Introspective reports of reaction times in dual-tasks reflect experienced difficulty rather than timing of cognitive processes

Donna Bryce, Daniel Bratzke

Angaben zur Veröffentlichung / Publication details:

Bryce, Donna, and Daniel Bratzke. 2014. "Introspective reports of reaction times in dual-tasks reflect experienced difficulty rather than timing of cognitive processes." *Consciousness and Cognition* 27: 254–67. https://doi.org/10.1016/j.concog.2014.05.011.

Introspective reports of reaction times in dual-tasks reflect experienced difficulty rather than timing of cognitive processes

Donna Bryce *, Daniel Bratzke

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

1. Introduction

The ability to introspect accurately about the time it takes to complete a task is important for successfully monitoring and controlling one's behavior in demanding situations (i.e. behaving metacognitively; Zimmerman, 2001). Furthermore, research into the accuracy of introspection has the potential to provide insights into current debates in consciousness (Smithies & Stoljar, 2012), for example regarding the link between attention and consciousness. While some theorists consider attention as a prerequisite for conscious awareness (Dehaene & Changeux, 2011; Dehaene, Kerszberg, & Changeux, 1998; Shallice & Burgess, 1996), others regard it merely as a confound (Kentridge, Nijboer, & Heywood, 2008).

Dual-task paradigms are especially useful for investigating the role of attention in conscious awareness because the temporal demands on attention vary. Two previous studies (Corallo, Sackur, Dehaene, & Sigman, 2008; Marti, Sackur, Sigman, & Dehaene, 2010) exploited this advantage by using an introspective version of a well-known dual-task paradigm – the psychological refractory period (PRP) paradigm. In the PRP paradigm, two stimuli (S1 and S2) are presented with a variable interval (stimulus onset asynchrony, SOA) and participants must respond to each task as quickly as possible. Responses to the second task are typically slower when the two stimuli are presented at short as compared to long SOAs, an effect called the PRP effect. It is proposed that each task is composed of (at least) three processing stages – a perceptual, a central and a motor stage. The PRP effect is said to be explained by a central processing bottleneck, in which only one task can be centrally processed at a time (McCann & Johnston, 1992; Pashler, 1994; Welford, 1952). Importantly, the perceptual and motor stages of one task can proceed in parallel with any stage of the other task, while the two central stages must be processed serially.

* Corresponding author. Address: Fachbereich Psychologie, Universität Tübingen, Schleichstrasse 4, 72076 Tübingen, Germany. Fax: +49 7071 29 2410.
E-mail addresses: donna.bryce@uni-tuebingen.de (D. Bryce), daniel.bratzke@uni-tuebingen.de (D. Bratzke).

Thus, at short SOAs the central and motor stages of Task 2 are delayed until the end of the Task 1 central stage (see Fig. 1). This processing delay for Task 2 is called the slack time.

Corallo et al. (2008) and later Marti et al. (2010) combined this classic PRP paradigm with a methodology named quantified introspection, in which participants gave estimates of their own reaction times for each task after every trial of a PRP experiment. They indicated their reaction time estimates (named introspective RTs, or iRTs) on a visual analogue scale (VAS). In these studies, it was found that participants failed to report the PRP effect in their reaction times. That is, while there was an objective PRP effect on RT2, participant's reports of their RT2 (iRT2s) were unaffected by SOA. Thus, participants appeared to be unaware of the dual-task cost on response speed for the second task. Marti et al. (2010) explained these findings by invoking "a single hypothesis: in a dual-task setting, introspection is tied up by the first task and cannot focus on the second target until decision on the first target is resolved" (p. 311). Thus, they support the idea that conscious awareness requires not only perceptual processes but also central (attentional) processes (Dehaene, Changeux, Naccache, Sackur, & Sergent, 2006; Del Cul, Baillet, & Dehaene, 2007; Sergent, Baillet, & Dehaene, 2005). Specifically, they posit that participants cannot be consciously aware of the second stimulus while they are centrally processing Task 1. Therefore, conscious awareness of the second stimulus is delayed at short SOAs, as it is linked to the end of the Task 1 central stage (see dotted lines in Fig. 1). This interpretation of the apparent unawareness of the PRP effect is based on the assumption that iRTs reflect accurate estimates of the time between conscious awareness of the stimulus and the related response. Therefore, we refer to this model as the *Temporal* model.

In order to test this Temporal model, Marti et al. (2010) asked their participants for estimates of RT1, RT2, SOA, and the slack or free time (i.e. the time between their decision on Task 1 and S2 onset) at the end of each trial. In addition to replicating the iRT results of Corallo et al. (2008), they found that the SOA was overestimated by about 250 ms when the two stimuli were presented simultaneously (i.e. at a SOA of 0 ms). This, as well as the finding that SOA estimates were non-linear across objective SOAs, is consistent with the assumption that the conscious awareness of the second stimulus is delayed at short SOAs while Task 1 is centrally processed. However, Corallo et al. (2008) found a similar overestimation of SOAs when participants only estimated the SOA without responding to the stimuli. Thus, overestimation of short SOAs may reflect a central tendency in time estimates (i.e. Vierordt's law; Bausenhart, Dyjas, & Ulrich, 2014; Lejeune & Wearden, 2009) rather than a delay of conscious awareness of S2. Marti et al. (2010) also found that while the objective slack/free time becomes negative at short SOAs (as S2 is presented before the end of Task 1 central stage), participants' estimates remained close to zero, supporting the idea that participants are not aware of S2 onset until the end of the Task 1 central stage. However, these slack/free time estimates may also be influenced by central tendency. While this result is especially difficult to interpret because there is no objective measure of the end of Task 1 central stage, the authors found that at short SOAs the slack/free time estimates did not significantly correlate with a measure that was intended to reflect objective slack/free time (the difference between RT1 and SOA). In summary, although Marti et al. (2010) provide some compelling evidence in support of the Temporal model, their data are not unequivocal and some key predictions of the model are missing. The most consistent data in support of Marti et al.'s (2010) idea that conscious awareness of S2 is delayed while Task 1 is centrally processed is the lack of a SOA effect on iRT2 in both Corallo et al. (2008) and Marti et al. (2010).

However, other explanations can also account for this null effect of SOA on iRT2 – for instance, participants might not base their iRTs on temporal information, but rather on the difficulty they experience in each task. In the PRP task, according to the central bottleneck model, RT2s are longer at short than long SOAs because central processing of the second task must wait for the central stage of Task 1 to be completed, and not because the task is more difficult. Indeed, Task 2 always remains

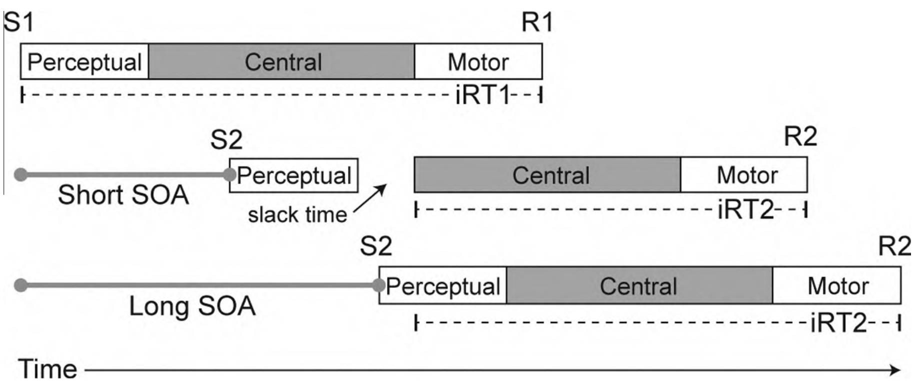

Fig. 1. Illustration of how introspective reaction times (iRTs, represented by the dotted lines) are produced according to the Temporal model, an example of both a short and a long SOA. iRT1 is composed of the Task 1 perceptual, central and motor stages. Participants are not consciously aware of the second stimulus while they are centrally processing Task 1. Therefore, the length of SOA determines which Task 2 stages are included in iRT2. At short SOAs, when Task 2 perceptual stage is completed before the end of Task 1 central stage, iRT2 contains only Task 2 central and motor stages. When SOA is such that there is some overlap between Task 2 perceptual stage and Task 1 central stage ('Long SOA' in this figure), iRT2 is composed of part of Task 2 perceptual stage, and all of Task 2 central and motor stages. At very long SOAs, when Task 1 central stage is completed before the second stimulus is presented, iRT2 is composed of Task 2 perceptual, central and motor stages.

the same regardless of SOA, and therefore its difficulty is also assumed to remain constant across all SOAs. While there is no objective measure of the difficulty of each trial, we can consider error rates to be a somewhat objective indicator of a condition's difficulty, following the logic that more difficult conditions probably elicit higher error rates. In line with the assumption that the difficulty of the two tasks are constant across SOAs, SOA does not usually affect error rates of either Task 1 or 2 (Kamienkowski & Sigman, 2008; Pashler, 1994), although see also Bratzke et al. (2008) and Schumacher et al. (2001, Experiment 2) for evidence of higher error rates at short SOAs. Thus, if the experienced difficulty (and error rate) of each task is unchanged across SOAs, and if participants base their iRTs on this experienced difficulty, a null effect of SOA on iRT2 would be produced. Otherwise, iRTs would follow the 'true' difficulty pattern indicated by error rates. This idea that participants base their iRTs on experienced difficulty is referred to as the *Difficulty* model. In order to further investigate such a model, in the present study we collected iRTs and difficulty estimates for each task.

While participants should use the time between stimulus onset and their response in order to estimate their reaction times, it is less clear what they base their difficulty estimates on. Delignières and Brisswalter (1996) investigated perceived difficulty in a range of tasks and found that estimates of physical exertion and of difficulty during cognitive tasks are closely related. This suggests that the feeling of difficulty could be based upon the amount of effort required to achieve success in a task. Indeed, while feeling of difficulty and task performance are typically related, one need not wait until the end of a task to evaluate performance and consequently estimate how much difficulty was experienced. Instead, this can be a rather immediate experience during task completion. Therefore, we posit that in the current context, various factors could affect a participant's feeling of difficulty in a task, e.g. trial-by-trial preparedness, automaticity of the stimulus–response pairing, and the perceptual complexity of stimuli.

To our knowledge there is only one other study that collected judgments of difficulty in a PRP task. Hartley, Maquestiaux, Brooks, Festini, and Frazier (2012) asked for difficulty estimates for the whole trial and combined data from Task 1 and Task 2 difficulty manipulations. It was found that trials with long SOAs were judged as easier than those with short SOAs, yet SOA only accounted for a very small proportion of the variance in judged difficulty (0.1%) in a linear regression. The present study complements these findings by examining the experienced difficulty of each task separately.

1.1. The present study

In order to test whether the Temporal or Difficulty model best described how iRTs are produced, we collected introspective difficulty estimates (iDiffs) and iRTs after each PRP trial in two experiments. In the first experiment, as in the study of Corallo et al. (2008; Experiment 2), the perceptual difficulty of Task 2 was manipulated. In Experiment 2, the perceptual difficulty of Task 1 varied. According to the central bottleneck model, the task in which perceptual difficulty is varied has different effects on RT2. When Task 2 perceptual difficulty is manipulated, the effect of Task 2 difficulty is absorbed during the slack time at short SOAs. That is, a more difficult S2 results in a longer perceptual stage, but since Task 2 perceptual processing and Task 1 central processing can occur in parallel, this prolongation extends into the slack time and is therefore not reflected in RT2. Across SOAs, this leads to an under-additive interaction of Task 2 difficulty and SOA on RT2 whereby the effect of Task 2 difficulty is reduced at short SOAs (Pashler, 1994; Fig. 2A). Conversely, when Task 1 perceptual difficulty is manipulated, the effect of Task 1 difficulty propagates onto Task 2 performance at short SOAs; this results in an over-additive interaction of Task 1 difficulty and SOA on RT2 whereby the effect of Task 1 difficulty is reduced at long SOAs (Fig. 2D).

As both iRT models in the context of the PRP task are rather complex, it is difficult to derive their predictions on a purely intuitive basis. Therefore, we used Monte-Carlo simulations to examine their predictions with respect to mean estimates. The length of each processing stage was selected with the aim of producing mean results similar to the objective RT data. Full details of the simulations are provided in Appendix A. iDiffs were derived from the objective difficulty levels. iRTs were calculated according to the Temporal model proposed by Marti et al. (2010) and the alternative Difficulty model. Since the Temporal model posits that conscious awareness of the second task is not possible while Task 1 is being centrally processed, Temporal iRTs were calculated according to this rule (Fig. 1; Eqs. (4)–(7) in Appendix A). In the Difficulty model, iRTs were essentially a multiplication of the iDiff for each task (Eqs. (8) and (9) in Appendix A). It should be noted that in these simulations, we assume that participants' experienced difficulty does not change with SOA as it was previously found that SOA has no effect on error rates (Kamienkowski & Sigman, 2008). Three SOAs were used as in the real experiments (50, 250, and 1250 ms) and 30,000 trials per SOA were simulated. Using these simulations, a series of predictions were obtained, five of which are diagnostic; that is, they are able to distinguish between the two models. Only diagnostic predictions are described here, and highlighted in Table 1 by an asterisk.

1.1.1. Introspective RTs

For both experiments, the Temporal model predicts increasing iRT2s with increasing SOAs (the reverse of a PRP effect, predictions 1.2t and 2.3t in Table 1; Fig. 2B and E), because at least part of the perceptual stage of Task 2 is included in iRT2 at long SOAs. It is important to note that neither Corallo et al. (2008) nor Marti et al. (2010) found such an increase in iRT2 at long SOAs. However, consistent with the idea that Task 2 perceptual stage contributes to iRT2 when the two tasks do not overlap, iRT2 showed an effect of Task 2 perceptual difficulty only at long SOAs in Experiment 2 of Corallo et al. (2008). This suggests that our interpretation of their explanation (here labeled the Temporal model) is correct. In contrast, the Difficulty model predicts flat iRT2 functions (predictions 1.2d and 2.2d in Table 1; Fig. 2C and F) if the assumption that SOA does not affect the difficulty of each task is correct. If this assumption is incorrect, and there is an effect of SOA on

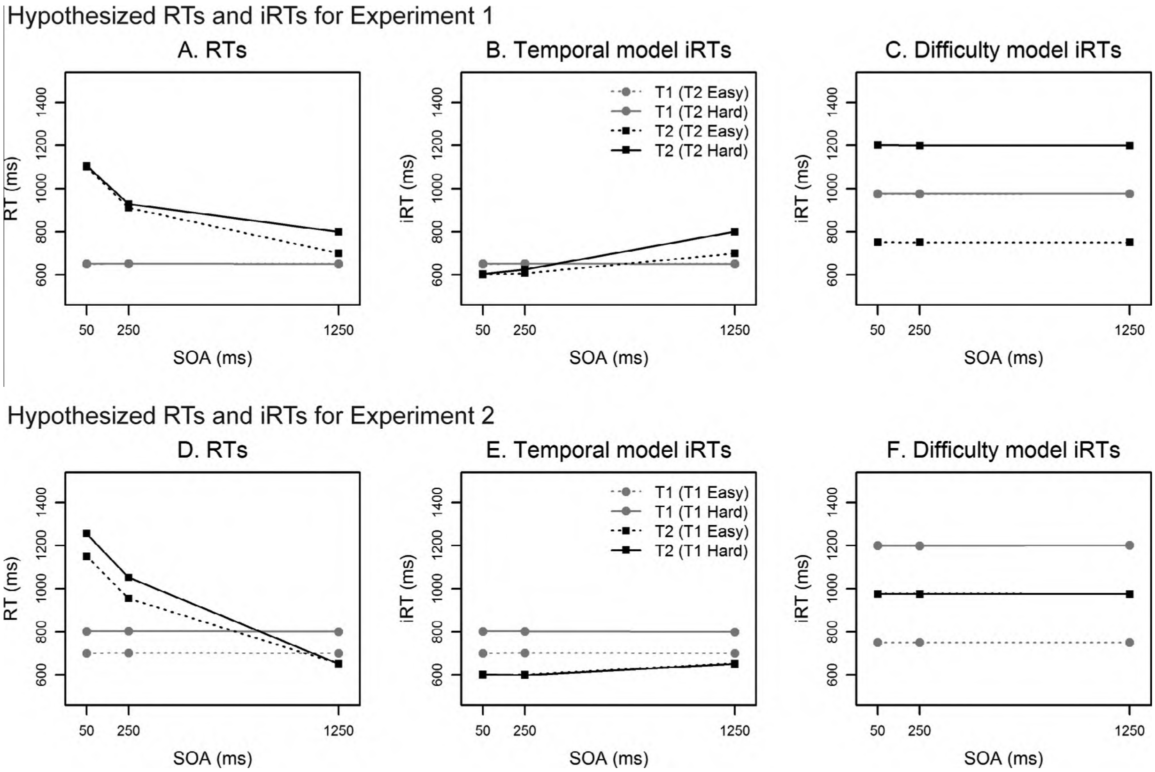

Fig. 2. Summary of model predictions for Experiment 1 (A–C) and Experiment 2 (D–F). The expected reaction time pattern (A and D) according to the central bottleneck model, the hypothesized introspective reaction time (iRT) pattern if they are generated according to the Temporal model (B and E) and the Difficulty model (C and F). Please note, the absolute values are not necessarily reliable, rather this figure simply illustrates the predicted patterns. Also, the predictions from the Difficulty model are based on the assumption that difficulty levels did not change across SOAs, hence the flat functions. *Note*, RT = reaction time; iRT = introspective reaction time; SOA = stimulus onset asynchrony.

Table 1
Summary of the predictions from the Temporal and Difficulty models. Diagnostic predictions (those that differ between the models) are highlighted with an asterisk (*).

Dependent variable	Temporal model		Difficulty model	
Experiment 1				
Mean iRT1	1.1t	Flat across SOAs	1.1d	Flat across SOAs
Mean iRT2	1.2t	Increasing with SOA*	1.2d	Flat across SOAs*
	1.3t	Greater effect of Difficulty at long than short SOAs*	1.3d	An effect of Difficulty that is unaffected by SOA*
Predictors	1.4t	SOA, RT2 & iDiff2 all contribute to iRT2. Model 1 (full) best*	1.4d	iDiff2 contributes to iRT2. Model 4 (excluding SOA) best*
Experiment 2				
Mean iRT1	2.1t	Flat across SOAs	2.1d	Flat across SOAs
	2.2t	An effect of Difficulty that is unaffected by SOA	2.2d	An effect of Difficulty that is unaffected by SOA
Mean iRT2	2.3t	Increasing with SOA*	2.3d	Flat across SOAs*
Predictors	2.4t	SOA & RT2 contribute to iRT2. Model 2 (excluding iDiff2) best*	2.4d	iDiff2 contributes to iRT2. Model 4 (excluding SOA) best*

Note. SOA = stimulus onset asynchrony; iRT = introspective reaction time; iDiff = introspective difficulty estimate; RT = reaction time.

experienced difficulty, this would be reflected in the error rates. In this case, the Difficulty model would predict that iRTs follow the same pattern as error rates. Our simulated data also showed that in Experiment 1, when Task 2 difficulty is manipulated, as well as an increase in iRT2 with increasing SOA, the Temporal model predicts that the Task 2 difficulty effect is visible in iRT2 but only at long SOAs (1.3t; Fig. 2B). The effect of Task 2 difficulty would only become visible at long SOAs because the perceptual stage is only included in iRT2 at long SOAs. In contrast, the Difficulty model predicts only a main effect of Task 2 difficulty on iRT2 (1.3d) as hard trials are experienced as more difficult and therefore reported as being longer than easy trials, regardless of SOA. In Experiment 2, the Temporal model predicts only the increasing iRT2 with increasing SOA already described (2.3t), and to a lesser extent than in Experiment 1 (Fig. 2E). This reduced SOA effect is predicted because when Task 1 difficulty is increased, the end of Task 1 central stage is delayed, and therefore less of the perceptual stage of Task 2 enters the iRT2 calculation at long SOAs than in Experiment 1. In contrast, the Difficulty model predicts only a flat function for iRT2 (i.e. no effect of SOA) or a pattern that mirrors the error rate pattern, and no effect of difficulty (2.3d).

1.1.2. iRT2 predictors

To establish what information contributes to iRT2, linear mixed effects (LME) models were conducted in which a full model was compared to three reduced models (see Section 2.2.1 for details). For Experiment 1, in which Task 2 difficulty was manipulated, the Temporal model predicts that SOA, RT2 and iDiff2 should contribute to iRT2 (1.4t). The inclusion of iDiff2 in the best model could be counterintuitive as the Temporal model assumes that iRTs are solely based on temporal information. However, iDiff2 could still contribute to iRT2 because RT2 and iDiff2 are related, at least at long SOAs (i.e. Task 2 difficulty affects RT2 via the duration of Task 2 perceptual processing and also more directly affects iDiff2). The Difficulty model predicts that iDiff2 should be the only significant predictor, and the model with SOA excluded is the best (1.4d). In Experiment 2, the Temporal model predicts that SOA and RT2 should significantly contribute to iRT2 (2.4t). However, in contrast to Experiment 1, for Experiment 2 iDiff2 should not contribute to iRT2 because RT2 and iDiff2 are no longer related when Task 1 difficulty is manipulated; therefore, the model with iDiff2 excluded is predicted to be the best model. The Difficulty model predicts that iDiff2 should be the only significant predictor in this model, and again that the model with SOA excluded is the best (2.4d).

2. Experiment 1: Task 2 difficulty manipulated

2.1. Method

2.1.1. Participants

Thirteen females and three males, aged between 19 and 35 years (M = 22.3 years) participated in one 1 h session. Participants reported normal hearing and normal or corrected-to-normal vision, and received either course credit or payment. All sixteen participants were right-handed.

2.1.2. 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. Two external response panels were used to record responses with the index and the middle finger of the left and right hand. 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 a plus or a minus symbol, degraded by a random square pattern, presented in the center of the screen for 500 ms. The degradation pattern was produced by small squares ($0.05 \times 0.05°$) randomly distributed over a quadratic area of $1.53 \times 1.53°$. Within each square, the colors of fore- and background were inverted (Leonhard, Bratzke, Schroeter, & Ulrich, 2012). There were 12 levels of degradation for each symbol (25–300 small squares, in steps of 25).

2.1.3. Procedure and design

Each trial began with a fixation point in the center of the screen for 250 ms. Next, S1 was presented, followed by S2 after a SOA of 50, 250 or 1250 ms. The task instruction was 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 minus symbol and with their right middle finger to the plus symbol. After both responses were collected, iRTs and iDiffs were collected. iRTs were prompted by the question "how long was your reaction time for the first/second task?"; iDiffs were prompted by the question, "how difficult was the first/second task?" Participants gave judgments of Task 1 in the first half and Task 2 in the second half of the experiment, or vice versa (counter-balanced across participants). The order of judgment question was also counter-balanced across participants (iDiff followed by iRT, or iRT followed by iDiff). In order to fully balance this design, there were 4 groups of participants. iRT judgments were given by a mouse click on a horizontal VAS ranging from 0 to 1500 ms. iDiff judgments were given by a mouse click on a vertical VAS ranging from "very easy" to "very difficult"; these were transformed into numerical values between 1 and 12 to echo the 12 levels of degradation in the visual stimuli. In case of an incorrect response in one of the two reaction time tasks, a feedback message was then displayed. Each half of the experiment began with a practice block (18 trials) followed by four experimental blocks (36 trials each). Every combination of stimuli was presented once in each half of the experiment: 3 SOA (50, 250, and 1250 ms) $\times$ 24 visual stimuli (plus or minus symbols, each with 12 levels of degradation) $\times$ 2 auditory stimuli (low or high tones), resulting in 144 experimental trials.

2.2. Results

2.2.1. Analysis

For ANOVA analyses, the visual stimuli were categorized as 'easy' (degradation levels 1–6) or 'hard' (levels 7–12). Error rates in Task 1 and Task 2 were analyzed in SOA (50, 250, and 1250 ms) $\times$ Task 2 difficulty (easy, hard) repeated measures

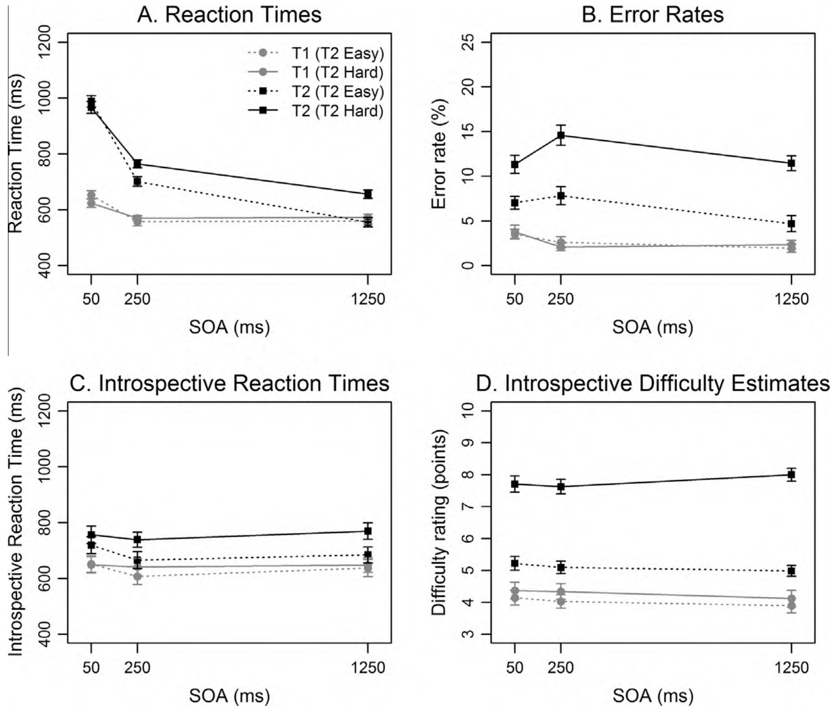

Fig. 3. Experiment 1 results. Reaction times (A), error rates (B), introspective reaction times (C) and introspective difficulty estimates (D) for Task 1 and 2 as a function of Task 2 difficulty and stimulus onset asynchrony (SOA). Error bars represent ±1 within-subjects SE.

ANOVAs. For all further analyses, all trials that contained an error in Task 1 or Task 2 were discarded (11.7%). Trials in which RT1 or RT2 deviated more than three standard deviations from the individual mean in each condition were excluded (3.2% of correct trials), as were trials in which the responses were grouped (within 100 ms of each other, 0.89% of remaining trials). RTs were analyzed by SOA (50, 250, 1250 ms) × Task 2 difficulty (easy, hard) × Judgment Type (judgment of Task 1 vs. judgment of Task 2) repeated measures ANOVAs. iRTs and iDiffs were analyzed by SOA (50, 250, 1250 ms) × Task 2 difficulty (easy, hard) repeated measures ANOVAs. To enhance readability of the ANOVA results, only significant results are reported unless $p < .1$ or the effect was of theoretical importance. 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 Cousineau (2005).

LME model analyses were conducted to investigate which information contributed to iRT2. Means of dependent variables (iRT2, RT1 and RT2) were calculated for each participant, SOA and difficulty level. These data were then used to create four models with iRT2 as the dependent variable: the full model included the fixed effects RT1, iDiff2, RT2, and three levels of SOA, and the random effect of participant as predictors. The three reduced models excluded iDiff2, RT2, and SOA, respectively. Descriptively, better models have a lower corrected AIC (AICc); chi-squared tests determined whether each reduced model was significantly different from the full model. If the test was non-significant, the reduced model with fewer variables was considered superior to the full model.

2.2.2. Reaction times

Fig. 3A shows RT1 and RT2 as a function of SOA and Task 2 difficulty. RT1 decreased with increasing SOA, $F(2,30) = 10.71$, $p = .001$, $\eta_p^2 = .15$ (post hoc tests revealed RT1 at the 50 SOA was significantly longer than the 250 SOA, $p < .001$, and the 1250 SOA, $p < .001$). RT2 showed the pattern predicted by the central bottleneck model. There was a classic PRP effect – RT2 was 366 ms longer at the shortest than the longest SOA, $F(2,30) = 117.48$, $p < .001$, $\eta_p^2 = .87$ (all comparisons significant $p < .001$). Furthermore, RT2 was 49 ms longer in hard trials than easy trials, $F(1,15) = 10.81$, $p = .005$, $\eta_p^2 = .42$, and this effect of Task 2 difficulty was reduced at short SOAs, $F(2,30) = 10.23$, $p < .001$, $\eta_p^2 = .41$.[1]

[1] There were some effects of Judgment Type on RTs. When Task 2 was being judged, RT1 was 49 ms slower than when Task 1 was being judged, $F(1,15) = 12.40$, $p = .003$, $\eta_p^2 = .45$. The greatest effect of Judgment Type (83 ms) was at the shortest SOA, $F(2,30) = 5.21$, $p = .011$, $\eta_p^2 = .26$. Importantly, the main effect of SOA on RT1 remained when data were analyzed separately for each Judgment Type, $F(2,30) = 6.31$, $p = .013$, $\eta_p^2 = .30$, and $F(2,30) = 10.41$, $p < .001$, $\eta_p^2 = .41$. The PRP effect on RT2 was 73 ms greater when Task 2 was judged than when Task 1 was judged, $F(2,30) = 7.27$, $p = .007$, $\eta_p^2 = .33$. Considering these results together, it appears that participants responded faster to Task 1 when it was the focus of the introspective report. In turn, when Task 2 was the focus of the introspective report, participants produced a slower RT1, which led to a longer slack time for Task 2 and a greater PRP effect. Importantly, these effects are considered small and the task which was being judged did not substantially change the general RT pattern across SOAs, so the previous results from the SOA × Task 2 difficulty ANOVA are considered reliable.

Table 2
Estimates for the fixed effects of the parameters entered into the linear mixed effect models predicting iRT2 in Experiment 1, and the corrected AIC (AICc) of each model.

Model	Intercept	RT1	iDiff2	RT2	SOA	AICc
1	397***	0.06	21.37***	0.16***	0.05***	3335
2	494***	0.00		0.26***	0.07***	3417
3	424***	0.24***	24.03***		0.01	3348
4	438***	0.15***	23.46***	0.05*		3344

Note. RT = reaction time; iDiff = introspective difficulty estimate; SOA = stimulus onset asynchrony.
* $p < .05$; ** $p < .01$; *** $p < .001$.

2.2.3. Error rates

Fig. 3B depicts the error rates in Task 1 and Task 2 as a function of SOA and Task 2 difficulty. Unexpectedly, Task 1 error rates were higher at short than long SOAs, $F(2,30) = 3.42$, $p = .046$, $\eta_p^2 = .19$. Tukey contrasts indicated that the 50 ms SOA had significantly higher Task 1 error rates (3.6%) than the 1250 ms SOA (2.1%, $p = .043$). Task 2 error rates were higher in hard (10.2%) as compared to easy trials (4.0%), $F(1,15) = 27.16$, $p < .001$, $\eta_p^2 = .64$, and higher at short than long SOAs, $F(2,30) = 5.02$, $p = .013$, $\eta_p^2 = .25$. Tukey contrasts indicated that the 250 ms SOA had significantly higher Task 2 error rates (11.2%) than the 1250 ms SOA (8.1%, $p = .005$).

2.2.4. Introspective RTs

Fig. 3C depicts iRT1 and iRT2 as a function of SOA and Task 2 difficulty. iRT1 was 14 ms longer for trials in which Task 2 was hard compared to easy, $F(1,15) = 7.70$, $p = .014$, $\eta_p^2 = .34$. The greatest Task 2 difficulty effect on iRT1 was at the 250 ms SOA (34 ms), followed by the 1250 ms (11 ms) and 50 ms (3 ms) SOA, $F(2,30) = 3.43$, $p = .046$, $\eta_p^2 = .19$. These iRT1 effects were not predicted by either the Temporal or Difficulty model. Similarly, iRT2 was 66 ms longer for hard trials than easy trials, $F(1,15) = 16.54$, $p = .001$, $\eta_p^2 = .52$. There was a trend toward a SOA $\times$ Task 2 difficulty interaction on iRT2, $F(2,30) = 2.53$, $p = .097$, $\eta_p^2 = .14$, indicating that the Task 2 difficulty effect may be larger at the long SOA (85 ms) than at the shortest SOA (41 ms).

2.2.5. Estimated difficulties

Fig. 3D depicts iDiff1 and iDiff2 as a function of SOA and Task 2 difficulty. iDiff1 was 0.25 points higher when Task 2 was hard than when Task 2 was easy, $F(1,15) = 5.22$, $p = .037$, $\eta_p^2 = .26$, an effect not predicted by either model. However, iDiff2 was 2.7 points higher in hard trials than easy trials, $F(1,15) = 65.15$, $p < .001$, $\eta_p^2 = .81$, and there was a greater difference between the difficulty levels at the long SOA than the short SOAs, $F(2,30) = 4.33$, $p = .048$, $\eta_p^2 = .22$.

2.2.6. iRT2 predictors

The full model (1; Table 2) was considered superior to model 2 without iDiff2, $\chi^2(1) = 82.55$, $p < .001$, model 3 without RT2, $\chi^2(1) = 20.32$, $p < .001$, and model 4 without SOA, $\chi^2(1) = 18.68$, $p < .001$.

2.3. Discussion

This experiment and Experiment 2 of Corallo et al. (2008) had a key feature in common – Task 2 perceptual stage difficulty was manipulated. The RT patterns in both their experiment and the present experiment were as predicted by the central bottleneck model. As in the study of Corallo and colleagues, participants did not report the objective PRP effect in their iRT2s. Neither the present nor Corallo et al.'s results meet the Temporal model's prediction that iRT2 increases with increasing SOA (the reverse of a PRP effect). In line with the predictions of the Temporal model, Corallo et al. found an interaction of SOA $\times$ Task 2 difficulty on iRT2. However, we observed only a trend toward such an effect. While a marginally significant result should be interpreted cautiously, a significant interaction could be interpreted as support for the Temporal model. The only clear support for the Temporal model was provided by the LME results of Experiment 1, which indicated that all predictors contributed to iRT2.

The Difficulty model was also only partly supported by the results of Experiment 1. Harder tasks had longer iRTs than easy tasks across all SOAs, as predicted by the Difficulty model. Further, the Difficulty model predicts that SOA and Task 2 difficulty effects on iDiffs should be reflected in iRTs. In line with this prediction, we found a SOA $\times$ Task 2 difficulty interaction pattern on iDiff2 which was also observed in iRT2. However, while this interaction was significant in iDiff2, it did not reach significance in iRT2. Thus, it could be that iRTs are not direct transformations of iDiffs, and that other information contributes to them, an interpretation that is also supported by the LME results.

Contrary to our expectation that error rates would be stable across SOAs, we observed SOA effects on both Task 1 and Task 2 error rates. As already mentioned in Section 1.1, it is reasonable to assume that experienced difficulty and error rates are related. Conditions that are experienced as being more difficult will likely elicit more errors and participants might infer the

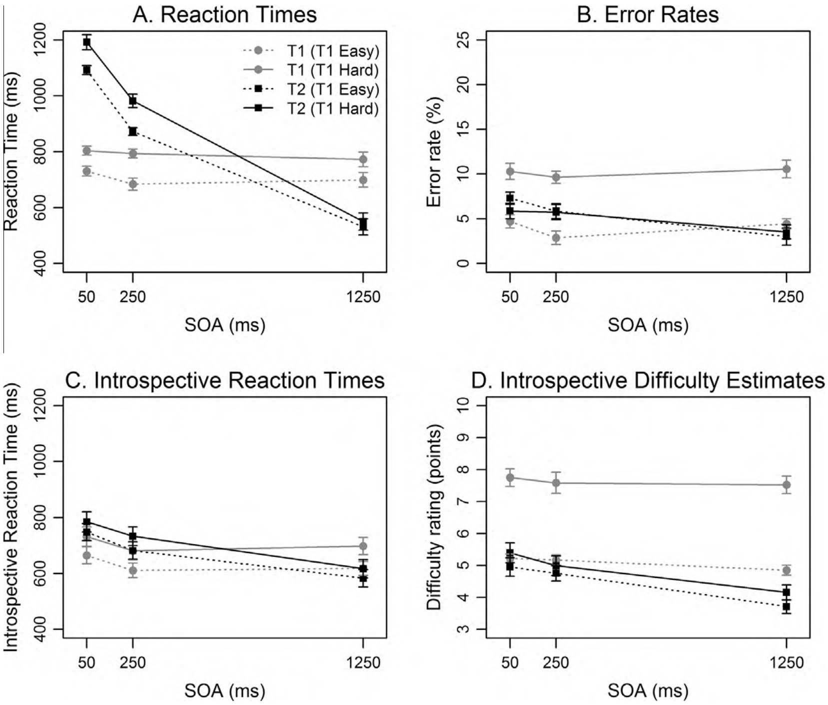

Fig. 4. Experiment 2 results. Reaction times (A), error rates (B), introspective reaction times (C) and introspective difficulty estimates (D) for Task 1 and 2 as a function of Task 1 difficulty and stimulus onset asynchrony (SOA). Error bars represent ±1 within-subjects SE.

difficulty of a particular condition from how many errors they commit. If error rates reflect experienced difficulty, we would expect that the iDiffs pattern would mirror the pattern of error rates. This, however, was not observed in Experiment 1.

In summary, the results from Experiment 1 were rather equivocal in their support for the Temporal and Difficulty models. It seems that participants may use both difficulty-based information and temporal information when producing their iRTs. To further test the predictions of the two models, in Experiment 2 we manipulated the perceptual difficulty of Task 1 which creates several further diagnostic predictions. Importantly, the Temporal model assumes that participants are able to time Task 1 fairly reliably, as Task 1 can be processed without any limitations due to the central bottleneck. Therefore, manipulating Task 1 difficulty allows us to evaluate the veracity of this assumption.

3. Experiment 2: Task 1 difficulty manipulated

3.1. Method

3.1.1. Participants

Twelve female and four male naïve participants, aged between 19 and 44 years (M = 22.4 years) participated in one 1 h session. Participants reported normal hearing and normal or corrected-to-normal vision, and received either course credit or payment. Fourteen participants were right-handed, two were left-handed.

3.1.2. Apparatus and stimuli

The experimental set up was identical to Experiment 1, apart from the order of stimuli.

3.1.3. Procedure and design

The design was the same as Experiment 1 except that S1 was the visual stimulus and S2 was the tone. In Task 1, participants had to respond with their left middle finger to the minus symbol and with their left index finger to the plus symbol. In Task 2, they had to respond with their right index finger to the low tone and with their right middle finger to the high tone. iRTs and iDiffs were collected in the same manner as in Experiment 1.

3.2. Results

3.2.1. Analysis

All trials that contained an error in Task 1 and/or Task 2 were discarded (11.7%). Trials with RT1 or RT2 that deviated more than three standard deviations from the individual mean in each condition were excluded (2.9% of correct trials), as were

Table 3
Estimates for the fixed effects of the parameters entered into the linear mixed effect models predicting iRT2 in Experiment 2, and the corrected AIC (AICc) of each model.

Model	Intercept	RT1	iDiff2	RT2	SOA	AICc
1	277***	0.05	63.12***	0.11**	−0.01	3463
2	519***	0.12*		0.14**	−0.06*	3553
3	326***	0.13***	64.55***		−0.06***	3462
4	258**	0.04	64.13***	0.13***		3454

Note. RT = reaction time; iDiff = introspective difficulty estimate; SOA = stimulus onset asynchrony.
* $p < .05$; ** $p < .01$; *** $p < .001$.

trials in which the responses were grouped (within 100 ms of each other, 1.2% of remaining trials). Further analyses were conducted in the same manner as in Experiment 1.

3.2.2. Reaction times

Fig. 4A shows RT1 and RT2 as a function of SOA and Task 1 difficulty. RT1 was 85 ms longer in hard trials than easy trials, $F(1,15) = 42.78$, $p < .001$, $\eta_p^2 = .74$. RT2 showed a classic PRP effect of 602 ms, $F(2,30) = 130.48$, $p < .001$, $\eta_p^2 = .90$ (all contrasts $p < .001$). As predicted by the central bottleneck model, the effect of Task 1 difficulty propagated onto Task 2 performance. RT2 was 77 ms longer in trials with a hard Task 1 compared to those with an easy Task 1, $F(1,15) = 21.82$, $p < .001$, $\eta_p^2 = .59$, and this effect of Task 1 difficulty on RT2 decreased with increasing SOA, $F(2,30) = 8.04$, $p = .002$, $\eta_p^2 = 35.$[2]

3.2.3. Error rates

Error rates are depicted in Fig. 4B. Task 1 error rates were higher in hard (10.2%) than easy trials (4.0%), $F(1,15) = 37.68$, $p < .001$, $\eta_p^2 = .72$. As in Experiment 1, Task 2 error rates were lower at long than short SOAs, $F(2,30) = 6.20$, $p = .006$, $\eta_p^2 = .29$. Tukey contrasts indicated that participants made fewer errors in the 1250 ms SOA condition (3.3%) than in both the 50 ms (6.6%, $p = .002$) and the 250 ms SOA conditions (5.8%, $p = .027$).

3.2.4. Introspective RTs

Fig. 4C depicts iRT1 and iRT2 as a function of SOA and Task 1 difficulty. iRT1 was affected by SOA, $F(2, 30) = 8.80$, $p < .001$, $\eta_p^2 = .37$. Post hoc tests revealed that iRT1 at the 50 ms SOA was significantly longer than the 250 ms SOA, $p < .001$, and the 1250 ms SOA, $p = .005$. Further, iRT1 was 72 ms longer in hard trials than easy trials, $F(1, 15) = 24.80$, $p < .001$, $\eta_p^2 = .62$. iRT2 also decreased with increasing SOA, $F(2,30) = 12.46$, $p = .001$, $\eta_p^2 = .45$ (iRT2 was significantly shorter at the 1250 ms SOA than the 250 ms SOA, $p = .004$, and the 50 ms SOA, $p < .001$). Additionally, iRT2 was 41 ms longer in trials with a hard Task 1 than trials with an easy Task 1, $F(1,15) = 14.43$, $p = .002$, $\eta_p^2 = .49$. That both iRT1 and iRT2 decreased with increasing SOA does not meet the original predictions of either the Temporal or the Difficulty model (predictions 2.1t/d, 2.2t/d and 2.3t/d in Table 1).

3.2.5. Estimated difficulties

Fig. 4D shows iDiff1 and iDiff2 as a function of SOA and Task 1 difficulty. iDiff1 was 2.5 points higher for hard trials than easy trials, $F(1,15) = 65.74$, $p < .001$, $\eta_p^2 = .81$. iDiff2 decreased with increasing SOA, $F(2,30) = 13.02$, $p = .002$, $\eta_p^2 = .46$ (post hoc tests revealed iDiff2 was significantly lower at the 1250 ms SOA than at the 250 and 50 ms SOAs, *ps* <.001), and was 0.4 points higher for trials with a hard Task 1 than with an easy Task 1, $F(1,15) = 12.09$, $p = .003$, $\eta_p^2 = .45$.

3.2.6. iRT2 predictors

Model 4 (with SOA removed; Table 3) was the only one not significantly different from the full model, $\chi^2(1) = 0.54$, $p = .461$; therefore it is considered the best model. The full model was significantly different from model 2 without iDiff2, $\chi^2(1) = 88.95$, $p < .001$, and model 3 without RT2, $\chi^2(1) = 7.55$, $p = .006$.

3.3. Discussion

The most striking feature of Experiment 2 results is that there was in fact a decrease in iRT2 with increasing SOA. This could be considered to be a PRP effect on iRT2, although it was much smaller than the PRP effect on RT2. This finding is clearly in contrast with previous results (Corallo et al., 2008; Marti et al., 2010) and the assumptions of the Temporal model. Further, the prediction of the Temporal model that SOA should contribute to iRT2 in the LME model was not supported.

[2] Judgment Type did not affect RT1, $F(1,15) = 0.18$, $p = .676$, $\eta_p^2 = .01$, but RT2 was 86 ms slower when Task 1 was being judged than when Task 2 was being judged, $F(1,15) = 7.48$, $p = .015$, $\eta_p^2 = .33$. As the task which was being judged did not affect the RT pattern of interest, the main analyses (above) are considered reliable.

On first examination, an apparent PRP effect on iRT2 is also not in line with the Difficulty model. However, as previously noted, in simulating these models, the experienced difficulty of each task was assumed to be stable across different SOAs. In fact, the results of Experiment 2 revealed that error rates were affected by SOA. Participants made more errors at short than long SOAs, indicating that Task 2 was more difficult at short SOAs. If we take into account that conditions with more errors are experienced as being more difficult, the results of Experiment 2 largely support the Difficulty model. This adjustment to how iDiffs were calculated was simulated and the detailed predictions are presented in Appendix B. When the calculation of iDiffs was adjusted to take into account observed error rates, and therefore more realistic feelings of difficulty, the main effects of Task 1 Difficulty and SOA on iRT1, and the main effect of SOA on iRT2 are correctly predicted. Indeed, it is clear from Fig. 4 that error rates in Task 2, iDiff2 and iRT2 all follow the same pattern – the longest SOA is significantly different (easier and faster) than the two short SOAs, suggesting that participants did indeed base their iRTs on iDiffs. For Task 1 performance, however, these three measures are not in such strong agreement. While Task 1 error rates and iDiff1 were unaffected by SOA, iRT1 decreased across SOAs.

Another explanation for the decreasing iRT2 with increasing SOA is that iRTs reflect accurate timing of RTs. However, two results argue against this possibility. First, participants failed to report the interaction of Task 1 difficulty and SOA in RT2. Second, we observed a SOA effect on iRT1 that was not present in the objective RT1 data. Thus, it appears that in Experiment 2, iRTs were based less on accurate time information, and more on a feeling of difficulty. This can perhaps explain the discrepancy between the current and previous (Corallo et al., 2008; Marti et al., 2010) results; it could be that participants were also basing their iRTs on experienced difficulty in previous studies, but SOA affected the feeling of difficulty differently. As neither Corallo et al. (2008) nor Marti et al. (2010) reported SOA effects on error rates we are unfortunately unable to further evaluate this explanation.

In summary, the results of Experiment 2 mostly support the Difficulty model – iRTs appear to be largely based on a participant's experience of difficulty which, in this experiment, was affected by SOA and reflected in higher error rates at short SOAs.

4. General discussion

The primary aim of the present study was to test Marti et al.'s (2010) explanation for the PRP effect not being reported by participants – that participants are not consciously aware of the second task while Task 1 is being centrally processed (here named the Temporal model). The results of Experiments 1 and 2 do not support this explanation. The secondary aim was to investigate whether the lack of a PRP effect in iRTs could also be explained by iRTs being strongly influenced by the experience of difficulty. While Experiment 1 results did not unequivocally support this explanation, Experiment 2 results largely concurred with it.

Only one finding could be interpreted as support for the Temporal model – in Experiment 1, the LME model results indicated that the best model of iRT2 included all variables. In contrast, the Temporal model cannot explain the mean iRT results of either experiment – iRT2 did not increase with SOA as predicted by the Temporal model (indeed iRT2 even decreased with SOA in Experiment 2).

In support of the Difficulty model, SOA did not enter the best fitting LME model in Experiment 2. The mean iRT2 results were also as predicted by the Difficulty model in Experiment 1 – iRT2 was only affected by Task 2 difficulty, and was unaffected by SOA. When the probable experienced difficulty was taken into account, the apparent PRP effect on iRT2 in Experiment 2 could also be explained by the Difficulty model. That is, in our Difficulty model we assumed that the 'real' difficulty of the tasks was unaffected by SOA. In fact, it is fair to suppose that Task 2 (in Experiment 2) or both tasks (in Experiment 1) were experienced as more difficult at short than long SOAs, based on the error rates.

Indeed, we adapted the Difficulty model to use the probable experienced difficulty to produce iRTs (Appendix B) and here we evaluate this revised model's ability to account for the observed data. For Experiment 1, both the main effect of Task 2 difficulty, and the trend for a SOA $\times$ Task 2 difficulty interaction on iRT2 (which could also be interpreted as support for the Temporal model) were correctly predicted by the revised Difficulty model. However, the effects on iRT1 were not predicted. With respect to Experiment 2, the revised Difficulty model predictions look strikingly similar to the observed data pattern. That is, the main effects of Task 1 difficulty and SOA on iRT1, and the main effect of SOA on iRT2 were correctly predicted. What was not predicted was the Task 1 difficulty effect on iRT2. Thus, when the calculation of iDiffs is adjusted to take into account the effect of SOA on error rates, and therefore more realistic feelings of difficulty, the majority of the data can be explained by participants basing their iRTs on their feeling of difficulty in each task. The major exception to this is an apparent cross-task transfer of difficulty effects. That is, when Task 2 difficulty was manipulated (Experiment 1) there was an effect of Task 2 difficulty (and an interaction with SOA) on iRT1. Likewise, when Task 1 difficulty was manipulated (in Experiment 2) there was an effect of Task 1 difficulty on iRT2. These results question an important assumption of the quantified introspection method – that participants are able to provide their judgments about each task independently. Instead, it seems that the perceptual stage difficulty of one task affects the introspective judgment of the other.

The SOA effects on error rates and difficulty estimates that we observed were not initially expected but are not without example. In a study investigating the source of dual-task interference, Hartley et al. (2012) also found SOA effects on two measures thought to reflect aspects of difficulty: the electrodermal response (EDR, considered a measure of top-down executive processing) and difficulty estimates. However, there are some crucial differences between their study and ours: first, they did not collect the two measures separately for each task within a PRP trial; and second, they manipulated more than one processing stage in each experiment and reported aggregated data, making it impossible to evaluate effects within each condition. In their study, Task 2 perceptual difficulty (numbers as digits vs. words) affected only EDR, while SOA affected both EDR and difficulty estimates. In contrast, we found effects of Task 2 difficulty on iDiff1 and iDiff2 as well as a SOA $\times$ Task 2 difficulty interaction on iDiff2. One possible reason for the divergent findings is that our stimulus degradation manipulation may have affected feelings of difficulty to a greater degree than did their notation manipulation. Indeed, our task difficulty effects on RTs were greater. Similar to our results, Hartley et al. (2012) found that both Task 1 (central stage) difficulty and SOA affected difficulty estimates, but the two did not interact (EDR was not collected in this condition). Hartley and colleagues concluded that EDR and difficulty estimates reflect different types of difficulty, as task difficulty affected EDR and difficulty estimates similarly, while SOA affected EDR to a much lesser extent than difficulty estimates. However, this conclusion seems somewhat premature as they did not collect EDR and difficulty estimates within one experiment. Nevertheless, an experiment in which both measures are collected within one trial could provide some insight regarding the information that contributes to difficulty estimates. However, our own findings suggest that participants' experiences of difficulty are affected by which of the two tasks is manipulated. Therefore, it seems important to separate these experimental conditions in future experiments.

There are some possible limitations of the present and previous studies related to the use of VASs for collecting introspective measures. First, asking for reports of both iDiff and iRT within one trial could encourage similar scoring across the two measures. However, to alleviate this limitation, the iRT scale was horizontal while the iDiff scale was vertical. It appears that this manipulation was at least partly successful, as the effects of SOA and Task difficulty on iRT and iDiff were not identical, indicating that participants were able to give independent reports of each measure. Second, collecting two introspective measures at the end of each trial could place memory demands on the participants, which may impair the accuracy of their judgments. However, Marti et al. (2010) asked four questions at the end of each trial, and found no effect of question order on accuracy of reports. Third, the use of VASs to report time durations (such as RT) could be flawed, as participants could be inclined to treat it as a timeline (as the mental timeline typically goes from left to right, Santiago, Lupiáñez, Pérez, & Funes, 2007; Ulrich & Maienborn, 2010). This would mean that participants would never place RT1 later in the scale than RT2 (because R2 follows R1). However, there is evidence that such a bias is unlikely here, as in Experiment 2 iRT1 was placed later in the scale (reported as longer) than iRT2 in trials with a long SOA (even though R2 always followed R1). Further, in these experiments, participants never gave an iRT1 and an iRT2 judgment within the same trial, which may have reduced the effects of such a bias.

The present results suggest that participants neither accurately time their RTs, nor portions of their RTs, as predicted by the Temporal model. Perhaps time information is accurately encoded at first but later lost, or perhaps in attentionally demanding dual-tasks the time information is never successfully encoded. Several previous studies have provided evidence that performing a timing task concurrent with a non-temporal task (as the PRP task is in these experiments) leads to a shortening of perceived time, and results in more variable estimates (Brown, 1997; Ruthruff & Pashler, 2010). Hence, it remains unclear whether participants are unable to time these durations accurately, or whether they simply choose to utilize difficulty information instead, as it is typically strongly related to RT. Further research into whether people can accurately judge the time intervals present in a PRP task when they are not simultaneously completing the task could shed light on this issue.

Previous studies (Corallo et al., 2008; Marti et al., 2010) interpreted iRT results in the PRP paradigm as evidence that participants' awareness of Task 2 processes was delayed while Task 1 was centrally processed. However, the present study indicates that iRTs are not based only on time information, which raises doubts about the conclusions of previous studies. Note that these data do not disprove the fundamental idea that attention is required for conscious awareness. Other evidence consistent with this notion is provided by Marti, Sigman, and Dehaene (2012), who found that late frontal brain activity (thought to represent conscious access) was delayed in PRP trials with short SOAs and completely absent in Attentional Blink trials.

In conclusion, the present results argue against the previous notion that iRTs only reflect timing of mental processes that are available to conscious awareness. Instead, they indicate that iRTs are primarily based on experienced difficulty. The present findings thus indicate that people are probably not able to accurately introspect about the timing of cognitive processes that develop in the millisecond range, especially in complex and attentionally demanding task situations. This could represent an important limitation in our metacognitive abilities.

Acknowledgments

Thanks to Rolf Ulrich for advice on previous drafts of this article, and to Anna-Lena Kraft for assistance with data collection. This research is supported by the Deutsche Forschungsgemeinschaft (BR 4297/1-1).

Appendix A. Simulating the Temporal and Difficulty models

In each trial, the duration of each processing stage (perceptual, central, motor) was sampled from a normal distribution with a particular mean, and a constant coefficient of variation of 0.3. This coefficient of variation was chosen to produce variance in RTs that was similar to that observed in these experiments, but it should be noted that additional simulations using coefficients of 0.2 and 0.4 made the same predictions as are reported here. In the real experiment, there were 12 levels of degradation (difficulty) of the visual stimuli, these were grouped for analyses into two levels 'easy' and 'hard'. For simulation purposes, we reduced this to six levels of difficulty, and therefore there were six corresponding perceptual stage distributions with means of 67, 100, 133, 167, 200 and 233 ms. The perceptual stage of the other task (a simple auditory judgment task) had a mean of 50 ms (perceptual stages were named 'P1' or 'P2'). The central stage (C1 or C2) had a mean of 500 ms. The motor stage (M1 or M2) had a mean of 100 ms.

RTs were calculated according to formulas derived from the central bottleneck model. RT1 was always calculated as the sum of all Task 1 processing stages:

$$\mathrm{RT1} = \mathrm{P1} + \mathrm{C1} + \mathrm{M1} \tag{1}$$

RT2 calculation was dependent upon the relationship between SOA and RT1. At short SOAs, when there was some slack time (i.e. when the second stimulus was presented such that the perceptual stage of Task 2 was not completed before the end of Task 1 central stage), RT2 was calculated as the sum of P2, C2, M2 and the slack time:

$$\begin{aligned} &\textit{If}\ \ SOA < P1 + C1 - P2, \\ &\mathrm{RT2} = \mathrm{P1} + \mathrm{C1} - \mathrm{SOA} + \mathrm{C2} + \mathrm{M2} \end{aligned} \tag{2}$$

In contrast, at long SOAs in which C1 was completed before the end of P2, so that there was no slack time, RT2 was calculated as the sum of all Task 2 processing stages:

$$\begin{aligned} &\textit{Else}, \\ &\mathrm{RT2} = \mathrm{P2} + \mathrm{C2} + \mathrm{M2} \end{aligned} \tag{3}$$

iDiffs were calculated based on the real difficulty levels of stimuli. That is, iDiffs for the visual task were sampled from normal distributions with means from 4 to 9 (relative to real difficulty levels 1 to 6). iDiffs for the auditory task were sampled from a distribution with a mean of 6.5. iDiff distributions had a constant coefficient of variation of 0.1. Again, it should be noted that increasing this coefficient of variation to 0.2 did not affect the predictions reported here. iRTs were calculated differently depending on the model, and in each case additional noise (denoted T) was added after the calculation to represent the noise involved in translating the introspective value to the VAS. The additional noise added at this stage was randomly sampled from a distribution ranging from –45 to 45 ms, in order to maintain a coefficient of variation of 0.6 for the Difficulty model iRT calculation.

In the Temporal model, iRT1 was always calculated as:

$$\mathrm{iRT1} = \mathrm{P1} + \mathrm{C1} + \mathrm{M1} + \mathrm{T} \tag{4}$$

iRT2 varied according to the length of the SOA. At very short SOAs, when Task 2 perceptual stage was completed before the end of Task 1 central stage (the 'Short SOA' scenario in Fig. 1):

$$\begin{aligned} &\textit{If}\ \ SOA < P1 + C1 - P2, \\ &\mathrm{iRT2} = \mathrm{C2} + \mathrm{M2} + \mathrm{T} \end{aligned} \tag{5}$$

When SOA was such that there was some overlap between Task 2 perceptual stage and Task 1 central stage, some but not all of Task 2 perceptual stage would be included in the iRT2 calculation. That is, the portion of Task 2 perceptual stage that was left after Task 1 central stage was completed was included in iRT2 (the 'Long SOA' scenario in Fig. 1).

$$\begin{aligned} &\textit{If}\ \ P2 > P1 + C1 - SOA, \\ &\mathrm{iRT2} = \mathrm{C2} + \mathrm{M2} + \mathrm{P2} - \mathrm{C1} - \mathrm{P1} + \mathrm{SOA} + \mathrm{T} \end{aligned} \tag{6}$$

Finally, at very long SOAs, when Task 1 central stage was completed before the second stimulus was presented, iRT2 was the sum of P2, C2, M2, and noise.

$$\begin{aligned} &\textit{Else}, \\ &\mathrm{iRT2} = \mathrm{P2} + \mathrm{C2} + \mathrm{M2} + \mathrm{T} \end{aligned} \tag{7}$$

In the Difficulty model simulations, iRTs were a multiplication of iDiffs (calculated as described above) plus some additional noise to represent the translation of the introspective value to the VAS (T, also described above).

$$\mathrm{iRT1} = \mathrm{iDiff1} * 150 + \mathrm{T} \tag{8}$$

$$\mathrm{iRT2} = \mathrm{iDiff2} * 150 + \mathrm{T} \tag{9}$$

This was independent of SOA, as it was assumed difficulty levels would not be affected by SOA.

Appendix B. Simulating revised Difficulty model

As there were unexpectedly higher error rates at short than long SOAs in each experiment, the calculation of iDiffs was adjusted to reflect the supposed greater feeling of difficulty at short than long SOAs. That is, trials in short SOAs had higher error rates, were assumed to be more difficult, and were therefore given an 'error rate' weighting to reflect this. These error

Table B1
Observed error rates and error rate weighting multipliers for each task at each SOA in each experiment.

	50 SOA		250 SOA		1250 SOA	
	Error rate	Multiplier	Error rate	Multiplier	Error rate	Multiplier
Experiment 1						
Task 1	3.65	0.35	2.34	0.25	2.15	0.20
Task 2	9.18	0.90	11.20	1.10	8.07	0.80
Experiment 2						
Task 1	7.49	0.75	6.25	0.65	7.49	0.75
Task 2	6.58	0.70	5.79	0.60	3.26	0.35

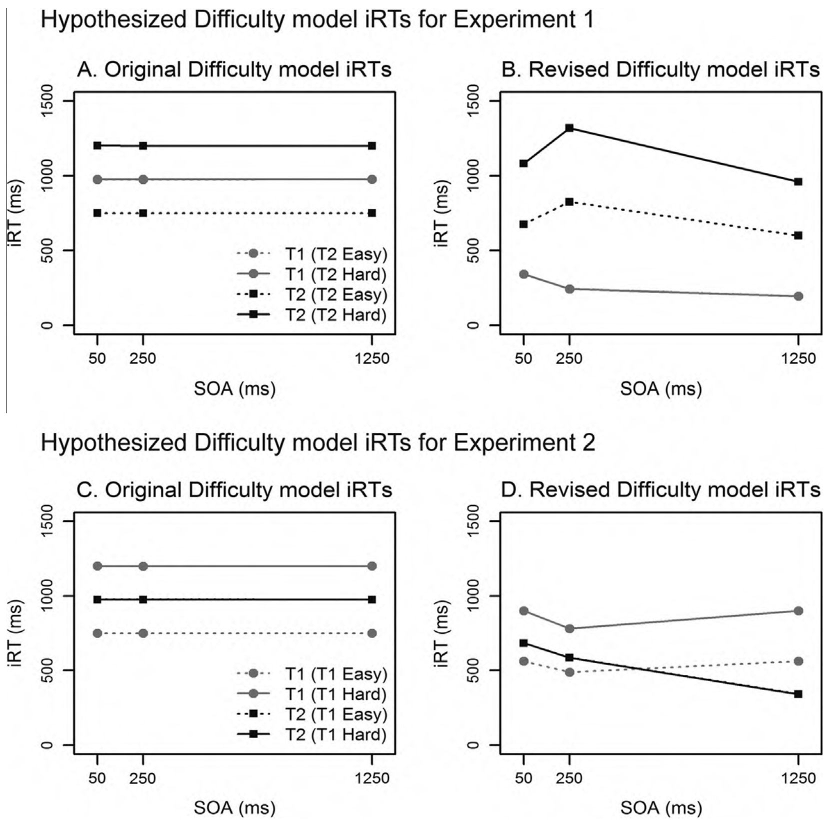

Fig. B1. The iRT predictions from the original Difficulty model (A and C) and the revised Difficulty model (B and D) for each experiment.

rate weightings were specific to each experiment and based on the observed error rate (Table B1 below). The error rate weighting (Multiplier) was rounded to the nearest 0.5.

Accordingly, iDiffs were calculated as:

$$\text{iDiff} = \text{Diff level} * \text{relevant error weighting} \tag{1}$$

iRTs were calculated in the same manner as described previously (see Appendix A) for each model. However, iRTs from the Difficulty model now reflected the true feeling of difficulty via the error rate weighting.

Fig. B1 below shows the predictions of the Difficulty models when iDiffs did not take error rates into account (A and C) and when they did (B and D). It is important to note that the absolute values of the hypothesized iRTs are not necessarily accurate, rather the reader should focus on the result patterns. The pattern of the mean iRT predictions now largely resemble the error rate patterns observed in each experiment.

References

Bausenhart, K. M., Dyjas, O., & Ulrich, R. (2014). Temporal reproductions are influenced by an internal reference: Explaining the Vierordt effect. *Acta Psychologica, 147*, 60–67. http://dx.doi.org/10.1016/j.actpsy.2013.06.011.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. http://dx.doi.org/0.1163/156856897X00357.

Bratzke, D., Ulrich, R., Rolke, B., Schroeter, H., Jentzsch, I., & Leuthold, H. (2008). Motor limitation in dual-task processing with different effectors. *The Quarterly Journal of Experimental Psychology, 61*(9), 1385–1399. http://dx.doi.org/10.1080/17470210701536856.

Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics, 59*(7), 1118–1140. http://dx.doi.org/10.3758/Bf03205526.

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.

Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method. *Tutorial in Quantitative Methods for Psychology, 1*(1), 42–45.

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., Kerszberg, M., & Changeux, J. P. (1998). A neuronal model of a global workspace in effortful cognitive tasks. *Proceedings of the National academy of Sciences of the United States of America, 95*(24), 14529–14534. http://dx.doi.org/10.1073/pnas.95.24.14529.

Del Cul, A., Baillet, S., & Dehaene, S. (2007). Brain dynamics underlying the nonlinear threshold for access to consciousness. *PLoS Biology, 5*(10), 2408–2423. http://dx.doi.org/10.1371/journal.pbio.0050260.

Delignières, D., & Brisswalter, J. (1996). The perception of difficulty and exertion in motor tasks: What can be known about perceptive continua through individual psychophysical exponents? *Journal of Human Movement Studies, 30*, 213–239.

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.

Kamienkowski, J. E., & Sigman, M. (2008). Delays without mistakes: Response time and error distributions in dual-task. *PLoS ONE, 3*(9). http://dx.doi.org/10.1371/Journal.Pone.0003196.

Kentridge, R. W., Nijboer, T. C. W., & Heywood, C. A. (2008). Attended but unseen: Visual attention is no sufficient for visual awareness. *Neuropsychologia, 46*, 864–869. http://dx.doi.org/10.1016/j.neuropsychologia.2007.11.036.

Lejeune, H., & Wearden, J. (2009). Vierordt's the experimental study of the time sense (1868) and its legacy. *European Journal of Cognitive Psychology, 21*(6), 941–960. http://dx.doi.org/10.1080/09541440802453006.

Leonhard, T., Bratzke, D., Schroeter, H., & Ulrich, R. (2012). Time-course analysis of temporal preparation on central processes. *Psychological Research Psychologische Forschung, 76*, 236–251. http://dx.doi.org/10.1007/s00426-011-0364-1.

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.

Marti, S., Sigman, M., & Dehaene, S. (2012). A shared cortical bottleneck underlying attentional blink and psychological refractory period. *Neuroimage, 59*(3), 2883–2898. http://dx.doi.org/10.1016/j.neuroimage.2011.09.063.

McCann, R. S., & Johnston, J. C. (1992). Locus of the single-channel bottleneck in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance, 18*(2), 471–484. http://dx.doi.org/10.1037//0096-1523.18.2.471.

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.

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

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

Schumacher, E. H., Seymour, T. L., Glass, J. M., Fencsik, D. E., Lauber, E. J., Kieras, D. E., et al (2001). Virtually perfect time sharing in dual-task performance: Uncorking the central cognitive bottleneck. *Psychological Science, 12*(2), 101–108. http://dx.doi.org/10.1111/1467-9280.00318.

Sergent, C., Baillet, S., & Dehaene, S. (2005). Timing of the brain events underlying access to consciousness during the attentional blink. *Nature Neuroscience, 8*, 1391–1400. http://dx.doi.org/10.1038/nn1549.

Shallice, T., & Burgess, P. (1996). The domain of supervisory processes and temporal organization of behaviour. *Philosophical Transactions of the Royal Society B-Biological Sciences, 351*(1346), 1405–1411. http://dx.doi.org/10.1098/rstb.1996.0124.

Smithies, D., & Stoljar, D. (2012). Introspection and consciousness: An overview. In D. Smithies & D. Stoljar (Eds.), *Introspection and consciousness* (pp. 3–26). New York: Oxford University Press.

Ulrich, R., & Maienborn, C. (2010). Left-right coding of past and future in language: The mental timeline during sentence processing. *Cognition, 117*(2), 126–138. http://dx.doi.org/10.1016/j.cognition.2010.08.001.

Welford, A. T. (1952). The "psychological refractory period" and the timing of high-speed performance – A review and a theory. *British Journal of Psychology, 43*, 2–19. http://dx.doi.org/10.1111/j.2044-8295.1952.tb00322.x.

Zimmerman, B. J. (2001). Theories of self-regulated learning – An overview and analysis. In B. J. Zimmerman & D. H. Schunk (Eds.), *Self-regulated learning and academic achievement: Theoretical perspectives*. NJ: Erlbaum.