Cognitive Psychology 110 (2019) 16–29

Contents lists available at ScienceDirect

Cognitive Psychology

journal homepage: www.elsevier.com/locate/cogpsych

Caution in decision-making under time pressure is mediated by timing ability

Steven Miletić*, Leendert van Maanen

Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129-B, 1018 WT Amsterdam, the Netherlands

ARTICLE INFO

Keywords:
Decision-making
Interval timing
Response caution
Urgency
Time pressure

ABSTRACT

The time available to inform decisions is often limited, for example because of a response deadline. In such circumstances, accurate knowledge of the amount of time available for a decision is crucial for optimal choice behavior. However, the relation between temporal cognition and decision-making under time pressure is poorly understood. Here, we test how the precision of the internal representation of time affects choice behavior when decision time is limited by a deadline. We show that participants with a precise internal representation of time respond more cautiously in decision-making. Furthermore, we provide an empirical test of theoretical accounts of decision-making that argue that it is optimal to commit to a decision based on increasingly less evidence as the deadline approaches (so-called 'collapsing decision bounds'). These theories entail that the speed of collapse of the decision bound should depend on the precision of the internal representation of the deadline. However, although we find evidence that participants collapse decision bounds, we found no relation between the amount of collapse and the internal representation of time.

1. Introduction

Real-life decision-making often comes with a cost for long deliberations: Typically, there is only a certain amount of time available to inform a decision, after which not choosing at all is the most expensive option. Imagine, for example, driving a car on a busy highway. When you notice that the brake lights of the car in front of you have lit up, you need to quickly decide to either brake yourself, or turn the wheel to switch lanes and evade the car. Which choice is safest depends on many aspects, such as whether there is a car right behind you or in the other lanes, and an informed decision should involve considering all the available information. However, waiting to decide for too long imposes the high cost of hitting the car in front of you (Xue, Markkula, Yan, & Merat, 2018).

Time pressure in the form of such a response deadline has a critical effect on decision-making: Typically, short response deadlines decrease both response time and response accuracy in decision-making tasks (e.g., Busemeyer, 1985; Link, 1971). Decision-making models (e.g., Brown & Heathcote, 2008; Ratcliff, 1978; Ratcliff, Smith, Brown, & McKoon, 2016) shed light on the possible relationships between response time and accuracy. These models propose that people collect evidence for each choice alternative until a threshold level of evidence is reached for one choice option, at which point people commit to a decision and initiate a motor response. This modelling framework accounts for many empirical phenomena in the decision-making literature. Specifically, it has been shown that when participants are asked to respond quickly, they commit to a choice on the basis of less evidence by decreasing the decision threshold, thereby sacrificing accuracy (Boehm, Van Maanen, Forstmann, & Van Rijn, 2014; Mulder et al., 2013; Rae, Heathcote, Donkin, Averell, & Brown, 2014; Van Maanen et al., 2011).

* Corresponding author.
E-mail address: s.miletic@uva.nl (S. Miletić).

https://doi.org/10.1016/j.cogpsych.2019.01.002
Accepted 23 January 2019
Available online 05 February 2019

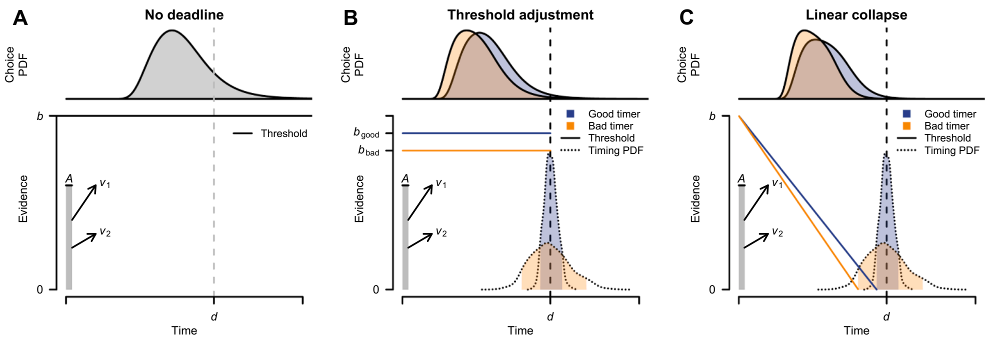

Fig. 1. Illustration of possible adjustments in decision-making behavior due to a response deadline, and their potential relation to time reproduction ability. All panels assume that the choice process can be approximated by the linear, ballistic accumulator (LBA) model. (A) Illustration of a decision-making process without a deadline. The arrows illustrate two accumulators that race towards a common threshold. Four LBA parameters are shown: threshold b, mean drift rates v_1 and v_2, and upper bound of start point distribution A. The probability density function (PDF) illustrates the response time (RT) distribution. In the absence of time pressure, the tail of the RT distribution includes responses that are slower than a potential deadline d (grey dashed line). (B) A possible strategy to respond before the deadline (d, dashed black line) is a reduction in threshold b. The panel shows thresholds for two decision-makers with different timing abilities. One decision-maker estimates the deadline with high precision ('good timer'; blue) relative to the less precise estimates of the deadline by the other decision-maker ('bad timer'; orange). In order to ensure responding before the deadline, the bad timer needs to lower its threshold more. (C) A second possible strategy to respond before the deadline involves collapsing bounds. Because of the precision of the deadline estimate, good timers require less collapse than bad timers. The possible mechanisms in Panels B and C both reduce overall decision times but lead to different shapes of decision time distributions. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

One potential strategy therefore to accommodate decisions before a deadline may be to decrease the decision threshold as well (Fig. 1B). Participants may thus commit to a decision on the basis of less evidence when a deadline is present than they would without such a deadline. Alternatively, participants may employ *collapsing* decision bounds: participants may dynamically adjust the amount of evidence required to commit to a decision over time during the course of deciding (Fig. 1C). In other words, as time passes and the deadline approaches, participants may grow impatient and may be willing to commit to increasingly less evidence.

Importantly, decision-making strategy adjustments due to a short deadline require participants to internally represent the deadline: It is crucial to know how much time is available to inform a decision, before (dynamically or fixed) thresholds can be set in a way that maximizes accuracy without risking missing the deadline. However, despite this key role of temporal cognition, little is known about how temporal cognition affects choice behavior under deadlines (but see Karşılar, Simen, Papadakis, & Balcı, 2014). In the current study, we aim to understand to what extent the precision of the internal representation of time affects decision-making.

A specific hypothesis about the effects of temporal cognition on decision-making under deadlines was brought forward by theoretical work on optimality of decision-making (Frazier & Yu, 2008; Karşılar et al., 2014). Their work extended the literature that argues that collapsing decision bounds are optimal with respect to reward rate under various conditions (Boehm, Hawkins, Brown, Van Rijn, & Wagenmakers, 2016; Deneve, 2012; Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012; Frazier & Yu, 2008; Malhotra, Leslie, Ludwig, & Bogacz, 2017a, 2017b; Standage, You, Wang, & Dorris, 2011; Thura, Beauregard-Racine, Fradet, & Cisek, 2012). That is, under such conditions, if participants make a sequence of decisions and earn points for each correct decision, employing collapsing bounds results in maximal points. Of interest for the current experiments, Frazier and Yu (2008) argued that, in order to maximize reward rate, the speed of collapse should depend on the precision of an individual's representation of the deadline: Having a precise representation of the deadline allows one to sample longer without risking reaching the deadline due to an overestimation of the amount of time available. Put more intuitively, if you are uncertain about where the deadline is, it is better to collapse decision bounds earlier to be safe (Fig. 1C).

Frazier and Yu (2008) elegantly showed that the optimal shape of a collapsing boundary – that is, the shape that maximizes rewards over a series of choices – is curvilinear (see also Deneve, 2012; Drugowitsch et al., 2012; Drugowitsch, Moreno-Bote, & Pouget, 2014; Malhotra, Leslie, Ludwig, & Bogacz, 2017b). Curvilinear functions require multiple additional parameters to be estimated that show significant correlations with other parameters, and cannot be estimated reliably (Voskuilen, Ratcliff, & Smith, 2016). Therefore, in order to be able to interpret parameter estimates, we use linear collapse as an approximation in the current paper. In the linear, ballistic accumulator model that we apply (LBA, Brown & Heathcote, 2008), a linear collapse is mathematically equivalent to an additive urgency signal (see the Supplementary Materials for a proof). This allows us to model linear collapse by incorporating an additive urgency signal in the model.

Here, we test the hypothesis that the precision of the internal representation of time affects decision-making behavior. In Experiment 1 we test whether timing ability, as measured in a separate task, determines choice behavior in decision-making. As detailed below, the results are supportive of our hypothesis. In Experiment 2 we further refine the results, and test to what extent the relationship between timing and decision-making depends on the perceived difficulty of the decision-making task, and to what extent timing ability predicts the amount of collapse of decision bounds.

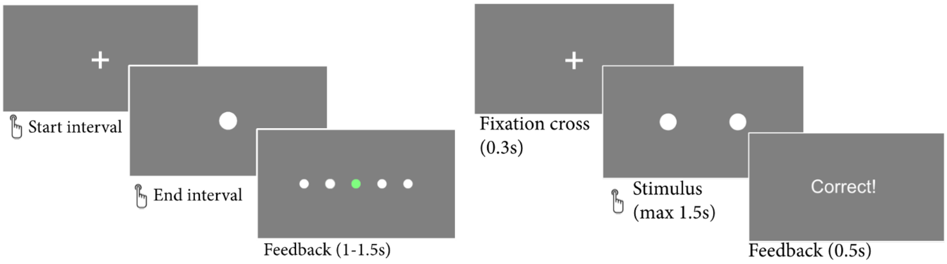

Fig. 2. Design of Experiment 1. Left: the time reproduction task. When presented with the fixation cross, participants can self-paced decide to start the time interval with a button press, upon which a circle appears. Participants stop the time interval by pressing the button again, after which feedback is presented. The feedback indicates how close the last estimate was to the actual time interval: the left circles indicate a major or minor underestimation (respectively), the right circles minor and major overestimation. The middle circle indicates a near-perfect estimate. Right: the decision-making task. After a short fixation cross, a stimulus consisting of two circles is presented. The circles appear and quickly disappear, which makes the impression of flashing. One of the circles flashes more often than the other, and the task is to decide which. Feedback is presented after every trial.

2. Experiment 1

2.1. Methods

2.1.1. Participants

The study was approved by the ethics committee of the Psychology Department, University of Amsterdam. 47 participants (mean age 21y [SD 1.9], 31 women, 43 right handed) were recruited from the subject pool of the Psychology Department of University of Amsterdam, and participated for course credit. As Experiment 1 was designed to be exploratory, sample size was not determined in advance, but we tested as many subjects as we could while the lab was available. All participants gave written informed consent prior to the experiment onset. Data from all participants were included in the analyses.

2.1.2. Tasks & Design

The participants performed two different tasks (Fig. 2): a time reproduction task (Kononowicz & Van Rijn, 2011; Macar, Vidal, & Casini, 1999) and a perceptual decision-making task (Brown, Steyvers, & Wagenmakers, 2009; Busemeyer & Rapoport, 1988; Hawkins, Brown, Steyvers, & Wagenmakers, 2012; Van Maanen, Fontanesi, Hawkins, & Forstmann, 2016). In the time reproduction task, participants reproduced a duration of 1.5 s by pressing the space bar twice – once to start the interval, and once to end the interval. Before the start of the interval, a fixation cross was presented, which changed to a circle upon the first keypress. After the second keypress, participants received feedback on how well they reproduced the interval, indicated by a colored circle that represented how close they were to the target duration (Kononowicz & Van Rijn, 2011). This feedback was presented for a random duration sampled uniformly between 1 s and 1.5 s. Participants were explicitly instructed not to count or tap their fingers or feet to reproduce the duration, as is standard practice in this paradigm (Balcı, Simen, et al., 2011; Kononowicz & Van Rijn, 2011).

The perceptual decision-making task requires participants to decide which of two visually presented circles flashes most often. Each trial starts with a fixation cross (0.3 s), after which a stimulus is presented. The stimulus consists of two circles drawn left and right from the fixation cross. Each circle has a fixed probability of flashing (quickly appearing and disappearing) on each moment in time during a trial. Each flash consists of a circle appearing for 0.050 s and subsequently disappearing for 0.0667 s (corresponding to 3 and 4 frames on a 60 Hz computer monitor). The probability that the circle corresponding to the correct answer flashed was set to $p_{correct} = 0.7$, and the probability that the incorrect circle flashes to $p_{incorrect} = 0.3$. Participants are instructed to respond as fast as possible without making mistakes and received feedback on their performance after each trial ('Correct!', 'Wrong!', 'Too late!').

All participants started with a block of 300 trials of the time reproduction task. The participants were instructed to estimate 1.5 s, and were given the opportunity to learn how long 1.5 s takes in 10 practice trials by trial-and-error prior to the onset of the experimental trials. After the time reproduction task, participants answered a questionnaire about risk taking behavior that was not part of the current study. Finally, participants performed the perceptual decision-making task. The number of trials depended on the participants' ability to respond accurately yet in time. Each correct answer was rewarded with 1 point, while each incorrect answer cost 1 point. Participants were instructed that there was a deadline of 1.5 s. If no decision was made before the deadline, a penalty of -3 points was imposed. Thus importantly, guessing in time had a higher expected reward (0) than missing the deadline (-3). The task ended if participants reached 200 points or if they performed 500 trials.

2.1.3. Data analysis: Time reproduction task

We analyzed the time reproduction data using the time-adaptive opponent Poisson diffusion model (TopDDM; Balcı & Simen, 2016; Simen, Balcı, deSouza, Cohen, & Holmes, 2011a, 2011b; Simen, Vlasov, & Papadakis, 2016). The TopDDM proposes that people estimate time intervals by integrating the difference in activity of two neural populations: One excitatory population, and one inhibitory population. These neural populations generate exponentially distributed pulses ('ticks'). Together, this opponent Poisson process forms an internal clock. Time intervals are estimated by integrating the net sum of ticks of the internal clock until a fixed threshold is reached, at which point the time interval has passed. By increasing or decreasing the firing rate of the neural populations, shorter and longer durations (respectively) can be estimated.

TopDDM predicts an accumulation process of ticks over time, that can be modeled as a diffusion process with a single bound. The first passage times of this process are Wald (also known as inverse-Gaussian) distributed, parametrized by drift rate μ, diffusion coefficient σ, and threshold α (Anders, Alario, & Van Maanen, 2016). In the TopDDM, the threshold is assumed to be constant, while drift rate μ and diffusion coefficient σ are functions of the parameters of the underlying Poisson processes. The drift rate is equal to the difference in spike rates of the populations, and the amount of noise is proportional to the square root of the drift rate: $\sigma = m\sqrt{\mu}$. In this specification, m is the balance between firing rates of the two neural populations, which is assumed to be constant within participants. Individual differences in temporal precision are thought to be a result of individual differences in m (Simen et al., 2016).

In the present study, we use the TopDDM as a model to infer individual differences in temporal precision by estimating m, where a small m indicates a high temporal precision. Note that m is strongly related to the coefficient of variation (CV), which is a more traditional measure of estimating temporal precision. We briefly return to their relation in the Discussion section.

To estimate the temporal precision parameter m, we fitted a shifted Wald distribution to the time reproduction data. Strictly, this model is not identical to the TopDDM as proposed in (Simen et al., 2016), because it adds a non-decision time shift τ to explicitly model processes unrelated to the accumulation of time ('non-decision' time). The reason for including this shift is that the target durations in our task are relatively short compared to typical target durations in other experiments (e.g., Kononowicz & Van Rijn, 2011; Simen et al., 2016), and as a consequence, processes unrelated to the accumulation of clock ticks may have a relatively larger influence on the final time reproduction time distributions.

The first 50 trials of the time reproduction task were discarded to prevent possible learning effects from influencing the results. For scaling purposes, it is necessary to fix one timing-related parameter to a constant value. In the TopDDM, within-subject adjustments in drift rate μ are assumed to underly the ability to estimate different target intervals, whereas between-subject differences in temporal precision are caused by interindividual variability in m. Therefore, we estimate both μ and m as free parameters (as well as non-decision time τ), and fix the threshold α to a value of 1 for all participants. In an initial attempt, parameters were fitted using maximum likelihood estimation. However, maximum likelihood estimation is sensitive to the presence of outliers in data (Ratcliff & Tuerlinckx, 2002), and presumably because of this, the resulting fits were poor. Therefore, we instead estimated parameters using χ^2-based difference minimization; a method known to be relatively insensitive to outliers in data (Ratcliff & Tuerlinckx, 2002), which led to much better fits (see Supplementary Figure B1 for an overview). The model was fitted on the seconds scale. Parameters were optimized using differential evolution optimization (Price, Storn, & Lampinen, 2006) with default parameters as implemented in the software package 'DEoptim' (Ardia, Arango, & Gomez, 2011; Ardia, Boudt, Carl, Mullen, & Peterson, 2011; Mullen, Ardia, Gil, Windover, & Cline, 2011) in the R programming language (R Core Team, 2017). Optimization was stopped when no improvement in likelihood was found after 300 iterations.

2.1.4. Data analysis: Decision-making task

We analyzed the decision-making data by calculating mean reaction times and accuracy, and fitting the LBA model. The LBA model is a formal decision-making model that proposes that people make decisions by collecting evidence for each choice alternative (Fig. 1A). For every choice alternative, an independent accumulator linearly increases until one of the accumulators reaches a shared threshold. At this point, the decision maker commits to that choice and initiates the motor response. The response time is the decision time plus the time to perform initial perceptual encoding of the stimulus, and the time for execution of the motor response. These two components are collapsed together in the non-decision time. To account for variability in choices and response times, evidence accumulation speed varies from trial to trial, as does the amount of evidence at the start of the accumulation process.

The LBA model is parametrized here by seven parameters: the upper bound of the uniform distribution of start points of evidence A, threshold b, the means and standard deviations of the normal distributions describing the drift rates for both accumulators v_1, sd_1, v_2, and sd_2, and finally, non-decision time $t0$. Drift rate v_1 and standard deviation sd_1 always correspond to the accumulator collecting evidence for the correct choice option. Because the expected starting point of the evidence accumulation process is $A/2$ (Brown & Heathcote, 2008), the best measure of *response caution* is b-$A/2$. Parameter sd_1 was fixed to 1 for scaling. Similar to TopDDM, fixing one decision-related parameter to a constant value for all participants is necessary to ensure that all remaining parameters are expressed on a common scale (Donkin, Brown, & Heathcote, 2009). Parameter sd_2 was estimated as a free parameter; assuming sd_2 is equal to sd_1 may overconstrain the model (see Donkin et al., 2009 for more detail on scaling and overconstraining). Reaction time data were fitted on the seconds scale. Model fitting was performed using maximum likelihood estimation. Parameters were optimized using particle swarm optimization (Clerc, 2010) with default settings as implemented in software package 'pso' (Bendtsen, 2012) in the R programming language (R Core Team, 2017). Optimization was restarted 10 times with random start points.

2.1.5. Multiple comparison correction

In analyzing the data of Experiment 1, we tested for correlations between temporal precision and 8 other variables: mean RT, accuracy, and 6 LBA model parameters (parameter sd_1 is fixed to 1 for all participants and therefore was not used for correlation analyses). To correct for Type 1 error rate inflation, we report Holm-Bonferroni-adjusted p-values (Holm, 1979).

2.2. Results

Descriptive statistics of the data are presented in Table 1. The key question of Experiment 1 is whether temporal precision, estimated as the m parameter in the TopDDM, predicts choice behavior. Therefore, we first tested whether temporal precision

Table 1
Descriptives of the data in both experiments.

	Time reproduction	Decision-making			
	Mean temporal precision m (SD)		Mean RT (SD)	Accuracy (SD)	Deadline time in seconds (SD)
Experiment 1 (n = 47)	0.108 (0.049)		0.632 (0.150)	0.815 (0.064)	1.5 (–)
Experiment 2 (n = 56)	0.169 (0.105)	Short deadline	0.508 (0.221)	0.726 (0.069)	1.06 (0.44)
		Long deadline	0.804 (0.327)	0.802 (0.076)	5 (–)

correlates with mean reaction times and accuracy across participants. The results indicated that precision correlated significantly with mean reaction time ($r(45) = -0.395, p = 0.036$, corrected, two-sided; Fig. 3A) and choice accuracy ($r(45) = -0.51, p = 0.002$, corrected, two-sided; Fig. 3B). Thus, participants with a higher temporal precision made slower and more accurate decisions in the choice task.

To better understand which aspects of decision-making are influenced by time reproduction ability, we modeled the decision-making data with the LBA model, and tested which model parameters correlate with temporal precision (Supplementary Figure B1 provides the fit of the model to the choice data). We found evidence for a correlation between m and v_1 ($r(45) = -0.415, p = 0.026$, corrected, two-sided, Fig. 3C), and between m and response caution (i.e. b-$A/2$; $r(45) = -0.395$, $p = 0.036$, corrected, two-sided, Fig. 3D). The correlations with other LBA parameters were not significant.

3. Interim discussion

So far, the hypothesis that timing ability affects decision-making seems supported by a correlation with mean response time and accuracy, and by correlations with specific parameters of a model of decision-making. Since we aim for a mechanistic understanding of which aspects of decision-making are affected by time reproduction ability, we focus further on the model-based analyses. These indicate that participants that have a precise internal clock, also have high drift rates as well as high response caution in the decision-making task. Interpreting the correlations we observed in however Experiment 1 not straightforward because in our sample the decision-making parameters drift rate and response caution themselves correlate positively and strongly ($r(45) = 0.966$, $p < 0.001$, two-sided, uncorrected). The correlation between drift rate and caution can intuitively be understood in terms of the costs of being cautious: participants with a high drift rate accumulate much evidence in each time step, so the costs (in terms of time and the probability of reaching the deadline) of using high thresholds are relatively low. Therefore, in the current task set-up, participants with a high drift rate can afford a high threshold, leading to both faster and more accurate decisions than participants with a low drift rate.

Because of the high correlation between drift rate and response caution, the correlation between timing precision and drift rate could be mediated by response caution, or – vice versa – the correlation between temporal precision and response caution could be mediated by drift rate. Arguably, a correlation between temporal precision and drift rate is plausible to be direct (i.e., not mediated by threshold). This is because between-subject variability in drift rates is often taken to reflect variability in the ability to extract relevant information from the stimulus (Donkin & Van Maanen, 2014; Donkin, Brown, Heathcote, & Wagenmakers, 2011; Mulder & Van Maanen, 2013; Palmer, Huk, & Shadlen, 2005; Ratcliff et al., 2016; Voss, Rothermund, & Voss, 2004), which correlates across multiple types of tasks (Ratcliff, Thapar, & McKoon, 2006). Sometimes the drift rate parameter is more generally interpreted in terms of cognitive ability (Schulz-Zhecheva, Voelkle, Beauducel, Biscaldi, & Klein, 2016; Van Ravenzwaaij, Brown, & Wagenmakers, 2011). Similarly, temporal precision in the time reproduction task can be interpreted as a measure of ability: Participants with a highly precise internal clock repeatedly estimate the interval with high precision (cf. CV, Wearden, 2016). Therefore, a possible interpretation is that the correlation between drift rate and temporal precision reflects task-general ability: Participants that are efficient at extracting relevant stimulus information in the choice task may also be precise at reproducing time intervals.

Alternatively, the correlation between drift rates and temporal precision could be caused by covariation between the ability to

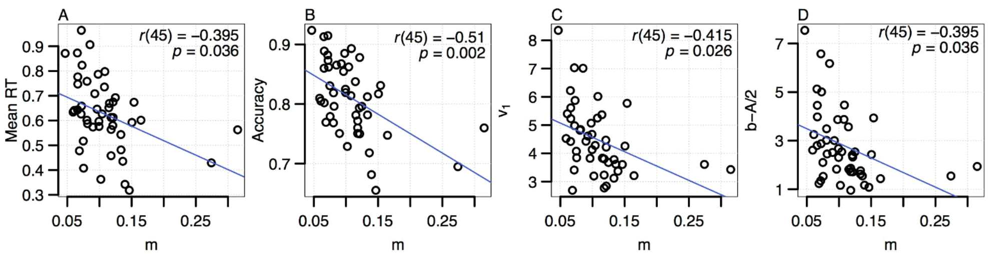

Fig. 3. Correlations between (A) temporal precision m and choice mean response time (B) m and choice accuracy, (C) m and v_1, and (D) m and response caution b-A/2 in Experiment 1. Note that two observations have a relatively high m, but we had no reason to suspect these participants did not understand the time reproduction task, nor any other reason to exclude these data points. Inclusion or exclusion of these data points has no consequences for the test results.

estimate a time interval and the collapse of decision bounds, which manifests as changes in drift rate. Higher drift rates can reflect a higher additive urgency signal, which is equivalent to stronger linearly collapsing bounds (see the Supplementary Materials).

In addition to these interpretations, the correlation between timing ability and response caution could also support the hypothesis that strategic adjustments of behavior due to a deadline (in the form of threshold settings) depend on the ability to precisely represent the deadline (cf. Fig. 1B). This hypothesis is however only supported by a direct correlation, that is not mediated by drift rate. If the correlation is mediated by drift rate, a fourth possible interpretation is that timing ability correlates with drift rate (reflecting task-general ability as above), and higher drift rates allow for higher thresholds (because of the task set-up to respond before the deadline).

To disentangle these possible interpretations of the results of Experiment 1, we need to understand whether the correlation between timing ability and response caution is direct or mediated by drift rate. Therefore, in Experiment 2, we first disentangle the correlational structure of timing ability, drift rate, and threshold. This will allow us to test whether timing directly correlates with response caution while controlling for the effect of drift rate. Secondly, in Experiment 2 we add a within-subject manipulation of time pressure to address the question of whether time estimation ability predicts the amount of collapse of response thresholds (Fig. 1C), as proposed by Frazier and Yu (2008). As explained below, contrary to the fully between-subject design of Experiment 1, the within-subject manipulation imposed in Experiment 2 allows us to approximate the amount of collapse by measuring the difference between choice behavior when faced with a short deadline compared to a long deadline.

4. Experiment 2

4.1. Methods

4.1.1. Participants

The study was approved by the ethics committee from the Psychology Department, University of Amsterdam. 64 participants (mean age 21y [SD 2.9], 55 women, 54 right handed) were recruited from the subject pool of the Psychology Department of University of Amsterdam, and participated for course credit. Since we had no a priori effect size estimate, we did not perform a power analysis in advance, but instead tested as many subjects as we could while the lab was available. All participants gave written informed consent prior to the experiment onset. One participant decided to not complete the experiment due to a headache. Another participant did not follow instructions but pressed the same response key on all trials of one block and was therefore excluded from analysis. Six other participants were excluded for reasons explained below (section "*Quality of subject-specific stimulus difficulty and deadline*"), leading to a final sample size of 56 participants (mean age 21y [SD 3.1], 47 women, 50 right handed).

4.1.2. Tasks

The participants performed the same two tasks as in Experiment 1, described above.

4.1.3. Design

The experiment consisted of three decision-making blocks and one time reproduction block. All participants started with a decision-making block (the *calibration block*), which was designed to tailor both task difficulty (Mulder, Wagenmakers, Ratcliff, Boekel, & Forstmann, 2012; Winkel, Keuken, Van Maanen, Wagenmakers, & Forstmann, 2014) and a deadline to each individual participant. In this block, participants performed 250 trials of randomly interleaved stimuli with five difficulty levels (50 trials each). Stimulus difficulty was defined as the ratio of the probabilities of flashing for the incorrect versus the correct circle, $p_{incorrect}/p_{correct}$. We kept $p_{correct}$ constant at 0.7 and varied $p_{incorrect}$ between 0.007, 0.21, 0.35, 0.49, and 0.63, which corresponds to five difficulty levels: 0.01, 0.3, 0.5, 0.7, and 0.9, respectively. The proportional-rate diffusion model (Palmer et al., 2005) was fitted to the mean reaction times and accuracies per stimulus difficulty using maximum likelihood estimation. From the model fit, we interpolated the stimulus *difficulty* at which the participant was expected to perform with 80% accuracy (Fig. 4, top). Using this interpolated stimulus difficulty, we estimated the subject-specific *deadline* by linearly interpolating between the 75th quantile RTs of the two nearest difficulty levels in the calibration block (Fig. 4, bottom).

After the calibration block, participants performed three additional blocks: one decision-making block with a long deadline, one decision-making block with a short deadline, and a time estimation block. We counterbalanced two block orders. In one, participants started with the long deadline choice task, then performed the time reproduction task, and ended with the short deadline choice block. In the other order, participants started with the timing task, then performed the short deadline choice task, and ended with the long deadline choice task. This ensured the time estimation block always directly preceded the short deadline decision-making block, for which it was most relevant.

Similar to Experiment 1, in both remaining decision-making blocks, participants were instructed to earn 200 points. The reward structure was the same as in Experiment 1 as well. In the long deadline block, the deadline was set to 5 s. Since the slowest response in the calibration block across all participants was 4.511 s, all participants were, in principle, able to fully perform the task well before the deadline even at the highest difficulty level. In the short deadline decision-making block, the deadline was subject-specific as explained above.

The time reproduction block consisted of 200 trials in which the participants reproduced their own deadlines for the short deadline decision-making block. The participants were instructed to estimate the subject-specific deadline, and were given the opportunity to experience how long this duration takes in 10 practice trials by trial-and-error before the onset of the experimental trials. The feedback was similar to Experiment 1 and consisted of colored circles that indicated how close the time estimate was relative to the target interval. The relative distances (i.e., the underestimation/overestimation proportional to the target interval) at which the

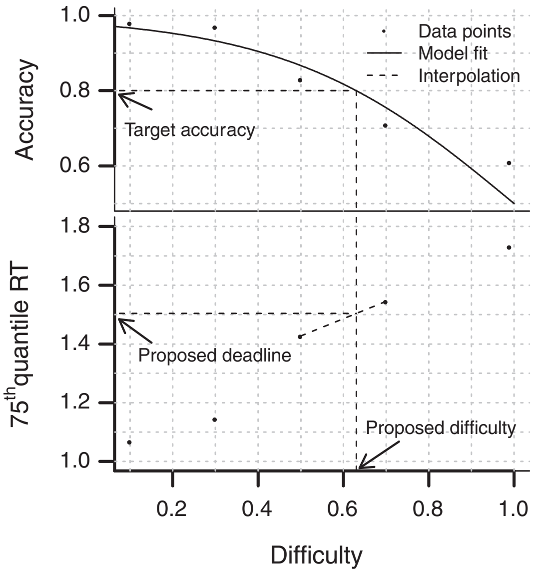

Fig. 4. Illustration of determining subject-specific difficulty and deadline based on the calibration block data. First, the proportional-rate diffusion model is fit to the mean reaction time (in seconds) and accuracy data from the calibration (top panel, only accuracy is shown). Using the model fit, we interpolate (dashed lines) the difficulty level at which the model predicts 80% accuracy. Then, using this proposed difficulty level, we linearly interpolate the 75th quantile reaction times of trials of the two nearest difficulty levels in the calibration block. The deadline is set at the estimated 75th quantile of the proposed difficulty. Note that while it is possible to use the proportional-rate diffusion model to predict the 75th quantile RT at the proposed difficulty level, we found in a pilot experiment that the quality of the model fit of the mean RTs was insufficient for some participants to accurately predict the 75th quantile at the proposed difficulty level (although it consistently fit the accuracy data well).

circles appeared were identical to Experiment 1.

4.1.4. Quality of subject-specific stimulus difficulty and deadline

Although the subject-specific determination of deadline and stimulus difficulty provides a way to level both perceived difficulty and perceived time pressure between participants, the procedure depends on two assumptions that did not hold for all participants. Firstly, it assumes that participants are able to do the task with at least 80% accuracy for the easiest stimuli in the calibration block. This was not the case for two participants, and as a consequence, their subsequent difficulty level was set to 0, making the task trivially easy (i.e., only the correct circle flashes). Data from these participants were excluded from further analyses.

Secondly, the procedure assumes an absence of time on task (e.g., learning) effects; i.e., it assumes that, apart from the imposed manipulations, participants' behavior stays comparable after the calibration block. For some participants, estimation of the 75th quantile reaction time based on the calibration block was highly inaccurate due to strong changes in behavior after the calibration block. Especially participants that performed the short deadline block before the long deadline block were relatively faster in the long deadline block compared to the calibration block. For this reason, we excluded a further four participants whose actual responses in the *long* deadline block were too fast to have reached the subject-specific deadline (had that deadline been imposed). We considered subject-specific deadlines too lenient if it was placed beyond the 97.5th percentile of the reaction times in the *long* deadline block.

4.1.5. Model fitting: Time reproduction data

As in Experiment 1, we used the TopDDM model to analyze the time reproduction data. Model fitting procedures were identical as in Experiment 1.

4.1.6. Model fitting: Choice data

We used the LBA model (Brown & Heathcote, 2008) to fit the choice task data. In all models discussed below, the between-trial standard deviation of drift rates for the accumulator corresponding to the correct answer (sd_1) was fixed to 1 across conditions to prevent scaling issues (e.g., Donkin et al., 2009). Models were fitted to data on the seconds scale.

The data from the short and long deadline blocks were fitted simultaneously. We fit 9 different model specifications (see Table 2), each supporting a hypothesis of the effect of the deadline on decision-making behavior. The first hypothesis is that participants base their decisions on less evidence when the deadline is short, which we model by allowing threshold b to vary between conditions (Fig. 1B; Models 1, 3, 5, 7, and 9). This hypothesis entails that imposing a deadline has a similar effect on choice behavior as a cue-based speed/accuracy trade-off in which participants are cued to act either fast or accurately (Boehm et al., 2014; Van Maanen et al., 2011).

The second hypothesis is that participants collapse decision bounds when the deadline is short (Fig. 1C). We model this by fitting an additive urgency signal (Cisek, Puskas, & El-Murr, 2009; Murphy, Boonstra, & Nieuwenhuis, 2016; Thura et al., 2012) in the short

Table 2
Model comparison.

Model	Interpretation	Parameters			Rank		
		Free between conditions	Fixed between conditions	k	Mean	n best	n top 3
1	Caution	b	$t0, sd_2, A, v_1, v_2$	7	3.839	10	27
2	Urgency	U	$t0, sd_2, A, v_1, v_2, b$	7	2.911	21	40
3	Urgency + caution	U, b	$t0, sd_2, A, v_1, v_2$	8	**1.734**	**23**	**54**
4	Attention	v_1	$t0, sd_2, A, b, v_2$	7	7.143	0	4
5	Attention + caution	v_1, b	$t0, sd_2, A, v_2$	8	3.911	1	21
6	Attention	v_1, v_2^*	$t0, sd_2, A, b$	8	8.661	0	0
7	Attention + caution	v_1, v_2^*, b	$t0, sd_2, A$	9	5.911	0	0
8	Attention	$v_1, t0$	sd_2, A, v_2, b	8	6.768	0	0
9	Attention + caution	$v_1, t0, b$	sd_2, A, v_2	9	4.143	1	22

Notes. Bold-faced entries indicate the winning values. k: number of free parameters, n best: number of subjects for which this was the best model, n top 3: number of subjects for which this model was in the top 3 of best models in this model comparison. Parameter abbreviations: v_1/v_2: drift rates for the correct/incorrect accumulator, respectively; v_2^*: drift rate for incorrect accumulator, but only free to decrease in short deadline condition compared to long deadline; sd_2: standard deviation of drift rates of incorrect accumulator; b: threshold; A: upper bound of uniform start point distribution; $t0$: non-decision time; U: additive urgency signal.

deadline condition, such that $v_{1,\text{short}} = v_{1,\text{long}} + U$ and $v_{2,\text{short}} = v_{2,\text{long}} + U$, where subscripts *short* and *long* indicate the specific experimental blocks. Importantly, we model linear collapsing decision bounds using an additive urgency signal because these are mathematically equivalent in the LBA (see the Supplementary Materials for a proof). Models 2 and 3 include this additive urgency signal.

The third hypothesis we test is that participants are more attentive when the deadline is short compared to when to deadline is long. In evidence accumulation models, attention effects are argued to influence a combination of drift rate and non-decision time (Mulder & Van Maanen, 2013; Nunez, Vandekerckhove, & Srinivasan, 2017; Smith & Ratcliff, 2009; Van Maanen, Forstmann, Keuken, Wagenmakers, & Heathcote, 2016). Here, we use various LBA model specifications that allow these two parameters to vary across conditions: we allow v_1 to increase in the short deadline block compared to the long deadline block (Models 4–9), possibly combined with a decrease in v_2 (Models 6–7), and possibly combined with a change in $t0$ (Models 8–9). Note that the crucial difference from the urgency models is that drift rate v_2 is never allowed to increase in the attention models. Since an increase in drift rate v_1 decreases decision-times and *increases* accuracy, attention by itself would not explain the typical effects of deadlines as mentioned earlier (i.e., decreased reaction times and accuracy; Busemeyer, 1985; Link, 1971). However, we include these models to exclude the possibility that a combination of attention and threshold effects provide the best explanation of the behavioral adjustments due to a response deadline.

All models were fitted using maximum likelihood estimation. Parameters were optimized using differential evolution optimization (Price et al., 2006) with default parameters as implemented in the software package 'DEoptim' (Ardia, Arango, et al., 2011; Ardia, Boudt, et al., 2011; Mullen et al., 2011) in the R programming language (R Core Team, 2017). Optimization was stopped when no improvement in likelihood was found after 1000 iterations.

To compare the quality of fit of different model specifications, penalized for the number of parameters, we used the Bayesian information criterion (BIC; Schwarz, 1978; Wagenmakers & Farrell, 2004), defined as $BIC = -2\ln(L) + k\ln(n)$, where k is the number of estimated parameters, n the number of observations (trials), and L the likelihood of the best-fitting parameters. Lower BIC values indicate a better trade-off between quality of fit and model complexity. Although BIC values could be summed across participants for each model specification, such summed BIC values can be highly influenced by the quality of fit of a few subjects, and by the differences in numbers of trials per participant (participants that performed more trials weighing in more heavily). Moreover, summed BIC values ignore possible individual differences in model fits, and we report below substantial individual variability in task strategies. Therefore, we provide multiple non-parametric statistics to provide an overview of the quality of fits of the model specifications. Firstly, we rank all model specifications (lowest BIC = 1, highest = 9), and average ranks across participants. Secondly, we report the number of participants for which each model specification has the highest rank. Thirdly, we report the number of participants for which each model is in the top 3 best-ranking models.

4.1.7. Bayesian model averaging

In this experiment, we test hypothesized relations between LBA model parameters and TopDDM parameter m. While it is possible to use the parameters from the winning LBA model specification, this would ignore individual variability in model fits and strategies in handling the deadline. Therefore, for all analyses using model parameters reported below, we used a Bayesian model averaged estimate of parameters. In Bayesian model averaging (Hoeting, Madigan, Raftery, & Volinsky, 1999; Van Maanen, Forstmann, et al., 2016), the parameter estimates of a model specification are weighted by the wBIC values for each model i, defined as (Wagenmakers & Farrell, 2004):

$$w_iBIC = \frac{e^{-\frac{1}{2}\Delta_i(BIC)}}{\sum_{k=1}^{K} e^{-\frac{1}{2}\Delta_k(BIC)}}$$

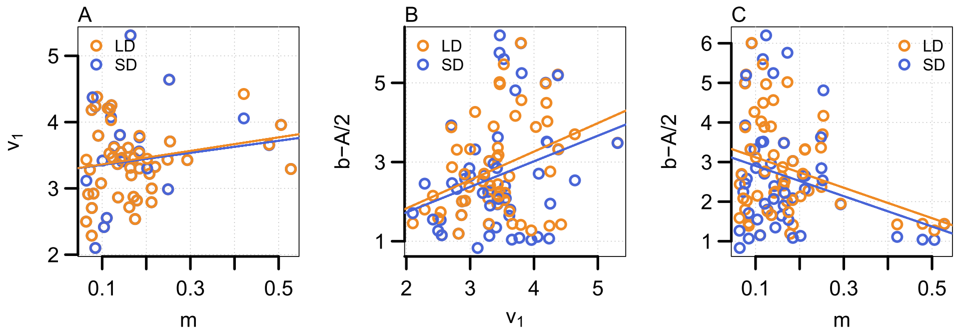

Fig. 5. Results of Experiment 2. (A) Scatterplot of temporal precision m and drift rate v_1, (B) scatterplot of v_1 and response caution $b\text{-}A/2$, and (C) scatterplot of m and $b\text{-}A/2$ in both conditions (SD: short deadline, LD: long deadline). Lines are linear regression lines.

where $\Delta_i(BIC) = BIC_i - \min(BIC)$. Thus, parameter values of each subject are weighted combinations of parameter estimates under all model specifications. The weighting factor is the quality of each model fit. The conclusions we report below would be the same if the parameters of the winning model (Model 3) were used instead of Bayesian model averaged parameters.

4.1.8. Multiple comparison corrections

We performed two confirmatory, one-sided null-hypothesis significance tests: One testing for a correlation between temporal precision and response caution, and one testing for a relation between temporal precision and urgency. All reported p-values are Holm-Bonferroni-adjusted (Holm, 1979) to correct for multiple comparisons, except for two tests that are included as manipulation checks.

5. Results

5.1. Disentangling the correlational structure of temporal precision, drift rate, and response caution

We aimed to disentangle the multivariate correlational structure reported in Experiment 1 by fixing the perceived difficulty level across participants, such that v_1 is roughly equal. As a result, the correlation between m and (the model-averaged) v_1 was not significantly negative in the short deadline condition ($r(54) = 0.151$, $p = 0.867$, uncorrected, one-sided, Fig. 5A). A one-sided Bayesian correlation test (Ly, Verhagen, & Wagenmakers, 2016) was used to quantify the amount of evidence against a negative correlation. This revealed that the data are 6.48 times as likely under the null hypothesis that no negative correlation exists than under the alternative hypothesis that a negative correlation exists (null hypothesis interval $[-1, 0]$, $BF_{0-} = 6.48$, 'substantial' evidence; Kass & Raftery, 1995).

Although fixing the perceived difficulty across participants seemed to strongly reduce the correlation coefficient between v_1 and response caution in the short deadline condition compared to Experiment 1, we found it was still significant ($r(54) = 0.290$, $p = 0.015$, one-sided, uncorrected, Fig. 5B). Since response caution and drift rate still covary, we used linear mixed effects models (LMEMs; Baayen, Davidson, & Bates, 2008; Barr, Levy, Scheepers, & Tily, 2013) to disentangle variance in response caution that can be explained by drift rate, by temporal precision, and by condition (long or short deadline). We first fitted a LMEM with these three main effects plus all interactions, and then removed interaction terms step-wise. In all LMEMs, participants were included as random intercepts (since there were only two repeated measures, this is the maximal model). If temporal precision influences response caution as a result of a behavioral adjustment to a short deadline, we expect a significant interaction effect between temporal precision and deadline condition: Temporal precision should then predict threshold only in the short deadline condition.

In contrast to this expectation, the simplest LMEM, including only the three main effects and the random intercept for subjects, was preferred (lower BIC values) over all LMEMs that included interactions. In this simplest model, all three terms were significant, indicating that drift rate was predictive of response caution ($t(58.27) = 3.620$, $p = 0.004$, two-sided, corrected), participants were more cautious in the long deadline condition ($t(55.06) = 2.730$, $p = 0.034$, two-sided, corrected), and temporal precision was negatively predictive of response caution ($t(53.04) = -3.124$, $p = 0.009$, one-sided, corrected). From this last term, we conclude that time estimation ability is related to response caution directly and independently from the influence of v_1 (Fig. 5C).

5.2. Modelling the deadline manipulation

In total, we compared 9 different LBA model specifications in their ability to explain the deadline manipulation. Table 2 provides a model comparison. Model 3 wins by all three metrics: it has lowest mean rank, is the best fitting model for the largest proportion of participants (41.1%), and occurs most often in the top 3 of best-fitting models across participants. Supplementary Figure B2 illustrates the quality of fit for this model. Model 3 is characterized by a combination of a threshold adjustment and urgency. Furthermore, Model 2 (only urgency) performs well, as it fits best for 21 participants (37.5%). For 10 participants (17.9%), Model 7 (only a change in threshold b) fitted best. All attention models (Models 4–9) perform poorly for almost all participants. This finding is in line with

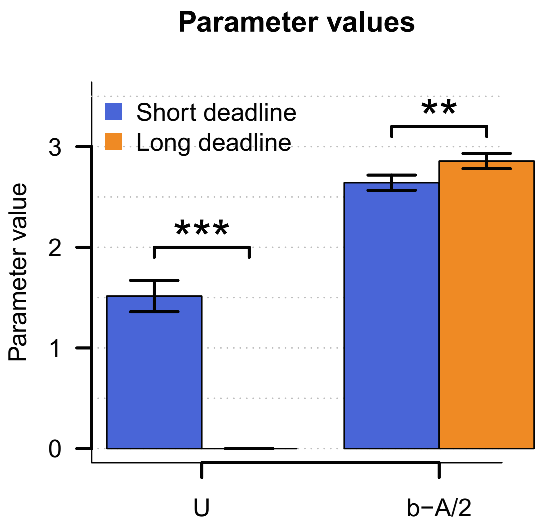

Fig. 6. Bayesian model averaged estimates of the urgency (U) and response caution (b-A/2) parameters in the short deadline (blue) and long deadline (orange) condition. Error bars indicate ± 1 SE of the within-subject differences in parameter estimates between conditions (Cousineau, 2005). Note that the urgency parameter was defined as the difference in both drift rates between conditions; it cannot be estimated for the long deadline condition. $^{**}p < 0.01$; $^{***}p < 0.001$, corrected two-sided p-values. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

expectations, since an increase in drift rate increases rather than decreases accuracy (which is the typical effect of a deadline, Busemeyer, 1985; Link, 1971). Only for 2 participants, an attention model (Models 5 and 9) fit the data best.

For most of the participants (78.6%), the winning model included an additive urgency signal (Models 2 and 3). More than half of the participants showed a change in threshold (62.5%; Models 1, 3, 5, 7, 9), and a substantial part (41.1%) incorporated both an additive urgency signal and changed threshold. To provide further detail on the change in these parameters due to the response deadline, Fig. 6 shows the Bayesian model averaged threshold and additive urgency parameter estimates in both conditions. The urgency parameter was significantly above zero in the short deadline condition (M = 1.52, SD = 1.16; $t(55) = 9.75$, $p < 0.001$, corrected, two-sided). Further, response caution was significantly higher in the long deadline condition (M = 2.86, SD = 1.25) than in the short deadline condition (M = 2.64, SD = 1.41; $t(55) = -2.84$, $p = 0.032$, corrected, two-sided).

Taken together, we conclude that although there is large variability in decision-making in the face of a short deadline, many participants seem to incorporate an additive urgency signal (which under the current LBA specification is equivalent to a linearly collapsing bound).

5.3. Testing for a correlation between additive urgency and temporal precision

The theoretical optimal amount of collapse is believed to depend on the temporal precision of the decision-maker (Frazier & Yu, 2008). We tested this hypothesis using the additive urgency parameter we fitted in the short deadline condition. To do this, we created linear regression models with temporal precision, response caution, and the location of the deadline as predictors, and included all their interactions. The latter two predictors were included because we hypothesized that these can both influence the degree to which participants feel urgency: Higher distances to bound and shorter times available to inform a decision both increase the chance of missing the deadline unless these effects are compensated by an urgency signal. A backwards step-wise regression procedure indicated that the best fitting model (lowest BIC) included the three main effects plus an interaction between response caution and deadline.

In this model, only response caution ($t(51) = 5.135$, $p < 0.001$, two-sided, corrected) and the interaction between caution and deadline ($t(51) = -2.677$, $p = 0.033$, two-sided, corrected) proved significantly predictive of the amount of collapse. Thus, higher response caution is related to a stronger sense of urgency – and this effect is stronger for participants with shorter subject-specific deadlines. These results support the idea that participants use an additive urgency signal as a way to respond before the deadline. However, the weight of the temporal precision term was non-significant ($t(51) = -0.733$, $p = 0.18$, one-sided, corrected), indicating that there is no evidence that temporal precision predicts the amount of collapse.

In light of the large variability in strategies to cope with a short deadline (as reported in the previous section), we repeated this analysis using only the subjects for which the urgency models fit the choice data best (44 out of 56 participants). The conclusions that would follow from this exploratory analysis are identical to the results including all subjects.

6. Discussion

Deliberation in decision-making takes time, which is especially costly in situations where time is limited by a response deadline. Strategical adjustments of choice behavior to a response deadline require decision-makers to internally represent the amount of time available to inform a decision. However, despite this reliance on temporal cognition, the effects of temporal cognition on decision-making under time pressure are not well understood. Here, we investigated how the precision of the internal representation of the response deadline affects decision-making.

In Experiment 1, we explored which decision-making characteristics, as identified by a decision-making model, correlate with the precision of the representation of the deadline as measured in a separate task. We found that information processing efficiency and response caution in decision-making correlate with timing ability. This suggests that good timers are also efficient in processing the relevant information to inform decisions, and furthermore respond more cautiously in choice tasks.

In Experiment 2, we showed that the precision of the internal representation of the deadline correlates with response caution even when the perceived difficulty of the task is kept constant across participants. This shows that the correlation between timing ability and response caution is not mediated by task-general ability.

Experiment 2 also showed evidence in favor of collapsing bounds, albeit not related to timing ability. The difference between decision-making under strict and lenient deadlines can be captured well by the proposal that participants collapse decision bounds under a strict deadline. Theory on optimal decision-making (Frazier & Yu, 2008) further predicts that the amount of collapse should depend on the temporal uncertainty surrounding the deadline. We did not find evidence to support this hypothesis. This result is in line with previous work testing the collapsing bound hypothesis (Karşılar et al., 2014). Using conditional accuracy functions (Heitz, 2014; Ridderinkhof, 2002; Van Maanen, Katsimpokis, & Van Campen, 2018) as a proxy for collapsing bounds, these researchers also studied the relationship between interval timing and decision-making under temporal deadlines. They observed a slight decline in accuracy over time (potentially indicating collapsing bounds), but this decline was not correlated with timing ability.

Taken together, the results indicate that *between* participants, timing ability predicts response caution: Participants with an accurate representation of the deadline accumulate more evidence to commit to a decision than participants with an inaccurate representation (Fig. 1B). Moreover, the behavioral adjustment *within* participants due to a strict deadline (compared to a lenient deadline) is explained by collapsing decision bounds, but the speed of this collapse was not mediated by timing ability.

If the response caution settings observed in Experiments 1 and 2 are a result of a strategic adjustment due to time pressure, we would expect that the relation between time estimation and response caution to disappear when a long deadline is imposed. Contrary to this expectation, we found no interaction effect between temporal precision and the deadline time (short vs long) in Experiment 2, suggesting that temporal precision and response caution are also correlated when a long deadline is imposed. However, this finding could be explained by considering that participants experience time pressure also when a long deadline is imposed. For example, it has been shown that even without imposing an explicit deadline or emphasizing speed, participants decide on the basis of increasingly less evidence as time passes (Van Maanen, Fontanesi, et al., 2016). This demonstrates that people adapt behavior based on *experienced* time pressure even when the task does not require participants to be fast. Instead, participants seem internally motivated to limit decision time. Various factors could contribute to this motivation, such as the desire to finish the experiment quickly and to reduce the amount of effort put into the task. Interestingly, instructing participants to prioritize accuracy over speed leads participants to use similar response boundaries as when no instruction is given (e.g., Forstmann et al., 2011; Forstmann, Dutilh, Brown, Neumann, von Cramon, et al., 2008). Thus, trading off speed and accuracy seems inherent in the experimental set-ups commonly used, which limits conclusions we can draw about the relation between timing and decision-making in situations without any experienced time pressure.

If participants experience time pressure in tasks without an explicit deadline, it may be possible to reduce this effect by penalizing errors stronger than correct answers are rewarded. In such a situation, using collapsing bounds or an additive urgency signal would reduce the average reward rate compared to a fixed bounds strategy, making it relatively costly to use urgency strategies. Such a paradigm might shed more light on the relation between timing ability and decision-making without time pressure.

In the present study, we used the TopDDM (Balcı & Simen, 2016; Simen et al., 2016; Simen, Balcı, deSouza, Cohen, & Holmes, 2011a, 2011b) to analyze the time reproduction data and estimate temporal precision. A more traditional measure of temporal precision is the coefficient of variation (CV) of time estimates, calculated as the ratio of the standard deviation of the estimates and their mean. As such it is important to understand how the measure we use (m) relates to CV. In the TopDDM, there is a theoretical relation between the m parameter and CV: Assuming that time estimates are a result of a one-sided diffusion process (i.e., a Wald distribution), then $CV = m/\sqrt{\alpha}$, where α is the threshold of the diffusion process. Since we fixed the threshold parameter to 1 for all participants (for scaling purposes), m directly reflects CV. Indeed, the temporal precision parameter m correlates with CV directly computed from the time estimates in both Experiment 1 ($r(45) = 0.45$, $p = 0.002$) and 2 ($r(54) = 0.66$, $p < 0.001$). These correlations are not 1 for two reasons. First, we did not fit a Wald distribution (as in the original TopDDM) but a *shifted* Wald distribution, to be able to account for processes unrelated to clock speed that do influence the response times. Second, we fitted the shifted Wald distribution using a method that is insensitive to the presence of outliers, whereas the standard deviation computed from the data is highly susceptible to the presence such data points. When performing all analyses mentioned in this paper using CV rather than TopDDM, the same conclusions would be reached. As such, the results therefore do not strictly depend on the theoretical stance regarding the (cognitive) mechanisms underlying temporal accumulation that we have adopted. Nevertheless, it is striking that a well-constrained model such as TopDDM accounts for the timing data as convincingly.

In modelling the data of both experiments, we assumed that the decision process underlying all observed responses can be approximated by the LBA model. This assumption entails that participants accumulate evidence until a static or linearly collapsing threshold is reached, at which point the motor response is initiated. An alternative modelling assumption is that participants accumulate evidence, but *guess* when the deadline draws near, either completely random (Van Maanen, 2016) or based on the information accumulated so far (Ratcliff, 1988, 2006). The observed data are then a mixture of LBA responses and guesses. In initial explorations of this proposal, we fitted a mixture model of LBA and random guesses to the data of Experiment 1. Although we found that the quality of fit was good, model comparisons (BIC) showed that the quality of fit relative to the pure LBA fit did not warrant the inclusion of the three additional parameters of the mixture model (i.e., a mixture proportion parameter and the mean and standard deviation of the Gaussian distribution of guess response times). Further, the parameter estimates of the mixture model showed a correlation between temporal precision and response caution ($r(45) = -0.36$, $p = 0.014$, two-sided, uncorrected), but not between temporal precision and drift rate ($r(45) = -0.22$, $p = 0.135$, two-sided, uncorrected). Thus, analyses of the parameters of this mixture model would have led to the same main conclusions.

It is interesting to see that the models with linear additive urgency (equivalent to linearly collapsing bounds) explain behavioral adjustments due to the deadline manipulation better than attention effects or overall threshold adjustments. In previous work, the empirical evidence for the hypothesis that participants employ collapsing bounds has been mixed (Cisek et al., 2009; Hawkins, Forstmann, Wagenmakers, Ratcliff, & Brown, 2015; Hawkins, Wagenmakers, Ratcliff, & Brown, 2015; Miletić, 2016; Milosavljevic, Malmaud, & Huth, 2010; Murphy et al., 2016; Thura et al., 2012; Thura & Cisek, 2016; Van Maanen, Fontanesi, et al., 2016; Voskuilen et al., 2016; Winkel et al., 2014). Various factors contribute to this mixed pattern of results. For one, the mixed results could partially reflect a measurement problem – models with collapsing bounds have no known likelihood functions, and fitting such models relies on noisy simulation-based methods (e.g., Turner & Sederberg, 2014; Turner, Sederberg, & McClelland, 2014). Secondly, it has been suggested that there may be significant individual differences in task strategies (Hawkins, Forstmann, et al., 2015). Our results support this suggestion, showing that a proportion of our subjects adjust overall thresholds rather than collapsing bounds. Finally, the strength of the deadline manipulation we used in Experiment 2 may provide a much stronger incentive to collapse decision bounds than the increase in reward rate in tasks without a strong deadline. Indeed, the increase in reward rate due to collapsing bounds relative to fixed bounds has been shown to be rather small in many situations (Boehm, 2018). Future research should shed additional light on the circumstances in which people employ collapsing bounds.

Possibly, our lack of evidence for a relation between timing ability and the speed of collapsing bounds may be related to the fact that our model includes a *linearly* collapsing bound, rather than curvilinear collapsing bounds suggested by the optimality literature (Boehm et al., 2016; Deneve, 2012; Drugowitsch et al., 2012; Frazier & Yu, 2008; Malhotra et al., 2017a, 2017b; Standage et al., 2011; Thura et al., 2012). It could be that our measure of collapse therefore is too crude to pick up small effects. However, curvilinear collapsing bounds require estimating more parameters using simulation-based methods (Turner & Sederberg, 2014; Turner et al., 2014). Estimating more parameters to capture the relatively small behavioral changes may lead to unwanted covariations between parameters (Miletić, Turner, Forstmann, & Van Maanen, 2017). Moreover, the necessary simulation-based methods introduce more uncertainty in the parameter estimates potentially decreasing or hiding crucial relations between the decision-making process and temporal cognition. Finally, it has been shown that parameters specifying the shape of curvilinear collapsing bounds correlate strongly with parameters specifying other parts of the decision process, and may be impossible to estimate reliably (Voskuilen et al., 2016).

The results we report shed additional light on the relations between decision-making and temporal cognition in general. Earlier work (Balcı, Freestone, et al., 2011; Balcı, Simen, et al., 2011) argued that, in decision-making without response deadlines, temporal cognition can explain deviations from optimality in decision-making: People that are not able to precisely represent time intervals, are also not able to precisely estimate reward rate. Instead, inaccurate interval timers are biased towards favoring accuracy over reward rate, and respond more cautiously than the optimal policy would prescribe. Here, we show that under response deadlines, this pattern reverses: people that are unable to precisely reproduce time intervals, make *less* cautious decisions under deadline stress. Therefore, both in situations with and without a response deadline, temporal cognition seems paramount in decision-making (Balcı & Simen, 2016; Simen et al., 2016).

Declarations of interest

None.

Code and data availability

Experiment code, data, and data analysis code are available online at https://osf.io/smj2u.

Appendix A. Supplementary material

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cogpsych.2019.01.002.

References

Anders, R., Alario, F., & Van Maanen, L. (2016). The shifted wald distribution for response time data analysis. *Psychological Methods, 21*(3), 309–327.

Ardia, D., Arango, J. O., & Gomez, N. G. (2011). Jump-diffusion calibration using differential evolution. Retrieved from *Wilmott Magazine, 55*, 76–79. http://www.wilmott.com/.

Ardia, D., Boudt, K., Carl, P., Mullen, K. M., & Peterson, B. G. (2011). Differential evolution with DEoptim: An application to non-convex portfolio optimization. Retrieved from *The R Journal, 3*(1), 27–34. https://journal.r-project.org/archive/2011-1/RJournal_2011-1_Ardia~et~al.pdf.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390–412. https://doi.org/10.1016/j.jml.2007.12.005.

Balcı, F., Freestone, D., Simen, P., DeSouza, L., Cohen, J. D., & Holmes, P. (2011). Optimal temporal risk assessment. *Frontiers in Integrative Neuroscience, 5*(September), 1–15. https://doi.org/10.3389/fnint.2011.00056.

Balcı, F., & Simen, P. (2016). A decision model of timing. *Current Opinion in Behavioral Sciences, 8*, 94–101. https://doi.org/10.1016/j.cobeha.2016.02.002.

Balcı, F., Simen, P., Niyogi, R., Saxe, A., Hughes, J. A., Holmes, P., & Cohen, J. D. (2011). Acquisition of decision making criteria: Reward rate ultimately beats accuracy. *Attention, Perception, and Psychophysics, 73*(2), 640–657. https://doi.org/10.3758/s13414-010-0049-7.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278. https://doi.org/10.1016/j.jml.2012.11.001.

Bendtsen, C. (2012). pso: Particle swarm optimization. Retrieved from < https://cran.r-project.org/package=pso > .

Boehm, U. (2018). *Optimal bounds, bounded optimality: Models of impatience in decision-making.* University of Groningen.

Boehm, U., Hawkins, G. E., Brown, S. D., Van Rijn, H., & Wagenmakers, E. J. (2016). Of monkeys and men: Impatience in perceptual decision-making. *Psychonomic Bulletin and Review, 23*(3), 738–749. https://doi.org/10.3758/s13423-015-0958-5.

Boehm, U., Van Maanen, L., Forstmann, B., & Van Rijn, H. (2014). Trial-by-trial fluctuations in CNV amplitude reflect anticipatory adjustment of response caution. *NeuroImage, 96*, 95–105. https://doi.org/10.1016/j.neuroimage.2014.03.063.

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology, 57*(3), 153–178. https://doi.org/10.1016/j.cogpsych.2007.12.002.

Brown, S. D., Steyvers, M., & Wagenmakers, E.-J. (2009). Observing evidence accumulation during multi-alternative decisions. *Journal of Mathematical Psychology, 53*(6), 453–462. https://doi.org/10.1016/j.jmp.2009.09.002.

Busemeyer, J. R. (1985). Decision making under uncertainty: A comparison of simple. *Journal of Experimental Psychology. Learning, Memory, and Cognition, II*(3), 538–564. https://doi.org/10.1037/0278-7393.11.3.538.

Busemeyer, J. R., & Rapoport, A. (1988). Psychological models of deferred decision-making. *Journal of Mathematical Psychology, 32*(2), 91–134.

Cisek, P., Puskas, G. A., & El-Murr, S. (2009). Decisions in changing conditions: The urgency-gating model. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience, 29*(37), 11560–11571. https://doi.org/10.1523/JNEUROSCI.1844-09.2009.

Clerc, M. (2010). Particle swarm optimization (Vol. 93). John Wiley & Sons, Inc.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorials in Quantitative Methods for Psychology, 1*(1), 42–45. https://doi.org/10.20982/tqmp.01.1.p042.

Deneve, S. (2012). Making decisions with unknown sensory reliability. *Frontiers in Neuroscience, 6*(JUN), 1–13. https://doi.org/10.3389/fnins.2012.00075.

Donkin, C., Brown, S. D., & Heathcote, A. (2009). The overconstraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin and Review, 16*(6), 1129–1135. https://doi.org/10.3758/PBR.16.6.1129.

Donkin, C., Brown, S. D., Heathcote, A., & Wagenmakers, E. J. (2011). Diffusion versus linear ballistic accumulation: Different models but the same conclusions about psychological processes? *Psychonomic Bulletin and Review, 18*(1), 61–69. https://doi.org/10.3758/s13423-010-0022-4.

Donkin, C., & Van Maanen, L. (2014). Piéron's Law is not just an artifact of the response mechanism. *Journal of Mathematical Psychology, 62–63*, 22–32. https://doi.org/10.1016/j.jmp.2014.09.006.

Drugowitsch, J., Moreno-Bote, R., Churchland, A. K., Shadlen, M. N., & Pouget, A. (2012). The cost of accumulating evidence in perceptual decision making. *Journal of Neuroscience, 32*(11), 3612–3628. https://doi.org/10.1523/JNEUROSCI.4010-11.2012.

Drugowitsch, J., Moreno-Bote, R., & Pouget, A. (2014). Optimal decision-making with time-varying evidence reliability. *Advances in Neural Information Processing Systems,* 1–9.

Forstmann, B. U., Dutilh, G., Brown, S. D., 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*(45), 17538–17542. https://doi.org/10.1073/pnas.0805903105.

Forstmann, B. U., Tittgemeyer, M., Wagenmakers, E.-J., Derrfuss, J., Imperati, D., & Brown, S. D. (2011). The speed-accuracy tradeoff in the elderly brain: A structural model-based approach. *Journal of Neuroscience, 31*(47), 17242–17249. https://doi.org/10.1523/JNEUROSCI.0309-11.2011.

Frazier, P. I., & Yu, A. J. (2008). Sequential hypothesis testing under stochastic deadlines. *Advances in Neural Information Processing Systems,* 1–8. https://doi.org/10.1.1.304.7658.

Hawkins, G. E., Brown, S. D., Steyvers, M., & Wagenmakers, E.-J. (2012). Context effects in multi-alternative decision making: Empirical data and a Bayesian model. *Cognitive Science, 36*(3), 498–516. https://doi.org/10.1111/j.1551-6709.2011.01221.x.

Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R., & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *Journal of Neuroscience, 35*(6), 2476–2484. https://doi.org/10.1523/JNEUROSCI.2410-14.2015.

Hawkins, G. E., Wagenmakers, E.-J., Ratcliff, R., & Brown, S. D. (2015). Discriminating evidence accumulation from urgency signals in speeded decision making. *Journal of Neurophysiology, 114*(1).

Heitz, R. P. (2014). The speed-accuracy tradeoff: History, physiology, methodology, and behavior. *Frontiers in Neuroscience, 8*(8 JUN), 1–19. https://doi.org/10.3389/fnins.2014.00150.

Hoeting, J. J. A., Madigan, D., Raftery, A. E. A., & Volinsky, C. T. (1999). Bayesian model averaging: A tutorial. *Statistical Science, 14*(4), 382–401. https://doi.org/10.2307/2676803.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics, 6*(2), 65–70.

Karşılar, H., Simen, P., Papadakis, S., & Balcı, F. (2014). Speed accuracy trade-offunder response deadlines. *Frontiers in Neuroscience, 8*, 1–18. https://doi.org/10.3389/fnins.2014.00248.

Kass, B., & Raftery, A. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795.

Kononowicz, T. W., & Van Rijn, H. (2011). Slow potentials in time estimation: The role of temporal accumulation and habituation. *Frontiers in Integrative Neuroscience, 5*(September), 1–10. https://doi.org/10.3389/fnint.2011.00048.

Link, S. W. (1971). Applying RT deadlines to discrimination reaction time. *Psychonomic Science, 25*(6), 355–358. https://doi.org/10.3758/BF03335904.

Ly, A., Verhagen, J., & Wagenmakers, E. J. (2016). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology, 72*, 19–32. https://doi.org/10.1016/j.jmp.2015.06.004.

Macar, F., Vidal, F., & Casini, L. (1999). The supplementary motor area in motor and sensory timing: Evidence from slow brain potential changes. *Experimental Brain Research, 125*(3), 271–280. https://doi.org/10.1007/s002210050683.

Malhotra, G., Leslie, D. S., Ludwig, C. J. H., & Bogacz, R. (2017a). Overcoming indecision by changing the decision boundary. *Journal of Experimental Psychology: General, 146*(6), 776–805. https://doi.org/10.1037/xge0000286.

Malhotra, G., Leslie, D. S., Ludwig, C. J. H., & Bogacz, R. (2017b). Time-varying decision boundaries: Insights from optimality analysis. *Psychonomic Bulletin and Review, 1–26*. https://doi.org/10.3758/s13423-017-1340-6.

Miletić, S. (2016). Neural evidence for a role of urgency in the speed-accuracy trade-off in perceptual decision-making. *The Journal of Neuroscience, 36*(22), 5909–5910. https://doi.org/10.1523/JNEUROSCI.0894-16.2016.

Miletić, S., Turner, B. M., Forstmann, B. U., & Van Maanen, L. (2017). Parameter recovery for the Leaky Competing Accumulator model. *Journal of Mathematical Psychology, 76*, 25–50. https://doi.org/10.1016/j.jmp.2016.12.001.

Milosavljevic, M., Malmaud, J., & Huth, A. (2010). The Drift Diffusion Model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *October, 5*(6), 437–449. https://doi.org/10.2139/ssrn.1901533.

Mulder, M. J., Keuken, M. C., Van Maanen, L., Boekel, W., Forstmann, B. U., & Wagenmakers, E.-J. (2013). The speed and accuracy of perceptual decisions in a

random-tone pitch task. *Attention, Perception & Psychophysics, 75*, 1048–1058.
Mulder, M. J., & Van Maanen, L. (2013). Are accuracy and reaction time affected via different processes? *PLoS One, 8*(11), https://doi.org/10.1371/journal.pone.0080222.
Mulder, M. J., Wagenmakers, E.-J., Ratcliff, R., Boekel, W., & Forstmann, B. U. (2012). Bias in the brain: A diffusion model analysis of prior probability and potential payoff. *Journal of Neuroscience, 32*(7), 2335–2343. https://doi.org/10.1523/JNEUROSCI.4156-11.2012.
Mullen, K. M., Ardia, D., Gil, D., Windover, D., & Cline, J. (2011). "DEoptim": An R package for global optimization by differential evolution. *Journal of Statistical Software, 40*(6), 1–26.
Murphy, P. R., Boonstra, E., & Nieuwenhuis, S. (2016). Global gain modulation generates time-dependent urgency during perceptual choice in humans. *Nature Communications, 7*, 1–14. https://doi.org/10.1038/ncomms13526.
Nunez, M. D., Vandekerckhove, J., & Srinivasan, R. (2017). How attention influences perceptual decision making: Single-trial EEG correlates of drift-diffusion model parameters. *Journal of Mathematical Psychology, 76B*. https://doi.org/10.1016/j.jmp.2016.03.003.
Palmer, J., Huk, A. C., & Shadlen, M. N. (2005). The effect of stimulus strength on the speed and accuracy of a perceptual decision. *Journal of Vision, 5*(5), 1. https://doi.org/10.1167/5.5.1.
Price, K. V., Storn, R. M., & Lampinen, J. A. (2006). *Differential evolution – A practical approach to global optimization.* Springer-Verlag.
R Core Team (2017). *R: A language and environment for statistical computing.* Vienna: Austra Retrieved from https://www.r-project.org/.
Rae, B., Heathcote, A., Donkin, C., Averell, L., & Brown, S. D. (2014). The hare and the tortoise: Emphasizing speed can change the evidence used to make decisions. *Journal of Experimental Psychology. Learning, Memory, and Cognition,* 1–39. https://doi.org/10.1037/a0036801.
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. https://doi.org/10.1037/0033-295X.95.2.238.
Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology, 53*(3), 195–237. https://doi.org/10.1016/j.cogpsych.2005.10.002.
Ratcliff, R., Smith, P. L., Brown, S. D., & McKoon, G. (2016). Diffusion decision model: Current issues and history. *Trends in Cognitive Sciences, 20*(4), 260–281. https://doi.org/10.1016/j.tics.2016.01.007.
Ratcliff, R., Thapar, A., & McKoon, G. (2006). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin and Review, 13*(4), 626–635. https://doi.org/10.3758/BF03193973.
Ratcliff, R., & Tuerlinckx, F. (2002). Estimating parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin & Review, 9*(3), 438–481.
Ridderinkhof, K. R. (2002). Activation and suppression in conflict tasks: Empirical clarification through distributional analyses. *Common Mechanisms in Perception and Action, 68335*, 26.
Schulz-Zhecheva, Y., Voelkle, M., Beauducel, A., Biscaldi, M., & Klein, C. (2016). Predicting fluid intelligence by components of reaction time distributions from simple choice reaction time tasks. *Journal of Intelligence, 4*(3), 8. https://doi.org/10.3390/jintelligence4030008.
Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464. https://doi.org/10.1214/aos/1176344136.
Simen, P., Balcı, F., deSouza, L., Cohen, J. D., & Holmes, P. (2011a). A model of interval timing by neural integration. *Journal of Neuroscience, 31*(25), 9238–9253. https://doi.org/10.1523/JNEUROSCI.3121-10.2011.
Simen, P., Balcı, F., deSouza, L., Cohen, J. D., & Holmes, P. (2011b). Interval timing by long-range temporal integration. *Frontiers in Integrative Neuroscience, 5*(July), 1–2. https://doi.org/10.3389/fnint.2011.00028.
Simen, P., Vlasov, K., & Papadakis, S. (2016). Scale (In)variance in a unified diffusion model of decision making and timing, 123(2), 151–181. https://doi.org/10.1037/rev0000014.
Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review, 116*(2), 283–317. https://doi.org/10.1037/a0015156.
Standage, D., You, H., Wang, D.-H., & Dorris, M. C. (2011). Gain modulation by an urgency signal controls the speed-accuracy trade-off in a network model of a cortical decision circuit. *Frontiers in Computational Neuroscience, 5*(February), 1–14. https://doi.org/10.3389/fncom.2011.00007.
Thura, D., Beauregard-Racine, J., Fradet, C.-W., & Cisek, P. (2012). Decision making by urgency gating: Theory and experimental support. *Journal of Neurophysiology, 108*(11), 2912–2930. https://doi.org/10.1152/jn.01071.2011.
Thura, D., & Cisek, P. (2016). Modulation of premotor and primary motor cortical activity during volitional adjustments of speed-accuracy trade-offs. *Journal of Neuroscience, 36*(3), 938–956. https://doi.org/10.1523/JNEUROSCI.2230-15.2016.
Turner, B. M., & Sederberg, P. B. (2014). A generalized, likelihood-free method for posterior estimation. *Psychonomic Bulletin & Review, 21*(2), 227–250. https://doi.org/10.3758/s13423-013-0530-0.
Turner, B. M., Sederberg, P. B., & McClelland, J. L. (2014). Bayesian analysis of simulation-based models. *Journal of Mathematical Psychology*. https://doi.org/10.1016/j.jmp.2014.10.001.
Van Maanen, L. (2016). Is there evidence for a mixture of processes in speed-accuracy trade-off behavior? *Topics in Cognitive Science, 8*(1), 279–290. https://doi.org/10.1111/tops.12182.
Van Maanen, L., Brown, S. D., Eichele, T., Wagenmakers, E.-J., Ho, T. C., Serences, J. T., & Forstmann, B. U. (2011). Neural correlates of trial-to-trial fluctuations in response caution. *Journal of Neuroscience, 31*(48), 17488–17495. https://doi.org/10.1523/JNEUROSCI.2924-11.2011.
Van Maanen, L., Katsimpokis, D., & Van Campen, A. D. (2018). Fast and slow errors: Logistic regression to identify patterns in accuracy-response time relationships. *Behavior Research Methods*. https://doi.org/10.3758/s13428-018-1110-z.
Van Maanen, L., Fontanesi, L., Hawkins, G. E., & Forstmann, B. U. (2016). Striatal activation reflects urgency in perceptual decision making. *NeuroImage, 139*, 294–303. https://doi.org/10.1016/j.neuroimage.2016.06.045.
Van Maanen, L., Forstmann, B. U., Keuken, M. C., Wagenmakers, E. J., & Heathcote, A. (2016). The impact of MRI scanner environment on perceptual decision-making. *Behavior Research Methods, 48*(1), 184–200. https://doi.org/10.3758/s13428-015-0563-6.
Van Ravenzwaaij, D., Brown, S., & Wagenmakers, E. J. (2011). An integrated perspective on the relation between response speed and intelligence. *Cognition, 119*(3), 381–393. https://doi.org/10.1016/j.cognition.2011.02.002.
Voskuilen, C., Ratcliff, R., & Smith, P. L. (2016). Comparing fixed and collapsing boundary versions of the diffusion model. *Journal of Mathematical Psychology, 73*, 59–79. https://doi.org/10.1016/j.jmp.2016.04.008.
Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition, 32*(7), 1206–1220. https://doi.org/10.3758/BF03196893.
Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review, 11*(1), 192–196. https://doi.org/10.3758/BF03206482.
Wearden, J. H. (2016). Methods commonly used in time perception research. *The psychology of time perception* (pp. 223–232). London: Palgrave Macmillan.
Winkel, J., Keuken, M. C., Van Maanen, L., Wagenmakers, E.-J., & Forstmann, B. U. (2014). Early evidence affects later decisions: Why evidence accumulation is required to explain response time data. *Psychonomic Bulletin & Review, 21*(3), 777–784. https://doi.org/10.3758/s13423-013-0551-8.
Xue, Q., Markkula, G., Yan, X., & Merat, N. (2018). Using perceptual cues for brake response to a lead vehicle: Comparing threshold and accumulator models of visual looming. *Accident Analysis and Prevention, 118*(May), 114–124. https://doi.org/10.1016/j.aap.2018.06.006.