Underestimation of Dual-Task Response Time is Not Caused by Delayed Perception

by

Michael David Klein

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Arts

in

Psychology

Waterloo, Ontario, Canada, 2015

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

In the Psychological Refractory Period (PRP) paradigm, the stimuli for two separate tasks are presented in rapid succession. The stimulus onset asynchrony (SOA) is typically varied, resulting in slower responses to Task 2 for shorter SOA. This slowing is referred to as PRP interference. Previous results have suggested that introspective estimates of response time are not sensitive to PRP interference. That is, although response time to Task 2 is slowed substantially at shorter SOAs, estimates of this response time typically do not vary with SOA. One interpretation of these findings is that perception of the stimulus for Task 2 is delayed during processing of Task 1. Therefore, according to this interpretation, introspective response times (IRTs) are not sensitive to PRP interference because the second stimulus (S2) is not perceived while the interference is occurring. Across two experiments using different approaches, I found that IRTs are still not sensitive to PRP interference under conditions in which delayed perception of S2 is precluded. The lack of sensitivity is therefore not dependent on perception being delayed, and must have had some other cause in my experiments. My results are consistent with an existing account in which IRTs are based on the perceived difficulty of each trial rather than on mental timing. However, the results may be best explained by an account in which people simply are not very good at estimating durations while under cognitive load.

Acknowledgements

I would like to thank my supervisor, Dr. Jennifer Stolz. This work would not have been possible without her ideas, support, and encouragement. I also thank the readers, Dr. Jonathan Fugelsang and Dr. Evan Risko, as well as the rest of the University of Waterloo department of Psychology Cognition and Perception division and my undergraduate research assistants. Finally, I would also like to thank my family and Melissa Meade for their support and encouragement throughout this process.

Table of Contents

Author's Declaration ii
Abstract iii
Acknowledgments iv
Table of Contents v
List of Figures vi
List of Tables vii
Overview of Previous Research 1
Experiment 1 6
    Methods 6
        Participants 6
        Apparatus 6
        Procedure 6
    Results 10
        Data analysis 10
        RT and IRT as a function of SOA 10
        Distance and notation manipulations 13
        Error percentage 17
        Regression analysis 17
    Discussion of Experiment 1 20
Experiment 2 23
    Methods 23
        Participants 23
        Apparatus 23
        Procedure 23
    Results 25
        Data analysis 25
        RT and IRT as a function of SOA 25
        Distance and notation manipulations 27
        Error percentage 31
        Regression analysis 31
        Response grouping 34
    Discussion of Experiment 2 34
General Discussion 37
    Alternative Accounts 37
    The Use of Time Estimation in Quantified Introspection 39
References 41

List of Figures

Figure 1 9
Figure 2 12
Figure 3 14
Figure 4 16
Figure 5 24
Figure 6 26
Figure 7 28
Figure 8 30

List of Tables

Table 1 19
Table 2 33

Overview of Previous Research

The use of introspection may be necessary to fully answer certain questions about the human mind, especially those regarding the nature of conscious experience. Unfortunately, introspective reports are highly subjective and often inaccurate about which mental processes are occurring (Nisbett & Wilson, 1977). As a result, researchers have increasingly endeavoured to use introspection in a more controlled way (e.g. Corallo, Sackur, Dehaene, & Sigman, 2008; Reyes & Sackur, 2014).

Corallo et al. (2008) designed a methodology that they called quantified introspection, wherein research participants produce quantitative reports of their introspective experience after each trial of a task. For example, they asked participants to estimate their own response time to every trial of a simple task and compared these subjective but quantitative introspective response times (IRTs) to the actual response times (RTs). They argued that this method can be used to identify which cognitive processes are accessible to introspection, reasoning that processes that are accessible to introspection should influence both RTs and IRTs, whereas those that cannot be accessed should influence only RTs.

Corallo et al. (2008) found that introspective and actual RTs followed the same pattern when only one task had to be performed, suggesting complete introspective access, but diverged in an overlapping tasks paradigm. The overlapping tasks, or Psychological Refractory Period (PRP), paradigm (see Pashler, 1994; Pashler & Johnston, 1989; Telford, 1931) involves the presentation of stimuli for two separate tasks in rapid succession. The general result is the PRP interference effect: Response to Task 2 is slower when the time between onset of the two stimuli (i.e. the stimulus onset asynchrony; SOA) is shorter. This is because most tasks involve a 'central' stage of processing that requires central resources, which process information serially (Pashler, 1994). This creates a bottleneck during which Task 2 central processing is queued while central resources are used for Task 1. Corallo et al. (2008) found that although decreasing SOA caused the expected increase in RT to Task 2, the corresponding introspective measure was completely insensitive to SOA. They interpreted this result as indicating that Task 2 queue time is inaccessible to introspection.

Marti, Sackur, Sigman, and Dehaene (2010) conducted a thorough follow-up study, asking four quantitative introspection-based questions after each trial of a PRP task. In addition to the RT estimates for the two tasks, they had participants estimate the SOA as well as the amount of time between their response to Task 1 and the onset of the second stimulus (S2), which they referred to as "slack time" when S2 occurs first and "free time" when the response occurs first. Their main finding, in addition to replicating Corallo et al.'s (2008) results, is that participants tended to overestimate short SOAs and to under-report slack time. Their data are consistent with an account in which, regardless of SOA, S2 is not consciously perceived until central resources are released from Task 1. This interpretation is consistent with dual-stage models of conscious access, which hold that conscious access requires central resources (Dehaene, Sergent, & Changeux, 2003; Chun & Potter, 1995), and fits in with other evidence that perception of S2 is delayed in the PRP paradigm (e.g. Marti, Sigman, & Dehaene, 2012).

Ruthruff and Pashler (2010), however, provided some evidence that the insensitivity of introspection to PRP interference does not depend on delayed perception. They asked whether time estimation itself is subject to the central bottleneck, and had participants perform time estimation as Task 2 in a PRP paradigm. In one of their experiments, for example, participants judged the brightness of a stimulus and then indicated when they thought a given duration had passed from the onset of that stimulus. The duration of the bottleneck was varied by manipulating the difficulty of Task 1 rather than the magnitude of an SOA. If time estimation is subject to the central bottleneck, then time should not be estimated during Task 1 central processing, and overall durations should therefore be underestimated by the duration of this processing stage. When Task 1 was more difficult, and therefore took longer, participants did underestimate duration to a greater degree, however the magnitude of underestimation was smaller than what would be expected if mental timing processes were queued behind Task 1 processing. Ruthruff and Pashler (2010) concluded that time estimation is only partially subject to the central bottleneck. Importantly, because the two tasks used the same stimulus, these results cannot be explained by delayed perception of S2. Therefore something other than delayed stimulus perception must have contributed to the underestimation of time found by Ruthruff and Pashler (2010), and may also have contributed to the results found by Corallo et al. (2008) and Marti et al. (2010). One possibility, as suggested by Ruthruff and Pashler (2010), is that mental timing, rather than, or in addition to, conscious access, is subject to the central bottleneck. Another possibility is simply that performance on timing tasks is poor under concurrent load. There is a large literature showing that concurrent loads interfere with duration judgments, often resulting in underestimation (e.g. Macar, Grondin, & Casini, 1994; Brown, 1997; Block, Hancock, & Zakay, 2010). This is usually interpreted as evidence that time estimation draws on the same pool of resources required for other tasks (Brown, 1997; Ogden, Salominaite, Jones, Fisk, & Montgomery, 2011). Performing two tasks concurrently could therefore tax resources that are required for time estimation.

Three points leave the door open for a delayed perception account of Corallo et al. (2008) and Marti et al.'s (2010) findings. First, Ruthruff and Pashler (2010) found only *partial* underestimation of PRP interference time. Therefore delayed stimulus perception may still be required to explain the *complete* underestimation found by Corallo et al. (2008) and others. Second, allocating a larger proportion of attention to a concurrent non-temporal task can lead to greater underestimation of duration (Macar et al., 1994). By increasing the difficulty of the non-temporal task, Ruthruff and Pashler (2010) may have influenced participants to allocate more attention to that task in addition to increasing the duration of the bottleneck. Therefore it is not clear whether their results would generalize to experiments in which the bottleneck duration is manipulated using only changes in SOA. Finally, participants in Ruthruff and Pashler (2010) estimated an external interval rather than their own response time to a stimulus. These two types of judgments do not necessarily rely on the same processes. In fact, IRTs need not rely on mental timing at all. Bryce and Bratzke (2014) suggested that they instead depend on perceived task difficulty, which is presumably a good predictor of actual RT. The degree to which delayed perception of S2 contributes to RT underestimation is therefore unclear.

The aim of the current study is to test the delayed perception account of temporal underestimation in the PRP paradigm. Although Marti et al. (2010) make persuasive arguments, their data cannot completely rule out alternative accounts (e.g. Ruthruff & Pashler, 2010; Bryce

& Bratzke, 2014). A more direct demonstration would be to show that underestimation of time during the bottleneck depends on perception of S2 being delayed. If IRTs are sensitive to PRP interference under conditions in which delayed perception is precluded, it can be concluded with more confidence that delayed perception is the cause of the temporal underestimation. However, if IRTs remain insensitive to interference under these conditions, alternative explanations must be identified. Such conditions were achieved in Experiment 1 by masking S2 after 100 ms, so that it would need to be perceived within 100 ms of onset in order to be perceived at all.

Experiment 1

Methods

Participants.

Twenty University of Waterloo undergraduate students (9 male, 11 female, M_{age} = 19.8, SD_{age} = 5.3) participated in the experiment for course credit. All reported normal or corrected-to-normal visual acuity and normal colour vision.

Apparatus.

Displays were generated by an Intel Core i7 computer connected to a 24-in. LCD monitor with 640 × 480 resolution and a 60 Hz refresh rate. Audio was played through Logitech over-ear headphones. Responses were collected via keypress on the computer's keyboard. Participants viewed the display approximately 50 cm from the monitor and set the audio volume to a comfortable level.

Procedure.

All participants completed Experiments 1 and 2 within the same session, in counterbalanced order. Each experiment took approximately 30 minutes, beginning with 10 practice trials followed by six blocks of 32 experimental trials, for a total of 202 trials. No feedback was given during either the practice or experimental blocks.

Experiment 1 was similar to Corallo et al. (2008) Experiment 2. Task 1 was an auditory discrimination task in which participants determined whether a tone was high- or low-pitched. Task 2 was a number comparison task in which participants determined whether a visually presented number was higher or lower than 45 (Figure 1).

Each trial began with the presentation of a black fixation cross subtending 0.6° × 0.6° on a white background. After 1000 ms, either a 500 Hz or 1500 Hz tone was played for 100 ms. The pitch was chosen randomly on a trial-by-trial basis, and was equally likely to be high or low. After either 0, 100, 200, 300, 400, 500, 750, or 1000 ms, a number between 21 and 69 (inclusive, and never 45) replaced the fixation cross in black, Times New Roman, size 24 bold font. The number was masked after 100 ms by a row of 12 percentage signs that remained on screen until response. Notation type was manipulated, with half of the numbers in each block presented in word form and half presented in digit form. For each notation type, half of the numbers were close to 45 (within 12), and half were far from 45. Each combination of distance and notation type occurred once at each of the eight SOAs in every block. The particular number and notation on a given trial was randomly chosen within these constraints.

Participants reported whether the tone was high or low pitched by pressing the "Q" key with their left middle finger following a high-pitched tone and by pressing the "A" key with their left index finger following a low tone. They were allowed to make this response as soon as the auditory stimulus was presented. Participants reported whether the number was higher or lower than 45 by pressing the "M" key with their right middle finger following a higher number and by pressing the "N" key with their right index finger following a lower number.

After the response to the number, the text "Auditory task duration?" appeared. Participants were instructed to report the amount of time that had elapsed between onset of the auditory stimulus and their "Q" or "A" button press by clicking on a scale ranging from 0 to 2400 ms. The scale was continuous, however 0, 600, 1200, 1800, and 2400 ms were explicitly marked, as well as one unlabeled tick mark between each of these values. Once a value was chosen, the text changed to "Visual task duration?" and participants reported their visual task duration estimate in the same manner. This response was followed by the text "press space" which remained on screen until the spacebar was pressed, inducing a 500 ms blank screen before the start of the next trial. Participants were reminded several times during the task instructions that they were to estimate the elapsed time from the onset of the relevant stimulus to their relevant response. They were informed that there would sometimes be a temporal gap between presentation of the auditory and visual stimuli, however they did not know the range or distribution of SOAs that would occur.

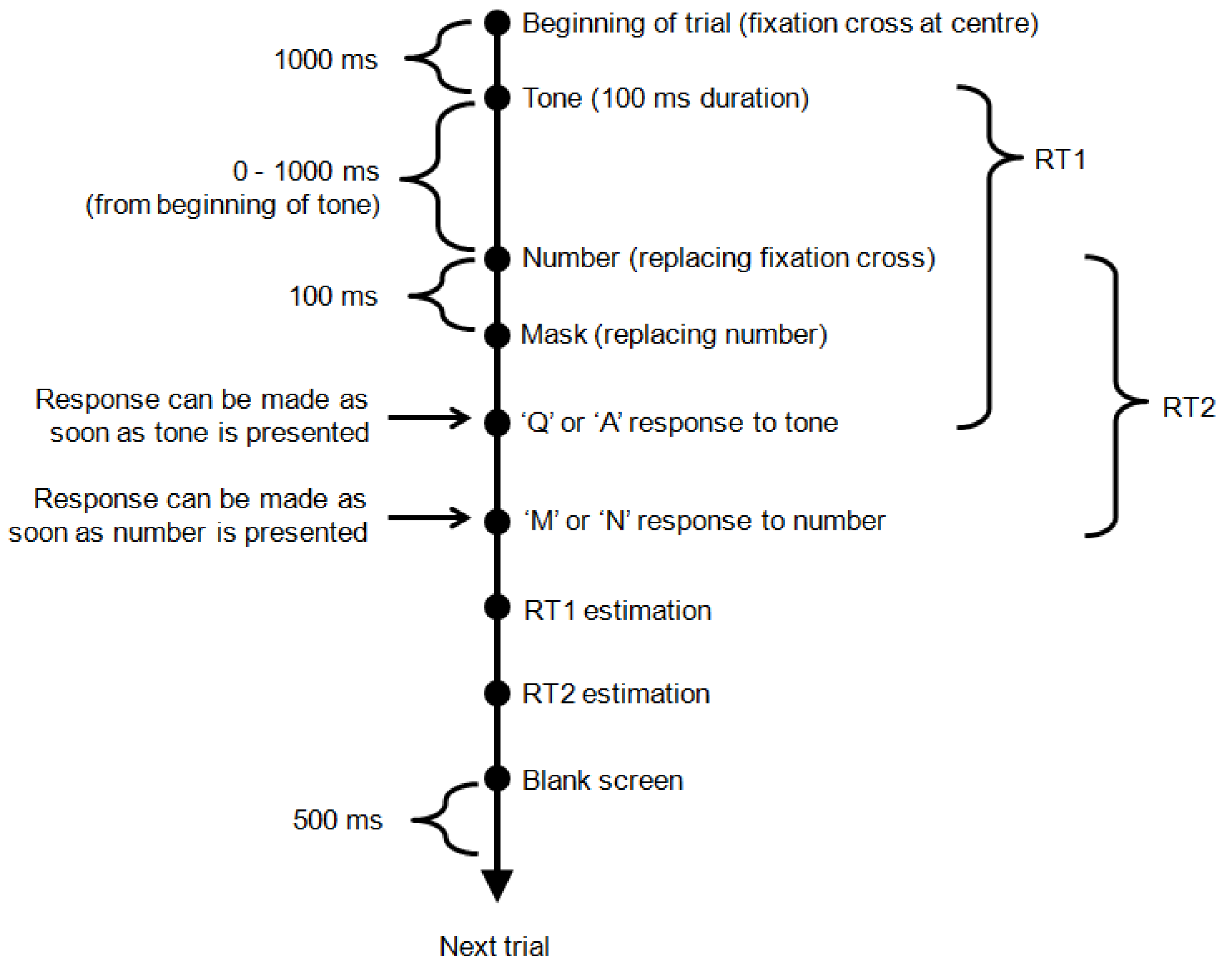

Figure 1. Outline of one trial of Experiment 1. RT1 = Response time to Task 1. RT2 = Response time to Task 2.

Results

Data Analysis.

All data from participants who failed to achieve an accuracy within 2.5 standard deviations of the mean accuracy on either of the tasks were removed from analysis. This procedure was performed recursively, resulting in the removal of data from 3 participants. The remaining participants averaged 94.2% (SD = 7.2%) accuracy on Task 1 and 88.6% (SD = 5.2%) accuracy on Task 2. Of the remaining data, trials with incorrect responses on either task (16.3%) as well as trials with RTs below 200 ms (0.7%) or more than 3 standard deviations above the mean for a given participant and SOA (2.2%) were discarded from the RT analysis.

RT and IRT as a function of SOA.

The RT data show the usual PRP effect, with longer responses to Task 2, but not to Task 1, at short SOAs (Figure 2). PRP interference occurred for SOAs shorter than about 300-400 ms. Therefore, for the following analyses I divided the data into trials with short SOA (< 300 ms), on which PRP interference occurred, and those with long SOA (> 400 ms), on which interference did not occur. Paired samples t-tests confirmed that RT2 was longer on short SOA trials ($M_{RT2,short}$ = 968 ms) than on long SOA trials ($M_{RT2,long}$ = 651 ms), $t(16) = 7.73$, $p < .001$. There was no significant difference in RT1 between short ($M_{RT1,short}$ = 812) and long ($M_{RT1,long}$ = 782) SOA trials, $t(16) = 0.52$, $p = .611$.

The pattern of IRTs matches that found by Corallo et al. (2008, Experiment 2). Participants seem not to have been aware of the large increase in RT2 at short SOAs, as there was no difference in IRT2 between short ($M_{IRT2,short}$ = 878 ms) and long ($M_{IRT2,long}$ = 909 ms)

SOA trials, $t(16) = 1.38$, $p = .187$. There was also no difference in IRT1 between short ($M_{IRT1,short}$ = 846 ms) and long ($M_{IRT1,long}$ = 835 ms) SOA trials, $t(16) = 0.90$, $p = .381$. [1]

[1] The order in which the two experiments were performed did not interact with the effect of SOA for any of the four measures (RT1, RT2, IRT1, or IRT2), all ps > 0.05.

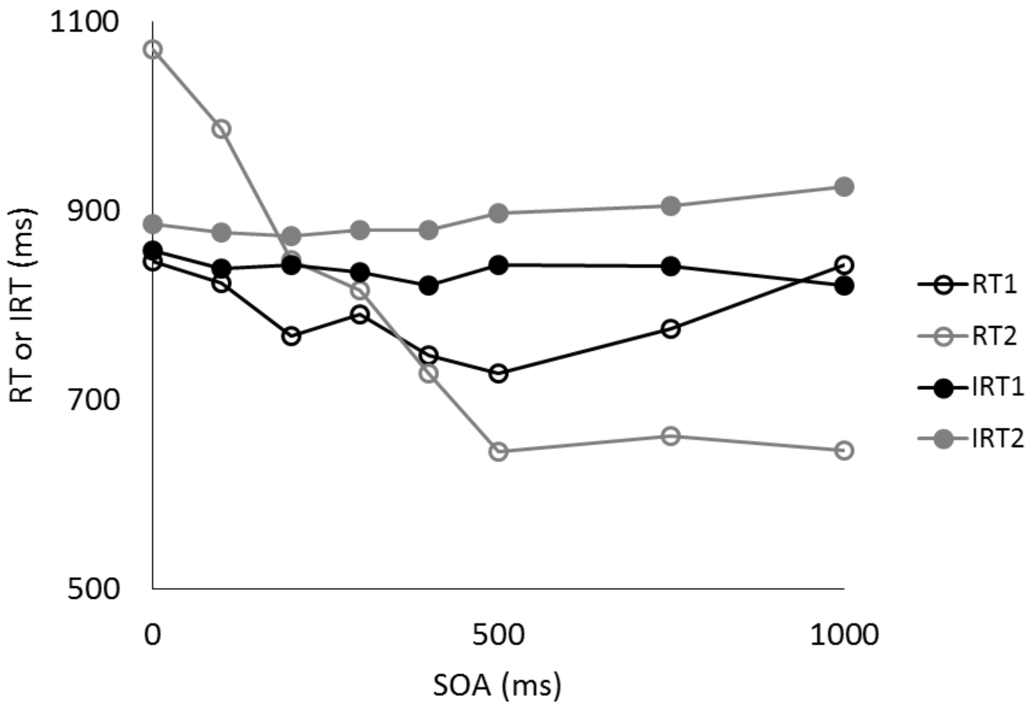

Figure 2. Response times and introspective response times to Tasks 1 and 2 as a function of stimulus onset asynchrony in Experiment 1.

Distance and notation manipulations.

I first analyzed the effects of the distance of the number from 45 on Task 2 RT using a 2x2 repeated measures analysis of variance (ANOVA) with Distance (Close vs. Far) and SOA (Short vs. Long) as factors (Figure 3). The slowing at shorter SOAs was reflected in a main effect of SOA, $F(1, 16) = 46.23$, $p < .001$, $n_p^2 = .74$. A main effect of Distance, $F(1, 16) = 24.47$, $p < .001$, $n_p^2 = .61$, confirmed that RTs were also slower for numbers that were closer to 45. As expected, the two factors did not interact, $F < 1$, suggesting that the distance manipulation influenced the central processing stage or later. The corresponding ANOVA on Task 2 IRTs revealed that the main effect of Distance was also present in the IRTs, $F(1, 16) = 28.07$, $p < .001$, $n_p^2 = .64$, with no effect of SOA, $F(1, 16) = 1.91$, $p = .186$, $n_p^2 = .11$, and no interaction, $F < 1$.

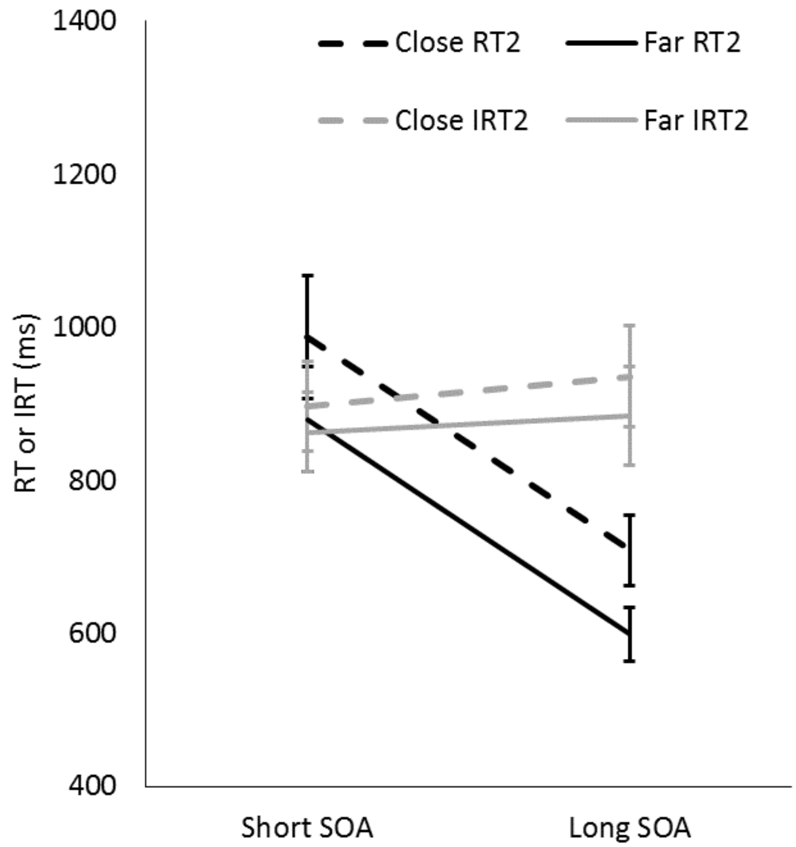

Figure 3. Average response time and introspective response time to Task 2 on trials with a number either close to (within 12) or far from (more than 12) forty-five, with a short (0 ms, 100 ms, or 200 ms) or long (500 ms, 750 ms, or 1000 ms) stimulus onset asynchrony, in Experiment 1. Error bars represent ± 1 standard error of the mean.

To parallel the analysis performed by Corallo et al. (2008), I conducted a separate 2x2 within-subjects ANOVA on Task 2 RTs with Notation (Word vs. Digit) and SOA (Short vs. Long) as factors (Figure 4). RTs were slower when the number was presented as a word, $F(1, 16) = 19.71$, $p < .001$, $n_p^2 = .55$. There was again a main effect of SOA, $F(1, 16) = 58.47$, $p < .001$, $n_p^2 = .79$, as well as an interaction between the two factors, $F(1, 16) = 6.67$, $p = .020$, $n_p^2 = .29$. The effect of notation was significant at long SOAs, $t(16) = 5.36$, $p < .001$, but only marginally so at short SOAs, $t(16) = 1.98$, $p = .065$, suggesting that this manipulation affected the perceptual stage of Task 2. The corresponding ANOVA on Task 2 IRTs did not reveal an effect of SOA, $F(1, 16) = 1.93$, $p = .184$, $n_p^2 = .11$. Although the pattern of IRT2s was numerically similar to the pattern of RT2s, neither the effect of notation, $F(1, 16) = 2.46$, $p = .136$, $n_p^2 = .13$, nor the interaction, $F(1, 16) = 1.76$, $p = .203$, $n_p^2 = .10$, reached significance.

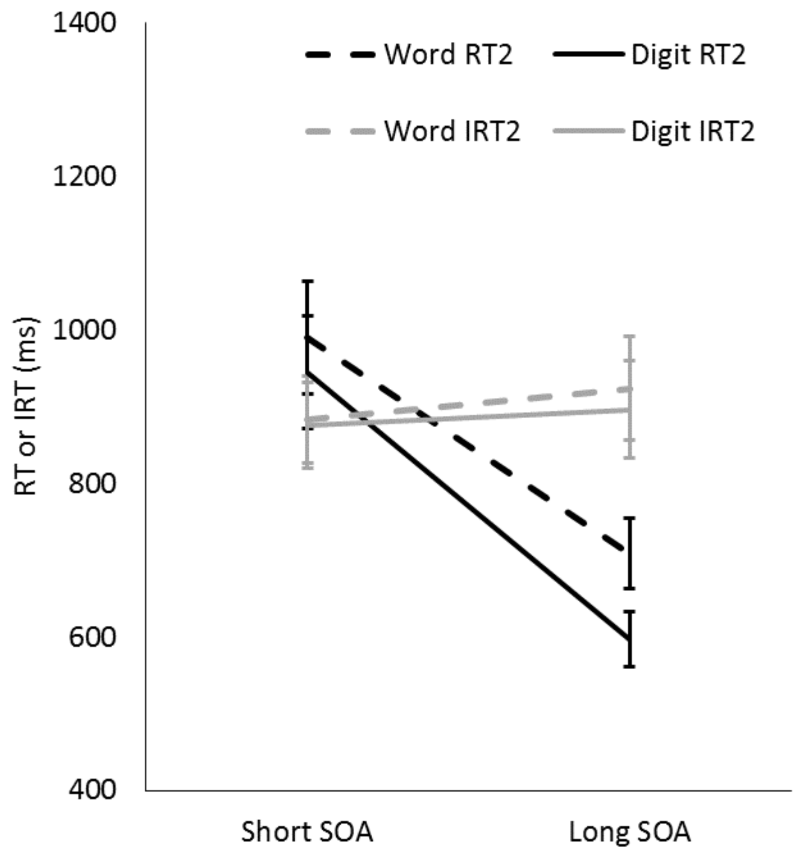

Figure 4. Average response time and introspective response time to Task 2 on trials with a number presented as a digit or spelled out as a word, with a short (0 ms, 100 ms, or 200 ms) or long (500 ms, 750 ms, or 1000 ms) stimulus onset asynchrony, in Experiment 1. Error bars represent ± 1 standard error of the mean.

Error percentage.

I repeated the preceding analyses using the error percentage data. Participants made significantly more errors on Task 1 at short SOAs than at long SOAs (M_{short} = 8.7% of trials, M_{long} = 3.8%), $t(16) = 3.71$, $p = .002$. Although Task 2 errors were also more frequent at short than at long SOAs (M_{short} = 12.3%, M_{long} = 10.0%), this effect did not reach significance, $t(16) = 1.75$, $p = .100$.

The effect of Distance on Task 2 error percentage was tested using a 2x2 repeated measures ANOVA with Distance (Close vs. Far) and SOA (Short vs. Long) as factors. There was a significant effect of Distance, $F(1, 16) = 48.28$, $p < .001$, $n_p^2 = .75$, wherein more errors were made for numbers close to 45. The effect of Distance did not interact with SOA, $F < 1$.

In a similar 2x2 repeated measures ANOVA with Notation (Word vs. Digit) and SOA (Short vs. Long) as factors, there was a significant effect of Notation, $F(1, 16) = 8.19$, $p = .011$, $n_p^2 = .34$, with more errors to words than to digits. The effect of Notation also did not interact with SOA, $F < 1$.

Regression analysis.

I conducted a regression analysis for each participant, including Distance, Notation, IRT1, RT1, RT2, and SOA as predictors of IRT2. To test whether each factor was a significant predictor of IRT2, I performed one-sample t-tests on each of the regression coefficients (Table 1). The only factor that did not significantly predict IRT2 was Notation, $t < 1$. Distance was a significant predictor, with longer estimates to Close than to Far numbers, $t(16) = 2.69$, $p = .016$. IRT2 also increased with increasing IRT1, $t(16) = 20.33$, $p < .001$, with increasing RT2, $t(16) =$

5.73, $p < .001$, and with increasing SOA, $t(16) = 3.40$, $p = .004$, but decreased with increasing RT1, $t(16) = 4.35$, $p < .001$. The intercept was significantly positive, $t(16) = 6.88$, $p < .001$.

A similar analysis including Distance, Notation, IRT2, RT1, RT2, and SOA as predictors of IRT1 revealed that neither Distance nor Notation predicted IRT1, *ts* < 1. IRT1 increased with increasing IRT2, $t(16) = 13.30$, $p < .001$, and with increasing RT1, $t(16) = 4.76$, $p < .001$, but decreased with increasing SOA, $t(16) = 2.41$, $p = .028$, and with increasing RT2, $t(16) = 2.30$, $p = .035$. The intercept was significantly positive, $t(16) = 5.64$, $p < .001$

Table 1. Regression coefficients from the regression analysis including Distance, Notation, introspective response time to Task 1 (IRT1), response time to Task 1 (RT1), response time to Task 2 (RT2), and stimulus onset asynchrony (SOA) as predictors of introspective response time to Task 2 (IRT2), and from the regression analysis including Distance, Notation, IRT2, RT1, RT2, and SOA as predictors of IRT1, for Experiment 1.

	Predicted variable	
Predictor	**IRT1**	**IRT2**
Constant	244.97**	191.83**
Distance†	7.04	-12.40*
Notation‡	1.10	-1.76
IRT1	-	0.74**
IRT2	0.61**	-
RT1	0.19**	-0.14**
RT2	-0.07*	0.21**
SOA	-0.05*	0.15*

* $p < 0.05$; ** $p < 0.001$

† Close coded as 0, Far as 1

‡Word coded as 0, Digit as 1

Discussion of Experiment 1

The goal of Experiment 1 was to test whether PRP interference would be apparent in introspection under conditions that preclude a delay in perception of S2. These conditions were achieved by masking S2 after 100 ms. Accuracy on Task 2 therefore required that perception of the stimulus was not delayed by more than 100 ms. The average accuracy of 88.6% is well above chance performance, suggesting that participants did perceive the stimulus within this time. Two alternative explanations for this high accuracy might be proposed. The first is that S2 was not perceived at all, and accurate responses resulted from implicit processing. Unconscious number primes can indeed influence response times to visible stimuli (e.g. Naccache & Dehaene, 2001). However, an 88.6% accuracy rate to the unseen stimulus itself is not realistic, especially given that pressing 'M' or 'N' is not a natural response to seeing a high or low number. Also, although I did not ask participants whether they perceived the number, I have run through the experiment myself and had no problem seeing the number on every trial. The second possibility is that the stimulus was 'on the road' to being perceived after 100 ms, but was not actually perceived until later. That is, processing that would eventually lead to perception had begun but had not been completed within 100 ms. Although it is possible that it takes some time for conscious access to occur, there is no reason for the process of achieving this access to be slower for S2. Dual-stage models of conscious access predict delayed access to S2 because processing is queued during the bottleneck, not because it is slowed. Therefore any processing eventually leading to conscious access and beginning within 100 ms should result in normal perception of S2, and any processing beginning after 100 ms should lead to perception of the mask instead (Enns & Di Lollo, 2000).

That performance was well above chance indicates that the former was true most often. One can therefore be confident that S2 was perceived within about 100 ms of onset. Note that this does not necessarily rule out the possibility that perception of S2 is usually delayed in the PRP paradigm, because usually ‘on-time’ perception of S2 is not necessary for accurate performance. It is likely that my design forced participants to adopt a strategy in which they divided attention between the two modalities so as not to miss S2.

Strikingly, despite the on-time perception of S2, my results match those of Corallo et al. (2008), with participants failing to report the large increase in RT2 with decreasing SOA. Could this have been caused by a failure to accurately estimate RT2 in general? My Distance manipulation influenced both RT2 and IRT2 in the same manner, suggesting that introspection was at least somewhat sensitive to response time. Conversely, the Notation manipulation influenced RT2 but did not have a significant effect on IRT2. The regression analysis also revealed Distance, but not Notation, as a significant predictor of IRT2, indicating that the effect of Distance on IRT2 may have been caused by a tendency to make longer estimates following numbers that were closer to 45, regardless of actual RT. However, the regression coefficients for RT2 as a predictor of IRT2 and for RT1 as a predictor of IRT1 were both significantly positive, demonstrating that introspection was indeed sensitive to actual RT. Although the coefficients were small, IRT2 should still have been expected to increase by about 70 ms between long and short SOA trials. The actual change was a decrease by 30.6 ms. It is therefore reasonable to conclude that the deficit in time estimation is caused by overlap of the two tasks rather than a general inability to estimate response times.

The regression analysis also provides insight into the strategy that participants used to generate their response time estimates. Longer responses to Task 1 led not only to higher IRT1 estimates but also to lower IRT2 estimates. Similarly, longer responses to Task 2 reduced IRT1 estimates. Finally, IRT1 and IRT2 were more strongly related to each other than to either actual response time. This suggests that participants used a strategy in which they first selected a value for IRT1, and then produced IRT2 by adjusting this value based on their perceived difference between RT1 and RT2. This strategy makes sense given that both estimates were inputted using the same visual analogue scale with no interruption between the two responses. The task may therefore be thought of as a mixture between duration labelling and duration comparison tasks, rather than strictly as a duration labelling task. This makes my results even more striking, as one would expect that a large increase in duration would be even more salient when compared to an unchanged duration instead of being perceived in isolation.

Overall, the results of Experiment 1 replicate the finding that IRTs are insensitive to PRP interference, and demonstrate that this insensitivity does not depend on whether perception of S2 is delayed until the end of Task 1 central processing. Experiment 2 provides additional evidence by achieving the same result using a different approach.

Experiment 2

Methods

Participants.

Experiment 2 used the same twenty University of Waterloo undergraduate students as Experiment 1.

Apparatus.

Experiment 2 used the same apparatus as Experiment 1, except for the headphones, as the experiment did not involve audio.

Procedure.

Experiment 2 followed the same procedure as Experiment 1 except for the following changes (Figure 5). The Task 1 stimulus was presented in the visual modality. Instead of a high or low pitched tone, the stimulus was a row of five percentage signs presented 1° above or below fixation. Participants responded by pressing the "Q" key with their left middle finger when the stimulus was above fixation and by pressing the "A" key with their left index finger when it was below fixation. The Task 2 stimulus remained the same, but instead of replacing the fixation cross, it replaced the Task 1 stimulus. That is, as soon as the SOA elapsed, the row of percentage signs disappeared and was replaced immediately by a number. This number remained on screen until response. The set of SOAs remained the same except for the 0 ms SOA, which was replaced by a 34 ms SOA. Finally, all text was presented in courier new font.

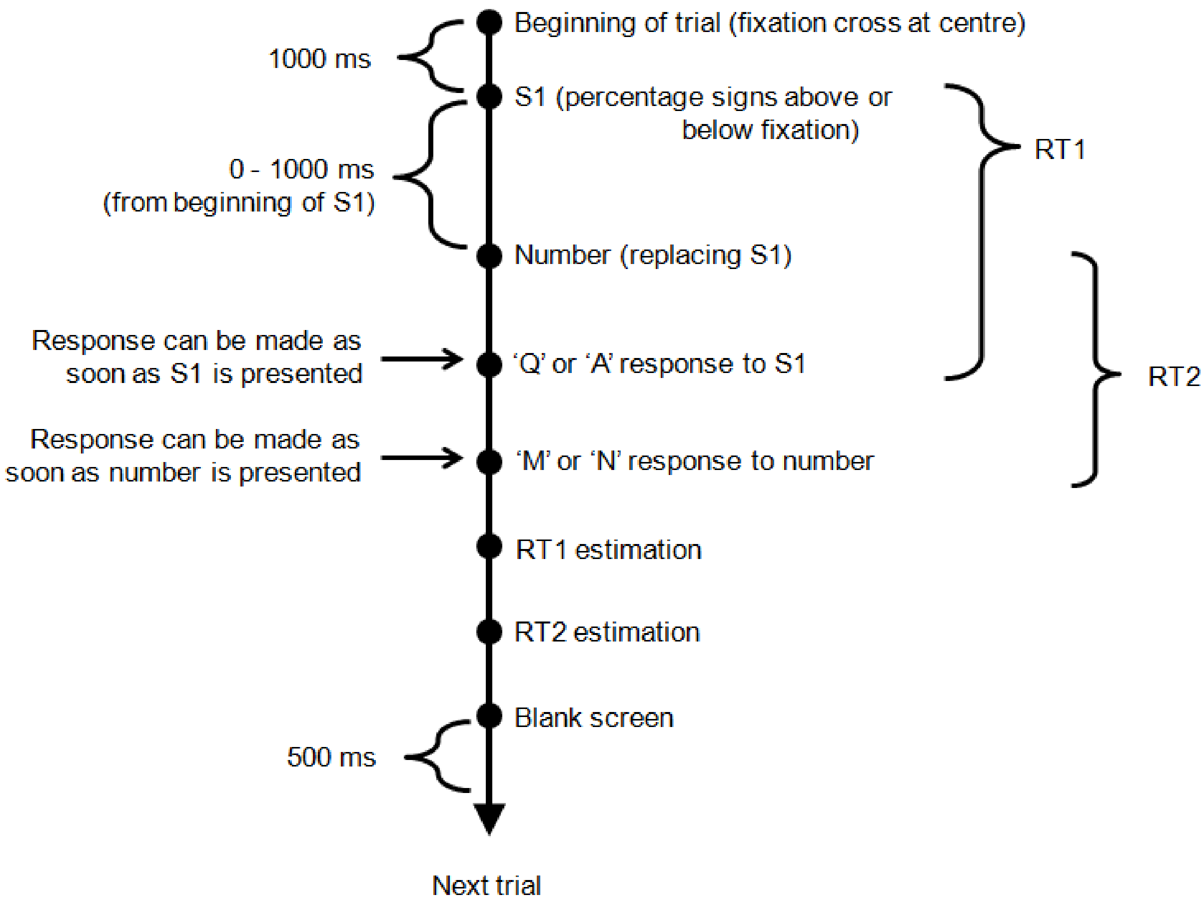

Figure 5. Outline of one trial of Experiment 2. S1 = Stimulus 1. RT1 = Response time to Task 1. RT2 = Response time to Task 2.

Results

Data analysis.

All data from participants who failed to achieve an accuracy within 2.5 standard deviations of the mean accuracy on either of the tasks were removed from analysis. This procedure was performed recursively, but resulted in the removal of data from only 1 participant. The remaining participants averaged 95.7% (SD = 6.9%) accuracy on Task 1 and 88.4% (SD = 8.9%) accuracy on Task 2. Of the remaining data, trials with incorrect responses on either task (15.2%) as well as trials with RTs below 200 ms (0.1%) or more than 3 standard deviations above the mean for a given participant and SOA (2.2%) were discarded from the RT analysis.

RT and IRT as a function of SOA.

As in Experiment 1, RT2 was longer on short SOA trials ($M_{RT2,short}$ = 1014 ms) than on long SOA trials ($M_{RT2,long}$ = 739 ms), $t(18) = 5.86$, $p < .001$. Unlike in Experiment 1, RT1 was also longer on short SOA trials ($M_{RT1,short}$ = 751 ms) than on long SOA trials ($M_{RT1,long}$ = 603 ms), $t(18) = 2.65$, $p = .016$. Neither IRT1 ($M_{IRT1,short}$ = 848 ms, $M_{IRT1,long}$ = 844 ms), $t < 1$, nor IRT2 ($M_{IRT2,short}$ = 884 ms, $M_{IRT2,long}$ = 921 ms), $t(18) = 1.22$, $p = .237$, differed between short and long SOA trials (Figure 6).[2]

[2] The order in which the two experiments were performed did not interact with the effect of SOA for any of the four measures (RT1, RT2, IRT1, or IRT2), all ps > 0.05.

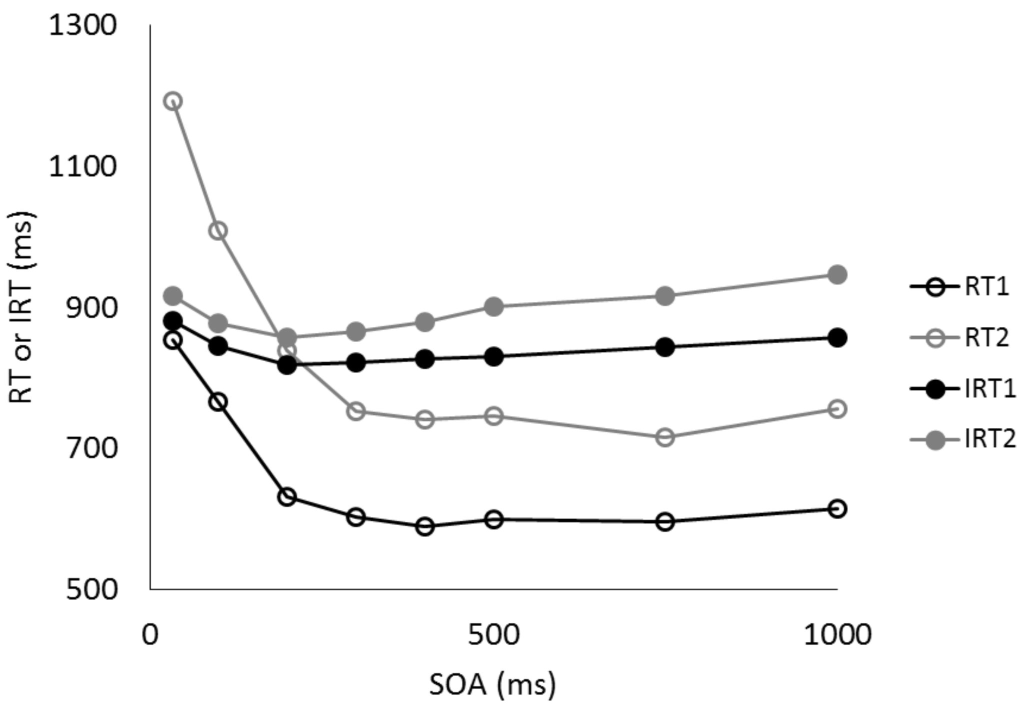

Figure 6. Response times and introspective response times to Tasks 1 and 2 as a function of stimulus onset asynchrony in Experiment 1.

Distance and notation manipulations.

A 2x2 repeated measures ANOVA with Distance (Close vs. Far) and SOA (Short vs. Long) as factors revealed the same pattern as in Experiment 1 (Figure 7). Responses to Task 2 were significantly longer for numbers that were close to 45, $F(1, 18) = 12.04$, $p = .003$, $n_p^2 = .40$, and for trials with short SOAs, $F(1, 18) = 34.78$, $p < .001$, $n_p^2 = .66$. The two factors did not interact, F < 1. The corresponding ANOVA on Task 2 IRTs revealed that the main effect of Distance was also present in the IRTs, $F(1, 18) = 10.65$, $p = .004$, $n_p^2 = .37$, with no effect of SOA, $F(1, 18) = 1.34$, $p = .262$, $n_p^2 = .07$, and no interaction, $F(1, 18) = 1.73$, $p = .205$, $n_p^2 = .09$.

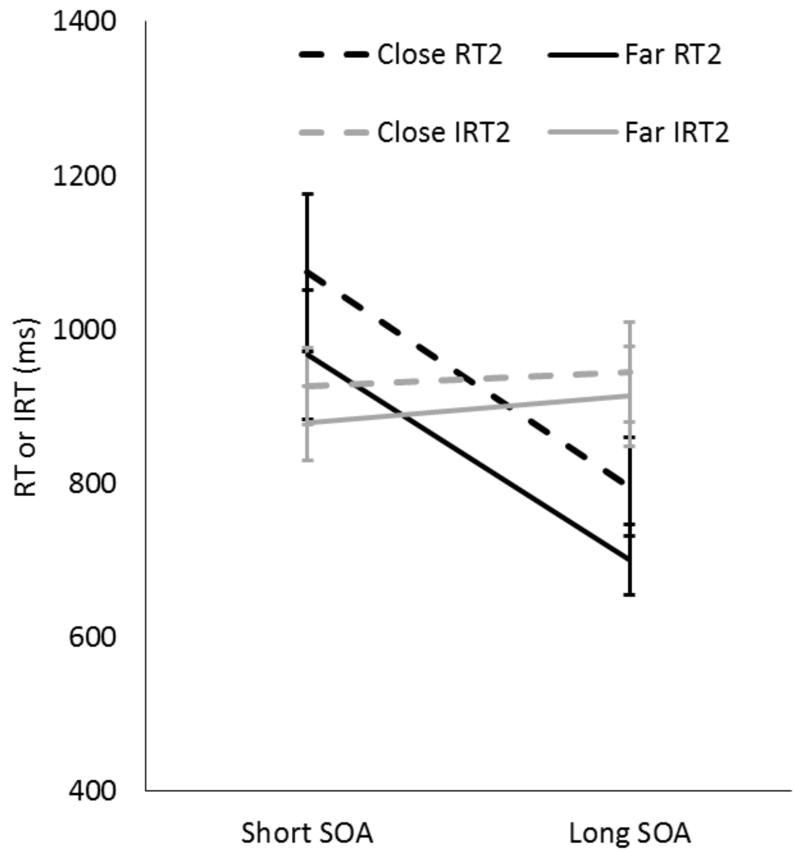

Figure 7. Average response time and introspective response time to Task 2 on trials with a number either close to (within 12) or far from (more than 12) forty-five, with a short (0 ms, 100 ms, or 200 ms) or long (500 ms, 750 ms, or 1000 ms) stimulus onset asynchrony, in Experiment 2. Error bars represent ± 1 standard error of the mean.

A 2x2 within-subjects ANOVA on Task 2 RTs with Notation (Word vs. Digit) and SOA (Short vs. Long) as factors also matched Experiment 1 (Figure 8). Responses to Task 2 were significantly longer for numbers presented as words than as digits, $F(1, 18) = 51.8$, $p < .001$, $n_p^2 = .74$, and for trials with short SOAs, $F(1, 18) = 35.17$, $p < .001$, $n_p^2 = .66$. There was a Notation × SOA interaction, $F(1, 18) = 4.62$, $p = .046$, $n_p^2 = .20$, with the Notation effect being significant at both short, $t(18) = 4.68$, $p < .001$, and long, $t(18) = 7.46$, $p = .001$, SOAs, but larger at long SOAs. The corresponding ANOVA on Task 2 IRTs revealed an effect of Notation, $F(1, 18) = 4.81$, $p = .042$, $n_p^2 = .21$, no effect of SOA, $F(1, 18) = 1.55$, $p = .229$, $n_p^2 = .08$, and a marginally significant Notation × SOA interaction, $F(1, 18) = 3.60$, $p = .074$, $n_p^2 = .17$ with a larger notation effect at long than at short SOAs.

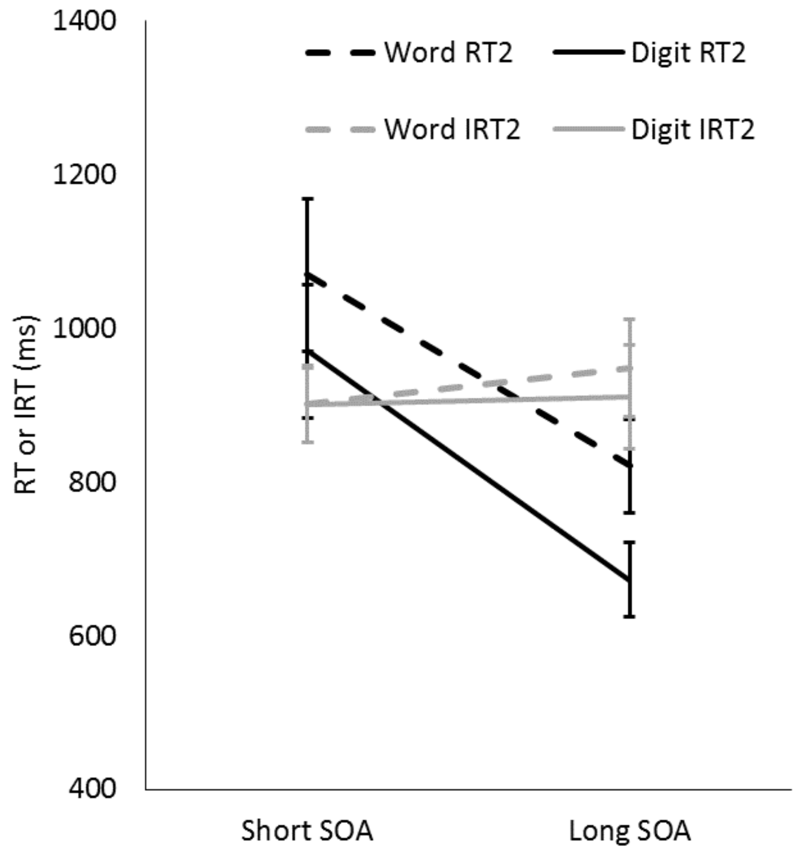

Figure 8. Average response time and introspective response time to Task 2 on trials with a number presented as a digit or spelled out as a word, with a short (0 ms, 100 ms, or 200 ms) or long (500 ms, 750 ms, or 1000 ms) stimulus onset asynchrony, in Experiment 2. Error bars represent ± 1 standard error of the mean.

Error percentage.

I repeated the preceding analyses using the error percentage data. Participants made significantly more errors on short than on long SOA trials for both Task 1 (M_{short} = 5.5% of trials, M_{long} = 3.3%), $t(18) = 2.32$, $p = .032$, and Task 2 (M_{short} = 13.1%, M_{long} = 9.3%).

The effect of Distance on Task 2 error percentage was tested using a 2x2 repeated measures ANOVA with Distance (Close vs. Far) and SOA (Short vs. Long) as factors. There was a significant effect of Distance, $F(1, 18) = 20.47$, $p < .001$, $n_p^2 = .53$, wherein more errors were made for numbers close to 45. The effect of Distance did not interact with SOA, $F < 1$.

In a similar 2x2 repeated measures ANOVA with Notation (Word vs. Digit) and SOA (Short vs. Long) as factors, there was no effect of Notation, $F(1, 18) = 0.01$, $p = .916$, $n_p^2 = .00$, and no interaction with SOA, $F(1, 18) = 2.79$, $p = .112$, $n_p^2 = .13$.

Regression analysis.

I conducted the same regression analysis as in Experiment 1, including Distance, Notation, IRT1, RT1, RT2, and SOA as predictors of IRT2 for each participant (Table 1). Neither Notation nor Distance, ts < 1, were significant predictors of IRT2. The rest of the results replicated those from Experiment 1. IRT2 increased with increasing IRT1, $t(18) = 14.89$, $p < .001$, decreased with increasing RT1, $t(18) = 5.72$, $p < .001$, and, importantly, increased with increasing RT2, $t(18) = 5.74$, $p < .001$. IRT2 also increased with increasing SOA, $t(18) = 3.30$, $p = .004$, and the intercept was significantly positive, $t(18) = 5.15$, $p < .001$.

A similar analysis including Distance, Notation, IRT2, RT1, RT2, and SOA as predictors of IRT1 revealed that only Distance did not significantly predict IRT1, $t < 1$. Longer estimates were made for IRT1 when S2 was a digit than when it was a word, $t(18) = 3.00$, $p = .008$. IRT1 also increased with increasing IRT2, $t(18) = 15.91$, $p < .001$, and with increasing RT1, $t(18) = 5.71$, $p < .001$, but decreased with increasing SOA, $t(18) = 2.28$, $p = .035$, and with increasing RT2, $t(18) = 5.62$, $p < .001$. The intercept was significantly positive, $t(18) = 5.33$, $p < .001$.

Table 2. Regression coefficients from the regression analysis including Distance, Notation, introspective response time to Task 1 (IRT1), response time to Task 1 (RT1), response time to Task 2 (RT2), and stimulus onset asynchrony (SOA) as predictors of introspective response time to Task 2 (IRT2), and from the regression analysis including Distance, Notation, IRT2, RT1, RT2, and SOA as predictors of IRT1, for Experiment 2.

	Predicted variable	
Predictor	**IRT1**	**IRT2**
Constant	210.90**	151.78**
Distance†	-2.30	-2.15
Notation‡	9.28*	-0.47
IRT1	-	0.74**
IRT2	0.69**	-
RT1	0.22**	-0.18**
RT2	0.12**	0.26**
SOA	0.03*	0.11*

* $p < 0.05$; ** $p < 0.001$
† Close coded as 0, Far as 1
‡Word coded as 0, Digit as 1

Response grouping.

Sometimes participants performing PRP tasks withhold their response to Task 1 in order to make both responses at the same time. If done selectively on short SOA trials, such behaviour could result in an increase in RT1 similar to that identified in the 'RT and IRT as a function of SOA' section. I analyzed the time between responses to the two tasks to determine whether participants were more prone to group responses on short SOA trials in Experiment 2 than in Experiment 1. For the sixteen participants whose data were included in the analyses for both experiments, the average interval between responses on short SOA trials was shorter for Experiment 1 (M = 266 ms) than for Experiment 2 (M = 387 ms), $t(15) = 3.16$, $p = .006$. Additionally, the average percentage of short SOA trials on which the interval between responses was less than 100 ms was larger for Experiment 1 (M = 31.6%) than for Experiment 2 (M = 6.3%), $t(15) = 4.26$, $p = .001$.

Discussion of Experiment 2

The goal of Experiment 2 was to use different methodology to replicate the finding from Experiment 1 that the insensitivity of introspection to PRP interference does not depend on perception of the second stimulus being delayed. This was accomplished by combining the stimuli for each task into a single stimulus. That is, instead of presenting a second stimulus after the SOA, the first stimulus was updated to include the information required to perform Task 2. Perception of this update should not have been delayed because the update constituted a change to a stimulus that was already being consciously processed, rather than a peripheral stimulus. Although it is theoretically possible that participants quickly encoded the stimulus and then performed central processing on an internal representation, thereby not perceiving the change to the external stimulus, there are three reasons to believe this did not occur. First, it would be inefficient to encode a complete internal representation of the stimulus and then to stop sampling from the freely available external information (e.g. Ballard, Hayhoe, & Pook, 1997). Second, the shortest SOA, 34 ms, is likely shorter than the time required to encode such a representation. There would therefore not be enough time to encode the representation before the change occurred. Finally, in a departure from usual PRP results, RT1 drastically increased at short SOA, suggesting that the stimulus change interfered with Task 1 processing. Although an increase in RT1 could be caused by response grouping, my analysis indicates that responses were not grouped. This suggests that participants were still processing the stimulus at the time of the change.

Despite a large increase in RT2 with decreasing SOA, no such change was evident for IRT2, confirming that delayed perception is not a prerequisite for temporal underestimation in the PRP paradigm. Surprisingly, introspection also failed to detect the large increase in RT1 on short SOA trials. This novel finding clearly conflicts with the delayed perception account, as this account does not predict any interference with estimation of RT1. Also of note is that both introspective measures seem to follow patterns similar to those of the corresponding RTs, with some curvature at shorter SOAs. Does this indicate some level of awareness of the passage of time during the bottleneck? Not necessarily. The regression analyses for Experiments 1 and 2 both demonstrate that IRT2s depend very strongly on IRT1s. Participants noticing that the stimulus change interfered with their performance on Task 1 may have produced IRT1s, and therefore also IRT2s, that followed this curved pattern. Consistent with this explanation, there was no such curvature in Experiment 1, in which RT1 did not vary as a function of SOA. However, it remains possible that the curvature was indeed caused by some awareness of time during the bottleneck. I will discuss this possibility in more depth in a later section.

General Discussion

Previous results have shown that introspective estimates of response time are not sensitive to PRP interference. That is, although response time to Task 2 is slowed substantially at shorter SOAs, estimates of this response time typically do not vary with SOA. One interpretation of these findings is that perception of the stimulus for Task 2 is delayed until the completion of central processing of Task 1. Therefore, according to this interpretation, IRT2s are not sensitive to PRP interference because S2 is not perceived while the interference is occurring.

Across two experiments using different approaches, I found that IRTs are still not sensitive to PRP interference under conditions in which delayed perception of S2 is precluded. The lack of sensitivity is therefore not dependent on perception being delayed, and must have had some other cause in my experiments. To be clear, my results do not show that delayed perception of S2 does not occur in the PRP paradigm, or that it would not cause similar results. I *have* shown that underestimation of time in the PRP paradigm cannot be taken as evidence that perception of S2 is delayed because underestimation occurs even in the absence of delayed perception. However, there is other evidence (e.g. Marti et al., 2012) that perception is delayed in normal PRP situations. Nonetheless, my results demonstrate that introspection is insensitive to PRP interference for reasons other than, or in addition to, this delay.

Alternative Accounts

One potential explanation is that introspective response times are based on perceived difficulty rather than on mental timing processes (Bryce & Bratzke, 2014). If this is true, one would only expect an increase in IRT2 at short SOA if the task is perceived to be more difficult at short SOA. Bryce and Bratzke (2014) provided support for this idea by showing that introspective response times are related to introspective reports of task difficulty, and suggested that error percentage may be useful for estimating perceived task difficulty. Unfortunately, the effect of SOA on error percentage has not been reported in other studies. My error data do not support the perceived difficulty account. Error percentage was significantly higher at short SOA for Task 1 in both Experiments 1 and 2, and for Task 2 in Experiment 2, whereas neither IRT1 nor IRT2 differed between long and short SOAs in either experiment. These results do not necessarily contradict the perceived difficulty account, because it could be that error percentage is not actually well coupled to perceived difficulty. However, it is also not clear how perceived difficulty could account for the curvature found in the IRTs in Experiment 2.

A potential explanation that can account for the curvature is that people simply are not very good at estimating time in multi-task situations. As discussed earlier, time estimation is very sensitive to disruption, with almost any concurrent task resulting in shortened and/or more variable estimates of time (Brown, 1997). Accordingly, it does not seem very surprising that duration is underestimated during performance of four overlapping tasks (i.e. two timing tasks and two non-temporal tasks). The result that IRTs are *completely* insensitive to PRP interference appears noteworthy because it seems to imply that the passage of time, rather than simply being poorly estimated during the bottleneck, is not perceived at all. However, this interpretation is based on the assumption that bottleneck time does not contribute at all to RT estimates for short SOA trials. This is presumably an attractive assumption because estimates on short SOA trials are roughly the same as those on long SOA trials, during which there is no bottleneck. Bryce and Bratzke (2014) revealed the flaw in this assumption by demonstrating that, if mental timing does not begin until the offset of Task 1 central processing, IRT2 should decrease with decreasing SOA. This is because the duration of the early processing stages of Task 2, which can occur in parallel with central processing of Task 1, would not be included in time estimates on short SOA trials, but would be included on long SOA trials. The finding of equal IRTs on short and long SOA trials is therefore consistent with an account in which the passage of time is perceived, albeit with poor accuracy, during the bottleneck.

The Use of Time Estimation in Quantified Introspection

The original intent of the work done by Corallo et al. (2008) was to determine which processes are accessible to introspection. I believe that their approach of quantified introspection is a clever and useful research tool. However, the usefulness of introspective response time as the quantitative measure may be limited to the study of time estimation itself. This is because time estimation processes are extremely sensitive to context, and break down with almost any concurrent task (Brown, 1997). For example, it cannot be concluded that Corallo et al.'s (2008) findings are caused by a lack of introspective access to the central bottleneck, because time estimation processes do not necessarily operate reliably and consistently whenever cognitive processes are accessible to introspection. The use of quantified introspection on problems such as 'Which cognitive processes are accessible to conscious report?' would therefore be best served by a measure that is relatively insensitive to context. Perhaps estimation of response time is more reliable than the estimation of external durations studied in most past research, however that remains to be seen. Future work should focus on identifying more potential quantifiable introspective measures, as well as uncovering how time estimation processes are influenced by the central bottleneck.

References

Ballard, D. H., Hayhoe, M. M., Pook, P. K., & Rao, R. P. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(04), 723-742.

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta psychologica*, *134*(3), 330-343.

Broadbent, D. E., & Broadbent, M. H. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & psychophysics*, *42*(2), 105-113.

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

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.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental psychology: Human perception and performance*, *21*(1), 109-127.

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.

Dehaene, S., Sergent, C., & Changeux, J. P. (2003). A neuronal network model linking subjective reports and objective physiological data during conscious perception. *Proceedings of the National Academy of Sciences*, *100*(14), 8520-8525.

Enns, J. T., & Di Lollo, V. (2000). What's new in visual masking?. *Trends in cognitive sciences*, *4*(9), 345-352.Macar, F., Grondin, S., & Casini, L. (1994). Controlled attention sharing influences time estimation. *Memory & Cognition*, *22*(6), 673-686.

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.

Marti, S., Sigman, M., & Dehaene, S. (2012). A shared cortical bottleneck underlying Attentional Blink and Psychological Refractory Period. *Neuroimage*, *59*(3), 2883-2898.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: verbal reports on mental processes. *Psychological review*, *84*(3), 231-259.

Ogden, R. S., Salominaite, E., Jones, L. A., Fisk, J. E., & Montgomery, C. (2011). The role of executive functions in human prospective interval timing. *Acta psychologica*, *137*(3), 352-358.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological bulletin*, *116*(2), 220-244.

Pashler, H., & Johnston, J. C. (1989). Chronometric evidence for central postponement in temporally overlapping tasks. *The Quarterly Journal of Experimental Psychology*, *41*(1), 19-45.

Reyes, G., & Sackur, J. (2014). Introspection during visual search. *Consciousness and cognition*, *29*, 212-229.

Ruthruff, E., & Pashler, H. (2010). Mental timing and the central attentional bottleneck. *Attention and time*, 123-135.

Telford, C. W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology*, *14*(1), 1-36.