CrossMark

Temporal discrimination of one's own reaction times in dual-task performance: Context effects and methodological constraints

Daniel Bratzke[1] · **Donna Bryce**[1]

Abstract In this study we used the method of constant stimuli to investigate introspective reaction times in the psychological refractory period (PRP) paradigm under different temporal contexts. Previous introspective PRP studies have mostly used visual analogue scales to assess introspective reaction times and found that participants were largely unaware of the typical dual-task costs that arise in this paradigm (PRP effect). This apparent limitation of introspection has been taken as evidence for a serial processing bottleneck that encompasses response selection as well as conscious perception. In our study, in each trial participants first performed the PRP task and were then presented with a comparison interval that they had to compare with their reaction time to the second task (RT2). Across three experiments, we observed that the subjective estimates of RT2 (i.e., the points of subjective equality) did not reflect the objective pattern but were almost completely biased toward the center of the comparison intervals (asymmetry effect). In a control experiment in which participants discriminated RT2s of other participants without performing the PRP task, this bias was largely reduced. We interpret these results as indicating that in dual-task performance participants acquire only poor temporal representations of their own reaction times, and the apparent unawareness of the PRP effect may reflect disturbed timing abilities rather than a conscious perception bottleneck.

Introspection has been one of the major tools of investigation since the early days of psychological research. Although it was never completely abandoned, with the rise of behaviorism in the beginning of the 20th century introspection fell into disfavor (Boring, 1953; Danziger, 1980). However, recently there has been renewed interest in the limitations of introspection as an object of experimental psychological research, mainly inspired by theoretical developments regarding the relationship between attention and consciousness (e.g., Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006).

To reveal the potential limitations of introspection, several studies used a typical discrete dual-task situation in which two tasks are presented in close succession. In this so-called psychological refractory period (PRP) paradigm, two stimuli (S1 and S2) are presented with varying stimulus onset asynchronies (SOAs), and participants are asked to provide separate speeded responses to each stimulus. The standard finding in this paradigm is that reaction time to the second stimulus (RT2) dramatically increases with decreasing SOA (i.e., the PRP effect). Previous introspective PRP studies reported a striking limitation of introspection in this paradigm: subjective estimates of RT2 (introspective RT2, IRT2) did not reflect the objective PRP effect but were constant across SOAs (Bryce & Bratzke, 2014, 2015; Corallo, Sackur, Dehaene, & Sigman, 2008; Marti, Sackur, Sigman, & Dehaene, 2010).

This apparent unawareness of the PRP effect has been explained by a central processing bottleneck that encompasses response selection as well as conscious perception (Marti, Sigman, & Dehaene, 2012; Tombu et al., 2011). That is, in the PRP task the conscious perception of S2 is disrupted by

✉ Daniel Bratzke
   daniel.bratzke@uni-tuebingen.de

[1] Fachbereich Psychologie, Universität Tübingen, Schleichstrasse 4, 72076 Tübingen, Germany

the central processing of Task 1 (Corallo et al., 2008; Marti et al., 2010). This explanation implies that timing is relatively intact under dual-task conditions and that the unawareness of the PRP effect occurs because only the conscious parts of Task 2 processing can be timed (see Marti et al., 2010). This implication of intact timing abilities, however, contrasts with the common notion that timing itself requires attentional resources (Block, Hancock, & Zakay, 2010; Brown, 1997; Ruthruff & Pashler, 2010), and also with recent findings that overlapping intervals cannot be timed independently even when there is no task other than timing (Bryce & Bratzke, 2016; Bryce, Seifried-Dübon, & Bratzke, 2015; van Rijn & Taatgen, 2008). Nevertheless, there is at least some indirect evidence for the implication of relatively intact timing abilities in the PRP paradigm. In previous studies, participants were sensitive to trial-by-trial variations in their RTs (RT1 and RT2), as indexed by positive correlations between their objective and introspective RTs at both long and short SOAs (Bryce & Bratzke, 2015; Marti et al., 2010), and there was no indication of an increased variability at short SOA as compared with long SOA either in IRTs (Bryce & Bratzke, 2014, 2015; Marti et al., 2010) or other measures related to the perceived onset of Task 2 (Bratzke, Bryce, & Seifried-Dübon, 2014; Marti et al., 2010).

To collect IRTs in the PRP paradigm, previous studies mostly used the so-called method of quantified introspection (Bryce & Bratzke, 2014; Corallo et al., 2008; Marti et al., 2010; for an exception, see Bryce & Bratzke, 2015, who used reproduction), which was established by Corallo et al. (2008) in their first introspective PRP study. In this method, participants indicate their RT estimate on a visual analogue scale (VAS) that represents a certain, more or less arbitrary, RT range and is labeled with millisecond values. Even though this method is not unlike the categorical speed rating scales that were used in a few earlier studies on the subjective perception of simple RTs (e.g., Sanford, 1970), it is certainly uncommon in timing research. Therefore, one major aim of our study was to reexamine IRT2s in the PRP paradigm using a classical psychophysical method, namely the method of constant stimuli. In research on time perception, this method is deemed preferable for relatively brief time intervals (i.e., in the millisecond range) whereas other methods (verbal estimation, production, and reproduction) are more frequently used for longer intervals (Grondin, 2008). The method of constant stimuli typically involves presenting two intervals in each trial (a standard and a comparison) and asking participants to indicate whether the second interval was shorter or longer than the first one. Based on these judgments, a psychometric function can be traced, and estimates of the point of subjective equality (PSE; in this case, this reflects perceived duration of the standard interval) and the discrimination threshold can be derived.

Another aim of our study was to examine the influence of temporal context on how participants perceive their RT to the second task (i.e., on IRT2s) in the PRP paradigm. A long history of timing research has shown that, similar to other physical properties, the subjective experience of the same physical duration can greatly vary depending on the context in which it is embedded (for reviews, see Bausenhart, Bratzke, & Ulrich, 2016; Shi, Church, & Meck, 2013). A famous example of such context effects is the Vierordt effect (Lejeune & Wearden, 2009; Vierordt, 1868): Within a certain range of presented intervals, relatively long intervals are underestimated and relatively short intervals are overestimated. Recent explanations for such context effects differ in their exact mechanism, but they all share the idea that not only the current trial information but also the temporal context is taken into account when participants provide duration estimates or comparison judgments (Bausenhart et al., 2016; Dyjas, Bausenhart, & Ulrich, 2012; Shi et al., 2013).

In the PRP paradigm, the temporal context is determined by the presentation mode of SOAs, which are usually presented in mixed blocks. In light of the ubiquitous contextual effects in time perception, it seems reasonable to assume that the subjective experience of RTs is shaped by this mixed context. In fact, it is possible to explain the previously observed null effect of SOA on IRT2 with recent contextual timing accounts. For example, when participants provide their RT2 estimate in a given trial they might rely on an internal reference (Bausenhart et al., 2016; Dyjas et al., 2012) of all their RT2s, instead of relying on only the currently produced RT2. If in the mixed SOA context this internal reference combines all RT2s irrespective of SOA, this would result in a central tendency effect on IRT2 similar to the Vierordt effect. That is, the relatively long RT2s in short SOA trials would be underestimated, and the relatively short RT2s in long SOA trials would be overestimated. One obvious way to investigate the influence of temporal context on IRTs would be to compare IRTs when SOAs are presented in mixed or in separate blocks, with the latter being atypical in the PRP paradigm. Only a few studies have directly compared RT performance observed in mixed with that observed in blocked SOA contexts in the PRP paradigm, but they found strikingly similar RT patterns for the two presentation modes (Bertelson, 1967; Burns & Moskowitz, 1971). This suggests that the presentation mode does not substantially alter the processing of the two tasks with respect to the bottleneck mechanism, and thus the above-mentioned comparison between mixed and blocked SOAs should be viable.

In our study, we conducted three introspective PRP experiments (and an additional control experiment) that all employed the method of constant stimuli but differed in their temporal context. In all three experiments, in each trial

participants performed the PRP task and were then presented with a comparison interval that could be either shorter or longer than their objective RT2 in that trial (see Fig. 1 for an illustration of the general experimental procedure). They had to indicate whether the comparison interval was longer or shorter than their preceding RT2. Figure 2 shows predicted psychometric functions for discrimination of RT2 at short and long SOA (plotted against absolute duration of comparison intervals) separately for the two hypotheses that participants are either unaware or aware of the PRP effect. If, as previous results suggest (Corallo et al., 2008; Marti et al., 2010), participants are unaware of the PRP effect, then psychometric functions for RT2 at short and long SOA should superimpose—that is, PSE and discrimination threshold should not differ between the two SOA conditions. Note that this prediction holds irrespective of whether the unawareness of the PRP effect is caused by a conscious perception bottleneck or reflects a temporal context effect. In contrast, if participants can fairly accurately time their RT2 and thus are aware of the PRP effect, the psychometric function should be shifted to the right for the short compared to the long SOA condition (because of the longer RT2s at short compared to long SOA). Accordingly, PSE should be larger at short than at long SOA. Note here that according to Weber's law (i.e., discrimination threshold proportionally increases with perceived duration), the psychometric function should be flatter for the short than for the long SOA condition.

In Experiments 1 and 2, comparison intervals were constructed proportionally to the RT2s in each SOA condition, and SOAs were presented either in mixed (Experiment 1) or in separate blocks (Experiment 2). To give a short preview of the results, the SOA effect on PSE differed depending on the presentation mode. Experiment 1 replicated the previously observed unawareness of the PRP effect in the mixed SOA context, showing superimposed psychometric functions for short and long SOA. In the blocked SOA context of Experiment 2, however, psychometric functions clearly differed between the two SOAs in a manner consistent with an awareness of the PRP effect. These first two experiments thus demonstrated the influence of the temporal context on the apparent unawareness of the PRP effect. Because in the first

two experiments the comparison intervals were confounded with the RT2s, Experiment 3a replicated the blocked SOA design of Experiment 2 but employed the same fixed comparison intervals for both SOA conditions. Contrasting with Experiment 2, the results again showed superimposed psychometric functions for both SOA conditions, as in Experiment 1. Taken together with the results of Experiments 1 and 2, this demonstrates that the PSE is strongly biased toward the center of the comparison intervals. In a control experiment (Experiment 3b) we observed that this bias was largely reduced when another group of participants had to discriminate the RT2s collected in Experiments 3a without performing the PRP task.

Experiment 1

In Experiment 1 we used the standard PRP paradigm—that is, trials with short and long SOAs were presented in mixed blocks. The objective RT2s served as standards for the comparison task, and comparison intervals were constructed proportionally to the RT2s in each SOA condition. If participants are unaware of the PRP effect, psychometric functions should not differ between short and long SOA. In contrast, if participants are aware of the PRP effect, the psychometric function for the short SOA should be shifted to the right compared to that of the long SOA.

Method

Participants Four males and 14 females, aged between 18 and 39 years ($M = 22.3$ years), participated in a 1-hr session. Participants reported normal hearing and normal or corrected-to-normal vision, and received either course credit or payment.

Apparatus and stimuli The experiment was run in a sound-attenuated, dimly illuminated room. The experiment was programmed in MATLAB using the Psychophysics Toolbox extension (Brainard, 1997), Version 3.0.8.
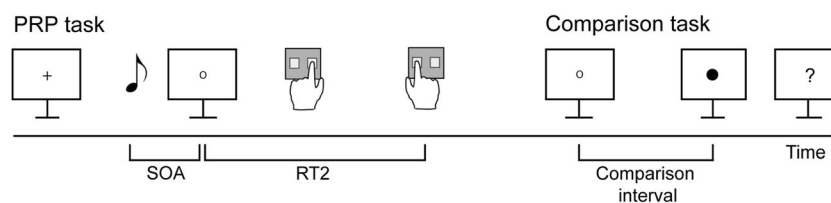
Fig. 1 Illustration of the general trial structure used in this study. The first part of each trial was a standard psychological refractory period (PRP) task with an auditory S1, a visual S2, and left- and right-hand responses to these stimuli, respectively (R1 and R2). In the second part of each trial, a variable comparison interval was presented starting with the reappearance of S2 and ending with a black dot that represented the participant's R2. Finally, participants had to indicate whether the comparison interval was shorter or longer than the RT2 in that trial. Note that the order of S2, R1, and R2 could vary between trials. SOA: Stimulus onset asynchrony; RT2: Reaction time to S2

Fig. 2 Predicted psychometric functions for the hypotheses that participants are either unaware (left) or aware (right) of the psychological refractory period (PRP) effect
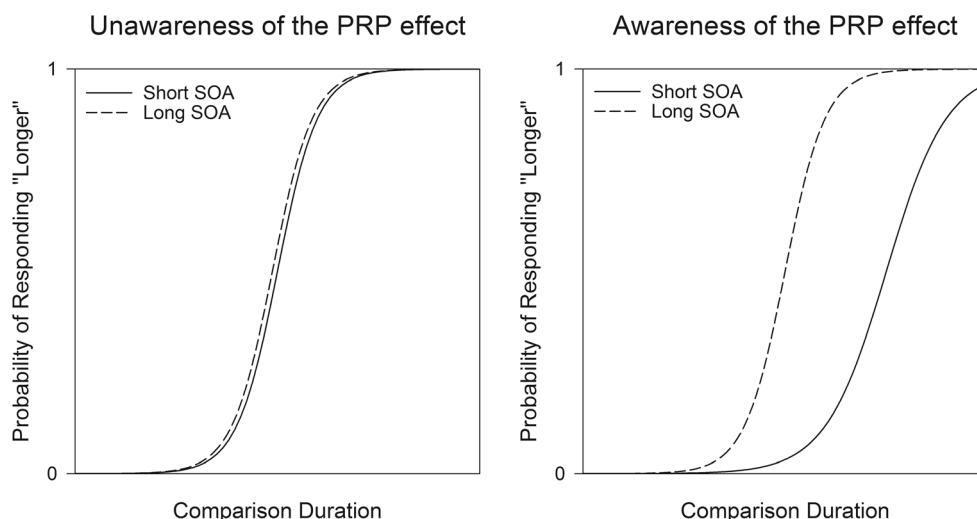
Participants sat in front of a CRT computer screen (150 Hz) with a viewing distance of about 60 cm. Two external response panels were used to record responses with the index and the middle finger of the left and right hand. S1 was a tone of either 440 or 880 Hz, presented via headphones (60 dB SPL, 80-ms duration). S2 was a letter (*O* vs. *X*, 0.3 × 0.5°) presented at the center of the screen for 200 ms. Comparison intervals were constructed by adding or subtracting a certain portion of the current RT2 to or from the current RT2 (±15 %, ±30 %, ±45 %, ±60 %). A black dot (0.7°) was used to indicate the end of the comparison interval.

Procedure and design Each trial started with the presentation of a small fixation point at the center of the screen for 1,000 ms. Then, S1 was presented for 80 ms. S2 presentation followed S1 onset according to one of two SOAs (50 vs. 800 ms). The instruction was to respond as quickly and as accurately as possible with the left index (middle) finger to the high (low) tone and with the right index (middle) finger to the letter *O* (*X*). In trials with correct responses, after another 1,000 ms the comparison interval was presented. This comparison interval started with a further presentation of S2 for 200 ms and was terminated by a black dot representing the participant's response. This dot was presented for the same duration as the participant had pressed down the response key in Task 2 in this trial. Participants were then asked to indicate whether they had perceived the comparison interval as being shorter (left index finger) or longer (right index finger) than their RT2 in this trial. In trials with an incorrect response, one of three possible feedback messages ("Incorrect response to the tone!", "Incorrect response to the letter!", or "Incorrect response to the tone and the letter!") appeared on the screen for 1,000 ms between the dual task and the presentation of the comparison interval. After an additional 1,000 ms, the next trial started.

Each experimental session comprised one practice block and five experimental blocks. Each block consisted of 64 trials (2 Tone × 2 Letter × 2 SOA × 8 Comparison intervals).

Analysis of comparison data As a first step, all trials that included an error in Task 1 and/or Task 2 were discarded. Then, for each factorial combination of participant and SOA, a psychometric function was constructed by plotting the eight comparison levels (±15 %, ±30 %, ±45 %, ±60 %) on the *x*-axis and the relative frequency of responding "longer" on the *y*-axis. A logistic function was fitted to the observed function to compute the maximum likelihood estimates of the point of subjective equality (PSE) and the difference threshold (difference limen; DL). The PSE is estimated as the 50 % point of the fitted function and reflects the comparison level that is perceived as being as long as the standard interval (in this case, mean RT2). That is, a PSE value that is smaller (larger) than mean RT2 would indicate an underestimation (overestimation) of RT2. The DL is estimated as being half the interquartile range of the fitted function. The steeper the psychometric function, and thus the smaller the DL, the more sensitive a participant is to differences between the standard (mean RT2) and the comparison intervals. To compare difference thresholds between different SOA conditions, Weber fractions (WF = DL/PSE) were calculated.

Results and discussion

All trials that included an error in Task 1 and/or Task 2 (9.1 %) were discarded from analyses. Figure 3 depicts RTs (left panel) and averaged psychometric functions for RT2 (right panel) as a function of SOA. RT performance showed the standard

Fig. 3 Reaction time performance in the psychological refractory period (PRP) task and discrimination performance in Experiment 1. Left panel: Mean reaction time in Task 1 and Task 2 as a function of stimulus onset asynchrony (SOA). Right panel: Relative frequency of responding "longer" as a function of mean comparison duration and SOA, and fitted psychometric functions (lines) for all participants' data. Note that statistical analyses were performed with individually fitted psychometric functions. Error bars represent ±1 within-subject standard error (Morey, 2008)
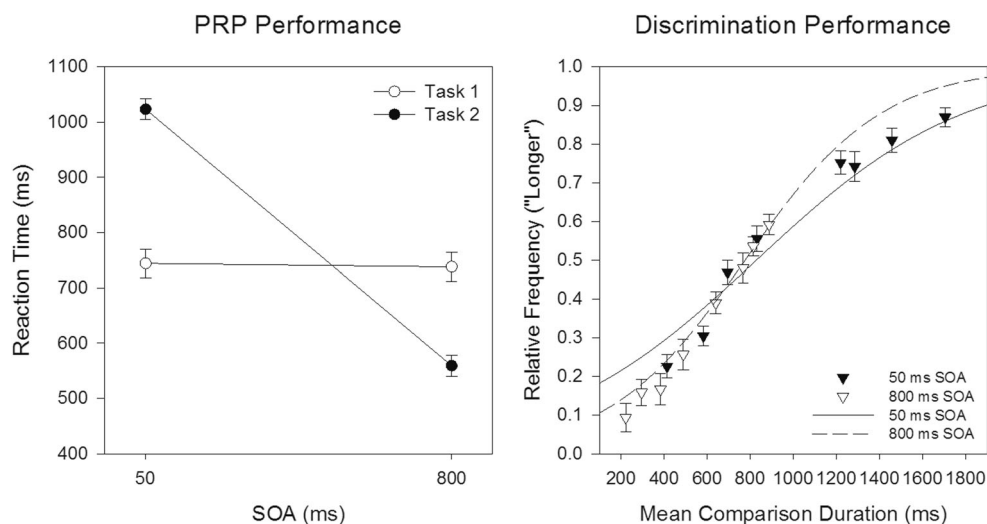
PRP pattern. RT2 was 464 ms longer at short than at long SOA, $t(17) = 16.88$, $p < .001$, Cohen's $d_z = 3.98$, and RT1 was very similar at short (744 ms) and at long SOA (738 ms), $t(17) = 0.17$, $p = .684$, Cohen's $d_z = 0.04$. Inspection of the psychometric functions suggested no difference between the two functions at short and at long SOA. Analysis of the individual fitted psychometric functions revealed that, even though mean PSE was 60 ms larger at short (954 ms) than at long SOA (894 ms), there was no significant effect of SOA on PSE, $t(17) = 1.27$, $p = .220$, Cohen's $d_z = 0.30$. Similarly, difference threshold as indexed by the WF was slightly larger at short (0.48) than at long SOA (0.38); however, this difference was also not significant, $t(17) = 1.98$, $p = .064$, Cohen's $d_z = 0.47$.

This result pattern essentially replicates the previous findings of unawareness of the PRP effect with the method of constant stimuli. However, even at long SOA the PSE was substantially larger than RT2 (i.e., participants overestimated RT2 by 335 ms). This finding contrasts with previous introspective PRP studies that reported IRT2 being generally smaller than RT2 (Bryce & Bratzke, 2015; Corallo et al., 2008; Marti et al., 2010; but see Bryce & Bratzke, 2014, for slight deviations from this pattern) and suggests that participants cannot accurately estimate their RT2 even when the two tasks are temporally separated and thus do not compete for access to central resources.

The apparent unawareness of the PRP effect is consistent with the previously proposed notion that participants underestimate RT2 at short SOA because Task 1 central processing disrupts the conscious perception of S2 (Corallo et al., 2008; Marti et al., 2010). However, as already outlined in the Introduction, this misperception can also be explained by an internal reference account. Accordingly, for the comparison task participants might rely not only on the RT2 from the current trial but on an internal reference based on the current

and all previous RT2s (e.g., Dyjas et al., 2012). Under mixed SOA conditions such an internal reference would integrate the information from all trials irrespective of SOA. This would lead to an internal reference somewhere in between the mean RT2s for short and long SOA trials. Consequently and in line with the results of Experiment 1, psychometric functions would not differ much between short and long SOAs and PSEs would reflect underestimation of RT2 at short SOA and overestimation of RT2 at long SOA. Experiment 2 aimed to test this alternative explanation.

Experiment 2

In Experiment 2 we used the same basic paradigm as in Experiment 1 but used a blocked instead of a mixed SOA design. That is, trials with short SOA were presented in one half of the experiment and trials with long SOA in the other half of the experiment. As in Experiment 1, comparison intervals were constructed proportionally to the RT2s in each SOA condition. If the apparent unawareness of the PRP effect is caused by a delayed conscious perception of S2, we should be able to replicate this finding also in a blocked SOA design. In contrast, if participants form an internal reference of their RT2s, we should no longer observe the *unawareness* effect in the blocked SOA context. This is because, when short and long SOAs are presented in separate instead of mixed blocks, the internal reference would no longer integrate the information from both SOA conditions but rather adapt to the blocked SOA context and thus tend toward the RT2 means for each SOA condition. Accordingly, we would expect a right-shifted psychometric function for the short compared to the long SOA condition (i.e., a larger PSE for short than long SOA).

Method

Participants Three males and 15 females, aged between 20 and 30 years ($M = 23.5$ years), participated in a 1-hr session. Four other participants were tested but had to be replaced because they produced virtually flat psychometric functions. Participants reported normal hearing and normal or corrected-to-normal vision, and received either course credit or payment.

Apparatus and stimuli The apparatus and stimuli were identical to Experiment 1.

Procedure and design The procedure was identical to Experiment 1 with the exception that the two SOA conditions were presented blocked in different halves of the experiment. The order of the SOA conditions was balanced across participants. Half of the participants started with the short SOA, the other half with the long SOA condition. Each experimental session comprised two practice blocks (one before each half of the experiment) and six experimental blocks for each SOA condition. Each block consisted of 64 trials (2 Tone × 2 Letter × 2 SOA × 8 Comparison intervals).

Analysis of comparison data Analysis of the comparison judgments followed the procedure of Experiment 1.

Results and discussion

All trials including an error in Task 1 and/or Task 2 (7.7 %) were discarded from analyses. Figure 4 depicts RTs (left panel) and averaged psychometric functions for RT2 (right panel) as a function of SOA. RT performance was very similar to Experiment 1. RT2 showed a PRP effect of 436 ms, $t(17) = 11.30$, $p < .001$, Cohen's $d_z = 2.66$, and RT1 was not significantly affected by SOA (846 ms at short SOA vs. 771 ms at long SOA), $t(17) = 1.34$, $p = .198$, Cohen's $d_z = 0.32$. In contrast to Experiment 1, however, PSE was affected by SOA, $t(17) = 4.15$, $p < .001$, Cohen's $d_z = 1.14$. In line with the SOA effect on RT2, PSE was 297 ms larger at short (946 ms) than at long SOA (649 ms). Even though participants still overestimated RT2 at long SOA (89 ms), this overestimation was much smaller than the one observed in Experiment 1 (335 ms). As in Experiment 1, WF was relatively large at both SOAs (0.51 at short vs. 0.52 at SOA) and was not affected by SOA, $t(17) = 0.05$, $p = .960$, Cohen's $d_z = 0.01$.

Taken together with Experiment 1, these results prompt the conclusion that the representation of RT2 is based not only on the current RT2 but also on all previous RT2s (i.e., an internal reference), irrespective of whether these RT2s were produced in trials with short or long SOA. Whether participants show an apparent unawareness of the PRP effect or not would thus be a consequence of mixing or blocking of the SOA conditions. However, such a conclusion would still be premature because in both previous experiments the comparison intervals were confounded with the RT2s of each SOA condition. It is possible that participants compared the comparison interval not with an internal reference based on their RT2s but rather with an internal reference based on the comparison intervals. This would mean that participants did not take into account their RT2s for their estimates but relied on the information provided by the comparison intervals. Thus, although Experiments 1 and 2 provide evidence that the temporal discrimination of RT2s is probably based on an internal reference that integrates the temporal information provided by the experimental context, the question remains: What is this reference composed of, the RT2s or the comparison intervals?
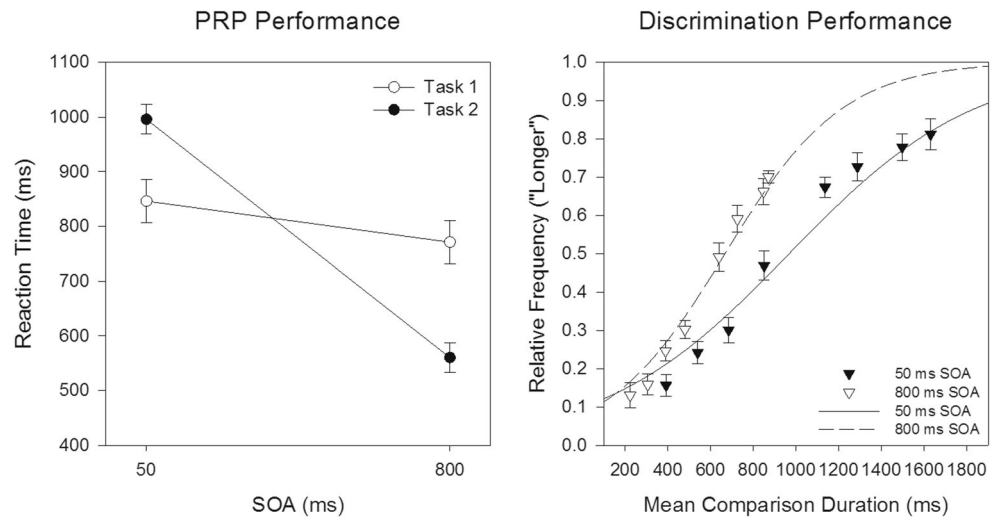
Experiment 3

In Experiment 3 we aimed to determine the contribution of the comparison intervals to the participants' comparison judgments. To this end, we deconfounded the comparison intervals and the RT2s by employing identical comparison interval distributions for short and long SOA trials in the same blocked SOA context as in Experiment 2. As in the previous experiments, in Experiment 3a in each trial participants performed the PRP task followed by the comparison task. Experiment 3b served as a control experiment and was identical to Experiment 3a with the exception that participants did not have to perform the PRP task but experienced the RT2s collected in Experiment 3a. If participants base their comparison judgments on an internal reference composed of their RT2s, the psychometric function should again differ between short and long SOA (as in Experiment 2). However, if the judgments are mainly based on the information provided by the comparison intervals, using the same comparison intervals should lead to similar psychometric functions for the two SOA conditions.

Method

Participants Two groups of 18 volunteers each participated in Experiments 3a and 3b (Experiment 3a: eight males and 10 females, aged between 20 and 33 years, $M = 24.8$ years; Experiment 3b: one male and 17 females, aged between 18 and 27 years, $M = 20.8$ years). Each experimental session lasted about 1 hour. In Experiment 3a, four other participants were tested but had to be replaced because they produced virtually flat psychometric functions. Participants reported normal hearing and normal or corrected-to-normal vision, and received either course credit or payment.

Fig. 4 Reaction-time performance in the psychological refractory period (PRP) task and discrimination performance in Experiment 2. Left panel: Mean reaction time in Task 1 and Task 2 as a function of stimulus onset asynchrony (SOA). Right panel: Relative frequency of responding "longer" as a function of mean comparison duration and SOA, and fitted psychometric functions (lines) for all participants' data. Note that statistical analyses were performed with individually fitted psychometric functions. Error bars represent ±1 within-subject standard error (Morey, 2008)

Apparatus and stimuli The apparatus and stimuli were identical to the previous experiments.

Procedure and design The procedure and design of Experiment 3a were identical to Experiment 2 with the exception that eight fixed comparison intervals (200, 400, 600, 800, 1,000, 1,200, 1,400, and 1,600 ms) were used in both SOA conditions.

For Experiments 3a and 3b we used a yoked design—that is, each participant in Experiment 3b experienced the same trials as another participant in Experiment 3a but without performing the PRP task (i.e., without providing speeded responses to S1 and S2). In Experiment 3b, the original response to S2 was represented by the same black dot that was also used to represent the key press at the end of the comparison interval. The time course of each trial in Experiment 3b was identical to the time course of the corresponding trial in Experiment 3a, with the exception that there was no stimulus representing the original response to S1. That is, in each trial the fixation point appeared at the center of the screen, followed by the auditory (S1) and the visual (S2) stimulus according to the original SOA. Then, after the original RT2, the black dot appeared and remained at the center of the screen for the same duration as the participant had pressed down the key in the corresponding trial. Finally, after another 1,000 ms, the same comparison interval was presented as in Experiment 3a, and participants were asked to indicate whether they had perceived the comparison interval as being shorter (left index finger) or longer (right index finger) than the first interval between the onsets of the letter and the dot. In both experiments, each experimental session comprised one practice block and five experimental blocks. Each block consisted of 64 trials (2 Tone × 2 Letter × 2 SOA × 8 Comparison intervals).

Analysis of comparison data Analysis of the comparison judgments followed the procedure of Experiment 1 and 2.

Results and discussion

For both experiments, all trials that included an error in Task 1 and/or Task 2 (6.3 %) in Experiment 3a were discarded from analyses. Figure 5 depicts RT2 in Experiment 3a (left panel) and averaged psychometric functions in Experiments 3a (middle panel) and 3b (right panel) as a function of SOA. Again, the standard PRP pattern was observed: RT2 was 980 ms at short and 587 ms at long SOA, yielding a PRP effect of 393 ms, $t(17) = 9.07$, $p < .001$, Cohen's $d_z = 2.14$. Even though RT1 was 76 ms longer at short (814 ms) than at long SOA (738 ms), this difference was not significant, $t(17) = 1.25$, $p = .228$, Cohen's $d_z = 0.29$.

Regarding comparison performance, in Experiment 3a the results of Experiment 2 were not replicated. Instead, we observed a similar result pattern as in Experiment 1. That is, when using the same comparison intervals for both SOA conditions, PSE did not significantly differ between short and long SOA (803 ms at short vs. 748 ms at long SOA), $t(17) = 0.93$, $p = .366$, Cohen's $d_z = 0.22$. Together with the results of Experiments 1 and 2, this suggests that PSE is strongly biased toward the center of the comparison intervals. It seems that when performing the comparison task, participants compare a current comparison interval with an internal reference of all comparison intervals instead of comparing it with the preceding RT2 as they are instructed, or with an internal reference of their RT2s.

In contrast to Experiment 3a, in Experiment 3b PSE was significantly affected by SOA $t(17) = 5.02$, $p < .001$, Cohen's $d_z = 0.32$. Mean PSE was 956 ms at short and 723 ms at long SOA. This SOA effect on PSE (233 ms) was much larger than that in Experiment 3a (55 ms), although still smaller than the objective PRP effect in RT2 (393 ms). An additional ANOVA on PSE with the between-subjects factor experiment (3a vs. 3b) and the within-subjects factor SOA revealed a significant
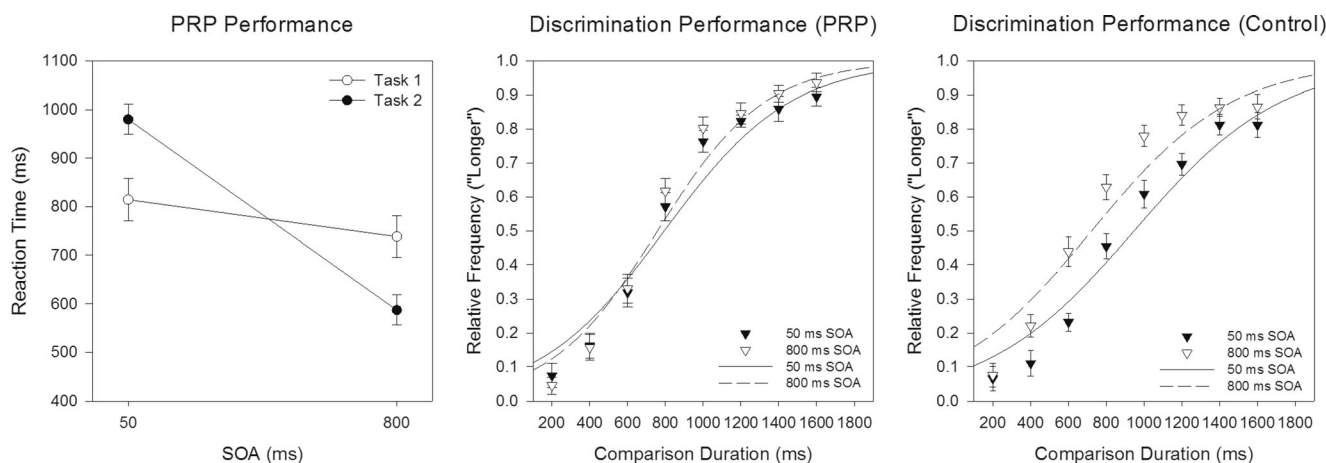
Fig. 5 Reaction-time performance in the psychological refractory period (PRP) task and discrimination performance in Experiment 3. Left panel: Mean reaction time in Task 1 and Task 2 as a function of stimulus onset asynchrony (SOA) in Experiment 3a (PRP). Middle panel: Relative frequency of responding "longer" as a function of comparison duration and SOA in Experiment 3a (PRP). Right panel: Relative frequency of responding "longer" as a function of comparison duration and SOA in Experiment 3b (control). Psychometric functions (lines) were fitted to all participants' data and are shown for illustration. Note that statistical analyses were performed with individually fitted psychometric functions. Error bars represent ±1 within-subject standard error (Morey, 2008)

main effect of SOA, $F(1, 34) = 14.66$, $p < .001$, $\eta_p^2 = .30$, and a significant interaction of experiment and SOA, $F(1, 34) = 5.60$, $p = .024$, $\eta_p^2 = .14$. The main effect of experiment was not significant, $F < 1$. The ANOVA thus confirmed that the participants who did not have to perform the PRP task were less biased by the comparison intervals than those who had to perform it.

In Experiment 3a, the WF was significantly larger at short (0.40) than at long SOA (0.30), $t(17) = 2.16$, $p = .048$, Cohen's $d_z = 0.51$, whereas in Experiment 3b WF did not differ between short (0.48) and long SOA (0.65), $t(17) = 1.37$, $p = .188$, Cohen's $d_z = 0.32$. This difference between the two experiments was again confirmed by a significant interaction between experiment and SOA on WF, $F(1, 34) = 4.16$, $p = .049$, $\eta_p^2 = .11$. The two main effects were not significant, both ps > .278.

General Discussion

In our study we introduced the method of constant stimuli to the field of introspective RT research and investigated how participants introspect about their own RT2s in the PRP paradigm under different temporal contexts. To manipulate the temporal context, we presented short and long SOA trials (in which participants typically produce longer RT2s in the former than in the latter case) either in mixed or in separate blocks. Our results revealed that in this situation, estimates of RT2s (i.e., PSEs) were strongly influenced by the temporal context. However, in contrast to what was initially hypothesized, it was not the temporal context of the RT2s but rather the temporal context of the comparison intervals that

influenced the participants' estimates. Specifically, when participants first performed the PRP task and then compared their RT2 with a comparison interval, RT2 estimates were biased toward the center of the comparison intervals. This bias was largely reduced when participants only perceived the RT2s without performing the PRP task. Overall, these results suggest that participants have only poor and unreliable representations of their RT2s and therefore rely almost completely on the information provided by the comparison intervals.

Our results again confirm that participants cannot truly introspect on their RT2s in the PRP paradigm (Corallo et al., 2008; Marti et al., 2010). However, they also suggest that this introspective limitation results from impaired timing abilities rather than a conscious perception bottleneck. We favor the former explanation because across the different experimental contexts of our study, RT2 estimates were mainly determined by the temporal context of the comparison intervals. Accordingly, PSEs showed both the previously observed null effect of SOA (Experiments 1 and 3a) and an SOA effect reflecting the objective RT2 pattern (Experiment 2) appearing as unawareness in the former case or as awareness of the PRP effect in the latter case. This interpretation is in line with the previous notion that timing itself requires attentional resources (e.g., Block et al., 2010; Brown, 1997), and that if these resources are not available, people might use a less precise, implicit timing mechanism (Ruthruff & Pashler, 2010).

Our results also agree with previous findings that introspective RTs can be biased by other temporal information, such as the interresponse interval (Bratzke et al., 2014), or even nontemporal information, such as the feeling of difficulty (Bryce & Bratzke, 2014). Whether and to what degree participants can time their own RTs in a prospective way and/or use other

information to provide an RT estimate probably depends on the availability of attentional resources as well as the saliency and reliability of other information in that particular context (e.g., the comparison intervals in this study). We suggest that even though in the introspective PRP task participants know in advance that they should time their RTs (i.e., it is a prospective timing task), the timing becomes more retrospective because of the high processing demands of the PRP task (see also Tobin, Bisson, & Grondin, 2010; Zakay & Block, 2004). Consequently, participants need to infer their RTs retrospectively on the basis of the episodic information encoded during performance of the PRP task (Zakay & Block, 2004). Apparently, in our study the comparison intervals provided the most salient episodic information that participants could use to infer their RT2s.

Experiment 3b demonstrated that the bias toward the center of the comparison intervals was largely reduced when participants did not have to perform the PRP task. This suggests that participants had better and more accurate representations of the to-be-timed intervals in this situation. However, even in this situation the bias was not completely eliminated, because the effect of SOA on PSE was still smaller than the PRP effect in RT2. In fact, that in the method of constant stimuli the PSE tends toward the center of the comparison stimuli was already described more than half a century ago (asymmetry effect; Guilford, 1954) and has been reported for nontemporal features of visual (line length; Levison & Restle, 1968; Masin, 1987) and auditory (pitch and loudness; Doughty, 1949) stimuli. The bias caused by the displacement of the center of the comparisons from the standard in these studies ranged from about 20 % in line discrimination to about 50 % in loudness discrimination. Principally, we assume that all biases that are known to affect judgments of sensory magnitude for other physical dimensions (e.g., Poulton, 1979) can also affect temporal judgments, and there may even be some biases that are specific to temporal judgments due to the unique unidirectionality of the time dimension (Riemer, 2015).

To our knowledge, up to now only one study has investigated the potential influence of the asymmetry effect on PSE in the temporal domain (Seifried & Ulrich, 2010). These authors were interested in how the placement of the comparison intervals influenced the so-called oddball effect (i.e., rare stimuli within a stream of frequent stimuli are overestimated). In one condition they used comparison intervals that were symmetrically placed around the standard interval; in another condition all comparisons were shorter than the standard. As a result, the PSE shifted by about 75 % of the difference between the two comparison interval distributions, demonstrating an even larger asymmetry effect than those previously observed with nontemporal stimuli (20 %–50 %). We nevertheless observed a reduction of the bias in Experiment 3b when participants did not have to process the PRP task, which agrees with

Doughty's (1949) suggestion that the influence of the comparison intervals gets stronger as stimulus discrimination becomes more difficult.

Compared with typically reported WFs (0.1–0.3) for the discrimination of brief visual and auditory durations (e.g., Grondin, 1993; Ulrich, Nitschke, & Rammsayer, 2006), the WFs in our study (0.30–0.65) were relatively large. This is perhaps not surprising given the highly complex and demanding character of the experimental situation. Even though WFs may be difficult to interpret in this case, some of our WF findings nevertheless deserve discussion, especially in light of our main conclusion that for the comparison task participants almost completely relied on the information provided by the comparison intervals. First, averaged across SOAs, the Experiments 3a and 3b yielded the smallest (0.35 in Experiment 3a) as well as the largest of all WFs (0.57 in Experiment 3b). The relatively small average WF in Experiment 3a can be explained by the fact that Experiment 3 provided the most stable temporal context due to the fixed comparison intervals. This explanation is based on the assumption that the variability of the temporal context determines the stability of the internal reference which is reflected in the slope of the psychometric function (see Dyjas et al., 2012). At first view, this explanation appears to be inconsistent with our observation that Experiment 3b yielded the largest average WF, although in this experiment the same relatively stable temporal context as in Experiment 3a was provided and participants did not have to process the PRP task. However, that the asymmetry effect was much less pronounced in Experiment 3b than in all other experiments suggests that in this experiment the comparison judgments were based on an internal reference composed not only of the comparison intervals but also of the (relevant) standard information, which might have complicated the comparison task in this case. Also, differences in overall task demands (lower in Experiment 3b than the other experiments) might have contributed to the differences in WFs. Second, Experiment 3a was the only experiment in which the WF was significantly affected by SOA (0.40 at short and 0.30 at long SOA; Experiment 1 showed a similar, albeit nonsignificant result pattern). Because Experiment 3a employed the same comparison intervals for the two SOA conditions, this suggests that the comparison judgments were at least not completely independent of what the participants experienced in the preceding PRP task.

Given the strong bias that the comparison intervals can induce, should one refrain from using the method of constant stimuli for assessing temporal judgments? Indeed, based on his research on pitch and loudness discrimination, Doughty (1949) concluded that this method should not be used for assessing the PSE or the constant error (i.e., standard minus

PSE) even though it may be suited to determine the difference threshold. However, the method of constant stimuli has been successfully used to investigate temporal phenomena related to perceived duration, for example, in studies on intentional binding (Nolden, Haering, & Kiesel, 2012), the temporal oddball effect (Birngruber, Schröter, & Ulrich, 2014; Tse, Intriligator, Rivest, & Cavanagh, 2004), the effects of stimulus repetition (Matthews, 2011), or the influence of global temporal contexts (Jones & McAuley, 2005). All these studies used comparison intervals that were symmetrically distributed around the physical standard duration, thereby avoiding a possible asymmetry effect. As Seifried and Ulrich (2010) pointed out, however, this advantage comes at the cost of a potential psychological asymmetry effect, that is, the risk of missing the true PSE that might be different from the standard.

Concerning the appropriate choice of method for assessment of perceived duration, it should be noted that other established methods (e.g., verbal estimation and reproduction) are also not free from biases. According to Zakay (1990), verbal estimates are prone to a whole number response bias (i.e., a tendency to round off estimates), and reproductions may be biased by attentional distraction and motor limitations. Furthermore, the specific reproduction method (i.e., start and/or termination with discrete key presses, or continuous reproduction; with or without sensory feedback) can influence the duration estimates (Bueti & Walsh, 2010; Mioni, Stablum, McClintock, & Grondin, 2014). In our view, the rather uncommon VASs that have been used in previous introspective PRP studies are most comparable to the verbal estimation method because both methods require a translation of perceived duration into conventional time units (the VASs are usually labeled with millisecond values). One could therefore assume that, similar to the whole number response bias of verbal estimates, estimates collected with VASs are also often "rounded off"—in this case, however, most probably to the tick markers provided on the scale. Another potential source of bias when using VASs is their restricted and rather arbitrarily chosen range. This may be especially problematic in situations in which the range of the to-be-estimated durations does not correspond to the range of the VAS and the same VAS is used in conditions that comprise different ranges of intervals (as in introspective PRP studies; see also Bryce & Bratzke, 2015).

In conclusion, in our study we used the method of constant stimuli to investigate RT2 estimates in an introspective PRP paradigm under different temporal contexts. Overall, our results revealed that RT2 estimates, as expressed by PSEs, were almost completely biased by the specific temporal context of the comparison intervals. This bias was substantially reduced when no PRP task had to be performed. We interpret this finding as indicating that participants acquire only poor representations of

their RT2s when they perform the PRP task. Our results thus confirm that introspection is substantially limited when participants introspect about their own RT performance in an attentionally demanding dual-task situation. The present results, however, suggest that this introspective limitation is more likely a consequence of disturbed timing abilities than the signature of a conscious perception bottleneck.

References

Bausenhart, K. M., Bratzke, D., & Ulrich, R. (2016). Formation and representation of temporal reference information. *Current Opinion in Behavioral Sciences, 8,* 46–52. doi:10.1016/j.cobeha.2016.01.007

Bertelson, P. (1967). The refractory period of choice reactions with regular and irregular interstimuli intervals. *Acta Psychologica, 27,* 45–56. doi:10.1016/0001-6918(67)90043-1

Birngruber, T., Schröter, H., & Ulrich, R. (2014). Duration perception of visual and auditory oddball stimuli: Does judgment task modulate the temporal oddball effect? *Attention, Perception, & Psychophysics, 76,* 814–828. doi:10.3758/s13414-013-0602-2

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

Boring, E. G. (1953). A history of introspection. *Psychological Bulletin, 50,* 169–189. doi:10.1037/h0021468

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

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. doi: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,* 1118–1140. doi:10.3758/BF03205526

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. doi: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. doi: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,* 317–323. doi:10.3758/s13423-015-0877-5

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. doi:10.1016/j.actpsy.2014.12.007

Bueti, D., & Walsh, V. (2010). Memory for time distinguishes between perception and action. *Perception, 39,* 81–90. doi:10.1068/p6405

Burns, M. M., & Moskowitz, H. (1971). Response time to a first signal as a function of time relationship to a second signal and mode of

presentation. *Perceptual and Motor Skills, 32,* 811–816. doi:10.2466/pms.1971.32.3.811

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

Danziger, K. (1980). The history of introspection reconsidered. *Journal of the History of the Behavioral Sciences, 16,* 241–262. doi:10.1002/1520-6696(198007)16:3<241::AID-JHBS2300160306>3.0.CO;2-O

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,* 204–211. doi:10.1016/j.tics.2006.03.007

Doughty, J. M. (1949). The effect of psychophysical method and context on pitch and loudness functions. *Journal of Experimental Psychology, 39,* 729–745. doi:10.1037/h0060307

Dyjas, O., Bausenhart, K. M., & Ulrich, R. (2012). Trial-by-trial updating of an internal reference in discrimination tasks: Evidence from effects of stimulus order and trial sequence. *Attention, Perception, & Psychophysics, 74,* 1819–1841. doi:10.3758/s13414-012-0362-4

Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Perception & Psychophysics, 54,* 383-394. doi:10.3758/BF03205274

Grondin, S. (2008). Methods for studying psychological time. In S. Grondin (Ed.), *Psychology of time* (pp. 51–74). Bingley, UK: Emerald Group.

Guilford, J. P. (1954). *Psychometric methods.* New York, NY: McGraw-Hill.

Jones, M. R., & McAuley, J. D. (2005). Time judgments in global temporal contexts. *Perception & Psychophysics, 67,* 398–417. doi:10.3758/BF03193320

Lejeune, H., & Wearden, J. H. (2009). Vierordt's the experimental study of the time sense (1868) and its legacy. *European Journal of Cognitive Psychology, 21,* 941–960. doi:10.1080/09541440802453006

Levison, M., & Restle, F. (1968). Invalid results from the method of constant stimuli. *Perception & Psychophysics, 4,* 121–122. doi:10.3758/BF03209522

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

Marti, S., Sigman, M., & Dehaene, S. (2012). A shared cortical bottleneck underlying attentional blink and psychological refractory period. *NeuroImage, 59,* 2883–2898. doi:10.1016/j.neuroimage.2011.09.063

Masin, S. C. (1987). Different biases in the methods of constant and single stimuli. *Bulletin of the Psychonomic Society, 25,* 379–382. doi:10.3758/BF03330373

Matthews, W. J. (2011). Stimulus repetition and the perception of time: The effects of prior exposure on temporal discrimination, judgment, and production. *PLoS ONE, 6,* e19815. doi:10.1371/journal.pone.0019815

Mioni, G., Stablum, F., McClintock, S. M., & Grondin, S. (2014). Different methods for reproducing time, different results. *Attention, Perception, & Psychophysics, 76,* 675–681. doi:10.3758/s13414-014-0625-3

Morey, R. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology, 4,* 61–64.

Nolden, S., Haering, C., & Kiesel, A. (2012). Assessing intentional binding with the method of constant stimuli. *Consciousness and Cognition, 21,* 1176–1185. doi:10.1016/j.concog.2012.05.003

Poulton, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin, 86,* 777–803. doi:10.1037/0033-2909.86.4.777

Riemer, M. (2015). Psychophysics and the anisotropy of time. *Consciousness and Cognition, 38,* 191–197. doi:10.1016/j.concog.2015.06.007

Ruthruff, E., & Pashler, H. (2010). Mental timing and the central attentional bottleneck. In A. C. Nobre & J. T. Coull (Eds.), *Attention and time* (pp. 123–136). Oxford, UK: Oxford University Press.

Sanford, A. J. (1970). Rating the speed of a simple reaction. *Psychonomic Science, 21,* 333–334. doi:10.3758/BF03335809

Seifried, T., & Ulrich, R. (2010). Does the asymmetry effect inflate the temporal expansion of odd stimuli? *Psychological Research, 74,* 90–98. doi:10.1007/s00426-008-0187-x

Shi, Z., Church, R. M., & Meck, W. H. (2013). Bayesian optimization of time perception. *Trends in Cognitive Sciences, 17,* 556–564. doi:10.1016/j.tics.2013.09.009

Tobin, S., Bisson, N., & Grondin, S. (2010). An ecological approach to prospective and retrospective timing of long durations: A study involving gamers. *PLoS ONE, 5,* e9271. doi:10.1371/journal.pone.0009271

Tombu, M. N., Asplund, C. L., Dux, P. E., Godwin, D., Martin, J. W., & Marois, R. (2011). A unified attentional bottleneck in the human brain. *Proceedings of the National Academy of Sciences, 108,* 13426–13431. doi:10.1073/pnas.1103583108

Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. *Perception & Psychophysics, 66,* 1171–1189. doi:10.3758/BF03196844

Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Crossmodal temporal discrimination: Assessing the predictions of a general pacemaker-counter model. *Perception & Psychophysics, 68,* 1140–1152. doi:10.3758/BF03193716

van Rijn, H., & Taatgen, N. A. (2008). Timing of multiple overlapping intervals: How many clocks do we have? *Acta Psychologica, 129,* 365–375. doi:10.1016/j.actpsy.2008.09.002

Vierordt, K. (1868). *Der Zeitsinn nach Versuchen [The time sense according to experiments].* Tübingen, Germany: Laupp.

Zakay, D. (1990). The evasive art of subjective time measurment: Some methodological dillemas. In R. A. Block (Ed.), *Cognitive models of psychological time* (pp. 59–84). Hillsdale, NJ: Erlbaum.

Zakay, D., & Block, R. A. (2004). Prospective and retrospective duration judgments: An executive-control perspective. *Acta Neurobiologiae Experimentalis, 64,* 319–328.