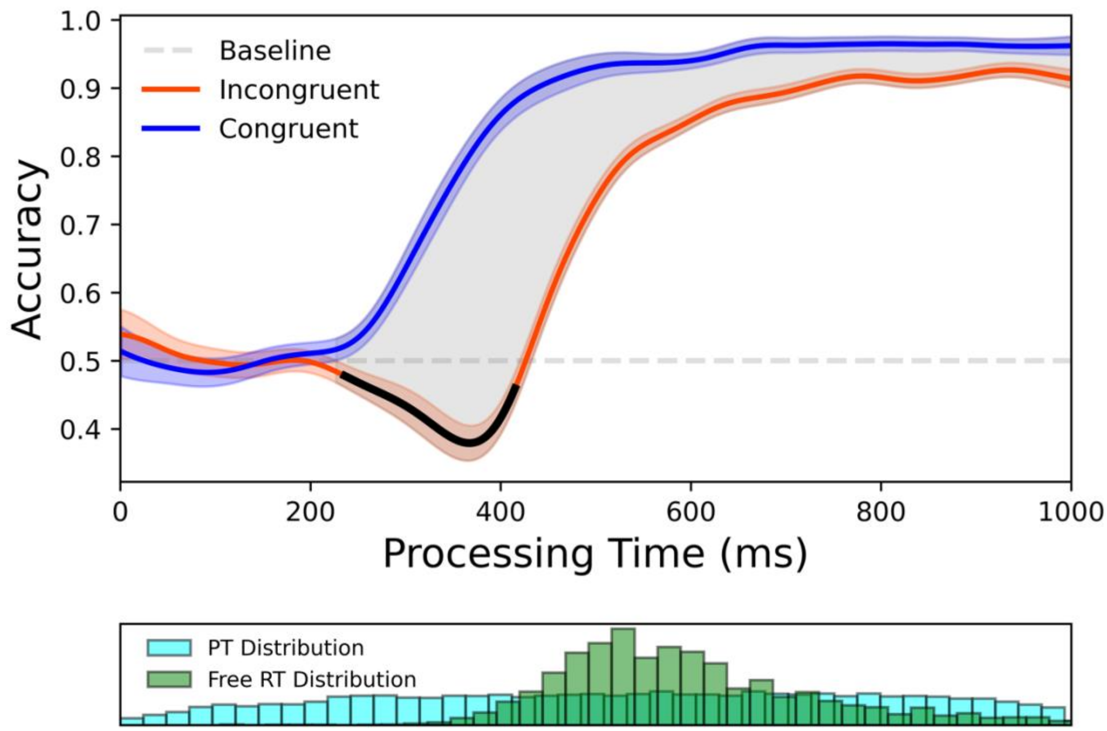

Figure 2. Response Accuracy as a Function of PT. The relationship between processing time (PT) and response accuracy in the forced-response flanker task. In the top panel, the shaded area between the two curves indicates the period of significant divergence in accuracies for congruent and incongruent trials. The black segment on the incongruent curve represents the period when accuracy for incongruent trials was significantly below 50%. Confidence bands indicate a 95% confidence level. In the bottom panel, the teal distribution represents PTs. For comparison, the green distribution shows RTs in the incongruent condition for the same participants completing a free-RT practice block.

The most intriguing feature of this plot appears at intermediate levels of PT. On congruent trials, response accuracy rose monotonically from chance as PT became longer. By contrast, on incongruent trials, response accuracy initially *decreased* from chance as PT increased, followed by a recovery in response accuracy at longer PTs. A cluster-based permutation test shows that response accuracy for incongruent trials was significantly below 50% from 236 milliseconds to 415 milliseconds, with a nadir at 367 milliseconds. In essence, for incongruent trials, there exists a period when accuracy worsened despite longer processing times. This time window reflects the influence of the prepotent, automatic response that needs to be inhibited and replaced with the goal-directed response.

Using a cluster-based permutation test, we also determined that the two curves diverged significantly starting at 227 milliseconds. If we define the flanker interference effect as the difference in accuracy between congruent and incongruent trials, then this point of divergence represents the earliest time point when the flanker interference effect emerges. At 391 milliseconds, the two curves reached their maximum divergence with a difference in accuracy rate of 44.9%. Following the same definition, this point represents when the flanker interference effect was the strongest.

To illustrate the advantage of the forced-response method compared to the free-response method, let us compare the distributions of PTs and free RTs. In the bottom panel, the teal distribution represents the distribution of PTs in this task, which is roughly uniformly distributed by design. For comparison, the green distribution shows RTs in the incongruent condition for the same participants completing a free-RT practice block. The bulk of the distribution of free-RT's is located after the largest interference effect, as revealed by the forced-response analysis. Indeed, the maximally divergent point (391 milliseconds) has a percentile rank of only 2.28% in this free-RT distribution. In other words, the free RT distribution almost entirely missed when the flanker effect was the strongest. But the forced-response method by its very design captures this effect and samples this time point relatively densely.

Finally, response accuracy on incongruent trials never recovered to the same level as that on congruent trials. For incongruent trials, response accuracy appears to have plateaued after roughly 750 milliseconds of PT, even though there is still “room for improvement” (accuracy at PT = 750 milliseconds was 91.2 % for incongruent trials and 96.4% for congruent trials). This suggests a lingering effect of response conflict despite sufficient time to process the stimuli. Of course, if we had extended PT beyond 1000 milliseconds, this difference between congruent and incongruent trials may have been reduced. Indeed, consider the thought experiment of giving participants as long as 10 sec of processing time. By that time, it is likely that accuracy for the congruent and incongruent trials would be nearly if not entirely identical.

To summarize, the forced response method offers a more nuanced understanding of response conflict dynamics compared to the traditional free-RT method. By systematically varying the time between stimulus onset and response, this approach uncovers the complete time course of the flanker interference effect, from its early emergence to its peak, and its lingering presence even with extended processing time.

Next, we assess several aspects of the forced-response paradigm that speak to the validity of the method.

Timing Accuracy

One might wonder to what extent participants can be forced to respond at a specific time. Figure 3 shows a distribution of the RTs in this task. The RTs were generally centered around the “go” signal (2000 milliseconds), with a mean of 2017 milliseconds. Of course, it is virtually impossible to force a response at precisely the “go” signal. There was variance around this value with a standard deviation of 128 milliseconds. Nevertheless, we consider this standard deviation to be quite small compared to that of a canonical free-RT distribution. Overall, 67.68% of the trials were judged as “on-time”, meaning that participants responded within an arbitrarily defined 100-millisecond margin of error. 12.70% of the trials were “too fast”, meaning that participants responded more than 100 milliseconds prior to the “go” signal. 19.62% of trials were "too slow", meaning that participants responded more than 100 milliseconds after the “go” signal.

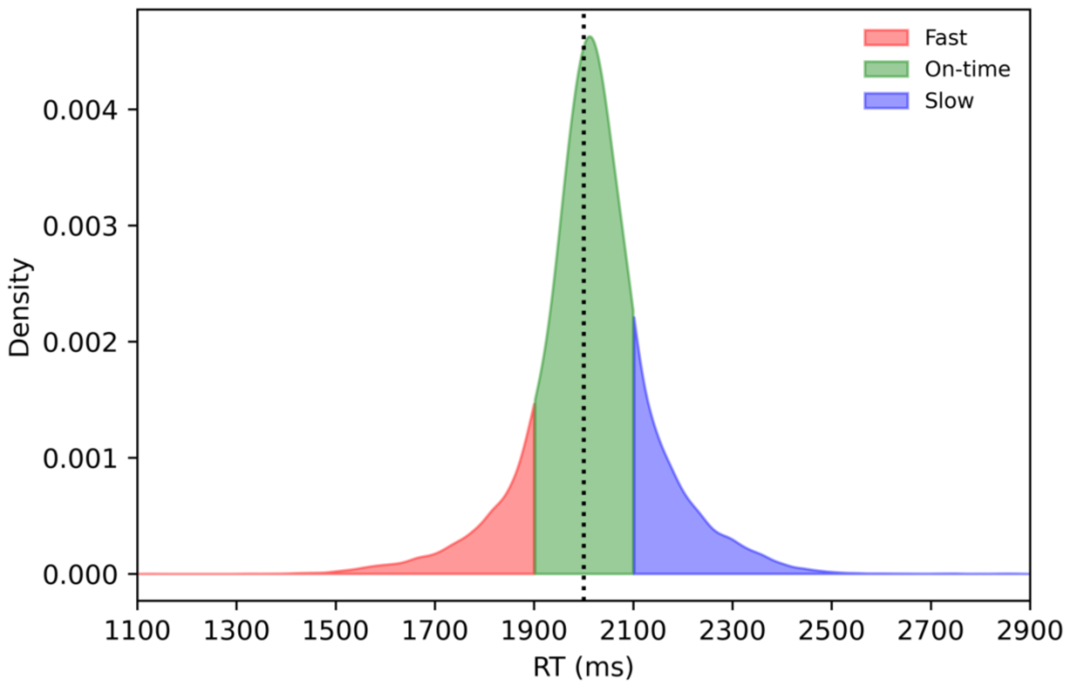

Figure 3. RT distribution in the forced-response flanker task. The green region indicates RTs judged as “on-time” (+/- 100 milliseconds). The red and blue regions indicate RTs judged as “too fast” and “too slow” respectively.

Recall that the calculation of PT is based on the time from stimulus onset to the actual RT rather than the ideal RT (i.e., 2000 milliseconds), which accounts for the variations in RTs around the timing goal. It is also important to emphasize that the decision to categorize responses as "fast", "slow", and "on-time" was arbitrary. The purpose of this was simply to display a feedback message after each trial so that participants could adjust their response timing accordingly.

Therefore, it may not be appropriate to simply discard "fast" and "slow" trials during data analysis. We recommend a sensitivity analysis to examine the potential impact of timing criterion on the results. We provide an example of such an analysis in the next section.

Sensitivity Analysis of the Timing Criterion

We varied the timing criterion from ± 25 milliseconds to ± 400 milliseconds in increments of 5 milliseconds. These criteria encompass a range from 22.8% to 98.8% of the entire RT distribution depicted in Figure 3. For each cut-off point, we constructed group-average time courses using the SMART method and calculated the nadir of the incongruent condition, referring to the point at which accuracy was lowest. The results are displayed in Figure 4. As illustrated in the figure, the time courses based on different timing criteria are strikingly similar. The nadir distribution centers around 367 milliseconds, which is the original value calculated based on "on-time" trials. These results reinforce the argument against simply discarding "too fast" or "too slow" trials: What truly matters is not whether participants were actually on time, but rather whether they were attempting to be on time.

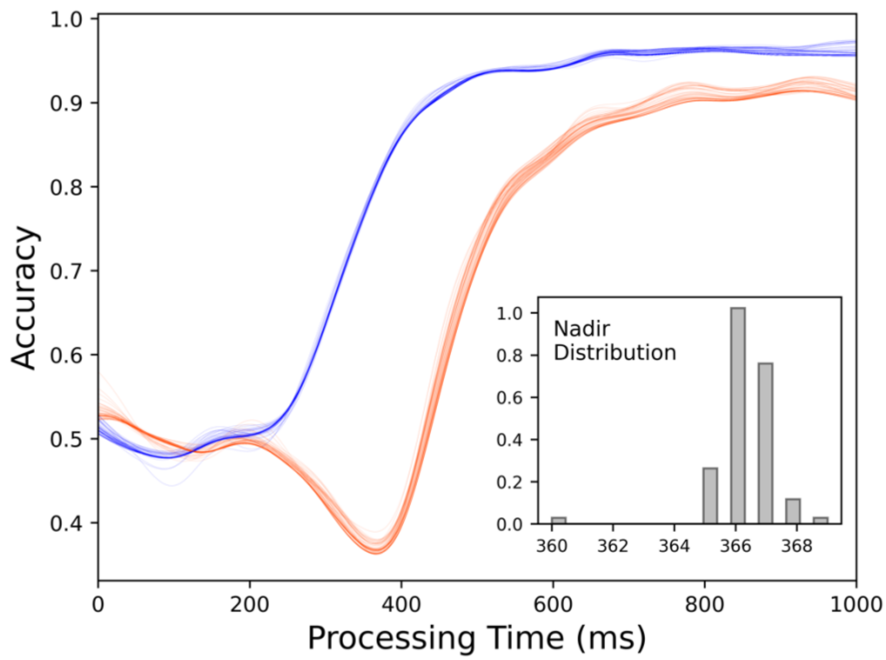

Figure 4. The time course of response accuracies for different timing criteria. Each line represents a distinct timing criterion (ranging from ± 25 milliseconds to ± 400 milliseconds in increments of 5 milliseconds). The inset panel displays the distribution of the nadir across various timing criteria.

Does the Forced-response Method Change the Nature of the Task?

The forced-response method requires participants to resolve response conflict while at the same time monitoring when to emit their responses at the proper time. One might wonder whether this dual-task situation yields performance in conflict tasks that is fundamentally different from performance that would be seen with free responding. To examine this issue, we selected three variables that are well-known to modulate the congruency effect in free-response tasks, and we examined whether qualitatively similar effects can be observed using the forced-response task. The three variables are distractor salience, conflict frequency, and previous-trial congruency.

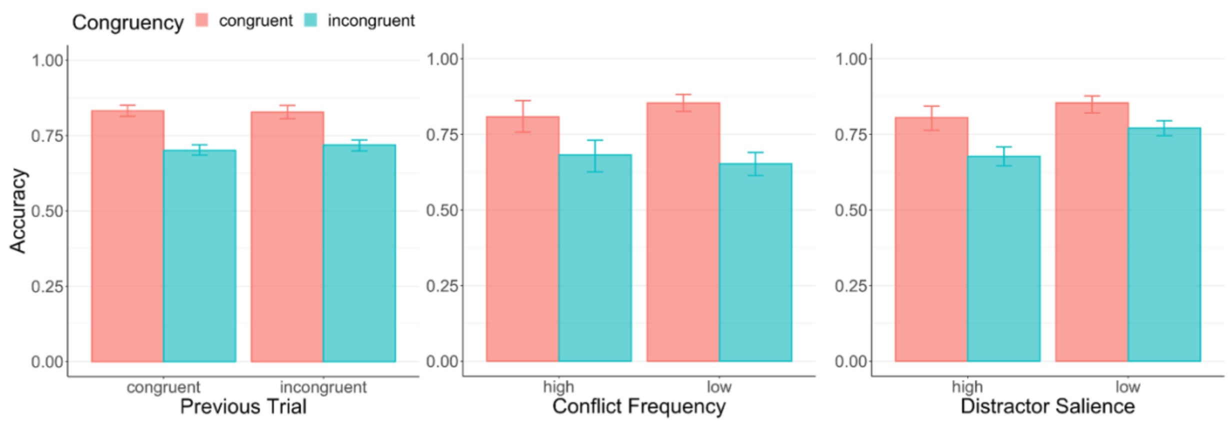

Figure 5. Effects of three variables on performance under conflict. **a.** Congruency sequence effect. The congruency effect is diminished following incongruent trials. **b.** Conflict frequency effect. Low: 30% incongruent and 70% congruent. High: 70% incongruent and 30% congruent. **c.** Distractor salience effect. High: flanking arrows bigger than the center arrow. Low: flanking arrows smaller than the center arrow. Columns and error bars reflect means and bootstrapped 95% confidence intervals, respectively.

Distractor Salience. In the flanker task, a larger interference effect can be obtained by increasing the salience of the flanking arrows relative to the center arrow (Zeischka et al., 2011). We recruited 45 participants from Prolific.co to complete a forced-response flanker task with a distractor salience manipulation. Data from 6 participants were discarded due to poor performance (accuracy less than 50%, less than 30% of trials on-time, or more than 15 trials in a row not being on-time). Participants were randomly assigned to one of two conditions: low salience (N = 20) or high salience (N = 19). In the low distractor-salience condition, the flanking arrows were smaller than the center arrow (flanking arrows: (.04, .04), center arrow: (.08, .08), unit: percentage of screen height). In the high distractor salience condition, the flanking arrows were larger than the center arrow (flanking arrows: (.08, .08), center arrow: (.04, .04), unit: percentage of screen height). In addition, there were two minor differences from the task described above. First, the feedback message was displayed only when participants were too slow or too fast. Second, the intertrial interval randomly varied between 0 to 1 second. All other aspects of the task were the same as the original task.

To facilitate comparison, we computed the average response accuracy for each condition, aggregating across all PTs (see Figure 5a). The data were then analyzed using a 2 (trial type: congruent/incongruent) * 2 (distractor salience: high/low) ANOVA. The results of the ANOVA show a statistically significant interaction between trial type and distractor salience (F (1, 37) = 5.90, p = 0.020). This was driven by a larger interference effect when distractor salience was high (0.045 difference in accuracy rate), a result that is similar to what the literature has shown for free response situations (Zeischka et al., 2011).

Conflict Frequency. Previous research has shown that increasing the proportion of incongruent trials leads to smaller interference effects (Logan, 1980). We recruited 61 participants from Prolific.co to complete a forced-response flanker task with a conflict frequency manipulation. Data from 7 participants were removed due to poor performance (accuracy less than 50%, less than 30% of trials on-time, or more than 15 trials in a row not being on-time). Participants were randomly assigned to one of two conditions: low conflict frequency (N = 36) or high conflict frequency (N = 18). In the low conflict-frequency condition, congruent trials had a 70% chance of appearing on each trial and incongruent trials had a 30% chance. In the high conflict-frequency condition, congruent trials had a 30% chance of appearing and incongruent trials had a 70% chance of appearing on each trial. All arrows had an equal size of (.06, .06). All other aspects of the task were the same as those of the distractor salience task.

We again computed the average response accuracy for each condition, aggregating across all PTs (see Figure 5b). A 2 (trial type: congruent/incongruent) * 2 (conflict frequency: high/low)

ANOVA show a statistically significant interaction between trial type and conflict frequency (F (1,52) = 6.48, p = 0.014). Again, this pattern of results mimics what the literature has documented for free response tasks (Logan, 1980): interference effects are smaller when conflict frequency is high (0.074 difference in accuracy rate).

Previous Trial Congruency. The effect of previous trial congruency on the congruency effect, or the “congruency sequence effect” (CSE), is the observation that congruency effects are typically smaller following incongruent trials compared to following congruent trials (Gratton et al., 1992). For this variable, we simply re-analyzed the forced-response flanker data reported in the previous section with the addition of previous trial congruency (N = 126).

We observed a CSE effect using the forced-response flanker task (see Figure 5c). A 2 (current trial congruency: congruent/incongruent) * 2 (previous trial congruency: congruent/incongruent) show a statistically significant interaction between current and previous trial congruency (F (1, 125) = 7.34, p = 0.008). Once again, this mimics the results from free response experiments (Gratton et al., 1992): interference effects are smaller following incongruent trials (0.023 difference in accuracy rate).

In short, we have compared results from the forced-response method to previously documented results using free-response methods, and we found qualitatively similar outcomes in all cases: Greater salience of the distractors resulted in more interference; lower distractor frequency resulted in more interference; and the CSE effect was demonstrated. We take these data to indicate that in spite of the forced-response method requiring a “dual-task” on the part of participants, this method does not do injustice to the underlying interference effect in the flanker task.

Discussion

The forced-response method we advance here can be fruitfully used to explicitly examine speed-accuracy tradeoff functions in a variety of conflict resolution tasks to gain greater insight into the mechanisms of cognitive control. We show that effects of classic manipulations that affect behavior in free-RT paradigms are preserved when accuracy is examined using the forced-response technique. We found that increasing the salience of the distractor led to worse accuracy, especially on incongruent trials. We also observed the classic conflict-frequency effect wherein a high proportion of incongruent trials leads to a reduction in interference and relatively improved performance on incongruent trials. Finally, we observed a congruency-sequence effect whereby an incongruent trial leads to improvement in performance if the subsequent trial is also incongruent, but a worsening in performance if the subsequent trial is congruent. These findings give us confidence that the forced-response method does not fundamentally change the nature of cognitive processing seen in conflict tasks with free RT and allows researchers to empirically observe how the speed-accuracy tradeoff function shifts with respect to independent variables of interest. Compared to free-RT tasks, the forced-response method is RT-irrelevant, accuracy-based, and provides a more nuanced look into the dynamics of conflict resolution.

The advantage of speed-accuracy tradeoff methods

Arguments for the superiority of explicitly measuring speed-accuracy tradeoffs relative to free RT experiments are not new (e.g., Wickelgren, 1977). However, it seems that the force behind these arguments has diminished over time, or perhaps they have just been forgotten as the vast majority of experimental psychology experiments over the last several years have employed free-RT measures. In addition to some of the issues we raised in the introduction, the basic issue is that participants can trade their accuracy for speed such that *any* mean value of reaction time is possible for *any* task depending on what level of accuracy participants set as their criterion. Unless accuracy is carefully measured, mean RT is not a good indicator of task difficulty or the time it takes for the cognitive processes of interest. Often researchers seek to have participants maintain an extremely high level of accuracy so that RT differences are more interpretable. Unfortunately, if one simply observes the form of a speed-accuracy tradeoff function, it is exactly at near ceiling levels of accuracy where one would expect to observe large variation in RT with negligible changes in accuracy (Pachella, 1974; Reed, 1973; Wickelgren, 1977) . That is, very small changes in accuracy may lead to RT differences that might be larger than RT differences due to independent variables of interest. The forced-response method we advance here wrests control of speed-accuracy criterion away from our participants back to the experimenter and obviates these concerns.

Application to other conflict and dual process tasks

While the results presented here combine the forced-response task with classic interference resolution tasks in cognitive psychology, we believe that this method can be fruitfully extended to any task that seeks to examine dual processes that have different latencies and competing responses. For example, many visual search tasks involve visually scanning an array of stimuli for a target stimulus with salient competing distractor stimuli. Researchers who are interested in the dynamics of distractor and target processing could adopt the forced-response methodology to fruitfully examine the time-course of both automatic attention capture and more endogenous, goal-oriented attention. Additionally, a recent study applied the forced-response method to look at how trained stimulus-response mappings affect goal directed processing when stimulus-response relationships are remapped (Hardwick et al., 2019). There are also many applications of this methodology outside of the strict domain of cognitive psychology. For example, dual process models in social psychology have been used to explain stereotyping using tasks such as the Implicit Association Task and dual process models in behavioral economics are often invoked to explain reasoning and decision making in a variety of contexts such as delayed discounting. The application of the forced response method might provide insights that would not be possible using standard tasks in these domains.

Limitations

Like any method used in cognitive psychology, there are still some potential limitations for the forced-response method. The method requires a large number of trials to produce the sort of speed-accuracy tradeoff curves shown in Figure 2 relative to free RT experiments that simply seek to measure a mean RT. This is because the entire time course of processing must be sampled to allow for inference. Trials that are not “on-time” either need to be thrown out or accounted for when drawing inferences about the nature of cognitive processing at a particular processing time (although see our analysis above showing that what criterion one uses has very modest effects on the outcome). Additionally, there might still be concern that the forced-response method changes the nature of the task. Since participants are required to resolve conflict while monitoring their timing, one might wonder if this yields a dual task that differs in performance observed using free response. However, this concern should be mitigated somewhat by our experiments above showing classic findings in conflict resolution using the forced-response method.

At very short processing times at which the stimulus appears just before a response must be emitted, responding accurately or responding on time is a special challenge. This is illustrated in the bottom panel of Figure 4, which shows a drop in the proportion of PTs at early time points. Although these responses should essentially leave participants at chance performance, a critical number of trials is required to establish this with a high level of confidence. There also may be concerns that when error rate across the experiment is high, participants might change their response strategy due to a high number of errors. Although this concern is again somewhat allayed by the fact that our method can reproduce classic effects in the cognitive control literature, researchers should bear in mind that the error rate in a forced response experiment may be much higher than tasks they typically use and may alter participants' strategies. If this is of concern, it can be mitigated by oversampling longer processing times where accuracy is expected to be higher. Finally, as is the case with response-signal methods, participants have to withhold their response until the cue at longer processing times. Researchers should again interrogate whether this situation is too distinct from the real-world situations in which they are interested, and whether using the forced-response method would fundamentally change the nature of the relevant cognitive processing.

Conclusion

As we reviewed, the response-signal method is a true advance in the study of cognitive control in that it provides a view of the entire chronometric course of processing in a task. However, its limitation is that there is uncertainty on the part of participants about when a response has to be emitted. What we have proposed here is a significant variant of the response-signal method that eliminates this uncertainty in response production. Participants learn through training to emit a response when it is demanded using this method, and by varying the timing of the onset of the stimulus that they must process, we preserve the value of the response-signal method in that we can trace out the entire course of processing from start to finish as opposed to free-RT methods that provide information at just one slice of time. We recommend application of the forced-response method to a variety of tasks for which it is appropriate, ones in which a prepotent response competes with a goal-directed response. Applying it to these tasks should enrich our understanding of the underlying cognitive processes that are engaged.

Declarations

Funding

This research was supported in part by the National Institute of Mental Health (NIMH) grant R21MH129909 to the University of Michigan (Principal Investigator: JJ) and in part by a grant from the College of Literature, Science, and the Arts to TL.

Competing Interests

The authors have no relevant financial or non-financial interests to disclose.

Ethics Approval

Approval was obtained from the ethics committee of the University of Michigan. The procedures used in this study adhere to the tenets of the Declaration of Helsinki.

Consent to Participate

Informed consent was obtained from all individual participants included in the study.

Consent to Publish

Not applicable.

Code and data availability

The datasets and code used for the current study are available in the OSF repository, https://osf.io/qa2uc/.

References

Cronbach, L. J., & Furby, L. (1970). How we should measure" change": Or should we? *Psychological Bulletin*, *74*(1), 68.

Dosher, B. A. (1976). The retrieval of sentences from memory: A speed-accuracy study. *Cognitive Psychology*, *8*(3), 291–310.

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

Edwards, J. R. (2001). Ten difference score myths. *Organizational Research Methods*, *4*(3), 265–287.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149.

Fitts, P. M. (1966). Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, *71*(6), 849.

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

Gratton, G., Coles, M. G., & Donchin, E. (1992). Optimizing the use of information: Strategic control of activation of responses. *Journal of Experimental Psychology: General*, *121*(4), 480.

Haith, A. M., Pakpoor, J., & Krakauer, J. W. (2016). Independence of Movement Preparation and Movement Initiation. *Journal of Neuroscience*, *36*(10), 3007–3015. https://doi.org/10.1523/JNEUROSCI.3245-15.2016

Hardwick, R. M., Forrence, A. D., Krakauer, J. W., & Haith, A. M. (2019). Time-dependent competition between goal-directed and habitual response preparation. *Nature Human Behaviour*, *3*(12), 1252–1262. https://doi.org/10.1038/s41562-019-0725-0

Hedge, C., Powell, G., Bompas, A., & Sumner, P. (2020). Strategy and processing speed eclipse individual differences in control ability in conflict tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. https://doi.org/10.31234/osf.io/vgpxq

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186.

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience*, *8*. https://doi.org/10.3389/fnins.2014.00150

Heitz, R. P., & Engle, R. W. (2007). Focusing the spotlight: Individual differences in visual attention control. *Journal of Experimental Psychology: General*, *136*(2), 217–240. https://doi.org/10.1037/0096-3445.136.2.217

Logan, G. D. (1980). Attention and automaticity in Stroop and priming tasks: Theory and data. *Cognitive Psychology*, *12*(4), 523–553.

Manohar, S. G., Chong, T. T.-J., Apps, M. A., Batla, A., Stamelou, M., Jarman, P. R., Bhatia, K. P., & Husain, M. (2015). Reward pays the cost of noise reduction in motor and cognitive control. *Current Biology*, *25*(13), 1707–1716.

Miller, J., & Ulrich, R. (2013). Mental chronometry and individual differences: Modeling reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin & Review*, *20*(5), 819–858. https://doi.org/10.3758/s13423-013-0404-5

Pachella, R. G. (1974). The Interpretation of reaction time in information-processing research. In B. Kantowitz (Ed.), *Human information processing* (1st ed., pp. 41–82). Routledge.

Pachella, R. G., & Pew, R. W. (1968). Speed-Accuracy Tradeoff in Reaction Time: Effect of Discrete Criterion Times. *Journal of Experimental Psychology*, *76*(1p1), 19.

Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science*, *181*(4099), 574–576.

Ridderinkhof, K. R. (2002). Activation and suppression in conflict tasks: Empirical clarification through distributional analyses. In W. Prinz & B. Hommel (Eds.), *Common mechanisms in perception and action* (1st ed., pp. 494–519). Oxford University Press.

Schouten, J., & Bekker, J. (1967). Reaction time and accuracy. *Acta Psychologica*, *27*, 143–153.

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology*, *81*(1), 174.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, *18*(6), 643.

Swensson, R. G., & Edwards, W. (1971). Response strategies in a two-choice reaction task with a continuous cost for time. *Journal of Experimental Psychology*, *88*(1), 67.

van Leeuwen, J., Smeets, J. B. J., & Belopolsky, A. V. (2019). Forget binning and get SMART: Getting more out of the time-course of response data. *Attention, Perception, & Psychophysics*, *81*(8), 2956–2967. https://doi.org/10.3758/s13414-019-01788-3

von Bastian, C. C., Blais, C., Brewer, G. A., Gyurkovics, M., Hedge, C., Kałamała, P., Meier, M. E., Oberauer, K., Rey-Mermet, A., Rouder, J. N., Souza, A. S., Bartsch, L. M., Conway, A. R. A., Draheim, C., Engle, R. W., Friedman, N. P., Frischkorn, G. T., Gustavson, D. E., Koch, I., … Wiemers, E. A. (2020). *Advancing the understanding of individual* *differences in attentional control: Theoretical, methodological, and analytical considerations* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/x3b9k

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics. *Acta Psychologica*, *41*(1), 67–85.

Wong, A. L., Goldsmith, J., Forrence, A. D., Haith, A. M., & Krakauer, J. W. (2017). Reaction times can reflect habits rather than computations. *Elife*, *6*, e28075.

Zeischka, P., Coomans, D., Deroost, N., Vandenbossche, J., & Soetens, E. (2011). Target–flanker discriminability affects conflict size but not sustained suppression. *Acta Psychologica*, *136*(1), 148–156.