Are introspective reaction times affected by the method of time estimation? A comparison of visual analogue scales and reproduction

Donna Bryce · Daniel Bratzke

Abstract In this study, we investigated whether the method of time estimation plays a role in the apparent limits of introspection in dual-task processing. Previous studies showed that when participants reported introspective reaction times after each trial of a dual task by clicking on a visual analogue scale, they appeared to be unaware of the dual-task costs in their performance. However, visual analogue scales have seldom been used in interval estimation, and they may be inappropriate. In the present study, after each dual-task trial, participants reported their introspective reaction times either via a visual analogue scale or via the method of reproduction. The results replicated the previous findings, irrespective of method. That is, even though responses to the second task slowed down with increasing task overlap, this slowing was only very weakly reflected in the introspective reaction times. Thus, the participants' failure to report the objective dual-task costs in their reaction times is a rather robust finding that cannot be attributed to the method employed. However, introspective reaction times reported via visual analogue scales were more closely related to the objective reaction times, suggesting that visual analogue scales are preferable to reproduction. We conclude that introspective reaction times represent the same information regardless of method, but whether that information is temporal in nature is as yet unsettled.

Keywords Psychological refractory period · Temporal processing · Dual-task performance · Introspection

D. Bryce · D. Bratzke
Department of Psychology, University of Tübingen, Schleichstrasse 4, Tübingen 72076, Germany

D. Bryce (✉)
Fachbereich Psychologie, Universität Tübingen, Schleichstrasse 4, 72072 Tübingen, Germany
e-mail: donna.bryce@uni-tuebingen.de

In three previous studies, participants failed to report the reaction time costs associated with performing two tasks in quick succession (Bryce & Bratzke, 2014; Corallo, Sackur, Dehaene, & Sigman, 2008; Marti, Sackur, Sigman, & Dehaene, 2010). In these studies, participants performed a commonly used dual-task called the *psychological refractory period (PRP) paradigm* and then estimated the interval between the stimulus and their response (i.e., their reaction time) for each task. Participants provided estimates by clicking on a horizontal line marked with temporal values (i.e., a visual analogue scale). While responses to the second task were much slower when the processing of the two tasks overlapped than when they were presented serially, participants' estimates were similar across these conditions. Thus, participants appeared to be unaware of the slowing of their responses, suggesting a severe limitation of introspection. However, although VASs have been widely used to measure subjective states such as mood, alertness, and pain (e.g., Butler, 1997; Monk, 1989), they have seldom been used in timing research and may indeed be inappropriate for introspective PRP tasks, which are particularly complex and demanding.

Three methods have traditionally been used in timing research to collect estimates of previously presented time intervals from participants—verbal estimation, reproduction, and the method of comparison (Grondin, 2010). Verbal estimation requires participants to freely estimate a time interval in temporal units (seconds or minutes). In reproduction, participants depress a button for the same duration as the interval, or terminate a second interval (via a buttonpress) when it reaches the same duration as the first interval. In the method of comparison, participants are presented with two intervals and are required to judge which is longer or shorter. Previous findings regarding which method is most sensitive and/or produces the most accurate estimates (in terms of deviation from the objective value) have been mixed

(Clausen, 1950; Kruup, 1961; McConchie & Rutschmann, 1971). Since some previous studies have shown different experimental effects depending on the timing method used (e.g., Gil & Droit-Volet, 2011; Matthews, 2011), it is important to assess the generality of previous introspective PRP results using different methods. Therefore, in the present study we compared the results obtained when participants reported their RTs via VASs and via reproduction.

Different strengths and weaknesses are associated with each of these two methods. Visual analogue scales are, of course, bound at both ends of the scale and labeled by values that are chosen by the experimenter rather than the participant. This means that although this is a socially agreed-upon scale (Matthews, 2013), the absolute values of the estimates should be interpreted with caution, since we cannot be sure how participants interpret the labels and use the scale (especially in the absence of training or calibration sessions). In contrast, interval reproductions are only bound at one end, and the absolute values can be interpreted better, since the intervals are generated by the participants themselves. On the other hand, Bindra and Waksberg (1956) highlighted a problem in interpreting over- or underestimations in reproductions; an "internal clock" is used twice in a reproduction trial (first for perceiving and then for reproducing the interval). Thus, in the case of reproductions being too short or too long, it is not clear whether the rate of one clock or the rates of both clocks are responsible. In contrast, it can be assumed that the VAS method only requires an internal clock for perceiving and not for reporting an interval. Another potential weakness of the reproduction method is that the estimates may be affected by motor limitations (Droit-Volet, 2010). However, it has been demonstrated that participants' use of VASs can be influenced by their mental states, since depressed participants generally give higher ratings on VASs than do non-depressed participants (Peet, Ellis, & Yates, 1981). In a previous study on how the time perception of two intervals was affected by their degree of overlap, we found the same experimental effects when either VASs or reproduction were used (Bryce, Seifried-Dübon, & Bratzke, 2014). However, whether the methods are also equivalent in the more challenging introspective PRP task, in which participants must time their own RTs while processing a dual-task, remains an open question.

One feature of VASs may be especially problematic for collecting estimates of RTs (also referred to as *introspective RTs*)—their restricted range. In all previous introspective PRP studies (Bryce & Bratzke, 2014; Corallo et al., 2008; Marti et al., 2010), the same range was used for all conditions, even though in the PRP paradigm the RT distributions vary considerably across conditions. That is, the RT to the second stimulus is typically longer and more variable when the two tasks overlap than when they are separated. If participants map these different RT distributions onto the same scale, the range of reported RTs to the second stimulus would contract in

conditions with strong task overlap and would expand in conditions with no task overlap (also referred to as, respectively, *short* and *long* stimulus onset asynchronies, SOAs). This would result in an underestimation of the second RT at short SOAs, and an overestimation at long SOAs. In other words, such an artifact introduced by the reporting methodology could be mistakenly interpreted as unawareness of the dual-task costs. Importantly, the method of reproduction should not introduce such an artifact, because the range of possible estimates is not restricted by the experimenter.

It is important to determine whether the method of reporting has contributed to the findings of previous introspective PRP studies, not only for methodological reasons, but also because the findings have been interpreted as reflecting limits of introspection. That is, on the basis of their observation that participants did not report the slowing of their responses in short-SOA conditions, Corallo et al. (2008) and Marti et al. (2010) concluded that response selection of the first task blocks conscious awareness of the second task. This, in turn, means that at short SOAs the second stimulus is perceived at a later time point than it was really presented at, resulting in an underestimation of the RT interval. However, there is also evidence that introspective RTs can be influenced by the experience of difficulty in each task (Bryce & Bratzke, 2014), and that reports of stimulus onsets can be biased by other events in the trial—namely the interval between the two responses (Bratzke, Bryce, & Seifried-Dübon, 2014). Thus, the interpretation of results from introspective PRP experiments is as yet unsettled, and an outstanding pertinent issue is whether the method has played a role in the observed result patterns.

To determine the reliability of previous findings in introspective PRP experiments, in this study we conducted a within-subjects comparison of two methods for reporting introspective RTs. As in previous introspective PRP experiments, participants completed a simple PRP task (with one auditory and one visual stimulus) before reporting their first and second introspective RTs. In half of the trials, they reported RTs by clicking on a VAS, and in the other half, by buttonpress reproduction.

Method

Participants

Thirty participants (23 female, seven male; 29 right-handed), between 20 and 38 years of age ($M = 23.6$ years), participated in one 40-min session. All of the participants had normal hearing and normal or corrected-to-normal vision, and all received course credit.

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; Pelli, 1997), Version 3.0.10, and was presented on a Mac desktop computer (OS X, VGA monitor, 150 Hz). The first stimulus (S1) was a tone of either 440 or 880 Hz, presented via headphones (60 dB SPL, 150 ms duration). The second stimulus (S2) was the letter O or X, presented in black on a white background in the center of the screen for 300 ms. Two external response panels were used to record responses with the index and middle fingers of the left and right hands. Introspective reports of each RT were collected via a mouse click on a VAS with extreme values of *0 ms* and *1,200 ms* presented on the screen (VAS condition), or by depressing and releasing the spacebar on the computer keyboard (reproduction condition).

Procedure and design

Each trial began with a fixation point in the center of the screen. After 1,000 ms S1 was presented, followed by S2 after an SOA of 50, 200, or 1,000 ms. The participants were instructed to respond as quickly and accurately as possible to each stimulus. In Task 1, participants had to respond with their left middle finger to the low tone and with their left index finger to the high tone. In Task 2, they had to respond with their right index finger to the letter O and with their right middle finger to the letter X. In the case of an incorrect response in one of the two RT tasks, a feedback message was displayed for 1,000 ms; in the case of two correct responses, a blank screen was presented for 1,000 ms. Immediately after this, introspective reports of RT1 and RT2 were collected (*iRT1* and *iRT2*), in that order. In the VAS condition, iRTs were prompted by the question "how long was the reaction to the tone/letter?," a VAS was presented on the screen, and participants clicked with the mouse on the scale to indicate their iRTs. In the reproduction condition, the instruction "please now press the spacebar for the same amount of time as the duration between the presentation of the tone/letter and your response to the tone/letter" was presented, and participants pressed and released the spacebar to indicate their iRTs. When the spacebar was pressed the instruction disappeared, and when it was released the next instruction appeared. Then, 500 ms after iRTs had been collected, participants could initiate the next trial by pressing the response key associated with the right index finger. Each judgment type (VAS or reproduction) was used for one half of the experiment, and the order of the judgment types was balanced across participants. Each half of the experiment consisted of a practice block and four experimental blocks (24 trials each). Every combination of stimuli was presented twice in each block of the experiment: 3 SOAs (50, 200, or 1,000 ms) ×2 auditory stimuli (low or high tone) ×2 visual stimuli (the letter O or X).

Results

Analysis

Due to technical problems, it was not possible to analyze 19 trials (0.003 % of all trials). The mean error rates in Task 1 and Task 2 were analyzed in SOA (50 vs. 200 vs. 1,000 ms) × Judgment Type (VAS vs. reproduction) repeated measures analyses of variance (ANOVAs). For all of the further analyses, trials that contained an error in Task 1 or 2 were discarded (9.77 %). Trials in which RT1 or RT2 deviated more than three standard deviations from the individual mean in each condition were also excluded (3.27 % of correct trials), as were trials in which the responses were grouped (within 100 ms of each other, 1.62 % of the remaining trials). The mean RTs and mean iRTs were analyzed via SOA (50 vs. 200 vs. 1,000 ms) × Judgment Type (VAS vs. reproduction) repeated measures ANOVAs. The Greenhouse–Geisser correction was used to adjust *p* values where appropriate, and partial eta-squared effect sizes are provided. Standard errors for within-subjects designs were calculated according to the Morey (2008) method. In order to examine which reporting method was most sensitive to trial-by-trial variation in objective RTs, Pearson product moment correlations were calculated between RT1 and iRT1 and between RT2 and iRT2 within each participant, judgment type, and SOA condition (see Marti et al., 2010). Separate one-sample *t* tests were then performed on the correlations for each SOA to test whether they differed from zero, and correlation coefficients were analyzed by SOA × Judgment Type repeated measures ANOVAs.

Error rates

Errors in response to Task 1 decreased with increasing SOA, $F(2, 58) = 9.80$, $p < .001$, $\eta_p^2 = .25$. Post-hoc tests indicated that more errors were made in the shortest SOA condition (mean error rate of 3.59) than in the 200 ms (2.19; $p = .005$) and 1,000 ms (1.67; $p < .001$) SOA conditions. We observed no significant effects of judgment type, $F(1, 29) = 1.56$, $p = .22$, $\eta_p^2 = .05$, nor a significant SOA × Judgment Type interaction, $F(2, 58) = 1.63$, $p = .21$, $\eta_p^2 = .05$, on Task 1 error rates. More errors were made in response to Task 2 in the VAS condition (7.22) than in the reproduction condition (6.21), $F(1, 29) = 4.49$, $p = .04$, $\eta_p^2 = .13$. Task 2 error rates were not affected by SOA, $F(2, 58) = 1.33$, $p = .27$, $\eta_p^2 = .04$, nor did the SOA × Judgment Type interaction reach significance, $F(2, 58) = 2.09$, $p = .13$, $\eta_p^2 = .07$.

Reaction times

Figure 1a depicts RT1 and RT2 as a function of SOA and judgment type. Responses to Task 1 were affected by SOA, $F(2, 58) = 6.51$, $p = .007$, $\eta_p^2 = .18$. Post-hoc tests indicated that RT1 was longer in the shortest SOA condition than in the 200 ms ($p = .002$) and 1,000 ms ($p = .02$) SOA conditions. Although RT1 did not change depending on judgment type, $F(1, 29) = 0.94$, $p = .34$, $\eta_p^2 = .03$, there was a significant SOA × Judgment Type interaction, $F(2, 58) = 3.51$, $p = .04$, $\eta_p^2 = .11$. This interaction reflected the different effects of SOA on RT1 in the two conditions- in the VAS condition, RT1 decreased across SOAs; in the reproduction condition, RT1 was shortest in the 200 ms SOA condition. Responses to Task 2 were much slower at short than at long SOAs, $F(2, 58) = 164.34$, $p < .001$, $\eta_p^2 = .85$. There was an overall PRP effect (difference in RT2 between short and long SOAs) of 449 ms, and post-hoc tests indicated that all comparisons were significant (all $ps < .001$). No significant main effect of judgment type was apparent, $F(1, 29) = 2.65$, $p = .11$, $\eta_p^2 = .08$, nor a significant SOA × Judgment Type interaction, $F(2, 58) = 0.24$, $p = .79$, $\eta_p^2 = .008$, on RT2.

Introspective reaction times

Figure 1b depicts iRT1 and iRT2 as a function of SOA and judgment type. Introspective reports of RT1 (iRT1) were affected by SOA, $F(2, 58) = 7.48$, $p = .004$, $\eta_p^2 = .21$. Post-hoc tests indicated that iRT1 was smallest in the 200 ms SOA condition (420 ms) and significantly different from iRT1 at the shortest (437 ms, $p = .047$) and longest (448 ms, $p < .001$) SOA conditions. Neither the judgment type main effect, $F(1, 29) = 0.003$, $p = .96$, $\eta_p^2 < .001$, nor the SOA × Judgment Type interaction, $F(2, 58) = 1.60$, $p = .21$, $\eta_p^2 = .05$, on iRT1 was significant. iRT2 decreased with increasing SOA, $F(2, 58) = 4.55$, $p = .03$, $\eta_p^2 = .14$, but post-hoc tests indicated that only the comparison of the 50 and 1,000 ms SOA conditions reached significance ($p = .01$). Neither the judgment type main effect, $F(1, 29) = 0.56$, $p = .46$, $\eta_p^2 = .02$, nor the SOA × Judgment Type interaction, $F(2, 58) = 0.24$, $p = .74$, $\eta_p^2 = .008$, on iRT2 were significant.

Correlations

All correlation coefficients were significantly different from zero (all $ps < .001$; Fig. 2). The RT1–iRT1 correlations were stronger in the VAS condition ($r = .51$) than in the reproduction condition ($r = .40$), $F(1, 29) = 8.20$, $p = .008$, $\eta_p^2 = .22$. Neither the main effect of SOA, $F(2, 58) = 2.82$, $p = .07$, $\eta_p^2 = .09$, nor the SOA × Judgment Type interaction, $F(2, 58) = 1.94$, $p = .15$, $\eta_p^2 = .06$, on RT1–iRT1 correlation coefficients reached significance. The RT2–iRT2 correlations were also stronger in the VAS condition ($r = .52$) than in the reproduction condition ($r = .41$), $F(1, 29) = 9.81$, $p < .001$, $\eta_p^2 = .25$. This correlation weakened with increasing SOA, $F(2, 58) = 12.98$, $p < .001$, $\eta_p^2 = .31$, and post-hoc tests indicated that the correlation coefficients were stronger in the 50 ms SOA condition than in the 200 ms and 1,000 ms SOA conditions (all $ps < .001$). Furthermore, we observed a significant SOA × Judgment Type interaction on the RT2–iRT2 correlation coefficients, $F(2, 58) = 5.76$, $p = .005$, $\eta_p^2 = .17$. This interaction reflects the different effects of SOA on the correlation coefficients in each condition; in the VAS condition, the correlations weakened with increasing SOA, but in the reproduction condition, the correlation was weakest in the 200 ms SOA condition.
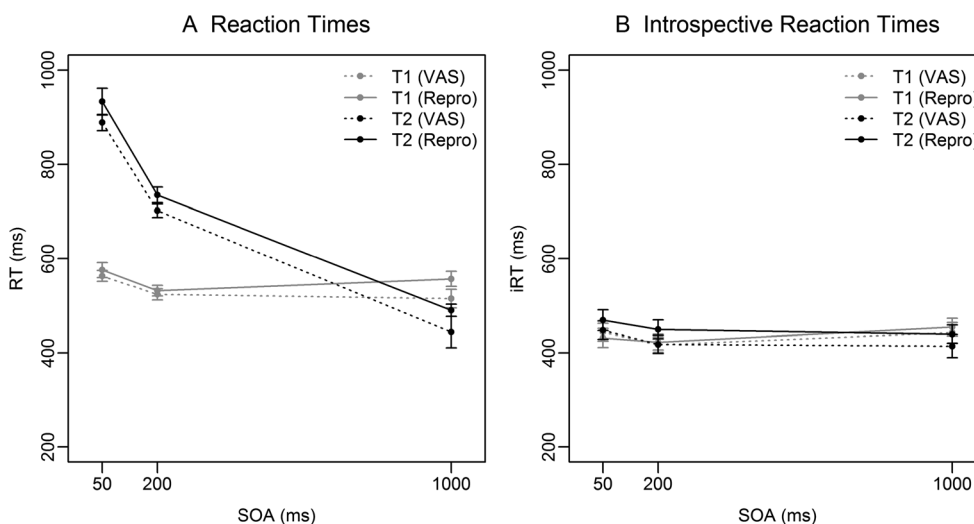

Fig. 1 Mean reaction times (**a**) and introspective reaction times (**b**) as a function of stimulus onset asynchrony (SOA), task, and judgment type. Error bars represent ±1 within-subjects *SE*. VAS = visual analogue scale method, Repro = reproduction method
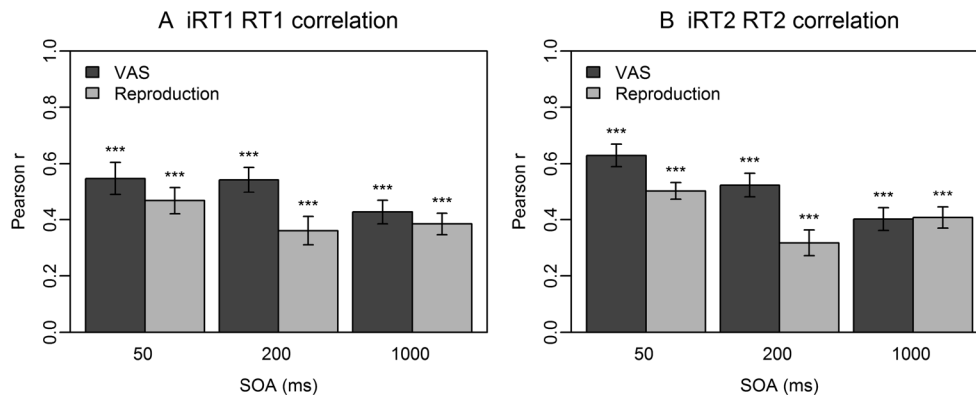
Fig. 2 Mean introspective reaction time–reaction time (iRT–RT) correlation coefficients for Task 1 (**a**) and Task 2 (**b**), as a function of stimulus onset asynchrony (SOA) and judgment type. VAS = visual analogue scale method. $^{***}p < .001$

Discussion

Overall, our result pattern is consistent with the findings of previous introspective PRP experiments. Indeed, the finding that iRT2 is unaffected or only very weakly affected by SOA has now been observed in four studies with two different methods; thus, it can be considered fairly robust. Importantly, our findings establish that this result pattern cannot be attributed to the use of VASs. We found strikingly similar mean introspective RTs when a more common method of interval estimation, reproduction, was used. Only one result distinguished between the two methods—when estimates were given via VASs, they were more strongly related to the equivalent objective measures than when estimates were given by reproduction.

It seems likely that the difference in correlation results is due to the reduced variability in estimates given via VASs, as compared to those given via reproduction. Although the limited range of the VAS may have contributed to the reduced variability in introspective RTs, the limits could also be useful and may offer some structure to the participants, which they use to make their estimates more precise. Indeed, it could be that the strong embodied link between space and time (Santiago, Lupiáñez, Pérez, & Funes, 2007; Ulrich & Maienborn, 2010) leads to VASs being more intuitive to use than reproductions. Furthermore, as mentioned in the introduction, limitations in the motor system and the fact that an internal clock is also needed for reporting may contribute to the variability in estimates given by reproduction. The correlations that we observed between subjective and objective RTs were stronger than those in Marti et al. (2010), possibly because Marti et al. asked for several other reports at the end of a trial. Given the issues highlighted in the introduction with regard to interpreting the absolute values of estimates given by VASs, it is particularly notable that the absolute values of introspective RTs were very similar across the two methods (but see Bryce et al., 2014, for different results in a pure timing

task). The fact that the mean introspective RTs in the VAS condition were not in the middle of the scale, but instead were around 500 ms, could indicate that rather than treating the extremes of the scale as *very fast* and *very slow*, participants did in fact use the temporal values of the labels. Alternatively, participants may tend to indicate most values in the lower half of the scale in order to have enough space to report rather long RTs. However, further controlled studies will be required before any conclusions can be reached.

Our results also differed from those of Corallo et al. (2008) and Marti et al. (2010) in other ways. As in Experiment 2 of Bryce and Bratzke (2014), we found that the mean iRT2 decreased very slightly with increasing SOA. However, there was one important methodological difference between the studies of Corallo et al. and Marti et al. and the present study—the method of outlier exclusion. Corallo and colleagues excluded trials with RTs longer than 1,500 ms, perhaps in order to address the problem of the VAS having a range of 0 to 1,200 ms. We find this approach to be inappropriate, since the use of a fixed criterion would lead to more trials being excluded from the short- than from the long-SOA conditions. In contrast, Marti and colleagues did not report excluding any trials—even those in which errors were committed in the PRP trial. In our previous studies and the present study, we excluded trials with errors, those in which responses were grouped, and those with RTs shorter or longer than three standard deviations from the individual and condition means. The aim of our outlier exclusion method was to identify and remove trials with either fast, impulsive responses or excessively slow responses that might reflect a lapse in attention— in other words, trials in which participants did not process the dual-task in the typical manner. Indeed, when the outlier exclusion method of Corallo et al. was applied to our data, the effect of SOA on iRT2 was no longer significant and reduced to 15 ms. Nevertheless, one cannot reasonably conclude from the present results that participants were aware of the dual-task costs in their performance (i.e., PRP effect),

since the objective effect was much larger (450 ms) than the subjective effect (32 ms).

In contrast to Bryce and Bratzke (2014), we cannot conclude that introspective RTs were influenced by a feeling of difficulty in the present study, since the introspective RTs did not show the same patterns as error rates (which can be considered an indicator of difficulty, on the basis that conditions in which more errors are committed are experienced as more difficult). It is perhaps not surprising that the feeling of difficulty played less of a role in the present experiment, as there was less variation in difficulty. In the Bryce and Bratzke study, the perceptual complexity, and therefore the difficulty of perceptual processing, of each task was manipulated. However, in the present experiment fewer factors could have contributed to fluctuations in the feeling of difficulty, such as trial-by-trial preparedness and the automaticity of stimulus–response pairings. Therefore, the impact of a feeling of difficulty on introspective RTs may have been negligible.

If we had to endorse one method as being superior for collecting introspective reports, on the basis of the present results it seems that VASs are slightly superior to reproduction. The reason for this is that although participants made more errors in the PRP task when they used the VASs, the correlations between the objective and introspective RTs were higher in the VAS than in the reproduction condition. The slightly increased error rate might indicate that giving estimates via a VAS required more effort than did giving estimates via reproduction. This, however, could be evaluated positively as indicating that participants really put effort into giving their introspective reports, rather than randomly clicking on the scale. Although the present results slightly favor VASs over reproduction, our previous findings prompt us to be cautious about the interpretation of introspective RTs, since they can be influenced by the difficulty experienced in a task (Bryce & Bratzke, 2014). Therefore, it could be that estimates represent the same information regardless of method, but the question of whether that information is the correct temporal information is as yet unsettled. Indeed, if Task 1 processing blocks awareness of the second stimulus at short SOAs, as has been proposed by Marti et al. (2010), this should affect the conscious awareness of Task 2 but not of Task 1. Therefore, we would expect that iRT1 should be rather accurate, which should be reflected in a small deviation from the objective RT values, as well as in high sensitivity to trial-by-trial variation. However, in the present study the iRT1 pattern did not reflect RT1, and the correlations between the RTs and iRTs were not stronger for Task 1 than for Task 2.

An important piece of evidence is still missing from all introspective PRP studies—evidence that participants are able to accurately report the time intervals that occur in a PRP trial when they are not simultaneously processing the dual-task. In our view, this is a crucial benchmark for any method used to draw conclusions about introspection. Corallo et al. (2008) did

provide evidence that participants were sensitive both to the effects of task difficulty on RT and to trial-by-trial variation in RTs in a single-task context. However, the timing demands in a single task are very different from those in the PRP task. Investigations into the various challenges facing participants when they are asked to time their own responses in a PRP task are necessary before we can really claim to understand the limitations of introspection.

In summary, the present results support the use of VASs to report estimates of time intervals. We observed only very small differences between the estimates given via VASs and reproduction, and these differences, if anything, favored VASs. Thus, these results are consistent with previous findings (Bryce et al., 2014) and suggest that researchers may use either of these methods free from concerns that their choice will introduce considerable artifacts. The choice of method will depend on the requirements and priorities of each experimental context (e.g., whether it is important to interpret absolute values of the estimates). Still, the precise information on which introspective RTs are based remains an open question.

References

Bindra, D., & Waksberg, H. (1956). Methods and terminology in studies of time estimation. *Psychological Bulletin, 53,* 155–159. doi:10.1037/h0041810

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

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., Seifried-Dübon, T., & Bratzke, D. (2014). How are overlapping time intervals perceived? Evidence for a weighted sum of segments model. *Acta Psychologica* (in press)

Butler, P. V. (1997). Linear analogue self-assessment and procrustean measurement: A critical review of visual analogue scaling in pain assessment. *Journal of Clinical Psychology in Medical Settings, 4,* 111–129. doi:10.1023/A:1026240322240

Clausen, J. (1950). An evaluation of experimental methods of time judgment. *Journal of Experimental Psychology, 40,* 756–761.

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

Droit-Volet, S. (2010). Stop using time reproduction tasks in a comparative perspective without further analyses of the role of the motor response—The example of children. *European Journal of Cognitive Psychology, 22,* 130–148. doi:10.1080/09541440902738900

Gil, S., & Droit-Volet, S. (2011). "Time flies in the presence of angry faces" … depending on the temporal task used! *Acta Psychologica, 136,* 354–362. doi:10.1016/j.actpsy.2010.12.010

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics, 72,* 561–582. doi:10.3758/App. 72.3.561

Kruup, K. (1961). The influence of method on time judgments. *Australian Journal of Psychology, 13,* 44–54. doi:10.1080/00049536108255988

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

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

Matthews, W. J. (2013). How does sequence structure affect the judgment of time? Exploring a weighted sum of segments model. *Cognitive Psychology, 66,* 259–282. doi:10.1016/j.cogpsych.2013.01.001

McConchie, R. D., & Rutschmann, J. (1971). Human time estimation: On differences between methods. *Perceptual and Motor Skill, 32,* 319–336.

Monk, T. H. (1989). A visual analogue scale technique to measure global vigor and affect. *Psychiatry Research, 27,* 89–99. doi:10.1016/0165-1781(89)90013-9

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

Peet, M., Ellis, S., & Yates, R. A. (1981). The effect of level of depression on the use of visual analogue scales by normal volunteers. *British Journal of Clinical Pharmacology, 12,* 171–178. doi:10.1111/j.1365-2125.1981.tb01197.x

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10,* 437–442. doi:10.1163/156856897X00366

Santiago, J., Lupiáñez, J., Pérez, E., & Funes, M. J. (2007). Time (also) flies from left to right. *Psychonomic Bulletin & Review, 14,* 512–516. doi:10.3758/BF03194099

Ulrich, R., & Maienborn, C. (2010). Left–right coding of past and future in language: The mental timeline during sentence processing. *Cognition, 117,* 126–138. doi:10.1016/j.cognition.2010.08.001