Are participants'reports of their own reaction times reliable? Re-examining introspective limitations in active and passive dual-task paradigms

Donna Bryce, Daniel Bratzke

Angaben zur Veröffentlichung / Publication details:

Bryce, Donna, and Daniel Bratzke. 2017. "Are participants'reports of their own reaction times reliable? Re-examining introspective limitations in active and passive dual-task paradigms." *Acta Psychologica* 172: 1–9. https://doi.org/10.1016/j.actpsy.2016.10.007.

Are participants' reports of their own reaction times reliable? Re-examining introspective limitations in active and passive dual-task paradigms

Donna Bryce *, Daniel Bratzke

Department of Psychology, University of Tübingen, Schleichstrasse 4, 72076 Tübingen, Germany

1. Introduction

Since introspective reports were first collected systematically by Wundt more than a century ago (e.g., Wundt, 1907), they have often been criticized for being unreliable and unsuitable as a research tool (Boring, 1953; Miller, Vieweg, Kruize, & McLea, 2010). However, people frequently use the information gained through introspection in their everyday lives, for example when deciding how much more study time is required before an exam, as well as during experiments. Since this information, even if flawed or limited, most certainly guides and influences human behavior, understanding the limits of introspection is important for understanding behavior. In the present study, we closely examine one previously reported limit of introspection in a dual-task situation, and offer a new introspective method that better reflects the participants' representation of an experimental trial.

The intriguing finding that people appear to be unaware of the substantial costs associated with processing two tasks simultaneously has been interpreted as evidence for a close dependent relationship between consciousness and attention (Dehaene & Changeux, 2011; Dehaene, Charles, King, & Marti, 2014; Marti, Sackur, Sigman, & Dehaene, 2010). However, the original experiments that established this effect (Corallo, Sackur, Dehaene, & Sigman, 2008; Marti et al., 2010) place considerable demands on participants to process two tasks as fast as possible, and then introspect about their reaction times to each task separately. This requires participants to transform the temporal relationships between four events, typically of three different modalities (auditory, visual, and tactile) into two interval estimates. Given these demands, in the present study we examine the apparent unawareness of dual-task costs in more detail. Specifically, we compare participants' introspective reports of their own reaction times with reports of the equivalent time intervals when participants have no dual-task processing demands, using two different reporting methods.

Introspective limitations in a dual-task context were first reported by Corallo et al. (2008) and Marti et al. (2010), and replicated by Bryce and Bratzke (2014, 2015) using the Psychological Refractory Period (PRP) paradigm. In this task, participants are presented two choice reaction time tasks with a variable temporal gap between them (called the stimulus onset asynchrony, SOA). Participants are instructed to respond as quickly as possible to each stimulus. The typical finding is that responses to the second task (RT2) are much slower when the two tasks are presented close together in time (short SOA condition) than when there is a longer interval between them (long SOA condition). In contrast, responses to the first task are unaffected by SOA (Pashler, 1994). This effect on RT2 is referred to as the PRP effect. Even though the PRP effect can be as large as 500 ms, introspective PRP experiments have found that participants' estimates of their reaction times (called

* Corresponding author at: Fachbereich Psychologie, Universität Tübingen, Schleichstrasse 4, 72076 Tübingen, Germany.

E-mail addresses: donna.bryce@uni-tuebingen.de (D. Bryce), daniel.bratzke@uni-tuebingen.de (D. Bratzke).

introspective reaction times, iRTs) do not reflect this large dual-task cost. That is, there is no SOA effect on iRT1 or iRT2.

This finding of an unawareness of the PRP effect was interpreted by Marti et al. (2010) as evidence that attention is required for conscious awareness. They reasoned that while a participant's attention is focused on the first task, they cannot become aware of another stimulus presented in that time. Thus, at short SOAs their conscious awareness of the second stimulus is delayed in time until they have reached a decision on Task 1 (i.e. the end of Task 1 central processing). This delayed awareness of the second stimulus leads to an underestimation of RT2 at short SOAs, and thus the lack of a PRP effect in iRT2.

This interpretation of a 'consciousness bottleneck' being at play in the PRP paradigm is in line with the Global Neuronal Workspace model (Dehaene et al., 2014), and with the associated taxonomy of conscious perception (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006). This taxonomy states that both sufficiently strong bottom-up stimulus strength and top-down attention are required for conscious awareness. According to this taxonomy, while a participant is centrally processing the first task in a PRP trial, the second stimulus is 'preconscious' and it can only enter conscious perception when attention is freed from Task 1 processing. This is also consistent with the dominant explanation for the related Attentional Blink phenomenon (Raymond, Shapiro, & Arnell, 1992) — when two target stimuli (e.g. letters) are presented in a stream of distractors (e.g. numbers), it is thought that the second target cannot be perceived when it follows the first one within a certain period of time because the first target has not yet been consolidated (i.e. attentional resources are not yet available again). While some researchers claim that this conscious awareness of the second target is 'all or nothing' (Sergent & Dehaene, 2004), other studies that additionally collected confidence judgements from participants (Pincham & Szücs, 2011) or used post-decision wagering (Nieuwenhuis & de Kleijn, 2011) support a more gradual change in conscious awareness in the Attentional Blink.

While the aforementioned explanation of the unawareness of dual-task costs has attracted a lot of interest within consciousness research, and is as yet the only proposed model of conscious access in the PRP paradigm, there are reasons to question it. For instance, we have evidence that iRTs in a PRP trial may be influenced by the feeling of difficulty regarding each task (Bryce & Bratzke, 2014; see also Hartley, Maquestiaux, Brooks, Festini, & Frazier, 2012). Additionally, participants' reports of stimulus onsets in a PRP trial seem to be influenced by the timing of their own responses (i.e. their inter-response interval; Bratzke, Bryce, & Seifried-Dübon, 2014) and other temporal information provided during the trial (Bratzke & Bryce, 2016). Further, the original method employed for providing iRTs, clicking on a visual analogue scale (VAS) labelled with millisecond values, is not a typical method in timing research (see also Bratzke & Bryce, 2016). However, counter to this criticism, the same result pattern has been observed with a more typical timing method, namely interval reproduction (Bryce & Bratzke, 2015). It should also be noted that in the original introspective PRP study (Corallo et al., 2008) an attempt was made to validate the use of VASs using a single task with varying difficulty levels. While participants could fairly accurately report the effect of the difficulty manipulation on their RTs, clearly a trial with only one task is much simpler than a PRP trial and therefore this finding does not validate this method for the dual-task situation.

Indeed, introspective PRP experiments present substantial timing challenges to participants — they must convert four events, which can be in different modalities and occur in different orders depending on the SOA condition, into two intervals to report their iRT1 and iRT2. Evidence from timing research indicates that even in simple timing tasks (when there are no additional task processing demands) these challenges can introduce inaccuracies. For instance, Grondin and Rousseau (1991) and Gontier, Hasuo, Mitsudo, and Grondin (2013) demonstrated that empty intervals marked by different modalities at the start and end (e.g. auditory and visual stimuli) are timed less accurately than those marked by the same modality (e.g. two auditory stimuli). Grondin, Kuroda, and Mitsudo (2011) also reported that temporal discrimination is poorer for intervals marked by tactile stimulation of different hands than those marked by tactile stimulation of the same hand. Additionally, timing can be distorted when participants have to time two intervals that temporally overlap (as is the case in most short SOA trials), as compared to when each interval presented one after the other (Brown & West, 1990; Bryce & Bratzke, 2016; Bryce, Seifried-Dübon, & Bratzke, 2015; van Rijn & Taatgen, 2008). Finally, we know that timing accuracy also deteriorates when attention must be shared between timing and some other non-temporal task (Block, Hancock, & Zakay, 2010; Brown, 1997). Such timing inaccuracies might play a role in introspective PRP experiments, and could be misinterpreted as a limitation in introspection.

Given the substantial timing challenges in introspective PRP experiments, it seems crucial to establish whether participants are able to time the intervals in a PRP trial even when there are no dual-task processing demands. This would inform us about whether the previously observed iRT result pattern truly reflects a limitation of introspection (i.e., conscious access) or whether it arises from limitations in our timing abilities. To this end, we first 'replayed' the trials from a previously conducted introspective PRP experiment (Experiment 1 in Bryce & Bratzke, 2014) to new naïve participants as a pure timing task (i.e. they did not respond to either stimulus) and collected their estimates of the intervals that represent RT1 and RT2 (referred to as replay RTs, rRT1 and rRT2) via the VAS (Experiment 1). In the previously conducted introspective PRP experiment, we observed no SOA effect on iRT1 and iRT2. By assessing the reported RTs in the replay experiment, we can infer the impact of timing demands on the iRTs in the original introspective PRP experiment. To elaborate, the original introspective and replay experiments make the same demands on timing, but differ in their introspective demands. Thus, any deviation of the rRTs from the iRTs can be attributed to introspective limitations, and any deviation of the rRTs from the objective RT pattern can be attributed to timing demands or the demands of the specific methodology used (in this case, the VAS). If timing demands do not contribute to the apparent unawareness of the PRP effect we expected participants to fairly accurately report the objective RT effects in Experiment 1.

2. Experiment 1

In Experiment 1, trials from a previously conducted introspective PRP experiment in which participants processed a typical PRP task and reported their iRTs on VASs (Experiment 1 in Bryce & Bratzke, 2014) were replayed to new participants. That is, for example, participant 1 of our Experiment 1 experienced exactly the same trials as participant 1 of the original experiment, presented in the same order. In this replay experiment, participants simply experienced a series of stimuli (auditory, visual, left-hand stimulation, right-hand stimulation) and then reported the interval between the auditory stimulus and the left-hand stimulation (rRT1), and the interval between the visual stimulus and the right-hand stimulation (rRT2), on a VAS.

2.1. Method

2.1.1. Participants

Data from 16 participants were analyzed; fifteen participants were female and one was male, fifteen were right-handed. Participants were aged between 18 and 26 years ($M = 21.3$ years).

2.1.2. Apparatus and stimuli

Visual and auditory stimuli were presented on a Mac computer via Matlab with the PsychToolbox extension (Brainard, 1997; Pelli, 1997). Stimulus 1 (S1) was a high or low frequency tone (440 or 880 Hz, 60 dB SPL) presented for 150 ms via headphones. Stimulus 2 (S2) was a plus or minus symbol presented in the center of the screen for

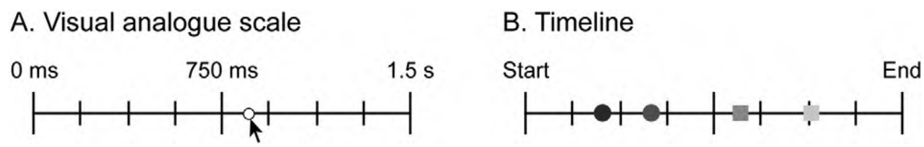

Fig. 1. Illustration of the visual analogue scale (A) used in Experiment 1, and the timeline (B) used in Experiments 2A and 2B. In the visual analogue scale experiment, participants provided their estimates of reaction times by clicking on the scale with the mouse. In the timeline experiments, participants moved markers representing each event in the trial with keyboard presses, and confirmed that their recreation of the trial was complete via the spacebar. S1: blue circle; S2: red circle; R1/left vibration: green square; R2/right vibration: yellow square.

500 ms, degraded by a square area (1.81 × 1.81°) of a random dot pattern. There were 12 levels of degradation for each symbol (25 to 300 dots, in steps of 25).[1] The stimuli and experimental set up were the same as in Bryce and Bratzke (2014), with the exception that in the original introspective experiment participants responded to the tone and symbol via external response boxes. In the current experiment, the stimuli representing original responses were tactile stimulation for 150 ms, delivered by vibrating hand held devices held in each hand (TheraTapper™). Reports of reaction times were collected via a horizontally presented VAS (29.5°) marked with vertical lines and labelled 0 ms (left end), 750 ms (center), and 1.5 s (right end; see Fig. 1A). Participants clicked with the mouse on the scale to indicate their rRT, and a marker appeared in the location they had clicked (again, this procedure was identical to the procedure used in Bryce & Bratzke, 2014).

2.1.3. Procedure and design

Each trial began with a fixation point for 250 ms. Next the tone was presented (S1) followed by the symbol (S2) separated by a stimulus onset asynchrony (SOA) of 50, 250 or 1250 ms. The time at which the original participant gave their left-hand response was represented by the left-hand device vibrating; the time at which the original participant gave their right-hand response was represented by the right-hand device vibrating.

After each PRP trial replay, participants were asked to indicate rRT1 and rRT2 on a VAS always in that order (consistent with the method of collecting iRTs in most previous introspective PRP experiments). In order to achieve the highest possible degree of comparability with our previous iRT results, in the current experiment participants experienced exactly the same trials in the same order as in Experiment 1 in Bryce and Bratzke (2014); this included a practice block at the beginning of each half of the experiment, although in the current experiment the second practice block was not labelled as such. Every combination of stimuli was presented twice: 3 SOA (50, 250, and 1250 ms) × 24 visual stimuli (plus or minus symbols, each with 12 levels of degradation) × 2 auditory stimuli (low or high tones), resulting in 288 experimental trials.

In creating the trials for the current experiment, extreme RTs (over 10 s) in practice trials were replaced with average RTs for that SOA condition. Otherwise, nothing was changed in the temporal structure of the trials between the original introspective PRP experiment (Experiment 1 in Bryce & Bratzke, 2014) and the current Experiment 1. Participants were not informed that the sequence of stimuli they experienced represented another participant's performance, nor were they informed of any response rules. Thus, we ensured that participants were unable to covertly process the PRP task.

2.1.4. Analysis

Data from practice blocks were not included in the analyses. For all analyses, the 12 levels of S2 degradation were categorized into two levels of Task 2 degradation (low vs. high; in the 'active' PRP processing experiment this was equivalent to two levels of Task 2 difficulty, easy vs. hard). All trials which were removed from the original dataset were also not included in the current analyses. These were trials in which the original participant made an error in response to Task 1 or Task 2, trials in which RT1 or RT2 deviated more than three standard deviations from the individual mean in each condition, and trials in which the responses were grouped (within 100 ms of each other). Mean rRTs were then analyzed in SOA × Task 2 degradation repeated measures ANOVAs. To enhance readability of the results, only significant results ($p < 0.05$) are reported. The Greenhouse–Geisser correction was used to adjust p-values where appropriate, Tukey contrasts were used for post-hoc comparisons, and partial Eta-squared effect sizes are provided. Standard errors for within-subjects designs were calculated according to Morey (2008).

2.2. Results

2.2.1. Summary of reaction times and introspective reaction times

In the original introspective PRP experiment (Experiment 1 of Bryce & Bratzke, 2014), reaction times to both tasks decreased with increasing SOA, but RT2 much more so than RT1 (see Fig. 3A of Bryce & Bratzke, 2014). Importantly, there was a PRP effect of 360 ms. Reaction times for the second task were also longer when Task 2 degradation was high ('hard') than when it was low ('easy'), and this effect was more pronounced at long than short SOAs (i.e. an underadditive interaction of SOA and Task 2 difficulty on RT2). Thus, aside from the small SOA effect on RT1, reaction times were as predicted by the central bottleneck model (Pashler, 1994). Crucially, when participants had to both process the PRP task and report their iRTs via a VAS, iRTs were not affected by SOA (see Fig. 3C in Bryce & Bratzke, 2014), demonstrating the standard unawareness of the PRP effect. The only introspective effect that matched the objective RT pattern was that iRT2 was longer when Task 2 degradation was high than when it was low (further details can be found in Bryce & Bratzke, 2014).

2.2.2. Replay reaction times

The time intervals presented to participants (analogous to RTs in Experiment 1 of Bryce & Bratzke, 2014) are presented here in Fig. 2A, and the rRTs provided by participants are presented in Fig. 2B. In the current Experiment 1, both rRT1 and rRT2 increased with increasing SOA (rRT1: $F(2, 30) = 12.99$, $p = 0.001$, $\eta_p^2 = 0.46$; rRT2: $F(2, 30) = 14.36$, $p = 0.001$, $\eta_p^2 = 0.49$). In both cases, post hoc tests indicated that participants' estimates were longer in the longest SOA condition than the two shorter SOA conditions ($ps < 0.001$). Importantly, this pattern is different from both the objective RT pattern and the iRTs collected in Experiment 1 of Bryce and Bratzke (2014). Further, rRT2 was 14 ms longer in the high than the low degradation condition, $F(1, 15) = 6.71$, $p = 0.020$, $\eta_p^2 = 0.31$. Similar to the iRTs observed previously, the Task 2 degradation effect on rRT2 was the only effect that reflected the objective RT pattern.

2.3. Discussion

The results of Experiment 1 indicate that even when participants do not have to process the PRP task, they do not accurately report the effects of SOA and Task 2 degradation on RT1 and RT2. That is, participants in fact reported that both RTs increased with increasing SOA, which is contrary to the objective RT pattern. However, they did correctly report that Task 2 degradation only affected RT2 and not RT1. Since participants cannot accurately report time intervals in PRP trials when timing

[1] Stimulus degradation was manipulated because estimates of 'difficulty' were also collected in Bryce and Bratzke (2014).

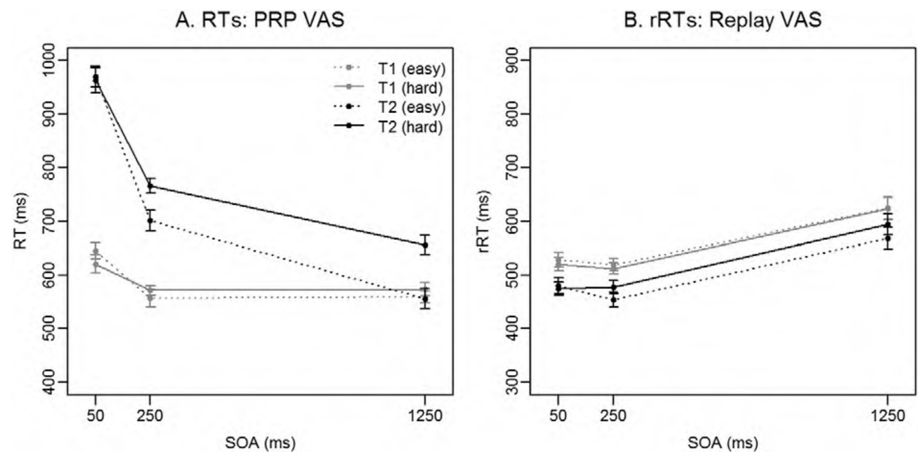

Fig. 2. Mean 'objective reaction times' (RTs) presented to participants (A), and mean replay reaction times (rRTs; B) in Experiment 1 for Task 1 and Task 2 as a function of stimulus onset asynchrony (SOA) and Task 2 degradation. Error bars represent ± 1 within-subjects SE.

is their only task, the timing demands may be at least partly responsible for the data pattern typically observed in introspective PRP experiments. This finding raises questions about the interpretation of the original introspective PRP experiments.

These results were rather unexpected. That is, if timing demands *were not* responsible for the unawareness of the PRP effect, we would have expected rRTs to show the same effects as were observed in objective RTs; if timing demands *were* responsible for the unawareness of the PRP effect, we would have expected rRTs to show the same effects as were observed in iRTs in the original experiment. Instead we observed a third result pattern. One possible explanation for this data pattern is that instead of timing the relevant intervals, participants timed the intervals between sequential events. Thus, at short SOAs iRT1 would be represented by the interval between S1 and S2, and iRT2 would be represented by the interval between the two hand vibrations; at long SOAs iRT1 would be (correctly) represented by the interval between S1 and the left-hand vibration, and iRT2 would be represented by the interval between S2 and the right-hand vibration. When rRTs are modelled according to such an approach using the current data set, an increase in each rRT with increasing SOA, and an interaction of SOA and Task 2 degradation on rRT2 are predicted. As such, it seems likely that this is the strategy participants employed in the current Experiment 1. Importantly, the fact that the rRTs also do not reflect the objective RT pattern raises the question of whether timing limitations are solely responsible for the apparent introspective limitation in the introspective PRP task, or whether the specific demands of the reporting method (i.e. the VAS) also contributes. This question was addressed in Experiment 2, with the introduction of a new method.

3. Experiment 2

Since the results of Experiment 1 showed that participants could not accurately estimate reaction times even when timing was their only task, in Experiment 2 we introduced a new method for collecting participants' introspective representations of the temporal course of each PRP trial. We speculated that in the original method, the demand to convert four events into two intervals to report them on a VAS was perhaps especially problematic for participants. Thus, in Experiments 2A and 2B, we presented participants with a horizontal line (a 'timeline') representing the entire trial, and four markers representing each event in the trial (S1, S2, right-hand, left-hand). Participants moved the markers along the timeline to recreate the temporal structure of the trial. iRTs/rRTs were then calculated as the difference between the relevant markers. In the same manner as Experiment 1, participants in Experiment 2A processed the PRP trial and recreated the trial on a timeline, while participants in Experiment 2B experienced 'replays' of the PRP trials from Experiment 2A and then recreated each trial on a timeline. We hypothesized that if the specific demands of the VAS were responsible for the inaccuracy of rRTs in Experiment 1, participants would more accurately report the timing of PRP trials using the timeline method, and therefore the rRTs in Experiment 2B would show the same data pattern as objective RTs.

3.1. Methods

3.1.1. Participants

In Experiment 2A, three participants had to be replaced due to a technical error, one was replaced because of excessively long responses, and one was replaced because they made too many errors (over 20%). Of the sixteen participants whose data were used, thirteen were female and three were male, fourteen were right-handed, and ages ranged from 20 to 32 years ($M = 24.1$ years). Experiment 2B had thirteen female and three male participants, fourteen of whom were right-handed, and participants were aged between 19 and 25 years ($M = 20.6$ years). All participants in each experiment reported normal hearing and normal or corrected-to-normal vision, and received either course credit or payment.

3.1.2. Apparatus and stimuli

The auditory, visual and tactile stimuli were as described in Experiment 1. In Experiment 2A responses were collected via external button boxes. Since participants took longer to complete one trial using the timeline method than using the VAS method, we used only 6 levels of degradation for each symbol (25 to 275 dots, in steps of 50) in order to maintain an experimental running time of 1.5 h. Additionally, in Experiments 2A and 2B a white frame (6.64 × 6.64°) appeared on the screen for the duration of the trial. The timeline was a horizontal line (29.5°) marked with vertical lines presented on the screen, labelled at each end with 'start' and 'end' (see Fig. 1B). Four markers appeared in the center of the timeline — a blue circle represented S1, a red circle represented S2, a green square represented the left-hand response/vibration, and a yellow square represented the right-hand response/vibration. A legend was provided for participants at the top of the timeline screen, indicating which event each marker represented. Participants could move the markers using the number keys on the computer keyboard which were labelled with stickers in the same

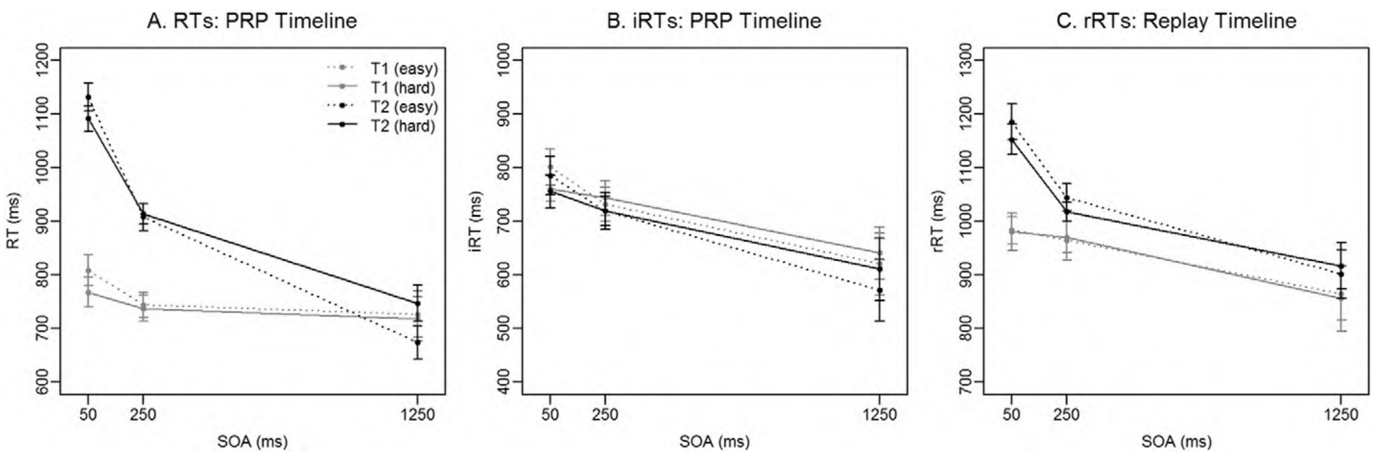

Fig. 3. Mean reaction times (RTs; A), mean introspective reaction times (iRTs; B), and mean replay reaction times (rRTs; C) in Experiments 2A and 2B for Task 1 and Task 2 as a function of stimulus onset asynchrony (SOA) and Task 2 degradation. Error bars represent ± 1 within-subjects SE.

color and shape as the markers. After the recreation of the trial on the timeline, we also collected total trial length estimates from participants, on a horizontal scale marked 2 to 6 s. However, participants reported very little confidence in these estimates, as the information was so degraded by the time they reached this question. Therefore, we do not present the analysis of these data.

3.1.3. Procedure and design

After a fixation point was presented for 250 ms, the start of the trial was indicated by the white frame appearing on the screen, followed by a 1000 ms foreperiod. Next followed the four trial events as described in Experiment 1: the tone and symbol separated by a SOA (50, 250, or 1250 ms) and two responses or vibrations of the hand held devices. After all four events occurred in the trial, there was again a 1000 ms endperiod before the white frame disappeared, representing the end of the trial. After each PRP trial (Experiment 2A) or PRP trial replay (Experiment 2B), participants were asked to recreate the trial on the timeline by moving the markers to the left or right. They pressed the spacebar to confirm the final positions of the markers.

In Experiment 2A participants completed two practice blocks of 18 trials each — in the first one they only responded to the PRP trial, and in the second one they responded to the PRP trial and then completed the timeline after every trial. In Experiment 2B, participants were instructed they had one practice block of 18 trials, but the trials from the second practice block were also excluded from further analysis. After this, participants experienced 6 blocks of 36 trials each. In each experiment, every combination of stimuli was presented three times: 3 SOA (50, 250, and 1250 ms) $\times$ 12 visual stimuli (plus or minus symbols, each with 6 levels of degradation) $\times$ 2 auditory stimuli (low or high tones), resulting in 216 experimental trials.

As before, in creating the trials for Experiment 2B extreme RTs (over 10 s) in practice trials were replaced with average RTs for that SOA condition, and participants in Experiment 2B were not informed that they were in a replay scenario, nor about any response rules.

3.1.4. Analysis

The final x-coordinate of each marker was converted to a millisecond value, based on the timeline representing the objective duration of that trial.[2] Subsequently, iRTs and rRTs were calculated as the difference (in ms) between the relevant marker positions. As in Experiment 1, data from practice blocks were not included in the analyses and the 6 levels of S2 degradation were categorized into two levels of Task 2 degradation (low vs. high). Error rates, mean RTs, iRTs and rRTs were analyzed in repeated measures ANOVAs (SOA $\times$ Task 2 degradation). Trials in which the original participant made an error in response to Task 1 or Task 2 (12%) were removed from both datasets, as were trials in which RT1 or RT2 deviated more than three standard deviations from the individual mean in each condition (2.83% of correct trials), and trials in which the responses were grouped (within 100 ms of each other, 0.30% of remaining trials).

3.2. Results

3.2.1. Error rates

In Experiment 2A, error rates in response to Task 1 were not affected by SOA or Task 2 degradation. However, participants made more errors in Task 2 when degradation was high (12% error rate) than when it was low (5%), $F(1, 15) = 20.58$, $p < 0.001$, $\eta^2_p = 0.58$.

3.2.2. Reaction times

In Experiment 2A, reaction times to Task 1 were not affected by SOA or Task 2 degradation (Fig. 3A). There was a PRP effect of 402 ms on RT2, $F(2, 30) = 57.38$, $p < 0.001$, $\eta^2_p = 0.79$. There was also a significant interaction of SOA and Task 2 degradation on RT2, $F(2, 30) = 5.57$, $p = 0.017$, $\eta^2_p = 0.27$, although it was not precisely as expected. That is, at short SOAs RT2 was even a little longer in the low than the high Task 2 degradation condition. This is probably a carryover from the same effect in RT1 at short SOAs, although this effect did not reach significance.

3.2.3. Introspective reaction times

When participants estimated their own reaction times after each trial via the timeline (Experiment 2A), there were no significant effects of SOA or Task 2 degradation on iRT1 or iRT2 (Fig. 3B). Mean iRTs both decreased with increasing SOA (iRT1 by 151 ms; iRT2 by 180 ms), but these effects did not reach significance (iRT1 $F(2, 30) = 3.32$, $p = 0.083$, $\eta^2_p = 0.18$; iRT2 $F(2, 30) = 4.06$, $p = 0.057$, $\eta^2_p = 0.21$).[3]

3.2.4. Replay reaction times

The rRT results from Experiment 2B are the only ones that at least partly resemble the objective reaction times (Fig. 3C). While rRT1 was

[2] This millisecond value could also have been calculated based on the timeline representing the *subjective* duration of that trial, i.e. the total trial length estimate given by the participant. The objective value was chosen because participants reported very low confidence in their total trial length estimates. In the following results section, a footnote will highlight where the results using this alternative method of calculation qualitatively differ from the main results.

[3] When the millisecond values of iRTs were calculated using the subjective duration of that trial, these effects did reach significance and were approximately the same size (iRT1 $F(2, 30) = 10.76$, $p = 0.004$, $\eta^2_p = 0.42$; iRT2 $F(2, 30) = 10.17$, $p = 0.005$, $\eta^2_p = 0.40$).

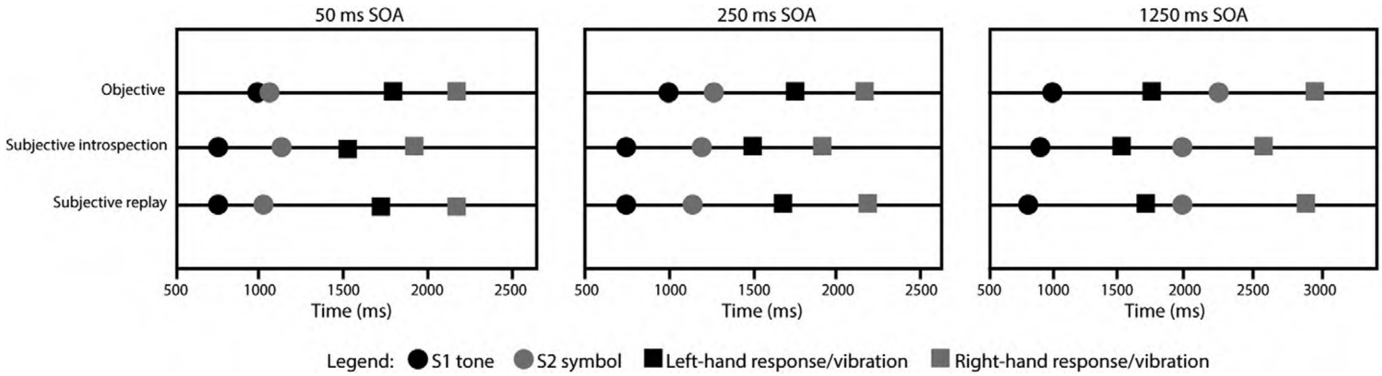

Fig. 4. Mean positions (in ms) of markers in each SOA condition. The first line depicts the objective positions, the second line depicts the positions placed by participants in the introspective condition (Experiment 2A), and the third line depicts the positions placed by participants in the replay condition (Experiment 2B). The x-axis is ±500 ms from the first/last objective marker positions.

not affected by SOA or Task 2 degradation, rRT2 decreased with increasing SOA, $F(2, 30) = 15.03$, $p = 0.001$, $\eta_p^2 = 0.50$.[4] Post hoc tests showed that all contrasts were significant ($ps < 0.05$). Although the reported 'replay' PRP effect (260 ms) was smaller than the objective PRP effect (402 ms), these reports are the only ones to correctly reflect that RT2 was much more affected by SOA than was RT1.

3.2.5. Exploratory analyses of timeline markers

In our planned analyses, iRTs were calculated from the marker positions in order to compare the VAS and timeline methods. However, more information can be gained from the timeline method. Fig. 4 depicts the mean positions of each marker at each SOA, for each condition (objective and subjective introspection positions from Experiment 2A, and subjective replay positions from Experiment 2B). Additional analyses of the marker positions allow us to test some hypotheses that logically emerge from Marti et al.'s (2010) notion that when processing the PRP task, the perception of S2 is delayed until the end of Task 1 processing at short SOAs. One such hypothesis is that the S2 marker should be selectively delayed at short SOAs in the introspection, but not in the replay condition. This was not observed — the SOA was overestimated at short SOAs in both conditions. However, visual inspection of Fig. 4 suggests that the reasons for this common effect may differ across conditions. It appears that in the introspection condition, markers are placed relatively evenly spaced, in the correct order, leading to an overestimation of SOA at short SOAs, and an underestimation at the long SOA. In the replay condition, participants still overestimate the gap between S1 and S2 (the SOA) at short SOAs, but they represent the relationships between markers more faithfully, leading to relatively accurate mean iRT results.

Another hypothesis that emerges from the original conceptualisation of the introspective blindspot is that in the subjective introspection condition, the positions of the S2 and R1 markers should be highly correlated, especially at short SOAs. This is because the conscious awareness of S2 should be linked to the end of Task 1 central processing, which is tightly linked to the Task 1 response. To test this hypothesis, correlations between the x-coordinates of S2 and R1 marker positions were calculated for each subjective condition (introspection and replay), participant, and SOA. Whether the correlation coefficient within each experiment and SOA condition differed significantly from zero was tested using t-tests. When participants first processed the PRP task and then placed the markers on the timeline (Experiment 2A), there were no significant correlations between S2 and R1 marker positions at any SOA; this is contrary to the hypothesis of a delayed conscious perception of S2 due to Task 1 central processing. When participants simply reported the events as they occurred in a replay of a PRP trial (Experiment 2B), S2 and R1 were positively correlated at the shortest SOA ($r = 0.18$, $p = 0.030$) and negatively correlated at the longest SOA ($r = -0.33$, $p < 0.001$). In summary, these exploratory analyses of the timeline marker positions did not provide support for the idea that conscious awareness of S2 is delayed when participants have to process two tasks at short SOA.

3.2.6. Within-method omnibus ANOVAs

In order to statistically verify that participants in Experiment 2B better distinguished the RT patterns for Task 1 and Task 2 than participants in any of the other current experiments and than participants in Experiment 1 of Bryce and Bratzke (2014), we additionally analyzed estimated reaction times (iRTs or rRTs) in larger omnibus ANOVAs (separately for each method). These ANOVAs had the within-subjects factors of SOA, Task 2 degradation, and Task (1 vs. 2) and the between-subjects factor of Experiment (Experiment 1 of Bryce & Bratzke, 2014 vs. current Experiment 1; current Experiment 2A vs. current Experiment 2B). These analyses confirmed our interpretation of the result patterns. That is, there was a significant SOA × Task × Experiment interaction on estimates of reaction times collected using the timeline method, $F(2, 60) = 5.68$, $p = 0.013$, $\eta_p^2 = 0.16$,[5] but not on estimates of reaction times collected using the VAS, $F(2, 60) = 0.02$, $p = 0.923$, $\eta_p^2 < 0.001$.

3.2.7. Across-method ANOVAs

To address whether the different methods (VAS and timeline) affected PRP performance differently, we analyzed mean RT1, mean RT2, Task 1 error rate and Task 2 error rate in additional ANOVAs. These ANOVAs included data from Experiment 1 of Bryce and Bratzke (2014) and the current Experiment 2A. Within-subjects factors SOA and Task 2 degradation, and the between-subjects factor Experiment (Experiment 1 of Bryce & Bratzke, 2014 vs. current Experiment 2A) were included. The only significant result that included the factor Experiment was on RT2 — there was a significant Experiment × Task 2 degradation interaction on RT2, $F(1, 30) = 4.58$, $p = 0.041$, $\eta_p^2 = 0.13$. This probably reflects that while there was a main effect of Task 2 degradation on RT2 in Experiment 1 of Bryce and Bratzke (2014), this main effect was not significant in the current Experiment 2A. In conclusion, there is no evidence that the different methods led to different performance in the PRP task.

[4] When the millisecond values of rRTs were calculated using the subjective duration of that trial, there was a significant main effect of SOA on each estimate indicating that rRTs decreased with increasing SOA (rRT1 $F(2, 30) = 12.62$, $p = 0.002$, $\eta_p^2 = 46$; rRT2 $F(2, 30) = 35.44$, $p < 0.001$, $\eta_p^2 = 0.70$). Importantly, the effect on rRT2 was much greater than the effect on rRT1 (see effect sizes). Thus, our interpretation of the data pattern remains the same even if these values are used — Experiment 2B is the only one in which participants report a much greater effect of SOA on RT2 than RT1 (see Section 3.2.6 for the relevant statistics).

[5] This result was also statistically significant when the millisecond values of estimates were calculated using the subjective trial length, $F(2, 60) = 7.17$, $p = 0.006$, $\eta_p^2 = 0.19$.

3.3. Discussion

When participants processed a PRP trial and then reported what happened in that trial via a timeline (Experiment 2A), they again seemed to be unaware of the PRP effect. This result is a replication of the findings of previous introspective PRP experiments such as Experiment 1 of Bryce and Bratzke (2014), even though the data pattern appears somewhat different. In the current Experiment 2A, both iRT1 and iRT2 slightly decreased with increasing SOA (although this effect did not reach significance). Importantly, however, we observed the same effect on iRT1 and iRT2; thus, there was no indication that the participants noticed that the SOA affected only their responses to the second task. Indeed, this decrease in iRTs with increasing SOA is most likely due to the objective trial length increasing with increasing SOA. A similar decrease with increasing SOA would be observed if participants had placed the markers evenly spaced across the timeline, only changing the order based on SOA condition (short SOAs: S1 S2 R1 R2, long SOA: S1 R1 S2 R2). Alternatively, participants may interpret the longer RT2 at short SOAs as an overall slowing of both reaction times, that is, they may be unable to discriminate the two tasks from each other. In either case, our interpretation of the data pattern observed in Experiment 2A is that participants are indeed unaware of the PRP effect.

The results of Experiment 2B gave the first evidence that participants can report the temporal relationships between events in a PRP trial when they have no dual-task processing demands. While the result pattern was not identical to the objective RT data pattern, participants correctly reported that RT1 and RT2 were differently affected by SOA. An additional analysis indicated that only participants in Experiment 2B were able to correctly distinguish between the SOA effects on RT1 and RT2. It should be noted that participants did not perfectly report the temporal structure of PRP trials in the replay context — the SOA $\times$ Task 2 degradation interaction present in objective RT2 was not reflected in rRT2. However, this is perhaps not surprising given the previously mentioned considerable timing demands in this task. Nevertheless, since performance on the PRP task was not unduly affected by the method of reporting, our results indicate that the timeline method can be used to better represent participants' introspective experiences of PRP trials, especially in comparison to the VAS method.

4. General discussion

The present study aimed to investigate the validity of previous introspective reports in dual-task contexts. The findings of Experiment 1 suggest that introspective reaction times reported via VASs are probably invalid, as participants could not even report these intervals correctly when there were no additional PRP processing demands. In contrast, when using a timeline method, participants who did not have to also process the PRP task could fairly accurately report what happened in a PRP trial (Experiment 2B). Overall, these results indicate that the timeline method presents fewer barriers (e.g., the need to extract two intervals from a temporal sequence of events) to the participants when reporting their experience of a trial, and support its use in future introspective PRP experiments.

The findings of Experiment 1 build on our previous findings that iRTs reported via VASs are biased by other influences such as feelings of difficulty (Bryce & Bratzke, 2014). In the replay context, where participants did not have a feeling of difficulty, rRTs reported via VASs were still not accurate. Unexpectedly, participants reported that both RTs increased with increasing SOA, and that RT2 was slightly longer in trials with highly degraded stimuli than those with less degraded stimuli. As previously suggested in the Discussion of Experiment 1, it seems likely that these rRTs were based on temporal information, but the wrong temporal information, namely the intervals between sequential events. Importantly, the finding that participants cannot time the intervals that represent RT1 and RT2 even when there are no dual-task processing demands, raises doubts about previous interpretations of introspective reports collected via VASs.

In contrast, participants were better able to describe what happened in replay trials using the timeline method. It seems likely that removing the demand to convert four events into two intervals was sufficient to more directly access the participant's representation of the trial. However, even using the timeline, participants were not able to report what happened in a PRP trial when they also had to process the dual-task. That is, the current study also observed a considerable limitation in introspection in a dual-task context, even with an improved reporting method that arguably better reflected participants' representation of the task. The reason for this inability to report the PRP effect when processing the dual-task remains unclear. The existing explanation for the unawareness of the PRP effect – that participants' conscious awareness of the second stimulus is delayed at short SOAs – was not supported by the data from Experiment 2A. That is, there was no evidence that the position of the marker representing S2 was selectively affected at short SOAs, nor that the perception of S2 was linked to the end of Task 1 central processing. Instead, the observed introspective limitation may result from participants never having formed any representation of the trial, or having formed an accurate representation which was then lost before it was reported. As a consequence, participants may recreate an inaccurate representation based on the wrong (i.e. not temporal) information in the trial (for instance the number of events in that trial, see the contextual change model, Block & Reed, 1978). Further evidence for such retrospective inferences in introspective PRP experiments has also been obtained in a recent study using the method of constant stimuli, in which participants had to report in each trial whether a comparison interval was shorter or longer than their RT2 (Bratzke & Bryce, 2016). With this method, it appeared that participants based their estimates almost completely on the temporal information provided by the comparison intervals, rather than their actual RT2 in each trial. Further research is needed to better characterize the limitations of and content of introspection in this attentionally demanding context.

Some limitations of the current study deserve consideration. On a theoretical level, the timeline provides only implicit measures of iRTs (iRTs are calculated based on the placement of markers), whereas the VAS method asks participants for explicit estimates of their reaction times. Thus, while the timeline method may be the better method for investigating the time-course of conscious access in the PRP paradigm, we are still reliant on other methods such as the VAS or temporal reproduction (Bryce & Bratzke, 2015) to assess explicit estimates of participants' own reaction times. It could also be questioned whether a comparison between PRP processing (Experiment 2A) and PRP replay (Experiments 1 and 2B) conditions is appropriate. The major difference between these conditions is that in the PRP processing conditions, participants knew the stimulus-response pairings and therefore that there was a meaningful relationship between the tone and the left-hand response, and the symbol and the right-hand response. Of course, one could argue that there was also a (perhaps weaker) meaningful relationship between these pairings in the replay VAS experiment (current Experiment 1), as the participants were explicitly instructed to estimate the intervals between these events. A further difference between processing and replay conditions is that when participants must process the PRP task, they engage in both sensory and motor processing, whereas in a replay context there is only sensory processing. Relatedly, the motor processing might make the button presses more salient than passive vibration. Counter to this argument, the saliency of the vibrations and the lack of motor processing is consistent across both replay conditions (Experiments 1 and 2B), but participants were only able to represent the temporal features of the trials using the timeline. Thus, our conclusion – that the timeline is a better method for collecting reports of participants' introspective representation of a trial – is nevertheless justified.

Overall, the results of the current study highlight the need to closely scrutinize the potential biases and influences on introspective reports.

They support other findings that introspective reports can be considerably biased. Biases that may play a role in introspective PRP experiments have already been discussed (Bratzke & Bryce, 2016; Bratzke et al., 2014; Bryce & Bratzke, 2014; and Experiment 1 of the present study), but introspective reports have also been found to be influenced by irrelevant information and false-feedback in other contexts. For instance, when participants are asked to estimate the timing of their intent to move in a Libet clock paradigm (in which a clock hand continuously rotates around a clock-face and participants report its location when certain internal or external events occur), their estimates are influenced by the perception of their own response (Banks & Isham, 2009), by the response being classified as a win or loss (Isham, Banks, Ekstrom, & Stern, 2011), and by whether or not they receive a reward (Isham & Geng, 2011). The risk of over-interpreting introspective reports that are influenced by irrelevant information is particularly high in consciousness research, as we have no objective measure of participants' introspection and therefore cannot easily identify biases. Indeed, when participants are asked to give a report, they will do so, even if they are very unsure about their introspection. This was evidenced in a study by Johansson, Hall, Silkström, and Olsson (2005). Participants chose which face they found most attractive out of two, and then sometimes received false feedback about their choice (i.e. they were informed that they had chosen the other face). Surprisingly, participants were mostly unaware of the mismatch between their choice and the feedback, and even offered explanations for the decision they had not made. Perhaps our interpretation of introspective judgements could be strengthened by adding confidence judgements to such tasks, in order to better evaluate the robustness of introspective information.

The field of introspection is certainly not restricted to the estimation of one's own processing time but also encompasses, for example, the study of what could be considered a verbal narrative of our thoughts. In such areas of research, the focus is often on the discrepancy between what people report to be their own attitude or perception, and what their behavior indicates is their true attitude or perception (e.g. see Nisbett & Wilson, 1977). The type of introspection that we investigate in the current experiments is distinct from this in numerous ways. First, we find that participants' introspective representation of the content of a PRP trial (i.e. what events occurred in what order) is more or less accurate; the limitation appears to be primarily regarding the temporal relationships between these events. Second, in relatively simple reaction time tasks it is rather unlikely that participants experience a verbal narrative that can be 'read out' via a method such as think-aloud (van Someren, Barnard, & Sandberg, 1994). Nevertheless, the non-verbal methods used to give introspective reports in the current experiments could perhaps be developed and applied to other contexts in order to help us understand especially the temporal aspects of our mental life that are difficult to translate into words.

Regardless of how introspective reports may be biased, we argue that introspection and its biases are interesting per se, as the information gained through (perhaps sometimes flawed) introspection has consequences for human behavior (for instance, based on an inaccurate feeling of difficulty, one may decide to stop studying for an exam prematurely). Of course, as researchers our aim should be to reduce the biases we introduce to introspective reports via our choice of methods, while striving to understand the ways in which introspective representations are formed and by what information they are influenced. To summarize, in the present study we replicated the finding that participants cannot accurately introspect about their own performance in a dual-task situation, and also argued that the original method for collecting introspective reports in previous introspective PRP experiments (VASs) is probably not a valid method for assessing the subjective temporal representation of a PRP trial. Further, exploratory analyses of data elicited from the timeline method indicate that the original explanation for the unawareness of the PRP effect may need to be reconsidered.

Acknowledgements

Thanks to Anna-Lena Kraft, Lisa Schreiber and Paul Feltes for assistance with data collection. This research is supported by the Deutsche Forschungsgemeinschaft (BR 4297/1-2).

References

Banks, W. P., & Isham, E. A. (2009). We infer rather than perceive the moment we decided to act. *Psychological Science*, *20*(1), 17–21. http://dx.doi.org/10.1111/j.1467-9280.2008.02254.x.

Block, R. A., & Reed, M. A. (1978). Remembered duration: Evidence for a contextual-change hypothesis. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 656–665.

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica*, *134*, 330–343. http://dx.doi.org/10.1016/j.actpsy.2010.03.006.

Boring, E. G. (1953). A history of introspection. *Psychological Bulletin*, *50*, 169–189.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*(4), 433–436. http://dx.doi.org/10.1163/156856897X00357.

Bratzke, D., & Bryce, D. (2016). Temporal discrimination of one's own reaction times in dual-task performance: Context effects and methodological constraints. *Attention, Perception, & Psychophysics*, *78*, 1806–1816. http://dx.doi.org/10.3758/s13414-016-1161-0.

Bratzke, D., Bryce, D., & Seifried-Dübon, T. (2014). Distorted subjective reports of stimulus onsets under dual-task conditions: Delayed conscious perception or estimation bias? *Consciousness and Cognition*, *30*, 36–47. http://dx.doi.org/10.1016/j.concog.2014.07.016.

Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, *59*(7), 1118–1140.

Brown, S. W., & West, A. N. (1990). Multiple timing and the allocation of attention. *Acta Psychologica*, *75*(2), 103–121. http://dx.doi.org/10.1016/0001-6918(90)90081-P.

Bryce, D., & Bratzke, D. (2014). Introspective reports of reaction times in dual-tasks reflect experienced difficulty rather than timing of cognitive processes. *Consciousness and Cognition*, *27*, 254–267. http://dx.doi.org/10.1016/j.concog.2014.05.011.

Bryce, D., & Bratzke, D. (2015). Are introspective reaction times affected by the method of time estimation? A comparison of visual analogue scales and reproduction. *Attention, Perception, & Psychophysics*, *77*, 978–984. http://dx.doi.org/10.3758/s13414-014-0804-2.

Bryce, D., & Bratzke, D. (2016). Multiple timing of nested intervals: Further evidence for a weighted sum of segments account. *Psychonomic Bulletin & Review*, *23*(1), 317–323.

Bryce, D., Seifried-Dübon, T., & Bratzke, D. (2015). How are overlapping time intervals perceived? Evidence for a weighted sum of segments model. *Acta Psychologica*, *156*, 83–95. http://dx.doi.org/10.1016/j.actpsy.2014.12.007.

Corallo, G., Sackur, J., Dehaene, S., & Sigman, M. (2008). Limits on introspection: Distorted subjective time during the dual-task bottleneck. *Psychological Science*, *19*(11), 1110–1117. http://dx.doi.org/10.1111/j.1467-9280.2008.02211.x.

Dehaene, S., & Changeux, J. P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*(2), 200–227. http://dx.doi.org/10.1016/j.neuron.2011.03.018.

Dehaene, S., Changeux, J. P., Naccache, L., Sackur, J., & Sergent, C. (2006). Conscious, preconscious, and subliminal processing: A testable taxonomy. *Trends in Cognitive Sciences*, *10*(5), 204–211. http://dx.doi.org/10.1016/j.tics.2006.03.007.

Dehaene, S., Charles, L., King, J. R., & Marti, S. (2014). Toward a computational theory of conscious processing. *Current Opinion in Neurobiology*, *25*, 76–84. http://dx.doi.org/10.1016/j.conb.2013.12.005.

Gontier, E., Hasuo, E., Mitsudo, T., & Grondin, S. (2013). EEG investigations of duration discrimination: The intermodal effect is induced by an attentional bias. *PloS One*, *8*(8). http://dx.doi.org/10.1371/journal.pone.0074073.

Grondin, S., & Rousseau, R. (1991). Judging the relative duration of multimodal short empty time intervals. *Perception & Psychophysics*, *49*(3), 245–256. http://dx.doi.org/10.3758/Bf03214309.

Grondin, S., Kuroda, T., & Mitsudo, T. (2011). Spatial effects on tactile duration categorization. *Canadian Journal of Experimental PsychologyRevue Canadienne De Psychologie Experimentale*, *65*(3), 163–167. http://dx.doi.org/10.1037/A0022785.

Hartley, A. A., Maquestiaux, F., Brooks, R. D., Festini, S. B., & Frazier, K. (2012). Electrodermal responses to sources of dual-task interference. *Cognitive Affective & Behavioral Neuroscience*, *12*(3), 543–556. http://dx.doi.org/10.3758/s13415-012-0094-x.

Isham, E. A., & Geng, J. J. (2011). Rewarding performance feedback alters reported time of action. *Consciousness and Cognition*, *20*(4), 1577–1585. http://dx.doi.org/10.1016/j.concog.2011.08.004.

Isham, E. A., Banks, W. P., Ekstrom, A. D., & Stern, J. A. (2011). Deceived and distorted: Game outcome retrospectively determines the reported time of action. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(5), 1458–1469. http://dx.doi.org/10.1037/A0023111.

Johansson, P., Hall, L., Sikstrom, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, *310*(5745), 116–119. http://dx.doi.org/10.1126/science.1111709.

Marti, S., Sackur, J., Sigman, M., & Dehaene, S. (2010). Mapping introspection's blind spot: Reconstruction of dual-task phenomenology using quantified introspection. *Cognition*, *115*(2), 303–313. http://dx.doi.org/10.1016/j.cognition.2010.01.003.

Miller, J., Vieweg, P., Kruize, N., & McLea, B. (2010). Subjective reports of stimulus, response, and decision times in speeded tasks: How accurate are decision time reports?

Consciousness and Cognition, *19*(4), 1013–1036. http://dx.doi.org/10.1016/j.concog.2010.06.001.
Morey, R. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64.
Nieuwenhuis, S., & de Kleijn, R. (2011). Consciousness of targets during the attentional blink: A gradual or all-or-none dimension? *Attention, Perception, & Psychophysics*, *73*, 364–373. http://dx.doi.org/10.3758/s13414-010-0026-1.
Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259.
Pashler, H. (1994). Dual-task interference in simple tasks — Data and theory. *Psychological Bulletin*, *116*(2), 220–244. http://dx.doi.org/10.1037/0033-2909.116.2.220.
Pelli, D. G. (1997). The video toolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437–442.
Pincham, H. L., & Szücs, D. (2011). Conscious access is linked to ongoing brain state: Electrophysiological evidence from the attentional blink. *Cerebral Cortex*, *22*(10), 2346–2353. http://dx.doi.org/10.1093/cercor/bhr314.
Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860.
Sergent, C., & Dehaene, S. (2004). Is consciousness a gradual phenomenon? Evidence for an all-or-none bifurcation during the attentional blink. *Psychological Science*, *15*(11), 720–728. http://dx.doi.org/10.1111/j.0956-7976.2004.00748.x.
van Rijn, H., & Taatgen, N. A. (2008). Timing ofmultiple overlapping intervals: How many clocks do we have? *Acta Psychologica*, *129*(3), 365–375. http://dx.doi.org/10.1016/j.actpsy.2008.09.002.
van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. London: Academic Press.
Wundt, W. (1907). Über Ausfrageexperimente und über die Methoden zur Psychologie des Denkens. *Psychologische Studien*, *3*, 301–360.