

Investigating the speed-accuracy tradeoff: Better use deadlines or response signals?

Michael Dambacher^{a,b} and Ronald Hübner^a

^a *Dept. of Psychology, Universität Konstanz, Germany*

^a *Zukunfts Kolleg, Universität Konstanz, Germany*

Running head: Investigating the speed-accuracy tradeoff

Corresponding author:

Michael Dambacher
Dept. of Psychology
Box D29
Universität Konstanz
78457 Konstanz
Germany
E-Mail: michael.dambacher@uni-konstanz.de
Fon: +49-(0)-7531 88-4376
Fax: +49-(0)-7531 88-3017

Abstract

Deadlines (DL) and response signals (RS) are two well-established techniques to investigate speed-accuracy tradeoffs (SATs). Methodological differences imply, however, that corresponding data do not necessarily reflect equivalent processes. Specifically, the DL procedure grants knowledge about trial-specific time demands and requires responses *before* a pre-specified period has elapsed. In contrast, RS intervals often vary unpredictably between trials, and responses must be given *after* an explicit signal. Here, we investigated the effects of these differences in a flanker task. While all conditions yielded robust SAT functions, a right-shift of the curves pointed to reduced performance in RS (Exp. 1: blocked; Exp. 2-3: randomized) compared to DL conditions (Exp. 1-3: blocked), indicating that the detection of the RS imposes additional task demands. Moreover, the flanker effect vanished at long intervals in RS settings, suggesting that stimulus-related effects are absorbed in a slack when decisions complete prior to the signal. In turn, effects of a flat (Exp. 2) vs. a performance-contingent payment (Exp. 3) indicated that susceptibility to response strategies is higher in the DL than in the RS method. Finally, the RS procedure led to a broad range of slow responses and high accuracies, whereas DL conditions resulted in smaller variations in the upper data range (Exp. 1-2); with performance-contingent payment (Exp. 3), though, data ranges became similar. Together, the results uncover characteristic procedure-related effects and should help to select the appropriate technique.

(230 words)

Keywords

Speed-accuracy tradeoff; deadline; response signal; flanker task

Investigating the speed-accuracy tradeoff:

Better use deadlines or response signals?

In situations requiring a rapid choice between different options, decision makers face the difficulty of trading speed for accuracy. Naturally, the best (subjective) outcome is desirable, but the acquisition of sufficient information for an accurate decision takes time and slows down responses. Conversely, a fast choice may meet time demands, but at the cost of accuracy. This *speed-accuracy tradeoff (SAT)*, i.e., the inverse relation between response times (RTs) and error rates, appears as a fundamental principle of human performance (e.g., Pachella, 1974; Wickelgren, 1977), and SAT research has provided important insights on the time course of information processing in a variety of tasks, such as memory retrieval (Doshier, 1976; Reed, 1973), visual search (e.g., Carrasco & McElree, 2001; McElree & Carrasco, 1999), or perceptual decision making (e.g., Kleinsorge, 2001; Wenzlaff, Bauer, Maess, & Heekeren, 2011).

Much of the current knowledge about SATs is based on data from two experimental techniques, the Deadline (DL) and the Response Signal (RS) procedure. Both set up a number of pre-specified response intervals and lead to increasing RTs and accuracies as these intervals increase. However, the procedures also differ in some aspects, and it is unclear to what extent the results depend on the applied method. For instance, in the DL procedure responses have to be given *prior* to a *known* temporal limit. The fact that the DL is known in advance might allow participants to adopt specific response strategies. In contrast, strategies can hardly be established in the RS procedure, because responses are required immediately *after* a mostly *randomly timed* external signal. Up to now, little is known about the behavioral consequences of such methodological differences, and it is open if one of the two techniques is generally better suited for SAT research.

In the present study, we address this issue by systematically examining how specific characteristics of the DL and RS procedures affect performance in different conditions. Accordingly, our results are helpful for the interpretation of corresponding data. Further, they can assist the selection of the appropriate method for specific research questions, as they unveil the respective

strengths and weaknesses of the procedures. In particular, our results show that the RS procedure provides accurate control of RTs over a large time range, and that it is more robust against strategic influences than the DL procedure. Yet, the required detection of an external response signal is resource-consuming and therefore impairs performance. The DL method, in contrast, permits a more flexible behavioral adjustment to task demands and accordingly enables participants to perform optimally. Furthermore, the DL procedure unveils effects of task difficulty over the entire SAT data range, whereas they can be masked in RS settings.

Before going into a detailed description and discussion of the present experiments, we provide some background on SAT research and elaborate the potential implications of the differences between the DL and RS procedures.

Delineating the speed-accuracy tradeoff

SATs have been widely used to examine the time course of stimulus processing. In typical SAT functions, a steep increase of accuracies in the range of fast responses documents a great benefit of relatively little additional processing time. For slower RTs, the growth rate decreases until accuracies eventually approximate an asymptote. Especially RS studies often describe SATs as a negatively accelerated exponential function, whose parameters inform about processing rates and asymptotic accuracy levels, as well as about the time when accuracy grows above chance (e.g., Doshier, 1976; Kumar, Rakitin, Nambisan, Habeck, & Stern, 2008; McElree & Carrasco, 1999).

Further, computational models of decision making have shaped our understanding of SATs. For instance, sequential sampling models for two-alternative choices, such as the diffusion model (Ratcliff, 1978), assume that noisy evidence about stimulus identity is accumulated over time until one of two decision criteria (or thresholds) is reached, and an associated response is initiated. They can account for SATs via adjustable decision criteria: A low criterion is rapidly reached by the accumulation process and therefore permits fast responses, but since evidence accumulation contains noise, the probability of hitting the incorrect threshold (i.e., triggering an erroneous response) is higher for low than for high criteria. Consequently, modulation of the criterion results in

a positive relation between RTs and accuracy (e.g., Ratcliff & Rouder, 1998). A second source of variability in such models is the rate of evidence accumulation that determines the speed of response selection: Higher rates for easy than for difficult tasks entail faster and more accurate responses (e.g., Ratcliff & Rouder, 2000).

Together, mathematical and computational models are useful for examining the dynamics of decision making. Naturally, however, they need to be empirically validated, because their theoretical assumptions do not necessarily reflect the actual processes. For instance, whereas many models account for SATs via variations of decision thresholds relative to baseline activation, recent neuroimaging studies suggest that SATs result (at least partly) from the modulation of baseline activity itself, which alters the distance to a threshold (Bogacz, Wagenmakers, Forstmann, & Nieuwenhuis, 2010). Thus, the disclosure of the exact mechanisms of SATs also relies on empirical evidence and makes the selection of appropriate experimental methods essential.

Differences between DL and RS procedures

In general, the DL and RS procedures appear as excellent techniques for the assessment of SATs: in contrast to standard RT tasks, which uncover only relatively small behavioral variations and sometimes tempt to disregard the interdependency of speed and accuracy (for a discussion, see Wickelgren, 1977)¹, DL and RS setups provide experimental control over a broad data range (e.g., RTs, accuracies, or correlates of neural activity) that reflects quite directly the dynamics of decision processes.

Specifically, the DL procedure requires responses to a stimulus *before* a time limit has exceeded. Timeliness of responses is often signaled explicitly by an additional stimulus (e.g., a tone) at the end of the DL interval. Alternatively, feedback after a response can indicate whether the DL

¹ Note that the SAT can also be assessed with a variant of the standard RT task, in which participants are instructed to emphasize either speed or accuracy in different trials. In fact, such settings are very similar to the DL procedure, since speed-instructed trials often use a relatively short deadline that must not be exceeded, whereas accuracy-instructed trials use a very long (or infinite) deadline (Forstmann et al., 2010; Ratcliff & Rouder, 1998). Yet, speed-accuracy instructions manipulate only two levels of the SAT, whereas classical DL procedures usually employ several temporal limits and provide a more fine-grained SAT course.

was missed. In any case, participants are informed about DL intervals in advance (Diederich & Busemeyer, 2006; Link & Tindall, 1971; Ratcliff & Rouder, 2000; Rinkenauer, Osman, Ulrich, Müller-Gethmann, & Mattes, 2004).

Responses in the RS procedure, in contrast, are given *after* a pre-specified interval. The end of this *RS lag* is indicated by a distinct signal, such as the on- or offset of an auditory or visual stimulus. After some practice, subjects are able to respond very quickly within a brief post-RS period (e.g., Carrasco & McElree, 2001; Göthe & Oberauer, 2008; Miller, Sproesser, & Ulrich, 2008; Ratcliff, 2008; Reed, 1973). Most often RS lags vary unpredictably from trial to trial, but it is also possible to use a fixed RS for blocks of several trials (Schouten & Bekker, 1967).

Up to now, differences between the two procedures together with critiques and suggestions have mainly been stated verbally (Ratcliff, 2006; Reed, 1976; Wickelgren, 1977). It is therefore not established if and under what conditions one method is favorable over the other. Moreover, it is unclear whether DL- and RS-based SATs reflect the same processes or if procedure-specific characteristics have to be taken into account for the interpretation of the data. In fact, the latter possibility receives support from fMRI studies on the localization of neural correlates of the SAT: with a DL procedure, the pre-supplementary motor area (pre-SMA) revealed higher activity under high than under low speed-stress, compatible with the assumption that baseline activity increases with temporal demands (Forstmann et al., 2008) (see also Ivanoff, Branning, & Marois, 2008; van Veen, Krug, & Carter, 2008). In contrast, the RS procedure yielded the opposite pattern, i.e., lower activation of the pre-SMA for short than for long response intervals (Blumen et al., 2011). These seemingly inconsistent effects of time pressure suggest that differences between the methods may well affect underlying processes. It is therefore important to explicate these differences and to scrutinize their consequences on SATs. Here, we focus on three issues differentiating the procedures: *response signal detection*, *waiting for the response signal*, and *response strategies*.

Response signal detection

An obvious difference between DL and RS procedures is that responses in DL settings are internally triggered and can be given any time between stimulus onset and the upper time limit (i.e., the deadline), whereas RS tasks – irrespective of randomized or blocked designs – require response execution after an external signal. Thus, participants in the RS procedure do not only perform the main task, but have also to continuously monitor for the occurrence of an imperative stimulus, a situation that imposes additional task demands. Indeed, there is ample evidence that divided attention can influence behavior (Gherri & Eimer, 2011; Pashler, 1994; Pashler & Johnston, 1998). Usually, RTs are slower when two tasks are processed simultaneously compared to their isolated execution (e.g., Ruthruff, Pashler, & Klaassen, 2001). It is therefore possible that overall performance is impaired in RS relative to DL settings.

Waiting for the response signal

As another issue, external response triggering in the RS procedure may alter theoretically relevant effects, such as the influence of task (or stimulus) difficulty on performance. According to sequential sampling models, short RS lags often require a response before accumulated evidence reaches a threshold, because unknown temporal demands (i.e., randomized RS lags) do not permit the adjustment of decision criteria. In an early study, Ratcliff (1978) even proposed an approach with no decision thresholds, so that RS-triggered responses are selected on the basis of the current state of accumulated evidence. Later, Ratcliff (1988) reinserted decision criteria and suggested that response distributions reflect a mixture of two possible states: (a) Incomplete processes that are interrupted by the RS before they hit a criterion. These responses are relatively low in accuracy, because they are based either on partial evidence or on guessing (Ratcliff, 2006, 2008). (b) Completed processes that have reached a threshold prior to the RS. In this case, responses are highly accurate. With these principles, diffusion models can account for the SAT in RS data even though thresholds are constant, because the proportion of completed decisions, and therefore accuracy [i.e., case (b) above], increases with RS intervals (Ratcliff, 2006).

Naturally, though, if decisions complete prior to the RS, then subjects have to wait, i.e., withhold an already chosen response until the RS occurs (Ratcliff, 2006). This situation generates a slack that can eliminate or at least reduce stimulus-related effects in the SAT. For instance, response selection usually completes faster in easy than in difficult tasks, but corresponding performance differences are absorbed when responses in the easy task are artificially delayed until RS occurrence. Further, especially at very long RS intervals, a large proportion of completed and therefore delayed responses in both the easy and difficult task result in highly similar latencies irrespective of processing rate. In contrast, the DL procedure does not involve waiting periods, because a response can be initiated as soon as it is selected. Accordingly, effects of stimulus difficulty should be preserved even at long DL intervals.

Response strategies

Finally, a major criticism against the DL procedure is the assumption that it is susceptible to strategic influences. The reason is that subjects are informed about the temporal demands on every trial, and therefore can adjust their behavior appropriately. For example, knowledge of a short DL may urge subjects to lower their decision threshold in order to respond rapidly; further, it is possible that they increase the proportion of fast guesses (Reed, 1976; Wickelgren, 1977). Critically, such subjective adaptation strategies may vary between experimental conditions. Considering motivational factors, for instance, participants may use the temporal knowledge to adjust decision thresholds or guessing rates differently when they receive performance-contingent compared to performance-independent rewards. Thus, DL data might not only reflect processes of evidence accumulation but also strategies that are mediated by specific experimental manipulations.

Such adaptive behavior is hardly possible in RS experiments. Especially randomly varying RS lags prevent temporal preparation, so that participants are essentially in the same state across all trials irrespective of temporal demands. Moreover, the imperative signal leaves little room for other factors, such as motivation. The RS procedure has therefore sometimes been considered as a more appropriate method to examine SATs (Ratcliff, 2006; Reed, 1976; Wickelgren, 1977).

Unfortunately, decisive evidence for strategic influences is missing. For example, Miller et al. (2008) recently used the RS procedure in a perceptual discrimination study to examine effects of temporal preparation. RS lags varied either randomly from trial to trial or were constant throughout several blocks. The results revealed longer RTs and higher accuracies for variable than for constant intervals, as well as more pronounced RS lag effects in the constant condition. Nevertheless, data of both conditions followed the same SAT function, suggesting that prior knowledge of processing time had no major impact on overall performance. Accordingly, it is unclear if and to what extent temporal preparation and associated response strategies affect DL and RS data differentially.

Present study

In the present study, we examined the implications of these procedural differences on SATs, i.e., effects of *response signal detection*, *waiting for the response signal*, and *response strategies*. In order to disclose methodological influences (i.e., *waiting for the response signal*) on effects of stimulus difficulty, we used a flanker paradigm as task (B. A. Eriksen & Eriksen, 1974) wherein participants judged the parity of a central target numeral in the presence of task-irrelevant flankers (cf., Dambacher, Hübner, & Schlösser, 2011; Hübner & Schlösser, 2010). These flankers modulate the overall item difficulty as they can be incongruent (e.g., odd flanker numerals when the target is even), neutral (e.g., non-numeric flankers), or congruent (e.g., even flanker numerals when the target is even). As a standard finding, such tasks reveal robust congruency or *flanker effects*, i.e., slower and more error prone responses for incongruent than for neutral or for congruent stimuli. This performance difference results from the co-processing of irrelevant flankers that – together with the target numeral – fall into the spatial focus of attention (often referred to as “spotlight” or “zoom lens” of attention) (C. W. Eriksen & St. James, 1986). Specifically, co-processing of congruent flankers supports the selection of the correct response, whereas incongruent flankers produce a conflict by activating the wrong response. Neutral flankers neither bias the correct nor the incorrect response. Importantly, depending on the stimulus type, the attentional focus can vary in size: for incongruent and neutral stimuli it is advantageous to minimize the focus because both flanker types carry no

useful information. Congruent stimuli, in contrast, benefit from a wide focus encompassing supportive flankers (Hübner, Steinhauser, & Lehle, 2010). Thus, to minimize influences of such cross-conditional variations of the attentional focus we used only incongruent and neutral stimuli. Importantly, this manipulation of stimulus congruency is suitable to uncover differential effects of task difficulty between DL- and RS-based SATs. As we will show, the congruency effect decreases with increasing response intervals in the RS but not in the DL procedure, indicating that theoretically relevant effects can depend on the method. Our flanker-task approach therefore unveils procedure-related effects that remain unobserved with a single stimulus type.

Overall, we conducted a series of three experiments. Experiment 1 established a baseline of the effect of internal vs. external response triggering in the DL and RS procedures, respectively. Specifically, we used constant (i.e., blocked) and therefore predictable DL and RS intervals, granting temporal preparation in both procedures. Thus, they differed only with respect to responding before (i.e., DL) or after (i.e., RS) the signal. The first purpose was to explore whether *response signal detection* produces costs in performance. In this case, SAT curves should be shifted to the right in RS compared to DL data. Furthermore, Experiment 1 tested effects of *waiting for the response signal*, i.e., the hypothesis that effects of stimulus difficulty are reduced when subjects withhold their response at long RS lags. We hypothesized that the flanker effect decreases with longer response intervals in the RS but not in the DL procedure.

As described above, blocked response intervals of Experiment 1 are fairly common in DL studies, but are rather unusual in RS experiments; instead, RS lags usually vary randomly (i.e., unpredictably) between trials to prevent systematic adaptation of response strategies to temporal demands. Experiment 2 adopted this procedure (i.e., blocked DL and randomized RS intervals) and examined the generalizability of the results of Experiment 1 (i.e., effects of *response signal detection* and *waiting for the response signal*) to the most commonly used DL and RS setups. We again expected reduced performance in the RS relative to the DL procedure, and an attenuation of the flanker effect at long RS intervals.

Finally, we addressed the assumption that influences of *response strategies* play a greater role in the DL than in the RS procedure. We used monetary incentives in Experiment 3 (in addition to blocked DL and randomized RS intervals, cf., Exp. 2) to encourage participants to optimize their response strategies in order to maximize performance-based rewards. If strategic confounds are less dominant in the RS procedure, improved performance in Experiment 3 relative to Experiment 2 should be observed especially in the DL procedure.

All these predictions are directly testable on the basis of empirical SAT curves. For reasons of clarity and simplicity, and because additional fits of mathematical or computational models would not add novel insights with respect to our present objectives, this work focuses on analyses of empirical SAT data².

General Methods

Stimuli

In all experiments, numerals from 2 to 9 served as target items in a parity-judgment task. Two identical flankers on either horizontal side of the target set up stimulus congruency. For incongruent stimuli, flankers consisted of response-incompatible numerals, i.e., flankers and targets differed in parity. For neutral stimuli the characters \$, &, ?, or # were used as flankers. The target was always presented at screen center. Each character extended a visual angle of approximately 0.9° horizontally and 1.27° vertically and the spacing between characters (center to center) was 1.27° of

² Another reason for not fitting theoretical functions was the fact that the commonly used exponential function for RS data is less suitable for our DL data: SAT curves in the DL procedure do not show a clear asymptote and feature a linear rather than a negatively accelerated course (cf., Fig. 1-3); the linearity is even stronger when accuracies are transformed into d' values, which are usually used as criterion in exponential fits. Consequently, exponential (nonlinear least squares) fits to the DL data are poor and often do not converge for individual subjects. Proper fits would therefore require different functions for the two methods, so that comparisons of the parameters are not meaningful. We also refrained from computational modeling, because our flanker-task data would have required the application of rather complex models (Hübner et al., 2010; White, Ratcliff, & Starns, 2011) (see also Hübner & Töbel, 2012), which was beyond the scope of this study.

visual angle. Stimuli were presented in white on a black background. Signals indicating the end of DL intervals as well as the imperative RS consisted of an 800 Hz sine tone with a duration of 100 ms.

Apparatus

Visual stimuli were displayed on an 18" color-monitor with a resolution of 1280 x 1024 pixels and a refresh rate of 60 Hz. A USB computer mouse served as response device. Stimulus presentation and response registration were controlled by the same PC. Auditory stimuli were presented binaurally via headphones.

Procedure

Participants were seated at a distance of approximately 50 cm from the monitor and received written instruction. Their task was to indicate the parity of the target numeral by pressing the corresponding mouse button with the index or middle finger of their right hand. After a central fixation cross a stimulus array was displayed for 165 ms and followed by a blank screen until participants' response (Figure 1). In the DL condition, subjects were instructed to respond before a tone signaled DL expiration after either 375, 450, 550, 650, or 750 ms relative to stimulus onset. In the RS session answers had to be given within a 300 ms interval after sound onset, which was presented 75, 150, 250, 350, or 450 ms post-stimulus. Thus, the maximum interval for timely reactions was identical for both response types (i.e., from 375 to 750 ms).

---- Insert Figure 1 about here ---

After each trial, feedback signaled whether the response was correct ("Korrekt", green color), incorrect ("Fehler!", red color), too slow ("Schneller antworten!", red color), or too fast ("Zu früh!", red color). In Experiments 1 and 2, long display times (i.e., 3000 ms) of feedbacks for too fast or too slow responses were supposed to encourage responses in the required time window (cf., Miller et

al., 2008). In Experiment 3, feedback was always displayed for 750 ms, together with the current balance of performance-contingent points (see below). In all experiments, mean RTs and the proportion of errors and of missed response intervals were presented after each block; in addition, the balance of points was shown in Experiment 3.

Participants worked through the DL and RS procedures in two separate sessions on consecutive days; the sequence of procedures was counterbalanced across subjects. In Experiments 1 and 2, each session comprised three practice blocks and ten main blocks of 64 trials; in Experiment 3, the number of practice blocks was increased to five.

Data processing and analyses

In the RS condition, RTs between 0 and 400 ms after sound onset entered statistical analyses. Thus, the upper bound of response intervals was extended by 100 ms relative to the 300 ms limit during the experiment (cf., Miller et al., 2008). Analogously, responses from stimulus onset up to 100 ms after the DL were considered. Practice trials as well as responses falling below or exceeding these intervals were excluded.

To examine the time course of the flanker effect, RTs and error rates in the DL and RS procedure were analyzed in separate two-way repeated measures ANOVAs with the factors *flanker type* (neutral, incongruent) and *response interval* (375, 450, 550, 650, 750)³. Levels of the latter factor describe the upper time limit of nominally valid responses during the experiment; note, however, that the RS occurred 300 ms earlier.

Further, to estimate differences in performance between the experimental conditions, we computed accuracy-referenced response times (ARRTs) that account for the correlation between RTs and error rates. Specifically, ARRTs adjust accuracies of different conditions to the same level on the

³ We conducted separate analyses for the DL and the RS procedure because their SAT functions covered partly different data ranges, such that the meaningfulness of a direct comparison between the response types is limited. Results of the full three-way ANOVAs on the factors *response type* (DL, RS), *flanker type* (neutral, incongruent), and *response interval* (375, 450, 550, 650, 750) for Experiments 1 to 3 are reported in the Appendix (Tab. A).

basis of their SAT functions, so that otherwise mutually confounded effects of latencies and accuracies fully translate into RT differences. In other words, ARRTs permit the estimation of performance in RTs at equalized accuracy levels (for a detailed description, see Dambacher et al., 2011). Here, ARRTs were computed concurrently for the four experimental conditions (i.e., DL-neutral, DL-incongruent, RS-neutral, RS-incongruent). For each response interval, averaged data were linearly interpolated to the accuracy level of the condition closest to the grand mean (i.e., least squared error). Individual subject data were then corrected accordingly, so that the overall range of accuracies shrunk, but that empirical and referenced data points sat on the same SAT function. Individual ARRTs of neutral and incongruent flankers were submitted to two-way ANOVAs on the within-subjects factors *response type* (DL, RS) and *response interval* (375, 450, 550, 650, 750)⁴. All analyses were conducted in the R-environment for statistical computing (2011). Data were visualized with the R-package ggplot2 (Wickham, 2009).

Experiment 1

Experiment 1 examined differences between DL and RS data, when both methods allow temporal preparation. That is, blocks of constant DL and RS intervals differed only with respect to answering before (DL) or after (RS) the external signal.

We expected reduced performance in RS compared to DL data, if RS detection interferes with the execution of the main task; in this case, SAT functions for the RS procedure should be shifted to the right. Further, if effects of stimulus difficulty are absorbed in waiting periods for RS, we should observe a smaller flanker effect in the RS compared to the DL condition especially at long intervals. In contrast, if RS and DL settings feature equivalent time courses of stimulus processing, corresponding SAT functions should largely overlap.

⁴ Results of the full three-way ANOVA with the factors *response type* (DL, RS), *flanker type* (neutral, incongruent), and *response interval* (375, 450, 550, 650, 750), on ARRTs in Experiments 1 to 3 are listed in the Appendix (Tab. A).

Methods

Twenty-six students (20 female; mean age: 23.2 y; range: 19 – 30 y) received course credit or 16 Euros for participation. DL and RS intervals were announced prior to a new block and remained constant throughout its trials. In either session, the five response intervals were presented twice in pseudo-randomized order, such that each interval occurred once in the first and once in the second half.

Results and Discussion

In total, 1.4% of the responses in the DL condition and 4.5% in the RS condition were outside the required response windows and were discarded (see *General Discussion*). In accordance with previous research (e.g., Carrasco & McElree, 2001; Hübner & Schlösser, 2010; Miller et al., 2008), both the DL and the RS procedure yielded robust SAT functions: RTs as well as accuracies increased with response intervals (Fig. 2A). As expected, though, the data also revealed a number of differences.

---- Insert Figure 2 about here ----

First, while SAT functions for the two response types largely overlapped for incongruent stimuli (especially at the three shortest intervals), RS data in the neutral condition were shifted to the right relative to the DL function (Fig. 2A). This visual impression was confirmed in analyses of ARRTs. For incongruent stimuli, ARRTs did not differ significantly between *response types*, $F(1,25) = 2.24$, $p = .147$ (mean ARRT differences for increasing response intervals: 7, 1, -2, 19, 15 ms; accuracy levels: .675, .798, .835, .889, .897), but neutral stimuli revealed longer ARRTs for RS than for DL data, $F(1,25)$

= 11.41, $p = .002$ (mean ARRT differences: 19, 19, 16, 16, 18 ms for increasing response intervals)⁵. Thus, for neutral stimuli, equivalent accuracy levels were reached faster in the DL than in the RS condition. The lower performance in the RS condition supports the assumption that the additional load of RS detection interfered with the execution of the main task (e.g., Gherri & Eimer, 2011).

Second, the course of the flanker effect differed between response types. In the DL procedure, significant interactions of *flanker type* \times *response interval* indicated that the flanker effect at long compared to short intervals was smaller in accuracies ($p = .040$; Fig 2C) but larger in RTs ($p < .001$; Fig 2B). Thus, it appears that the flanker effect translated from accuracy to RTs at longer intervals. In the RS procedure, in contrast, the flanker effect decreased across intervals for accuracies ($p = .026$; Fig 2C), but there was no difference for RTs ($p = .965$; Fig 2B). Accordingly, Figures 2B-C show a larger RT flanker effect at long compared to short intervals only for the DL condition whereas the effect dropped for accuracies in both the DL and RS condition. Notably, this drop is particularly marked in RS data: While the DL procedure produced a robust flanker effect even at the longest response interval (RT: $F(1,25) = 17.51$, $p < .001$; accuracy: $F(1,25) = 17.57$, $p < .001$), the flanker effect was not reliable at the longest RS lag (RTs: $F(1,25) = 1.73$, $p = .20$; accuracy: $F(1,25) = .01$, $p = .939$). This pattern is compatible with the view that effects of stimulus difficulty can be absorbed when stimulus processing completes prior to the RS. Particularly at long RS lags, rapid decisions (i.e., for easy stimuli) have to be withheld, so that slower processes (i.e., for difficult stimuli) catch up and eventually also complete. Accordingly, waiting for the signal results in similar RTs and accuracies for neutral and incongruent items.

As a third, yet not explicitly predicted, observation RTs and accuracies covered a broader range in RS (RT: 307 ms to 600 ms, $\Delta = 293$ ms; accuracy: 61.1% to 98.6%, $\Delta = 37.5\%$) than in DL sessions (RT: 294 ms to 449 ms, $\Delta = 155$ ms; accuracy: 60.2% to 94.1%, $\Delta = 33.9\%$). Especially at long intervals, responses were given well ahead of DL expiration, such that they hardly revealed

⁵ In addition, the ANOVA yielded increasing ARRTs with *response intervals* (neutral: $F(4,100) = 119.2$, $p < .001$; incongruent: $F(4,100) = 143.3$, $p < .001$). The interaction of *response type* \times *response interval* was not significant (neutral: $F(4,100) = .053$, $p = .995$; incongruent: $F(4,100) = 1.75$, $p = .146$).

behavioral effects (Fig. 2A). Apparently, subjects did not exploit the time for stimulus processing, but chose to respond rapidly even though this came at the cost of accuracy. In contrast, the RS procedure urged subjects to withhold responses until the auditory signal occurred. The higher accuracy level demonstrates that this additional time was at least partly used for stimulus processing. One reason for the relatively fast responses in the DL condition is that subjects may have underestimated the available time especially at long intervals, because temporal uncertainty increases with interval duration (e.g., Niemi & Näätänen, 1981). As a related issue, participants presumably put some effort in avoiding the three-second penalties for DL misses. Accordingly, DL-based SAT functions did not show a clear asymptote and were quite linear, whereas a negatively accelerated RS function featured superior control in the upper data range.

In summary, Experiment 1 shows that RS and DL procedures yield stable SATs, but that they do not reflect fully equivalent processes, even when temporal preparation is granted in both conditions.

---- Insert Table 1 about here ----

Experiment 2

In contrast to blocked RS lags of Experiment 1, most RS studies use randomized and therefore unpredictable intervals to prevent potential confounds with response strategies (Reed, 1976). To provide a comparison of the most established setups, Experiment 2 adopted randomized instead of blocked RS lags together with a blocked DL condition.

In view of the results of Experiment 1, the same procedure-related effects should hold under randomized RS settings. Specifically, given that additional demands of RS detection reduce performance relative to DL data, a similar (or even pronounced) rightward-shift of RS-based SAT functions can be expected when temporal preparation is prevented. Further, if the proportion of completed decisions increases with RS lags (cf., Ratcliff, 2006), a sizable flanker effect at short

intervals should be attenuated at longer RS intervals. In the DL procedure, in contrast, robust influences of stimulus difficulty can be expected across the entire data range.

Methods

Twenty-two students (17 female; mean age: 23.4 y; range: 19 to 35 y) received course credit or 16 Euros for participation. As in Experiment 1, DL intervals were blocked, but RS lags varied randomly between trials.

Results and Discussion

Records from three subjects were excluded because they missed more than 50% of the RS intervals in at least one condition. From the remaining data, a total of 0.8% and 9.0% in DL and RS sessions, respectively, fell out of the required response intervals and were discarded (see *General Discussion*).

As expected, varying RS lags as well as constant DL epochs yielded stable SAT functions. Compatible with the observations of Experiment 1, though, they also revealed differences.

---- Insert Figure 3 about here ---

First, Figure 3A illustrates that RTs were generally slower and accuracies were higher for RS than for DL responses (see also Appendix Tab. A). This pattern corresponds to findings from Miller et al. (2008) who compared randomized with constant RS lags. In contrast to Miller and colleagues, however, the present SAT functions did not lie on top of each other. Instead right-shifted RS curves indicated that equivalent accuracy levels were reached faster in the DL than in the RS condition. This was confirmed in ARRT comparisons between the *response types*: For neutral stimuli ARRTs were significantly shorter for DL than for RS data, $F(1,18) = 12.17$, $p = .002$ (mean ARRT differences for increasing response intervals: 14, 12, 18, 15, 18 ms; accuracy levels: .769, .858, .865, .882, .889). Incongruent stimuli showed a trend in this direction, $F(1,18) = 3.27$, $p = .087$ (mean ARRT differences:

15, 9, 5, 9, 2 ms for increasing response intervals)⁶. Thus, performance was reduced for RS data, suggesting that the detection of temporally unpredictable signals required capacities and impaired performance. Similar to Experiment 1, the shift was more expressed for neutral stimuli.

Second, we observed again different courses of the flanker effects (Tab. 1; Fig. 3B-C). In the DL condition, the interaction of *flanker type x response interval* attested an increasing flanker effect across intervals in RTs ($p = .032$, Fig. 3B), whereas variations were not significant in accuracies ($p = .628$, Fig. 3C). In contrast, longer intervals in the RS condition revealed a decreasing flanker effect in accuracies ($p = .004$, Fig. 3C), but no significant RT differences ($p = .232$, Fig. 3B). Thus, the RT flanker effect increased with intervals in the DL but not in the RS condition. For accuracies, the effect decreased in the RS but not in the DL procedure. As before, the different courses are particularly evident at the longest interval where the effect disappeared in RS (RT: $F(1,18) = .02$, $p = .882$; accuracy: $F(1,18) = .14$, $p = .714$) but not in DL data (RT: $F(1,18) = 36.51$, $p < .001$; accuracy: $F(1,18) = 9.97$, $p = .005$). Hence, the flanker effect translated from accuracies to RTs for longer DL intervals, whereas the decrease across RS lags pointed to an augmenting impact of waiting slacks.

Notably, as in Experiment 1, constant DL periods yielded responses well before the longest time limits. Subjects did not exploit the available time to maximize accuracy but responded earlier (RT range: 301 ms to 435 ms, $\Delta = 134$ ms; accuracy range: 64.6% to 92.7%, $\Delta = 28.1\%$). In contrast, the RS procedure generated long RTs with high accuracies (RT range: 358 ms to 572 ms, $\Delta = 214$ ms; accuracy range: 69.5% to 98.7%, $\Delta = 29.2\%$). The overall range of accuracies, however, was similar across response types and it was smaller than for RS data in Experiment 1 ($\Delta = 37.5\%$). The reason for this shrinkage in Experiment 2 is that especially at short intervals RTs and accuracies for RS responses were higher than in Experiment 1. Apparently, the randomization of RS lags increased the difficulty of promptly responding to rapid signals.

⁶ In addition, this ANOVA with the within-subject factors *response type* and *response interval* yielded increasing ARRTs with *response intervals* (neutral: $F(4,72) = 26.54$, $p < .001$; incongruent: $F(4,72) = 29.45$, $p < .001$). The interaction of *response type x response interval* was not significant (neutral: $F(4,72) = .19$, $p = .942$; incongruent: $F(4,72) = 0.75$, $p = .559$).

As a consequence, SAT functions of the RS and DL procedures in Experiment 2 covered quite different regions. DL data comprised rather short RTs and low accuracies, whereas RS curves captured higher data areas. These partly non-overlapping ranges complicate a direct comparison of SAT courses: the DL condition lacks data points in regions where RS-based curves show an asymptotic behavior and a decreasing flanker effect. Thus, it remains unclear whether effects of stimulus difficulty in the DL condition also survive at high accuracy levels. Alternatively, it is possible that DL data converge with the RS pattern and also do not show a flanker effects at high accuracy levels. Because the DL procedure does not require waiting periods for an external signal, such a result would argue against a slack. We return to this point in Experiment 3.

Experiment 3

The DL procedure has been criticized because prior knowledge of response intervals permits the adjustment of response strategies to current temporal demands (e.g., varying decision thresholds or the proportion of fast guesses). As a consequence, DL-based SATs may not only reflect dynamics of stimulus processing but may also be affected by secondary processes. In contrast, unpredictable RS lags are supposed to largely prevent a systematic adjustment of strategies (Reed, 1976; Wickelgren, 1977). Yet, there are also reports that temporal preparation alone has no major impact on performance (Miller et al., 2008). To add further evidence to this issue, Experiment 3 used performance-contingent incentives that – compared to the flat payment in Experiment 2 – encouraged participants to optimize performance in order to maximize their benefit. Thus, if temporal preparation makes the DL procedure more susceptible to strategic influences, the effect of monetary incentives should be stronger in DL than in RS data.

As another point, partly non-overlapping regions of RTs and accuracies in RS and DL conditions of the previous experiments left open whether the disappearance of stimulus-related effects is a specific consequence of the RS procedure or whether it is a general phenomenon that also translates to DL settings when accuracies reach high levels. To improve the comparability between the response types, monetary incentives were supposed to motivate participants in the DL

condition to exploit the time for stimulus processing and thus to increase accuracies at long intervals. We expected that the previous results generalize to common data regions, i.e., a rightward shift of SAT functions for RS relative to DL data as well as a robust flanker effect at high accuracy levels in the DL but not in the RS condition.

Methods

Twenty-two students (17 female; mean age: 23.3 y; range: 19 to 54 y) received a base payment of 10 EUR and, depending on their performance, earned an additional amount of up to 16 EUR. For performance-contingent payment, each trial was rewarded with 10 points for a correct response in the required interval, while errors, too fast, or too slow responses were not incentivized. In addition, participants received a bonus of 500 points after each block if they reached a pre-specified accuracy level (i.e., DL: 60%, 70%, 80%, 90%, or 95% for the five deadlines; RS: 79% for all blocks). Points were converted into money after the experiment. Written instructions explained that accuracy, and hence the overall profit, increases with the time spent for stimulus processing. Subjects were therefore advised to put effort in meeting time demands, but at the same time to exploit the available interval for accurate decisions. As in Experiment 2, DL intervals were blocked and RS intervals varied randomly between trials.

Results and Discussion

Data from two subjects were discarded because one missed more than 50% of the longest RS intervals, and the other was much older (i.e., 54 y) than the rest of the subject sample. From the remaining data, a total of 1.0% and 9.4% in DL and RS conditions, respectively, missed the required response intervals and were excluded (see *General Discussion*).

SAT functions (Fig. 4A) illustrate that RS data for both neutral and incongruent stimuli were situated to the right of DL curves. Accordingly, ANOVAs on ARRTs yielded slower responses for RS than for DL data for both neutral, $F(1,19) = 15.21$, $p < .001$ (mean ARRT differences for increasing response intervals: 18, 14, 18, 25, 24 ms; accuracy levels: .814, .883, .900, .929, .947) and

incongruent flankers, $F(1,19) = 20.87$, $p < .001$ (mean ARRT differences: 19, 20, 16, 32, 32 ms for increasing response intervals)⁷. Hence, the right shift of SAT functions clearly revealed reduced performance for RS data: Equal accuracy levels were reached faster in the DL condition, presumably because RS detection deducted capacities from the main task.

---- Insert Figure 4 about here ----

Second, and in line with the previous experiments, the flanker effect revealed different courses for the response types (Tab. 1; Fig. 4B-C). In the DL condition, an interaction of *flanker type* \times *response interval* attested an increasing flanker effect for RTs ($p = .011$; Fig. 4B) and a decreasing effect for accuracies ($p = .021$; Fig. 4C). In the RS condition, the flanker effect decreased for accuracies ($p = .016$; Fig. 4C), whereas RT differences between response intervals were not significant ($p = .718$; Fig. 4B). Thus, the flanker effect in accuracies revealed an overall drop for both response types, but in the DL condition, it translated progressively from accuracies into RTs and survived across the entire data range. Even at the longest DL, the flanker effect was reliable in RTs, $F(1,19) = 33.14$, $p < .001$, as well as in accuracies, $F(1,19) = 7.20$, $p = .015$. In contrast, it was not significant at the longest RS interval for RTs, $F(1,19) = 2.44$, $p = .135$, or for accuracies, $F(1,19) = .07$, $p = .791$.

Third, to test whether participants were able to modify their strategies under performance-contingent incentives relative to the flat payment in Experiment 2, individual ARRTs of each response type were submitted to a three-way ANOVA on the between-subjects factor *experiment* (Exp. 2 vs. Exp. 3) and the within-subjects factors *flanker type* (neutral, incongruent) and *response interval* (375, 450, 550, 650, 750); only reliable effects including the factor *experiment* are reported. For RS conditions, ARRTs revealed no reliable main effect of *experiment*, $F < 1$, that is, there was no general incentive-based improvement of performance. However, an interaction of *experiment* \times *flanker type*,

⁷ Further the ANOVA with the within-subject factors *response type* and *response interval* yielded increasing ARRTs with *response intervals* (neutral: $F(4,76) = 135.4$, $p < .001$; incongruent: $F(4,76) = 139.0$, $p < .001$). The interaction of *response type* \times *response interval* was not significant (neutral: $F(4,76) = .657$, $p = .624$; incongruent: $F(4,76) = 1.66$, $p = .169$).

$F(1,37) = 24.47, p < .001$, as well as the three-way interaction, $F(4,148) = 5.19, p < .001$, pointed to shorter ARRTs in Experiment 3 at short RS lags. Post-hoc analyses within flanker types, attested a trend of shorter ARRTs under monetary incentives for neutral, $F(1,37) = 3.95, p = .054$, but not for incongruent stimuli, $F < 1$.

For DL sessions, the main effect of *experiment*, $F(1,37) = 4.73, p = .036$, yielded shorter ARRTs for Experiment 3 than for Experiment 2, indicating that participants successfully improved performance in prospect of monetary rewards. While none of the interactions was reliable (all p 's $> .10$) exploratory post-hoc tests confirmed the pattern within flanker types: Compared to Experiment 2, ARRTs were shorter for incongruent items, $F(1,37) = 5.28, p = .027$, and the effect was marginally significant for neutral stimuli, $F(1,37) = 4.05, p = .052$. Thus, the positive reward effect was observable in both procedures, but it was more robust for DL data, where it held across flanker types and over a broad span of response intervals (see also Fig. 5A). Accordingly, performance-contingent credits⁸ revealed a significant increase in the DL condition, $F(1,42) = 5.56, p = .023$, (average points in Exp. 2: 6969 vs. Exp. 3: 8425), but not in the RS sessions, $F < 1$ (Exp. 2: 6980 vs. Exp. 3: 7194 points). The results therefore indicate that performance is indeed more susceptible to strategic influences (e.g., motivational factors due to payoffs) in DL than in RS procedures (Ratcliff, 2006; Reed, 1976; Wickelgren, 1977).

---- Insert Figure 5 about here ---

Notably, the payoff in Experiment 3 altered the relative data ranges covered by blocked DL and randomized RS modes. While both methods yielded comparable maximum accuracies at long

⁸ Note that the flat payment in Experiment 2 did not include performance-contingent points. For the cross-experimental comparison, we calculated virtual points post-hoc according to the same criteria as in Experiment 3 (i.e., +10 points for correct and timely responses and 0 points for errors or mistimed responses; in addition, a bonus of 500 points was added in blocks in which performance exceeded a pre-specified threshold; cf., Methods of Exp. 3). The two-way ANOVA on the between-subject factor *experiment* (Exp. 2, Exp. 3) and the within-subject factor *response type* (DL, RS) yielded a trend of *response type*, $F(1,42) = 3.69, p = .062$, as well as a marginally significant interaction of *experiment* \times *response type*, $F(1,42) = 3.82, p = .057$. The main effect of *experiment* was not significant, $F(1,42) = 1.66, p = .204$.

intervals, responses at short intervals were faster and accuracies were lower for DL (RT: 329 ms to 459 ms, $\Delta = 130$ ms; Accuracy: 69.2% to 96.6%, $\Delta = 27.4\%$) than for RS data (RT: 366 ms to 552 ms, $\Delta = 186$ ms; Accuracy: 75.5% to 97.6%, $\Delta = 22.1\%$). Thus, accuracies spanned a wider range in the DL than in the RS condition. Apparently, the prospect of monetary rewards motivated participants to exploit the available time at long DL intervals, so that accuracies increased relative to Experiment 2. In contrast, RS data virtually revealed the same maximum accuracy level as before. Advantageously, the coverage of similar accuracy values increased the comparability between the response modes and hence strengthened the validity of our conclusions.

General discussion

Deadlines (DL) and response signals (RS) are two commonly used techniques to investigate the fundamental relation between RTs and error rates, i.e., the speed-accuracy tradeoff (SAT). Yet the two procedures reveal a number of differences whose effects are not fully established: it is unclear whether data from the two procedures reflect equivalent processes or whether one procedure is generally preferable. It has sometimes been argued that the RS procedure is more suitable to track the time course of information processing but such evaluations remained rather superficial. Here, we aimed at providing an empirical basis and scrutinized the impact of methodological differences between the DL and RS technique in a flanker task.

As one methodological difference, responses in RS conditions have to be given immediately after an explicit signal, whereas the DL procedure permits responses at any time between stimulus onset and DL expiration. The RS procedure therefore exerts advanced control of RTs, but the detection of and the waiting for an imperative signal can interfere with performance of the main task. As another difference, RS intervals usually vary unpredictably between trials, whereas prior information about DL periods may open the door for strategic influences. Compatible with these hypotheses, the data revealed three major outcomes.

Response signal detection

First, longer RTs in the RS compared to the DL procedure did not translate into an equivalent increase in accuracy. That is, a right-shift of RS-based SAT functions pointed to reduced performance relative to DL data. This was especially evident in Experiments 2 and 3, where unpredictable RS intervals generally slowed responses. The pattern is compatible with the notion that dual-task situations can interfere with behavior. In particular, continuous monitoring for the occurrence of an auditory RS detracts capacities from the main task and impairs the efficiency of stimulus processing (cf., Gherri & Eimer, 2011; Pashler, 1994). The results therefore suggest that the RS procedure does not yield an unbiased estimate of perceptual evidence accumulation but involves processes that are specifically related to the imperative RS. In contrast, the absence of monitoring demands yielded superior performance in DL conditions throughout all experiments.

Waiting for the response signal

Second, task-relevant effects were modulated by response types. We were able to disclose this finding using a flanker task that varies the difficulty of perceptual categorization. In particular, a robust flanker effect showed up across the entire data range in the DL procedure. In contrast, the effect decreased over time in the RS condition where it eventually vanished at the longest intervals. This result confirms predictions derived from sequential sampling models (Hübner et al., 2010; Ratcliff & Rouder, 1998; Ratcliff & Smith, 2004; Usher & McClelland, 2001): responses in DL conditions are triggered as soon as evidence accumulation hits a decision criterion, so that different rates for easy and difficult items translate into stimulus-related effects even at long intervals. In contrast, participants in RS conditions have to withhold their answers when stimulus processing completes prior to the RS. This waiting for the signal generates a slack that attenuates or even absorbs influences of stimulus difficulty, such as the flanker effect (cf., Ratcliff, 2006). Thus, the observation that stimulus-related effects can be neutralized in RS but survive in DL conditions should be taken into account for the interpretation of SATs and the selection of the appropriate procedure.

Response strategies

Third, an influence of incentive-induced response strategies showed up particularly in the DL procedure. At equivalent accuracy levels, latencies for DL responses were shorter under a performance-contingent (Exp. 3) than under a flat payment (Exp. 2) (cf., Dambacher et al., 2011; Hübner & Schlösser, 2010). Likewise, performance-related points suggested a stronger influence of monetary rewards in the DL than in the RS condition (see also Fig. 5A). Although the motivation to optimize performance was presumably comparably high in both conditions, the realization of adequate response strategies was apparently more successful in the DL procedure. The results therefore support the notion that temporal preparation and internally triggered responses make DL data more susceptible to strategies, a characteristic that has been considered as disadvantage (Ratcliff, 2006; Reed, 1976; Wickelgren, 1977). At the same time, however, our data show that DL settings offer the possibility to encourage participants to optimize behavior. This property may be advantageous for research dealing with maximum performance in perceptual decisions.

Other differences

As an additional observation across experiments, RTs were generally longer and accuracies higher for RS than for DL data (Fig. 5A). Clearly, slow responses are expected in RS procedures, because the method demands responses after a pre-defined epoch. Accordingly, RS-based SAT functions yielded a steep increase in accuracy over short lags and an asymptotic behavior at longer intervals (McElree & Carrasco, 1999; Miller et al., 2008). In contrast, responses in the DL conditions of Experiments 1 and 2 were given well ahead of long intervals. Maximum accuracies therefore stayed at a relatively low level and did not approach an asymptote. In fact, DL-based SAT functions featured a linear rather than a negatively accelerated exponential function that is often used to fit RS data (cf., Footnote 2). Apparently, the possibility to respond any time between stimulus onset and the maximum interval in DL conditions encouraged participants to respond rapidly. One likely reason is a strong motivation to avoid DL misses because they entailed time penalties in Experiments 1 and 2. This was supported in Experiment 3, where performance-contingent monetary rewards fostered a

better exploitation of time in favor of higher accuracy. Here, the overall span of accuracies was even broader for DL than for RS conditions. Thus, one advantage of the RS procedure, namely performance control over a wide range of data points, was caught up by the DL procedure when participants were highly motivated to maximize accuracy. At the same time, considering brief response intervals, RTs were shorter and accuracies were lower in DL than in RS data of Experiments 2 and 3. Participants were more successful in responding very rapidly, when temporal requirements were known and when response execution was independent from an external signal. Tracking the course of evidence accumulation over the earliest intervals therefore appears to be more feasible with DL than with RS settings. This is important because especially short RTs display enormous variations in accuracy and are therefore indicative for rapid accumulation of stimulus information; in contrast, even large RT differences in the upper data range hardly cause modulations in error rates (Pachella, 1974; Wickelgren, 1977).

Finally, it seems noteworthy that in all experiments the proportion of responses outside the required intervals was higher in RS (Exp. 1: 4.50%, Exp. 2: 8.99%, Exp. 3: 9.42%) than in DL sessions (Exp. 1: 1.46%, Exp. 2: 0.82%, Exp. 3: 1.03%)⁹. This may not be too surprising because the interval for valid responses is smaller for RS than for DL conditions. Yet, Figure 5B also shows that ratios of excluded data were imbalanced across response intervals. That is, a large proportion of too slow responses at short RS lags and of too fast responses at long intervals pointed to difficulties not only to answer very rapidly, but also to withhold responses when the RS appears late. In comparison, fewer responses in the DL procedure missed time demands at short and hardly ever at long intervals. Therefore, the exclusion of mistimed responses poses a problem especially for RS data because it leads to strong and unequal truncations of underlying distributions. It should be noted, though, that other RS studies indicate that the number of mistimed responses can be reduced by excessive training; nevertheless, unequal truncations are likely to survive even with substantially more

⁹ An ANOVA with *response type* (DL, RS) as within- and *experiment* (1,2,3) as between-subject factors revealed significant effects of *experiment*, $F(2,62) = 7.5$, $p = .001$, of *response type*, $F(1,62) = 238.8$, $p < .001$, as well as of their interaction, $F(2,62) = 20.6$, $p < .001$.

practice. As an alternative, the DL procedure offers a viable method to reduce the imbalance of excluded data, at least with moderately experienced subjects.

Summary and conclusions

In summary, both the RS and the DL procedures confirmed their general usefulness for tracking dynamics of information processing, but our data also show that methodological differences affect performance, compatible with the notion that the two procedures involve partly different processes. Indeed, recent evidence suggests that the two methods can evoke qualitatively different effects on a neural level (Blumen et al., 2011; Forstmann et al., 2008). Future neuroimaging studies may therefore extend our behavioral approach and systematically investigate the neural underpinnings in DL and RS tasks. This is essential because the understanding of empirical data requires the comprehension of method-based influences; after all, such effects themselves can inform about the nature of the SAT.

Further, knowledge of procedure-related effects can help to advance formal models of decision making. Although the present work had an empirical focus, our data may be suited for model simulations. Good candidates are diffusion-type approaches that are able to account for data from conflict paradigms, such as the flanker task (Hübner et al., 2010; White et al., 2011). We refrained from presenting simulations here for the sake of simplicity and because our hypotheses were testable on an unbiased empirical basis. In upcoming studies, though, parameters from model fits can give quantitative estimates of the underlying processes (e.g., drift rates, threshold separation, proportion of completed decisions prior to RSs).

Apparently, more research is necessary for a comprehensive picture of the mechanisms of SATs under different conditions. The present data can support the next steps as they unveil procedure-based effects that may be preferable or detrimental for distinct experimental settings. In particular, the results suggest that the RS procedure is an excellent tool to investigate SATs when emphasis is put on

- precise control of RTs. Especially the assessment of responses with long latencies appears to be feasible with external triggers of RS settings. In contrast, participants in the DL procedure often respond markedly before the expiration of long intervals, at least in the absence of additional incentives.
- minimal influences of response strategies. Specifically, varying RS lags do not permit temporal preparation to trial-specific temporal demands, so that participants are virtually in the same state in all trials. Accordingly, the RS procedure shows limited susceptibility to motivating factors, such as payoffs. Hence, responses provide a solid estimate of the time course of evidence accumulation, although it should be noted that RS detection itself affects performance.

Moreover, and against previous criticism, our data also revealed characteristics that make the DL procedure an eligible technique for a number of experimental questions. This is the case when research focuses on

- response adaptation under different task demands. Knowledge of time demands and the possibility to initiate responses any time before DL expiration greatly permits the selection of adequate strategies. For instance, motivational factors induced by different payoffs can be reliably captured.
- maximal performance. The DL procedure does not impose explicit additional demands, so that processing efficiency is generally superior to the dual-task situation in RS settings. Again, manipulations of other factors (e.g., payoffs) can even foster this characteristic.
- effects of task difficulty. Those are preserved in the DL procedure over a broad data range and even at long intervals, where task-related effects in RS settings can be attenuated or absorbed in waiting slacks.
- minimization of mistimed responses. Especially when subjects have little practice, internal response triggering and knowledge of temporal demands in DL settings facilitate timely reactions for informative short as well as for long response intervals. In comparison,

unpredictable RS lags aggravate the problem of mistimed data, as they may be distributed unequally across intervals.

Taking into account the respective strengths of RS and DL methods may support the selection of the appropriate response procedure for distinct research questions and enhance the validity of SAT-based conclusions.

References

- Blumen, H. M., Gazes, Y., Habeck, C., Kumar, A., Steffener, J., Rakitin, B. C., & Stern, Y. (2011). Neural networks associated with the speed-accuracy tradeoff: Evidence from the response signal method. *Behavioral Brain Research*, 224(2), 397-402. doi: 10.1016/j.bbr.2011.06.004
- Bogacz, R., Wagenmakers, E. J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neurosciences*, 33(1), 10-16. doi: DOI 10.1016/j.tins.2009.09.002
- Carrasco, M., & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Sciences of the United States of America*, 98(9), 5363-5367.
- Dambacher, M., Hübner, R., & Schlösser, J. (2011). Monetary incentives in speeded perceptual decision: Effects of penalizing errors versus slow responses. *Frontiers in Psychology*, 2, 248. doi: 10.3389/fpsyg.2011.00248
- Diederich, A., & Busemeyer, J. R. (2006). Modeling the effects of payoff on response bias in a perceptual discrimination task: Bound-change, drift-rate-change, or two-stage-processing hypothesis. *Perception and Psychophysics*, 68(2), 194-207.
- Doshier, B. A. (1976). Retrieval of sentences from memory: A speed-accuracy study. *Cognitive Psychology*, 8(3), 291-310.
- Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics*, 16(1), 143-149.
- Eriksen, C. W., & St. James, J. D. (1986). Visual attention within and around the field of focal attention: A zoom lens model. *Perception and Psychophysics*, 40(4), 225-240.
- Forstmann, B. U., Anwander, A., Schafer, A., Neumann, J., Brown, S., Wagenmakers, E. J., . . . Turner, R. (2010). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proceedings of the National Academy of Sciences of the United States of America*, 107(36), 15916-15920. doi: 10.1073/pnas.1004932107

- Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, 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.
- Gherri, E., & Eimer, M. (2011). Active listening impairs visual perception and selectivity: An ERP study of auditory dual-task costs on visual attention. *Journal of Cognitive Neuroscience*, *23*(4), 832-844.
- Göthe, K., & Oberauer, K. (2008). The integration of familiarity and recollection information in short-term recognition: Modeling speed-accuracy trade-off functions. *Psychological Research*, *72*(3), 289-303. doi: 10.1007/s00426-007-0111-9
- Hübner, R., & Schlösser, J. (2010). Monetary reward increases attentional effort in the flanker task. *Psychonomic Bulletin and Review*, *17*(6), 821-826. doi: 10.3758/pbr.17.6821
- Hübner, R., Steinhauser, M., & Lehle, C. (2010). A dual-stage two-phase model of selective attention. *Psychological Review*, *117*(3), 759-784. doi: 10.1037/a0019471
- Hübner, R., & Töbel, L. (2012). Does attentional selectivity in the flanker task improve discretely or gradually? *Frontiers in Psychology*, *3*. doi: 10.3389/fpsyg.2012.00434
- Ivanoff, J., Branning, P., & Marois, R. (2008). fMRI evidence for a dual process account of the speed-accuracy tradeoff in decision-making. *PloS One*, *3*(7), e2635.
- Kleinsorge, T. (2001). The time course of effort mobilization and strategic adjustments of response criteria. *Psychological Research*, *65*(3), 216-223.
- Kumar, A., Rakitin, B. C., Nambisan, R., Habeck, C., & Stern, Y. (2008). The response-signal method reveals age-related changes in object working memory. *Psychology and Aging*, *23*(2), 315-329. doi: 10.1037/0882-7974.23.2.315
- Link, S. W., & Tindall, A. D. (1971). Speed and accuracy in comparative judgments of line length. *Perception and Psychophysics*, *9*, 284-288.

- McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), 1517-1539.
- Miller, J., Sproesser, G., & Ulrich, R. (2008). Constant versus variable response signal delays in speed-accuracy trade-offs: Effects of advance preparation for processing time. *Perception and Psychophysics*, 70(5), 878-886. doi: 10.3758/Pp.70.5.878
- Niemi, P., & Näätänen, R. (1981). Foreperiod and simple reaction time. *Psychological Bulletin*, 89(1), 133-162.
- Pachella, R. G. (1974). The interpretation of reaction time in information processing research. In B. H. Kantowitz (Ed.), *Human information processing: Tutorials in performance and cognition* (pp. 41-82). Hillsdale, NJ: Erlbaum.
- Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin*, 116(2), 220-244.
- Pashler, H., & Johnston, J. C. (1998). Attentional limitations in dual-task performance. In H. Pashler (Ed.), *Attention* (pp. 155–189): Psychology Press.
- R Development Core Team. (2011). R: A language and environment for statistical computing. (3-900051-07-0). from R Foundation for Statistical Computing <http://www.R-project.org>
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, 85(2), 59-108.
- Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling accumulation of partial information. *Psychological Review*, 95(2), 238-255.
- Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology*, 53(3), 195-237. doi: 10.1016/j.cogpsych.2005.10.002
- Ratcliff, R. (2008). Modeling aging effects on two-choice tasks: Response signal and response time data. *Psychology and Aging*, 23(4), 900-916. doi: 10.1037/a0013930
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, 9(5), 347-356.

- Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), 127-140.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, 111(2), 333-367. doi: 10.1037/0033-295x.111.2.333
- Reed, A. V. (1973). Speed-accuracy trade-off in recognition memory. *Science*, 181(4099), 574-576.
- Reed, A. V. (1976). List length and time course of recognition in immediate memory. *Memory and Cognition*, 4, 16-30.
- Rinkenauer, G., Osman, A., Ulrich, R., Müller-Gethmann, H., & Mattes, S. (2004). On the locus of speed-accuracy trade-off in reaction time: Inferences from the lateralized readiness potential. *Journal of Experimental Psychology: General*, 133, 261-282.
- Ruthruff, E., Pashler, H. E., & Klaassen, A. (2001). Processing bottlenecks in dual-task performance: Structural limitation or strategic postponement? *Psychonomic Bulletin and Review*, 8(1), 73-80.
- Schouten, J. F., & Bekker, J. A. M. (1967). Reaction time and accuracy. *Acta Psychologica*, 27, 143-153.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3), 550-592. doi: 10.1037//0033-295X.108.3.550
- van Veen, V., Krug, M. K., & Carter, C. S. (2008). The neural and computational basis of controlled speed-accuracy tradeoff during task performance. *Journal of Cognitive Neuroscience*, 20(11), 1952-1965.
- Wenzlaff, H., Bauer, M., Maess, B., & Heekeren, H. R. (2011). Neural characterization of the speed-accuracy tradeoff in a perceptual decision-making task. *Journal of Neuroscience*, 31(4), 1254-1266. doi: 10.1523/Jneurosci.4000-10.2011

- White, C. N., Ratcliff, R., & Starns, J. J. (2011). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*, 63(4), 210-238. doi: 10.1016/j.cogpsych.2011.08.001
- Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information-processing dynamics. *Acta Psychologica*, 41(1), 67-85.
- Wickham, H. (2009). *ggplot2: Elegant graphics for data analysis*. New York: Springer.

Author notes

This research was supported by the Center for Psychoeconomics and the Zukunftskolleg at the Universität Konstanz.

Tables

Table 1. Analyses of response times and accuracies in Experiments 1 to 3. Separate ANOVAs for the DL and the RS procedures comprised the within-subject factors *flanker type* (neutral, incongruent) and *response interval* (375, 450, 550, 650, 750 ms).

Exp.	df	Deadline				Response Signal			
		Response Time		Accuracy		Response Time		Accuracy	
		F	p	F	p	F	P	F	p
Exp.1									
Interval	4, 100	161.6	< .001	149.9	< .001	488.3	< .001	183.3	< .001
Flanker	1, 25	42.0	< .001	118.6	< .001	8.8	.007	43.9	< .001
Interval x Flanker	4, 100	6.3	< .001	2.6	.040	.1	.965	2.9	.026
Exp.2									
Interval	4, 72	90.6	< .001	63.6	< .001	526.1	< .001	98.6	< .001
Flanker	1, 18	31.6	< .001	33.8	< .001	12.7	.002	37.6	< .001
Interval x Flanker	4, 72	2.8	.032	.7	.628	1.4	.232	4.3	.004
Exp.3									
Interval	4, 76	131.1	< .001	93.1	< .001	448.6	< .001	75.9	< .001
Flanker	1, 19	67.9	< .001	29.3	< .001	24.8	< .001	32.2	< .001
Interval x Flanker	4, 76	3.5	.011	3.1	.021	.5	.718	3.3	.016

Figures

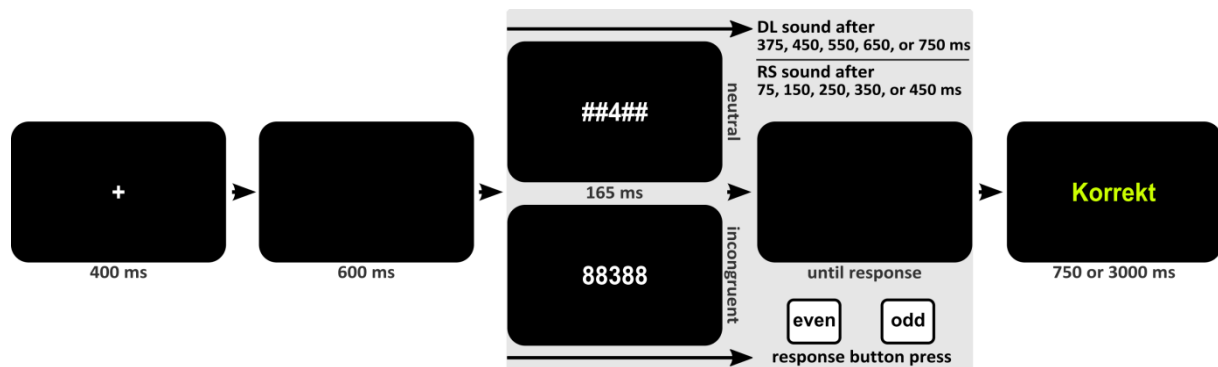


Figure 1. General Procedure. After fixation cross and blank screen, a neutral or incongruent stimulus was displayed at screen center. Participants were instructed to categorize the parity of the central target numeral by pressing the corresponding button either before (DL procedure) or within 300 ms after (RS procedure) a tone occurring with a delay relative to stimulus onset. Feedback was given after each trial.

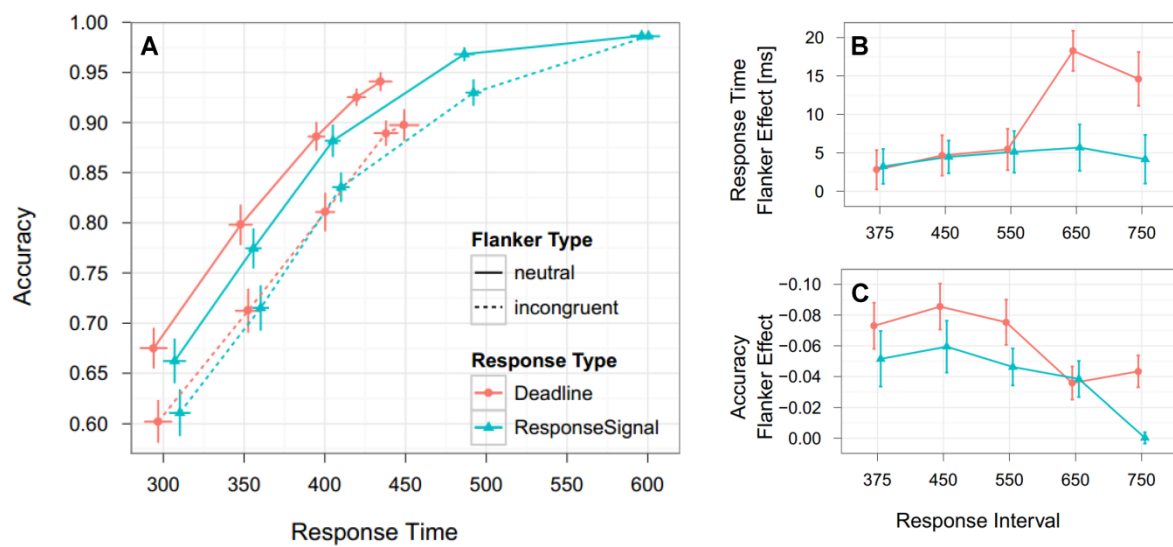


Figure 2. Mean response times and accuracies in Experiment 1; error bars reflect standard errors of means. (A) Speed-accuracy tradeoff (SAT) functions for blocked deadlines and blocked response signals. (B-C) Flanker effect (incongruent minus neutral) across response intervals for (B) response times and (C) accuracies.

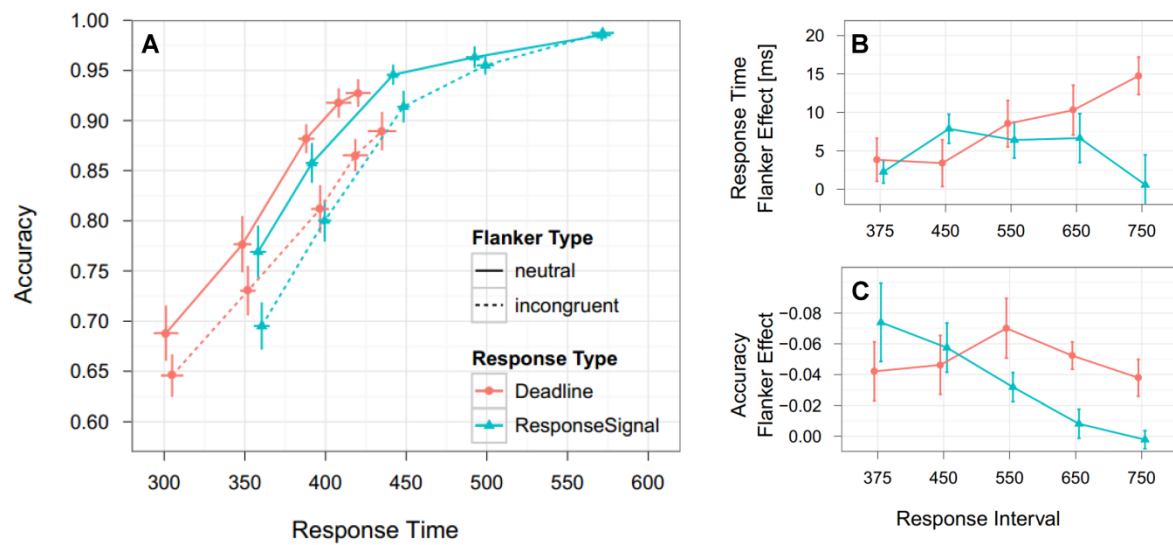


Figure 3. Mean response times and accuracies in Experiment 2; error bars reflect standard errors of means. (A) Speed-accuracy tradeoff (SAT) functions for blocked deadlines and randomized response signals. (B-C) Flanker effect (incongruent minus neutral) across response intervals for (B) response times and (C) accuracies.

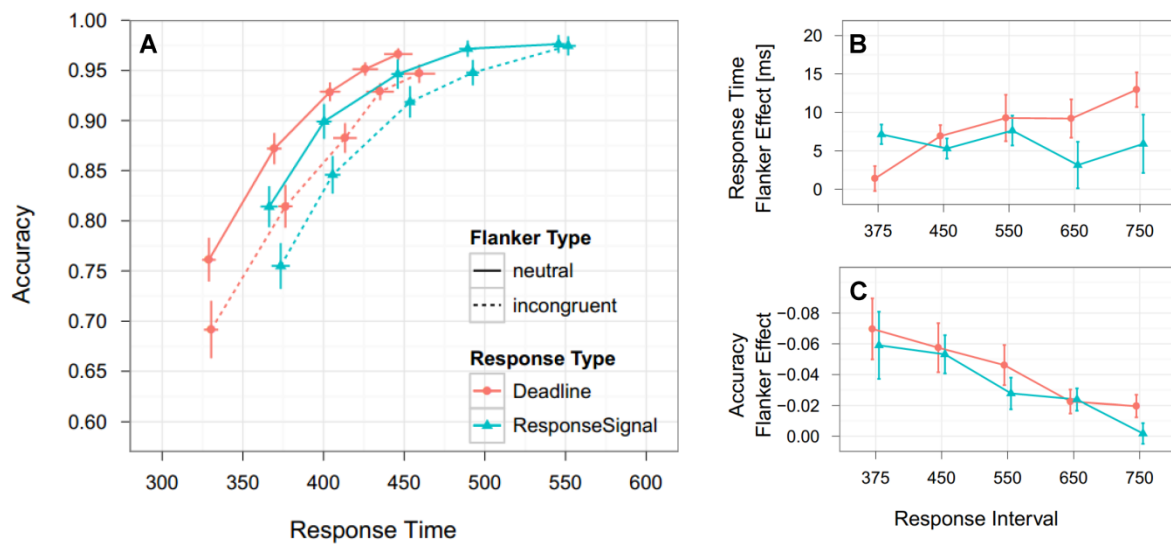


Figure 4. Mean response times and accuracies in Experiment 3; error bars reflect standard errors of means. (A) Speed-accuracy tradeoff (SAT) functions for blocked deadlines and randomized response signals under performance-contingent payment. (B-C) Flanker effect (incongruent minus neutral) across response intervals for (B) response times and (C) accuracies.

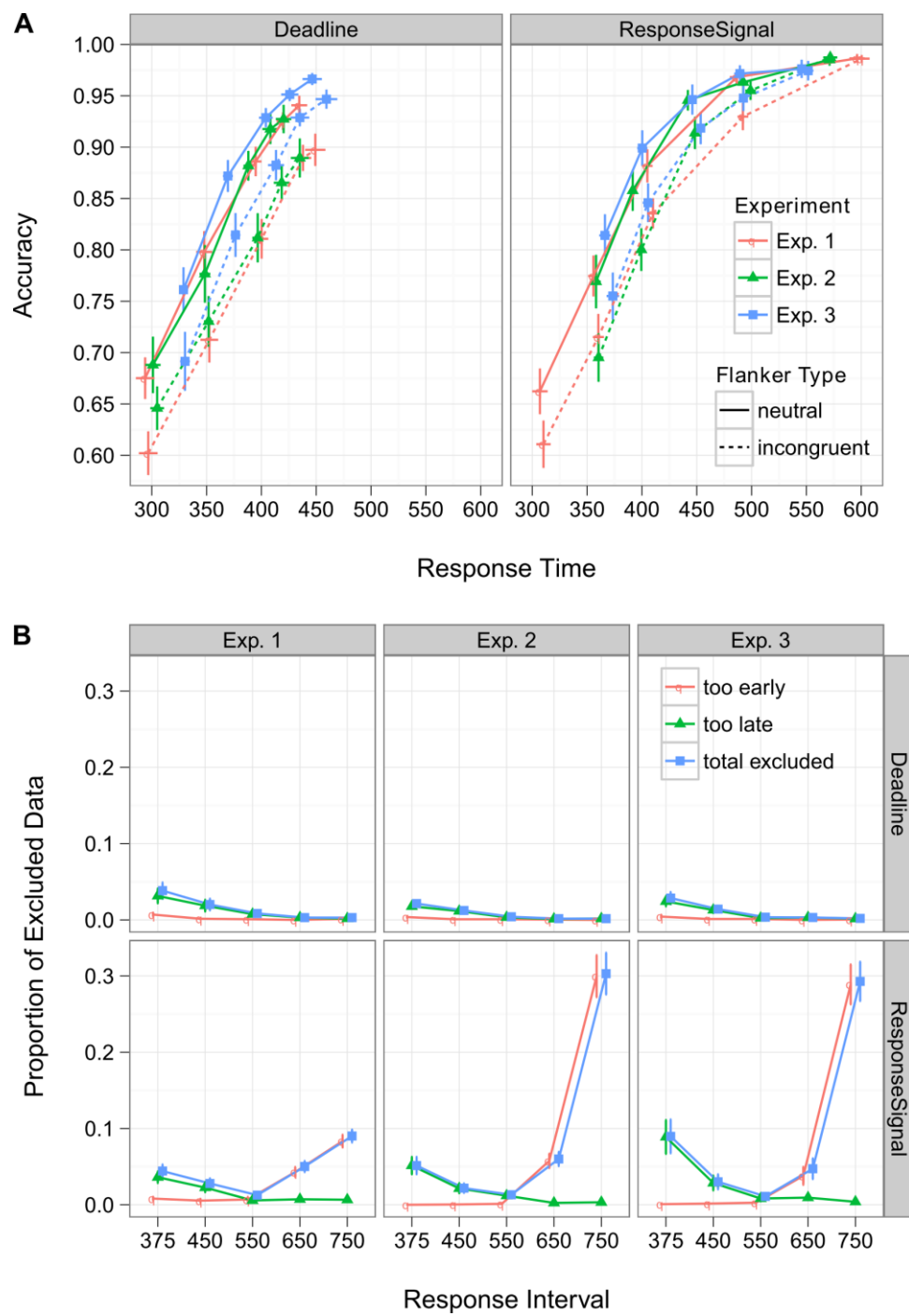


Figure 5. Comparison between the experiments; error bars reflect standard errors of means. (A)

Speed-accuracy tradeoff functions for the DL and the RS procedures across Experiments 1 to 3. (B) Proportion of excluded data due to response mistiming for DL and RS procedures across Experiment 1 to 3.

Appendix

Table A. Results of three-way repeated measures ANOVAs with the within-subject factors *response interval* (375, 450, 550, 650, 750 ms), *flanker type* (neutral, incongruent) and *response type* (DL, RS) for the dependent measures response times, accuracies, and accuracy-referenced response times (ARRTs).

<i>Exp.1</i>	df	Response Time		Accuracy		ARRTs	
		<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Interval	4, 100	491.9	< .001	240.6	< .001	139.0	< .001
Flanker	1, 25	36.1	< .001	101.3	< .001	1169.0	< .001
Response Type	1, 25	90.5	< .001	6.8	.016	6.0	.022
Interval x Flanker	4, 100	3.7	.007	4.5	.002	7.0	< .001
Interval x Response Type	4, 100	104.2	< .001	8.2	< .001	0.5	.731
Flanker x Reponse Type	1, 25	6.1	.021	16.2	< .001	24.6	< .001
Interval x Flanker x Response Type	4, 100	3.3	.014	0.8	.522	7.0	< .001
<i>Exp.2</i>							
	df	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Interval	4, 72	299.0	< .001	119.7	< .001	29.9	< .001
Flanker	1, 18	30.6	< .001	49.9	< .001	820.4	< .001
Response Type	1, 18	317.6	< .001	45.4	< .001	7.3	.015
Interval x Flanker	4, 72	1.2	.339	2.1	.090	1.3	.266
Interval x Response Type	4, 72	65.2	< .001	0.3	.891	0.1	.976
Flanker x Reponse Type	1, 18	5.0	.038	3.8	.068	24.3	< .001
Interval x Flanker x Response Type	4, 72	3.4	.014	2.9	.027	3.1	.020
<i>Exp.3</i>							
	df	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>	<i>F</i>	<i>p</i>
Interval	4, 76	531.6	< .001	131.7	< .001	150.8	< .001
Flanker	1, 19	82.1	< .001	42.0	< .001	2208.0	< .001
Response Type	1, 19	110.0	< .001	9.4	.006	18.3	< .001
Interval x Flanker	4, 76	1.5	.199	5.1	.001	8.3	< .001
Interval x Response Type	4, 76	21.3	< .001	2.1	.092	1.1	.343
Flanker x Reponse Type	1, 19	2.0	.178	1.8	.201	6.1	.023
Interval x Flanker x Response Type	4, 76	2.3	.066	0.2	.888	1.4	.246