Making time: Estimation of internally versus externally defined durations

Michael D. Klein¹ · Jennifer A. Stolz¹

Published online: 19 September 2017
© The Psychonomic Society, Inc. 2017

Abstract Most research on the topic of duration estimation has examined the mechanisms underlying estimation of durations that are demarcated by experimental stimuli. It is not clear whether the estimation of durations that are instead defined by our own mental processes (e.g., response times) is underlain by the same mechanisms. Across five experiments, we tested whether the pattern of interference between concurrent temporal and nontemporal tasks was similar for the two types of intervals. Duration estimation of externally defined intervals slowed performance on a concurrent equation verification task, regardless of whether participants were required to report their estimate by clicking within an analogue scale or by reproducing the duration. Estimation of internally defined durations did not slow equation verification performance when an analogue scale response was required. The results suggest that estimation of internally defined durations may not depend on the effortful temporal processing that is required to estimate externally defined durations.

Keywords Attention: Divided Attention and Inattention · attention and executive control

The number of recent review articles on the topic of time estimation demonstrates its current popularity as a topic of research (see Block & Grondin, 2014, for a review of reviews). Much of this research involves asking participants either to estimate the durations of intervals that are demarcated by experimental

stimuli or to produce their own intervals matching durations provided by the experimenter. This research helps us to understand our ability to estimate the timing of events in our environment (e.g. ‘How long has that traffic light been yellow?’), which is clearly an important skill if we are to efficiently interact with the world. However, our lives are driven not only by events in our outside environment but also by events within our minds. It is therefore equally important to understand how we estimate intervals that are defined by our own mental processes (e.g. ‘How quickly can I solve this math problem?’, ‘How quickly can I react to a stimulus?’). This topic has received far less attention than the estimation of ‘externally-defined’ intervals, that is, those that are demarcated only by external stimuli.

Some studies have demonstrated that estimation of ‘internally defined’ durations such as the amount of time taken to respond to a stimulus (i.e. response time; RT) can be reasonably accurate under some conditions (e.g. Bryce & Bratzke, 2014; Corallo, Sackur, Dehaene, & Sigman, 2008; Han, Viau-Quesnel, Xi, Schweickert, & Fortin, 2010). For example, Corallo et al. (2008) had participants perform a number-comparison task in which they reported whether a target number was larger or smaller than a comparison number. Manipulations that influenced actual RTs on this task, such as the numerical distance between the target and comparison numbers, also influenced estimated RTs in the same way, suggesting that participants had at least some awareness of the duration of their RTs. That RTs can be estimated accurately is not entirely surprising, however, because RTs are demarcated by external stimuli (e.g. a stimulus onset and a keypress) and are therefore not necessarily different from other externally demarcated intervals (e.g. a stimulus onset and offset). It is possible for RTs to be estimated in exactly the same way as any other duration. Here we ask whether the processes underlying RT estimation are indeed the same as those underlying estimation of externally defined intervals.

✉ Michael D. Klein
mdklein@uwaterloo.ca

¹ Department of Psychology, University of Waterloo, 200 University Ave. W., Waterloo, Ontario N2L 3G1, Canada

Although the exact processes behind time estimation are disputed, one consistent finding is the pattern of interference between concurrent temporal and nontemporal tasks. Brown (1997) found that the performance of a concurrent nontemporal task interfered with time estimation in 89% of a set of 80 experiments, which included a large variety of concurrent tasks, such as perceptual judgment tasks (e.g. Macar, Grondin, & Casini, 1994; Zakay, 1993), verbal tasks (e.g. Miller, Hicks, & Willette, 1978), and effortful cognitive tasks (e.g. Brown, 1985; Fortin & Breton, 1995), among others. Interference in the opposite direction, however, was much less likely (Brown, 2006). In a review of 33 experiments, Brown (2006) found that timing interfered with nontemporal task performance in only 52% of cases. He argued that timing should generally only interfere with performance on concurrent tasks that require executive resources, because timing itself is heavily based on executive resources (see also Brown, 1997). Subsequent research has demonstrated that timing does interfere with many concurrent tasks that are thought to require executive resources, such as randomization tasks (Brown, 2006) and working memory tasks (Rattat, 2010). Timing has also been related to supposed executive functions, such as updating, switching, and access (Ogden, Salominaite, Jones, Fisk, & Montgomery, 2011; Ogden, Wearden, & Montgomery, 2014; see also Fisk & Sharp, 2004; Miyake et al., 2000). Therefore, although the exact mechanisms underlying time estimation are not known for sure, one straightforward way to test whether these mechanisms are the same for externally defined and internally defined intervals is to test whether RT estimation interferes with the same pattern of concurrent tasks as does external duration estimation. If the two types of time estimation are the same, they should follow the same pattern of interference with other tasks. In contrast, the finding of a different pattern of interference between the two types would suggest that they differ.

How might estimation of externally defined and internally defined durations differ? The main difference is likely the amount of relevant information that is available. Thomas and Brown (1974; see also Thomas & Weaver, 1975) argued that the outcome of mental timing is both an encoding of the output from a timer and an encoding of the nontemporal information that was processed during the interval, and that perceived duration is a weighted average of these two encodings. A number of temporal illusions confirm that nontemporal information does influence duration judgements (e.g. Brown, 1995; Buffardi, 1971; Tse, Intriligator, Rivest, & Cavanagh, 2004; although these manipulations could arguably influence an internal timer in addition to nontemporal processes). Cognitive processing is often associated with metacognitive feelings, such as feelings of error (e.g. Wessel, 2012) and confidence (e.g. Boekaerts & Rozendaal, 2010), which may provide potentially useful nontemporal information that is not present when timing externally defined

intervals. Because RTs depend on one's level of cognitive performance, this information is directly relevant to RT estimation and may even be sufficient to generate accurate estimates in the absence of any explicit temporal processing (for a similar argument, see Bryce & Bratzke, 2014). Because temporal processing is resource intensive, whereas metacognitive feelings often seem to arise effortlessly, we expect that if the two types of timing do differ, it should be the case that RT estimation is less likely to interfere with concurrent nontemporal processing.

To test this hypothesis, we adopted a dual-task approach in which a nontemporal task requiring executive resources was performed either alone or concurrently with an (internally defined or externally defined) interval duration estimation task. To our knowledge, one previous study has also compared performance on a nontemporal task under single task and concurrent RT estimation conditions, finding that RT estimation did not interfere with nontemporal task performance (Marti, Sackur, Sigman, & Dehaene, 2010). However, because their goal was not to contrast timing of externally defined and internally defined durations, they (a) did not compare their results to a condition in which estimation of an externally defined duration was required, and (b) used nontemporal tasks (tone and letter discrimination) that probably did not strongly depend on executive resources and were not likely to be interfered with even by timing of externally defined intervals.

In the majority of the experiments reported here, the nontemporal task was an equation-verification task in which participants verified whether a simple mathematical equation was true. We selected the equation-verification task because mental arithmetic is cognitively demanding, has been argued to require executive functioning (e.g. Redding & Wallace, 1985; Reisberg, 1983), and, most importantly, is sensitive to interference from time-estimation tasks (Brown, 1997). As the temporal task, we had participants estimate either the screen time of the stimulus (e.g. the equation) or their own RT to the stimulus. They reported their estimates by clicking within an analogue scale that was displayed on screen. This task has been used in several other studies of RT estimation (e.g. Bryce & Bratzke, 2014; Corallo et al., 2008), and is similar to the method of verbal estimation (i.e. simply reporting an estimate verbally), which has been related to the executive function of 'access' (Fisk & Sharp, 2004; Ogden et al., 2014).

Experiments 1–3 demonstrate that, under the correct conditions, estimation of externally defined but not internally defined durations interferes with performance on a concurrent equation-verification task. Experiment 4 shows that this difference is not universal to all temporal tasks, as estimation of both types of durations interfered with equation verification when the estimates were reported using the method of duration reproduction. Experiment 5 shows that the findings from the first three experiments generalize to another nontemporal task (sequence reasoning).

Experiment 1

Method

Participants Participants were 20 University of Waterloo undergraduate students (11 female, nine male, $M_{\text{age}} = 21.1$ years, $SD_{\text{age}} = 1.4$ years) who participated for course credit. We planned to remove any participants who scored below 70% accuracy on the equation-verification task and to replace them in order to maintain counterbalancing; however, all participants achieved >70% accuracy. All reported normal or corrected-to-normal visual acuity and were naïve to the purpose of the experiment.

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. Responses were collected via keypress on the computer's keyboard and by mouse clicks. Participants viewed the display approximately 50 cm from the monitor.

Stimuli Stimuli consisted of subtraction problems displayed on the computer screen. The problems were presented on the computer screen in bold, black, size 18 courier new font against a white background. They were arranged with one number (the minuend) presented just above another number (the subtrahend). A subtraction symbol was displayed to the left of the subtrahend. Below the subtrahend was a horizontal black line, beneath which was the answer (the difference between the minuend and subtrahend). The minuend ranged from 20 to 99 and the subtrahend ranged from 1 to 19, excluding 10. The answer was either correct, meaning that it was equal to the difference between the minuend and subtrahend, or incorrect, in which case it differed from the correct answer by either ± 1 or by ± 10 .

Duration estimates were inputted by clicking within a visual analogue scale presented on screen. The scale ranged from 0 to 4 seconds, with 0, 1, 2, 3, and 4 s explicitly marked as well as one unlabeled tick mark midway between each of these values.

Design and procedure The experiment consisted of two back-to-back sessions. Both sessions included blocks of single task trials in which participants performed the equation-verification task only as well as blocks in which participants were also required to make duration estimations. The sessions differed in the nature of these duration estimations. In the equation-duration session, participants estimated the duration for which the equation was presented on screen, whereas in the response-time' session, they estimated the time between the onset of the equation and their response to the equation-verification task. Thus, there were two blocked manipulations: concurrent timing (i.e. timing vs. no timing) and session type (i.e. equation duration vs. response time).

Each session consisted of two blocks of five practice trials followed by six blocks of 32 experimental trials. Practice trials did not differ from experimental trials except that they were not included in data analyses. The pattern of timing and no-timing blocks was AB for the practice trials and ABBABA for the experimental trials, with order counterbalanced between participants. The order in which participants performed the two sessions was also counterbalanced. The entire process took less than 1 hour.

Depending on whether it was part of a timing or a no-timing block, each trial began with either the word *time* in green or the words *no time* in red, presented at the centre of the screen against a white background. After 500 ms the text was replaced by a black fixation cross which remained on-screen for 1,000 ms. This was followed by the equation. The value of the minuend was randomly generated on each trial. All possible subtrahends appeared twice, except for 3, 7, 13, and 17, which all appeared only one time within every experimental block. On each trial the answer had an 80% chance of being correct, a 10% chance of being too high, and a 10% chance of being too low. The equation duration depended on the magnitude of the subtrahend and the correctness of the answer because these variables were expected to also influence response times. The average duration of correct equations was set to 650 ms for a subtrahend of 1, and increased fairly linearly with increasing subtrahend, up to 1,767 ms for a subtrahend of 19. The average durations for incorrect equations ranged from 783 to 2,083 ms. The actual duration on each trial was drawn from a uniform distribution ranging 5% above and below the average duration. Overall, the equation durations paralleled response times but were about 35% shorter on average.

Participants responded by pressing *N* for correct equations and *M* for incorrect equations. If the response was made before offset of the equation, the equation remained on screen until its predetermined offset time, at which point the screen turned blank for 500 ms before displaying the visual analogue scale below the text *Equation Duration?* in the equation-duration session or the text *Equation Task Duration?* in the response-time session. If the response was not made before offset of the equation, the screen remained blank until a response was made, at which point the scale was displayed immediately. Once the estimate was inputted, participants were prompted to “press space” to begin the next trial. The scale was omitted on no-timing trials, with the “press space” prompt immediately following the blank screen.

Analysis of time estimation performance Time-estimation performance is often measured in two ways: (1) by calculating the mean of a set of duration estimates to determine how well the estimates compare to the actual durations on the whole, and (2) by computing some measure of the degree to which

participants were sensitive to trial-by-trial variations in duration. Allocating attention away from the passage of time often results in relatively shorter duration estimates or in less sensitivity to trial-by-trial variations of duration. Assuming that the requirement to perform the equation-verification task forces participants to allocate attention away from the passage of time, we might expect to see differences on these measures if the two duration types differ in terms of their attentional requirements. For example, if response-time estimation is less dependent on attention, then estimates of response times might be longer and less variable than estimates of equation durations. We therefore examined these two aspects of time-estimation performance. First, we compared the estimated durations for each interval type to determine whether one type of interval was perceived as being longer or shorter than the other. Next, we computed correlations between actual and estimated durations of each interval type for each participant in order to measure how well trial-by-trial variations in duration were picked up on by the participants.

These analyses were complicated by the fact that the to-be-timed response times were participant generated and could therefore not be controlled precisely, resulting in different distributions of durations for the two interval types. This affected both of our measures of time-estimation performance (i.e. the means and correlations), because (a) longer durations are likely to be underestimated relative to shorter durations (Lejeune & Wearden, 2009), and (b) the magnitude of the correlation between two variables depends on the variability of those variables (Cohen, Cohen, West, & Aiken, 2013). To address the issue, we matched individual trials from the equation-duration session with those from the response-time session based on the duration of the to-be-timed interval. That is, after outlier removal, the closest pair of trials was matched up, followed by the next closest, and so on until there were no remaining trials with to-be-timed durations within 10% of each other. Only matched trials were submitted to subsequent analyses, and only participants with at least 10 matched pairs of trials were included in these analyses. Because both of these requirements (i.e. that durations were within 10% of each other and that participants had at least 10 matched trials) are arbitrary, we repeated the analyses with all combinations of 5%, 10%, and 20% similarity, and 5, 10, and 20 matched pairs. We report the results of the 10%/10 pairs analysis but provide the range of p values and effect sizes for the other eight analyses in brackets.

Results

Outlier removal Trials were removed from the analyses if they had either incorrect equation verification responses (13.8% of trials) or correct responses that were more than 2.5 standard deviations below (0.0%) or above (2.4%) the participant's mean for a given condition.

Fig. 1 Response time to the equation verification task in Experiment 1. Numbers within each bar show the error percentage for each condition. Error bars represent ± 1 standard error, adjusted to reflect only within-subject variance (Cousineau, 2005)

Equation verification performance Results are shown in Fig. 1. We conducted a 2×2 repeated-measures ANOVA on the equation-verification task RTs, with session type and concurrent timing as factors. Equation-verification-task RTs were marginally longer during the equation-duration session, $F(1, 19) = 3.02$, $MSE = 205999.9$, $p = .099$, $\eta_p^2 = .137$, and when timing was required, $F(1, 19) = 2.82$, $MSE = 63934.1$, $p = .109$, $\eta_p^2 = .129$. The two factors interacted with one another, $F(1, 19) = 7.78$, $MSE = 30682.8$, $p = .012$, $\eta_p^2 = .291$, so we tested the effect of concurrent timing separately for each session type. Participants responded more slowly on timing trials than on no-timing trials during the equation-duration session, $t(19) = 2.38$, $SE = 85.9$, $p = .028$, $d_{rm} = 0.24$,¹ but not during the response-time session, $t < 1$. The same analysis using the error percentage data instead of RTs did not reveal any main effects or interactions (all F s < 1).

Time estimation performance *Mean of estimates.* Based on the raw data, response times (1,672 ms, $SE = 166$ ms) were longer than equation durations (1,141 ms, $SE = 6$ ms), $t(19) = 3.18$, $SE = 166.8$, $p = .005$, $d_{rm} = 1.08$. After applying the matching algorithm, all 20 participants met the criterion of having at least 10 matched pairs of trials (mean = 48.5 pairs, $SD = 16.0$), and amongst this set of trials, response times (1,206 ms, $SE = 42$ ms) and equation durations (1,206 ms, $SE = 41$ ms) had the same average duration. Participants' estimates of the two duration types differed significantly, $t(19) = 2.52$, $SE = 96.0$, $p = .021$ (.021–.043), $d_{rm} = 0.56$ (0.45–0.56), with shorter estimates for response times (1,350

¹ d_{rm} is a measure of effect size calculated by $d_{rm} = \frac{M_{diff}}{\sqrt{SD_1^2 + SD_2^2 - 2 \times r \times SD_1 \times SD_2}} \times \sqrt{2(1-r)}$, (Lakens, 2013).

ms, $SE = 95$ ms) than for equation durations (1,592 ms, $SE = 98$ ms).

Correlations. We calculated the Pearson's correlation coefficient between actual and estimated durations of each trial type for every participant, using the data from the matched pairs of trials. The correlation between actual and estimated response times (0.44, $SE = 0.04$) did not differ significantly from the correlation between actual and estimated equation durations (0.50, $SE = 0.04$), $t(19) = 1.34$, $SE = 0.05$, $p = .197$ (.197–.268), $d_{rm} = 0.33$ (0.29–0.38).

Discussion

Although correlations between actual and estimated durations did not differ between the two interval types, equation durations were estimated as being about 15% longer than equivalent response times. This suggests that equation durations were perceived as being relatively longer than response times, and more generally, that externally defined intervals might be perceived as being longer than internally defined intervals. However, although the trials that were analyzed were matched by actual duration, the matched trials were performed under different temporal contexts: The response-time session involved intervals that were on the whole longer than those in the equation-duration session. This means that the RTs that were chosen by the matching algorithm were among the shorter intervals in their session, whereas equation durations that were chosen were among the longer intervals in their session. The most likely effect of this difference, however, would be for the data to follow Vierdordt's law within each session (Lejeune & Wearden, 2009), which would predict that intervals that were presented alongside other longer intervals (i.e. the response times) would be relatively overestimated compared to those that were presented alongside shorter intervals (i.e. equation durations). Given that it was the equation durations that were estimated as longer, context is unlikely to explain this difference.

In line with previous work examining interference between time estimation and concurrent nontemporal tasks (e.g. Brown, 2006), participants performed the equation-verification task more slowly when required to provide an estimate of the equation duration compared to when no estimate was required. Response-time estimation, however, had no effect on equation verification, suggesting that the processing required to estimate response times differed from what was required to estimate the equation durations. In particular, the results are consistent with the idea that estimates of one's own response times can be accurately generated based only on nontemporal information, without the need for invoking executive resource-dependent temporal processing. However, there may be another explanation. Although one effortful aspect of temporal processing is likely to be a temporal information gathering process (e.g. monitoring the output of an internal clock), attentional processes

have been related to a decisional stage, at which point the outcome of temporal processing is compared to representations in memory in order to produce an estimate (e.g. Meck, 1984). Additionally, the executive function 'access', which has been related to several time-estimation tasks including verbal estimation (Ogden et al., 2014), refers to the access of information from semantic memory (Fisk & Sharp, 2004; Ogden et al., 2014), which is likely to be required during a decisional stage. In our task, which involves a visual analogue scale input at the end of each trial, the decisional stage may occur near or after the end of the to-be-timed interval. The results of Experiment 1 are as would be expected if the locus of interference was a decisional stage occurring at or near the end of the interval, even if effortful processing were required for both duration types. This is because the equation durations used in Experiment 1 were generally shorter than response times, such that on equation-duration trials the end of the interval generally occurred while participants were still performing the equation-verification task, whereas on response-time trials, the end of the interval necessarily occurred once participants were finished the equation-verification task. A late source of interference could therefore have only slowed equation verification on equation-duration trials and not on response-time trials.

Experiment 2

To distinguish between the two potential accounts of the Experiment 1 results (different processes for the different duration types vs. late locus of interference), we lengthened the equation durations in Experiment 2 so that equations generally offset after a response had been made. If the source of interference on equation-duration trials is a process that occurs throughout the interval, the results of Experiment 2 should be the same as those of Experiment 1. However, if interference occurs only at the end of a to-be-timed interval, there should be no interference on either equation-duration or response-time estimation trials, due to both interval types ending at or after the completion of nontemporal processing.

Method

Experiment 2 methods were the same as those of Experiment 1, except where noted.

Participants Twenty-four participants (17 female, seven male, $M_{age} = 20.0$ years, $SD_{age} = 1.8$ years) were included in the analysis of Experiment 2. One additional participant was tested but scored below 70% accuracy on the equation-verification task and was therefore replaced by a new participant.

Design and procedure There were only two procedural differences from Experiment 1. First, the average equation

durations were 2.5 times longer in Experiment 2 than in Experiment 1. The actual duration on each trial was again drawn from a uniform distribution ranging 5% above and below the average duration. To account for the increase in duration, the upper end of the analogue scale was increased to 6 seconds, with labels at 0, 1.5, 3, 4.5, and 6 seconds. Second, because each trial took longer in Experiment 2, we reduced the number of blocks for each duration type from six to four, following an ABBA pattern of timing and no-timing blocks.

Results

Outlier removal Trials were removed from the analyses if they had either incorrect equation verification responses (13.1% of trials) or correct responses that were more than 2.5 standard deviations below (0.0%) or above (2.0%) the participant's mean for a given condition.

Equation verification performance Results are shown in Fig. 2. A 2×2 repeated-measures ANOVA on the equation verification task RTs, with session type and concurrent timing as factors, revealed neither a main effect of concurrent timing, $F(1, 23) = 1.69$, $MSE = 34746.0$, $p = .207$, $\eta_p^2 = .068$, nor of session type ($F < 1$). The Session Type $\times$ Concurrent Timing interaction was marginally significant, $F(1, 23) = 4.16$, $MSE = 27625.8$, $p = .053$, $\eta_p^2 = .153$; however, the underlying cause of the interaction was not the same as in the previous experiment. Unlike in Experiment 1, there was not an effect of concurrent timing during the equation-duration session ($t < 1$), but there was during the response-time session, $t(23) = 2.39$, $SE =$

49.6, $p = .025$, $d_{rm} = 0.14$, with faster responses when concurrent timing was required. Repeating the analysis using error percentage instead of RTs did not reveal any significant main effects or interactions (all $ps > .17$).

Time estimation performance *Mean of estimates.* Based on the raw data, response times (1,870 ms, $SE = 171$ ms) were shorter than equation durations (2,857 ms, $SE = 15$ ms), $t(23) = 5.56$, $SE = 177.5$, $p < .001$, $d_{rm} = 1.88$. Of the 24 participants, 19 of them met the criterion of having at least 10 matched pairs of trials (mean for those 19 participants = 27.3 pairs, $SD = 10.2$), and amongst this set of trials, response times (2,457 ms, $SE = 87$ ms) were on average only 5 ms shorter than equation durations (2,462 ms, $SE = 86$ ms). Estimates of the two duration types differed significantly, $t(18) = 3.50$, $SE = 111.8$, $p = .003$ (.001–.019), $d_{rm} = 0.66$ (0.47–0.67), with shorter estimates for response times (2,027 ms, $SE = 146$ ms) than for equation durations (2,418 ms, $SE = 117$ ms).

Correlations. The correlations between actual and estimated response times (0.58, $SE = 0.03$) did not differ significantly from the correlations between actual and estimated equation durations (0.55, $SE = 0.05$), $t(18) = 0.76$, $SE = 0.05$, $p = .458$ (.119–.924), $d_{rm} = 0.19$ (0.03–0.51).

Discussion

As in Experiment 1, the correlations between actual and estimated durations did not differ between response times and equation durations, suggesting that participants were equally sensitive to changes in the two types of intervals. Also as in Experiment 1, estimated durations were shorter for response times than for equation durations. This was despite the fact that equation durations were longer overall, meaning that any effect of context (i.e. trials that participants completed but that were not included in the analysis) should have acted in the opposite direction from Experiment 1. Because this was not the case, it can be concluded that the difference in estimated duration of response times compared to equation durations was not caused by contextual differences.

In terms of nontemporal task performance, the pattern of results did not match the predictions from either of our potential accounts of Experiment 1. The ‘late locus of interference’ account correctly predicted that equation duration estimation would not slow responses on the verification task but incorrectly predicted that there would be no effect of concurrent timing during the response-time session as well as no interaction between concurrent timing and session type. Alternatively, the ‘different processes for different duration types’ account correctly predicted the interaction and its direction but incorrectly predicted that the effect of concurrent timing would occur in the equation-duration session instead of the response-time session. This discrepancy suggests that some aspect of our

Fig. 2 Response time to the equation verification task in Experiment 2. *Numbers within each bar* show the error percentage for each condition. *Error bars* represent ± 1 standard error, adjusted to reflect only within-subject variance (Cousineau, 2005)

design of Experiment 2 had a consequence that we did not account for.

One possibility that we did not consider is that when equation durations were reliably longer than response times, participants could have used a strategy in which they segmented the interval into two parts: equation onset to response and response to equation offset. Such a strategy would allow participants to use nontemporal information to estimate the duration of the first segment and then perform effortful temporal processing only during the second segment, after equation verification had already been completed. Evidence from research on duration estimation of segmented and overlapping intervals suggests that people are indeed capable of segmenting and reconstructing temporal intervals to produce accurate estimates (Bryce & Bratzke, 2015b; Bryce, Seifried-Dübon, & Bratzke, 2015; Matthews, 2013; van Rijn & Taatgen, 2008). This strategy would explain the lack of concurrent-timing effect in the equation-duration session but cannot explain the effect in the response-time session.

Another possibility is that doubling the durations of the equations simply made the single task blocks less engaging. Depending on how quickly participants made their equation-verification response, they had to wait up to several seconds for the equation to offset on single task trials, with nothing to do during this time. If the single task blocks were too boring, participants may have put less effort into this condition, thereby producing slower RTs which would have had the result of preserving the direction of the interaction but shifting the concurrent-timing effect from the equation-duration trials to response-time trials.

Finally, it is possible that the discrepancy between our predictions and our results was simply due to sampling error. Although we used a within-subjects design, a sample of 24 participants is not exceedingly large. However, if we assume that one of our two predictions for Experiment 2 was correct, the concurrent-timing effect during the response time must have been a false positive, which should rarely occur.

Experiment 3

In Experiment 3, we intermixed short and long equation-duration trials within blocks. This made interval segmentation a less viable strategy because participants did not know at the beginning of each trial whether the equation interval could be segmented by their response and also made the single task condition more engaging as there was always a possibility that the equation would offset quickly. If the lack of interference on equation-duration trials in Experiment 2 was caused by one of these two issues, and not because the locus of interference was a late decisional stage, then equation-duration estimation should interfere with equation verification on both short and long trials in Experiment 3. Otherwise, it should interfere only on short trials.

Method

Experiment 3 methods were the same as those of Experiment 1, except where noted.

Participants Forty-eight participants (28 female, 20 male, $M_{\text{age}} = 20.9$ years, $SD_{\text{age}} = 3.3$ years) were included in the analysis of Experiment 3. One additional participant was tested but scored below 70% accuracy on the equation-verification task and was therefore replaced by a new participant. Three additional participants scored below 70% accuracy on the equation-verification task and were therefore replaced by new participants.

Design and procedure Experiment 3 differed from Experiment 1 only in that on each trial the equation duration was drawn randomly and with equal probability from either the set of durations used in Experiment 1 or the set used in Experiment 2. The analogue scale was always the same as that used in Experiment 2 (i.e. ranging from 0 to 6 seconds).

Results

Outlier removal Trials were removed from the analyses if they had either incorrect equation verification responses (15.4% of trials) or correct responses that were more than 2.5 standard deviations below (0.0%) or above (2.3%) the participant's mean for a given condition.

Equation verification performance² Results are shown in Fig. 3. We conducted a $2 \times 2 \times 2$ repeated-measures ANOVA on the equation verification RTs, with equation length, session type, and concurrent timing as factors. RTs were longer on *long* than on *short* trials, $F(1, 47) = 24.93$, $MSE = 277393.5$, $p < .001$, $\eta_p^2 = .347$, but there was no main effect of concurrent timing, $F(1, 47) = 2.64$, $MSE = 114187.1$, $p = .111$, $\eta_p^2 = .053$,

² We initially collected a sample of 24 participants. In this sample, the Session Type $\times$ Concurrent Timing interaction, $F(1, 23) = 5.77$, $p = .025$, and the Session Type $\times$ Concurrent Timing $\times$ Equation Length interaction, $F(1, 23) = 5.55$, $p = .027$, were both significant. As predicted by the 'different processes for different duration types' account, the Session Type $\times$ Concurrent Timing interaction held for *long* trials, $F(1, 23) = 9.95$, $p = .004$, with an effect of concurrent timing in the equation-duration session, $t(23) = 2.48$, $p = .021$, but not the response-time session. However, the interaction surprisingly did not occur for *short* trials ($F < 1$). Because *short* trials differed from those in Experiment 1 (which showed the interaction) only in that they were interleaved with *long* trials, we conducted an exact replication to determine whether the lack of Session Type $\times$ Concurrent Timing interaction on *short* trials under these conditions was replicable or was simply a Type II error. In the replication, the Session Type $\times$ Concurrent Timing interaction was once again significant, $F(1, 23) = 4.63$, $p = .042$, but was not modified by any higher order interactions. The results presented in this section represent the combination of both samples. Importantly, the key result (the Session Type $\times$ Concurrent Timing interaction on *long* trials) was consistent (and significant) across both samples.

Fig. 3 Response time to the equation verification task in Experiment 3. Data are divided into trials on with the equation duration was drawn from either the pool of short durations or of long durations. *Numbers within*

each bar show the error percentage for each condition. *Error bars* represent ± 1 standard error, adjusted to reflect only within-subject variance (Cousineau, 2005)

or of session type ($F < 1$). There was a Session Type $\times$ Concurrent Timing interaction, $F(1, 47) = 10.51$, $MSE = 56564.6$, $p = .002$, $\eta_p^2 = .183$, and no other significant interactions (all $ps > .05$). The interaction was driven by an effect of concurrent timing during the equation-duration session, $t(47) = 2.74$, $SE = 47.6$, $p = .009$, $d_{rm} = 0.10$, with slower responses when duration estimation was required, and no effect of concurrent timing during the response-time session ($t < 1$).

Repeating the same analysis using the error data instead of RTs showed that more errors were committed on *short* trials than on *long* trials, $F(1, 47) = 16.75$, $MSE = 0.004$, $p < .001$, $\eta_p^2 = .263$. This effect was in the opposite direction from the effect in the RT data, suggesting a speed–accuracy trade-off. No other effects or interactions were significant ($ps > .09$).

Time estimation performance *Mean of estimates.* Based on the raw data, response times (1,864 ms, $SE = 118$ ms) were not significantly different from equation durations (2,015 ms, $SE = 15$ ms), $t(47) = 1.27$, $SE = 118.3$, $p = .209$, $d_{rm} = 0.26$, although the sample means differed by 151 ms. All 48 of the participants met the criterion of having at least 10 matched pairs of trials (mean = 49.3 pairs, $SD = 10.6$). Amongst this set of trials, response times (1,800 ms, $SE = 71$ ms) were on average only 3 ms shorter than equation durations (1,803 ms, $SE = 71$ ms). Estimates of the two duration types differed significantly, $t(47) = 4.52$, $SE = 57.9$, $p < .001$ (all $< .001$), $d_{rm} = 0.48$ (0.46–0.48), with shorter estimates for response times (2,012 ms, $SE = 80$ ms) than for equation durations (2,274 ms, $SE = 77$ ms).

Correlations. The correlations between actual and estimated response times (0.55, $SE = 0.03$) did not differ significantly from the correlations between actual and estimated equation

durations (0.57, $SE = 0.03$), $t(47) = 0.38$, $SE = 0.04$, $p = .705$ (.488–.714), $d_{rm} = 0.06$ (0.06–0.12).

Discussion

The time-estimation results of Experiment 3 matched those from the previous two experiments: Correlations between actual and estimated durations did not differ between the two interval types, but equation durations were perceived as being longer than equivalent response times. In contrast to the results of Experiment 2, estimation of equation durations interfered with performance on the equation-verification task, regardless of whether the equation duration was *short* or *long*. This was the result that was predicted by the ‘different processes for different duration types’ hypothesis, which holds that estimation of the duration of intervals relies on different types of processing, depending on whether the interval is demarcated internally (e.g. response times) or externally (e.g. a stimulus on a computer screen). Additionally, whereas the estimation of externally defined intervals requires effortful processing (Brown, 1997, 2006), estimation of internally defined intervals does not seem to share this requirement, as evidenced by the lack of interference between response-time estimation and equation verification in the three experiments presented so far.

Experiment 4

The results up to this point show that when required to provide a duration estimate by clicking within an analogue scale, estimating the duration of an externally defined stimulus slows equation verification, whereas estimating response times does

not. Although different timing tasks were generally developed for the same goal (i.e. to test time estimation ability), they clearly differ in many aspects (e.g. Ogden et al., 2014), and it is important to understand whether findings generalize across timing tasks or interact with the different aspects of each task. With this in mind, we tested in Experiment 4 whether our findings would generalize to a different time-estimation task: time reproduction. In time-reproduction tasks, instead of reporting estimates by clicking within a scale, participants attempt to produce a separate interval that has the same duration as the encoded interval.

Method

Experiment 4 methods were the same as those of Experiment 1, except where noted.

Participants Twenty-four participants (19 female, five male, $M_{\text{age}} = 19.9$ years, $SD_{\text{age}} = 1.5$ years) were included in analysis of Experiment 4. Three additional participants were tested but scored below 70% accuracy on the equation-verification task and were therefore replaced by new participants.

Design and procedure The procedure was the same as that of Experiment 1 except that instead of reporting their duration estimates using a scale, participants attempted to reproduce the interval. At the point at which the scale would have appeared in the previous experiments, the word *Begin* was displayed instead. Participants pressed the spacebar to mark the beginning of the reproduced interval and pressed the spacebar again to mark the end. The interval was filled with a small fixation cross at the centre of the screen. Following completion of the reproduction, the words *press space* appeared on-screen, and the experiment continued in the same fashion as the others.

Results

Outlier removal Trials were removed from the analyses if they had either incorrect equation verification responses (14.3% of trials) or correct responses that were more than 2.5 standard deviations below (0.0%) or above (2.3%) the participant's mean for a given condition.

Equation verification performance Results are shown in Fig. 4. A 2×2 repeated-measures ANOVA on the equation verification RTs, with session type and concurrent timing as factors, revealed a main effect of concurrent timing, $F(1, 23) = 9.31$, $MSE = 16856.0$, $p = .006$, $\eta_p^2 = .288$, with slower responses when timing was required. There was no main effect of session type ($F < 1$), nor was there a Session Type $\times$ Concurrent Timing interaction ($F < 1$). Repeating the analysis

Fig. 4 Response time to the equation verification task in Experiment 4. Numbers within each bar show the error percentage for each condition. Error bars represent ± 1 standard error, adjusted to reflect only within-subject variance (Cousineau, 2005)

using error percentage instead of RTs did not reveal any significant main effects or interactions (all $ps > .20$).

Time estimation performance *Mean of estimates.* Based on the raw data, response times (1,886 ms, $SE = 138$ ms), were longer than equation durations (1,148 ms, $SE = 5$ ms), $t(23) = 5.34$, $SE = 138.1$, $p < .001$, $d_{rm} = 1.57$. All 24 of the participants met the criterion of having at least 10 matched pairs of trials (mean = 37.0 pairs, $SD = 12.1$), and amongst this set of trials, response times (1,298 ms, $SE = 37$ ms) were on average only 4 ms longer than equation durations (1,294 ms, $SE = 36$ ms). Estimates of the two duration types differed significantly,³ $t(23) = 2.12$, $SE = 91.7$, $p = .045$ (.044–.176), $d_{rm} = 0.45$ (0.34–0.47), with shorter estimates for response times (1,377 ms, $SE = 101$ ms) than for equation durations (1,571 ms, $SE = 66$ ms).

Correlations. The correlations between actual and estimated response times (0.32, $SE = 0.05$) did not differ significantly from the correlations between actual and estimated equation durations (0.30, $SE = 0.05$), $t(23) = 0.43$, $SE = 0.04$, $p = .670$ (.670–.884), $d_{rm} = 0.08$ (0.03–0.08).

Discussion

Time estimation results using the reproduction task were very similar to those found in the previous experiments using the analogue scale response. The correlations between actual and

³ Only one of the nine analyses (the one requiring both a 5% similarity between trials in a matched pair and 20 matched pairs per participant) had $p > .05$. This was likely due to the smaller sample size and therefore lower power resulting from the stricter criteria.

estimated durations were once again equal for the two interval types, and equation durations were again perceived as being longer than response times. In corroboration of previous results that have demonstrated similarities between the two methods (Bryce & Bratzke, 2015a) the pattern of timing performance found in the previous three experiments seems not to be specific to time estimation using analogue scale reports (but see Mioni, Stablum, McClintock, & Grondin, 2014, who found differences between differing reproduction tasks).

Despite the similarities in timing performance, the equation-verification results from the previous experiments did not extend to the use of a duration-reproduction task. Whereas we expected an effect of concurrent timing for equation durations but not for response times, we instead found a main effect of concurrent timing that did not interact with session type. Potential explanations for the difference between reproduction and analogue scale results will be discussed in the General Discussion.

Experiment 5

In Experiment 5, we tested whether the previous results would generalize to a different nontemporal task. One shortcoming of the equation-verification task is that the amount of time required to respond on a given trial can be easily deduced a priori by simply checking the magnitude of the subtrahend. For example, it is clear that it would take more time to subtract 16 from 74 than it would to subtract 1 from 74. Participants could theoretically have produced reasonable estimates without making any effort to monitor their performance. For this reason, we selected a nontemporal task for Experiment 5 for which the amount of time required to respond on a given trial would be much less obvious. To this end, we changed the nontemporal task to a sequence reasoning task, in which participants judged whether two temporally related phrases were presented in the correct order. Brown and Merchant (2007; also Brown, 2014) have argued that sequence reasoning relies on similar executive processes to time estimation and have shown that repeated interval production interferes with concurrent sequence reasoning.

Method

Experiment 5 methods were the same as those of Experiment 1, except where noted.

Participants Twenty-four participants (14 female, 10 male, $M_{\text{age}} = 20.0$, $SD_{\text{age}} = 1.5$)

were included in analysis of Experiment 5. Four additional participants were tested but scored below 70% accuracy on the sequence reasoning task and were therefore replaced by new participants.

Design and procedure This experiment was similar to Experiment 1 except that the nontemporal task consisted of sequence reasoning instead of equation verification. Instead of an equation, two temporally related phrases of up to four words each were presented next to each other on-screen, with the logically earlier phrase presented either on the left (e.g. “start your car–drive away”) or on the right (e.g. “drive away–start your car”). Given that English is read left to right, the order was considered correct when the earlier phrase was presented on the left and the later phrase on the right, and incorrect when the order was reversed. The phrases remained on-screen for a random amount of time between 800 and 4,200 ms, and participants responded in the same manner as in the other experiments. The sequence of experimental blocks was also the same as in previous experiments; however, each block consisted of 18 trials.

Results

Outlier removal Trials were removed from the analyses if they had either incorrect equation verification responses (13.5% of trials) or correct responses that were more than 2.5 standard deviations below (0.0%) or above (2.4%) the participant’s mean for a given condition.

Sequence reasoning performance Results are shown in Fig. 5. A 2×2 repeated-measures ANOVA was conducted on the sequence reasoning task RTs, with concurrent timing (sequence reasoning only vs. sequence reasoning + concurrent timing) and session type (phrase duration vs. response time) as factors. Overall, response times were longer in the phrase-

Fig. 5 Response time to the sequence reasoning task in Experiment 5. Numbers within each bar show the error percentage for each condition. Error bars represent ± 1 standard error, adjusted to reflect only within-subject variance (Cousineau, 2005)

duration session than in the response-time session, $F(1, 23) = 5.92$, $MSE = 63907.3$, $p = .023$, $\eta_p^2 = .205$. There was no main effect of concurrent timing, $F(1, 23) = 2.44$, $MSE = 44094.5$, $p = .132$, $\eta_p^2 = .096$; however, concurrent timing interacted with session type, $F(1, 23) = 12.88$, $MSE = 29030.9$, $p = .002$, $\eta_p^2 = .359$. The interaction was caused by participants performing the sequence reasoning task more slowly when they were required to concurrently time the phrase durations, $t(23) = 2.79$, $SE = 68.8$, $p = .011$, $d_{rm} = 0.27$, but not when required to concurrently time response times, $t(23) = 1.57$, $SE = 36.8$, $p = .129$, $d_{rm} = 0.16$. Repeating the preceding analysis using the error data instead of RT data revealed that no other significant effects or interactions (all $ps > .10$).

Time estimation performance *Mean of estimates.* Based on the raw data, response times (1,666 ms, $SE = 77$ ms), were shorter than phrase durations (2,399 ms, $SE = 58$ ms), $t(23) = 7.46$, $SE = 481.6$, $p < .001$, $d_{rm} = 2.20$. Of the 24 participants, 23 met the criterion of having at least 10 matched pairs of trials (mean for those 23 participants = 22.3 pairs, $SD = 7.3$), and amongst this set of trials, response times (1,815 ms, $SE = 75$ ms) were on average only 5 ms shorter than phrase durations (1,820 ms, $SE = 74$ ms). Estimates of the two duration types did not differ significantly, $t(22) = 1.18$, $SE = 80.6$, $p = .250$ (.087–.228), $d_{rm} = 0.29$ (0.31–0.45), although estimates of response times (1,660 ms, $SE = 69$ ms) were numerically shorter than for phrase durations (1,755 ms, $SE = 69$ ms).

Correlations. The correlation between actual and estimated response times (0.57, $SE = 0.06$) did not differ significantly from the correlation between actual and estimated phrase durations (0.62, $SE = 0.05$), $t(22) = 0.79$, $SE = 0.06$, $p = .441$ (.272–.725), $d_{rm} = 0.19$ (–0.23–0.26),

Discussion

The results of Experiment 5 demonstrate that the previous results extend beyond equation verification. Whereas phrase-duration estimation interfered with concurrent sequence reasoning, response-time estimation did not. Importantly, participants could not easily estimate a priori how long they should take on a given trial and therefore had to base their response-time estimations on information gathered during the trial.

General discussion

The aim of the present research was to shed light on how people estimate the duration of intervals that are defined by their own mental processing (e.g. how long it takes to perform a mental task). We asked whether people estimate the duration of these intervals in the same way that they estimate the duration of externally defined intervals, such as the amount of time

for which a stimulus is displayed on a computer screen. Although there is debate over how exactly externally defined durations are estimated, dual-task experiments involving time estimation produce a fairly consistent pattern of results: Whereas the performance of almost any nontemporal task will interfere with concurrent time estimation, time estimation generally only interferes with concurrent nontemporal tasks that are thought to require executive resources (Brown, 1997). Thus, our approach was to test whether estimation of internally defined intervals would show this same pattern of results.

In five experiments, participants performed an executive-resource-demanding nontemporal task either alone or concurrently with an internally defined or externally defined time-estimation task. Generally, estimation of externally defined durations slowed performance on the nontemporal task compared to the single task condition, whereas estimation of internally defined durations did not affect nontemporal task performance. This suggests that the processes underlying the estimation of one's own response times may differ from those required to estimate externally defined durations. A potential difference is that the amount of informative nontemporal information that is available whenever we perform mental tasks could be sufficient for accurate estimation without the need to perform demanding temporal processing. In the following sections, we will review the results of both the time-estimation and nontemporal tasks and discuss their theoretical implications.

Time estimation performance

The means of, and correlations between, the actual and estimated durations from each of the five experiments are presented in Table 1. Of note is the fact that equation (or phrase) durations were consistently estimated to be longer than equivalent response times. One potential reason for this difference may be that some aspect of our manipulation influenced the way in which participants allocated attention between the temporal and nontemporal tasks. This possibility will be discussed in more detail in a later section. Another potential explanation is the fact that whereas externally defined intervals were terminated with the offset of a stimulus from the computer screen, internally defined intervals were terminated with a participant's keypress. This might have had two possible effects. First, it means that internally defined intervals were 'intermodal', in that the onset and offset of the intervals were demarcated in different modalities (i.e. visual and tactile). Although this has been shown to influence duration estimates (e.g. Rousseau, Poirier, & Lemyre, 1983), the most likely effect would be a decreased correlation for the intermodal (i.e. internally defined) durations, which we did not observe (see Table 1). One would also expect that the estimation of intermodal durations would be more demanding and therefore more likely to interfere with concurrent nontemporal tasks, which is the opposite of what we observed. The other potential

Table 1 Summary of the duration estimation task results from the five experiments

Exp.	Actual durations (ms)		Estimated durations (ms)		Correlations	
	Internal	External	Internal	External	Internal	External
1	1206 (42)	1206 (41)	1350 (95)	1592 (98)	0.44 (0.04)	0.50 (0.04)
2	2457 (87)	2462 (86)	2027 (146)	2418 (117)	0.58 (0.03)	0.55 (0.05)
3	1800 (71)	1803 (71)	2012 (80)	2274 (77)	0.55 (0.03)	0.57 (0.03)
4	1298 (37)	1294 (36)	1377 (101)	1571 (66)	0.32 (0.05)	0.30 (0.05)
5	1815 (75)	1820 (74)	1660 (69)	1755 (69)	0.57 (0.06)	0.62 (0.05)

Note. All values are based only on trials that were paired during the matching algorithm. Standard errors are presented in brackets

effect of the difference in interval termination modality is that because externally defined durations were terminated by the offset of a stimulus, they may be subject to the internal-marker hypothesis (Grondin, 1993), which assumes that more time is required to eliminate the internal trace of a signal than is required to generate an internal onset from the onset of a signal. In other words, externally defined intervals may have seemed longer because they were terminated by a stimulus offset (i.e. the equation or phrase) rather than a stimulus onset (i.e. the keypress).

Finally, if we are correct in our suggestion that different types of processes underlie the estimation of externally defined and internally defined durations, it may simply be the case that, at least under the conditions tested here, the processes underlying the estimation of externally defined intervals lead to longer estimates than those used to estimate internally defined intervals. Importantly, if internally defined durations are estimated using metacognitive information such as feelings of difficulty, it is likely that the magnitude of these estimates would be strongly dependent on the type of task from which the response times are generated. It therefore may be wise to avoid reading too much into the difference in estimates of the two duration types.

Nontemporal task performance

Effects of interval length Although we generally observed that the estimation of externally defined but not internally defined intervals interfered with nontemporal tasks, there were two notable exceptions to this general pattern of results. First, in Experiment 2, when the to-be-timed external stimuli were consistently longer in duration than the response times to the nontemporal task, duration estimation did not slow nontemporal task performance. Importantly, the same was not true for long-duration trials that were mixed with short-duration trials, indicating that the lack of slowing was not caused by the fact that the durations were longer than response times but rather by the fact that they were consistently longer. Therefore, the results of Experiment 2 cannot be taken as indicating that the locus of interference between the temporal and nontemporal

tasks was a decisional stage occurring at the end of the interval, which would preclude interference on response-time trials because performance of the nontemporal task is, by definition, completed by the end of the response time. Instead, the locus of interference was likely a process that occurred throughout the interval, such as one associated with monitoring the passage of time, or with updating a representation of the amount of time that has passed. Alternatively, decisional processes may occur throughout the interval instead of only at or near the end. Whatever the exact cause of this interference, it occurs only for estimation of externally defined, but not internally defined, durations.

Temporal reproduction versus analogue scale The second exception to the general pattern of results was that when duration estimates were reported by temporal reproduction instead of with the analogue scale, timing slowed nontemporal task performance, regardless of whether the interval was externally defined or internally defined. However, in line with some previous research (Bryce & Bratzke, 2015a; but see Mioni et al., 2014), time-estimation performance on the reproduction task was almost exactly equivalent to estimation using the analogue scale. There are two major ways in which the analogue scale and temporal reproduction tasks differ. First, reproduction, in addition to relying on temporal processing, relies on the ability to accurately control the motor system. Differences between time estimation by reproduction and other means (such as clicking within a visual analogue scale) can therefore be attributable either to differences in a participant's perception of time or to differences in their ability to produce an interval matching that perception (e.g. Droit-Volet, 2010; Wing & Kristofferson, 1973). Second, reproduction places greater demand on memory because a representation of the to-be-reproduced duration must be maintained while the duration is reproduced (e.g. Ogden et al., 2014). Neither of these differences can explain our results, because both occur during the reproduction itself, whereas the difference we observed, that is, the interference (or lack thereof) with a nontemporal task, occurred during the encoding of an interval. Instead, it may be the case that it is simply too difficult or effortful to

make accurate temporal reproductions on the basis of nontemporal information (e.g. feelings of difficulty). Whereas the analogue-scale task only required participants to make relative judgments by selecting a relative position within the scale, the reproduction task forced participants to form explicit representations of duration. If such representations were difficult enough to form based only on nontemporal information, participants may have resorted to temporal processing even on response-time estimation trials. Further work will be required to determine whether this is indeed the case. At any rate, the analogue-scale task probably more closely measures the type of ability that is useful in everyday life: Whereas knowing how long certain mental processes might take relative to other durations would help to optimize our interactions with the world, it is less clear how the ability to accurately reproduce the duration of our mental processing would be useful.

Potential alternative interpretations

Overall, the results are consistent with our initial suggestion that any difference between timing of externally defined and internally defined durations would likely be caused by the opportunity to avoid effortful temporal processing during internally defined intervals, and to instead estimate these intervals based on metacognitive information, such as feelings of difficulty. If this is the case, however, it might be considered surprising that the degree of correlation between actual and estimated durations was not significantly different between the two duration types. One might reasonably argue that the cause of the similarity is that the two duration types were in fact timed using the same processes. As discussed earlier, the magnitudes of duration estimates are generally proportional to the amount of attention that was paid to the passage of time during the interval (Block, Hancock, & Zakay, 2010). Our finding that response times were perceived as being shorter than externally defined durations is therefore consistent with participants having timed both duration types using similar processes but having paid more attention to the passage of time on externally defined trials. Such an account would further explain our finding that responses to the nontemporal tasks were slower when the concurrent timing task included externally defined rather than internally defined intervals, because increased attention to the passage of time on externally defined trials implies decreased attention to the nontemporal task on those same trials. However, this account cannot explain why there was no performance decrement on internally defined trials compared to single task trials. Internally defined trials clearly should require much more attention to the passage of time than should single task trials, and should therefore also result in much less attention to, and therefore slower responses to, the nontemporal task than on single task trials. In addition to this, if our findings were caused only by differences in the allocation of attention between externally defined

and internally defined trials, one might expect estimates of response times to be not only shorter than estimates of equation durations but also poorer, as measured by a weaker correlation between actual and estimated durations. Given that neither of these occurred, we favour an account in which response-time estimates were generated from processes that do not interfere with equation verification or sequence reasoning and therefore presumably do not require executive resources. One possibility is that response-time estimation is instead based on nontemporal metacognitive information, such as feelings of difficulty; however, in the next sections, we will discuss two alternative explanations which hold that response-time estimation may be based on largely the same set of processes as the estimation of externally defined durations.

A possible role for expectancy? One possibility is that the difference between timing of internally defined and externally defined intervals is related to expectancy. Generally, when a stimulus is expected to appear during a timed interval, estimates are shorter when the stimulus appears later (Casini & Macar, 1997; Macar, 2002). This has been interpreted as indicating that when a stimulus is expected, resources are allocated away from temporal processing and towards processing related to monitoring for the onset of the stimulus. It is conceivable that the reverse may be true, with the expectation of important events in a to-be-timed interval, such as its offset, resulting in increased attention directed towards the passage of time and away from concurrent nontemporal processing. If, in our experiments, interval-offset expectancy drew attention away from the nontemporal task, it may have produced the pattern of results found in the present study, with high expectancy (and therefore interference) early on, when there was a chance that the interval might end early, but no expectancy (and no interference) until later if the intervals were generally long. There might have been no interference on internally defined trials because participants knew when they were about to respond and therefore did not have to allocate resources towards monitoring for the end of the interval. However, this account rests on two major assumptions. First, that expectation can disrupt not only time estimation but also equation verification and sequence reasoning. Second, that estimation of externally defined durations (using the analogue scale) would not interfere with performance of the nontemporal tasks in the absence of expectancy. Further work will be required to determine whether end-of-interval expectancy plays any major role in time estimation.

A possible role for decentralized timing? A second alternative account is based on the idea of decentralized timing mechanisms. Although most time-estimation studies are interpreted within a single internal clock framework, it has been suggested that different types of intervals may be timed using different clocks (e.g. different clocks for separate modalities;

Wearden, Edwards, Fakhri, & Percival, 1998). If the timing of mental processing and of visual stimuli depend on different timing mechanisms, our findings might be explained if mental processing is always accompanied by a reading from a ‘mental processing clock’, with additional processing being required to read a separate ‘visual stimulus clock’. This account is similar to what we have already proposed (i.e. estimation based on nontemporal information that is always collected during mental processing), except that the estimation is instead based on a reading from a clock that is similar to that used to time visual stimuli but that runs slightly slower. The main drawback of this account is that it requires one to assume the existence of a ‘mental processing clock’ that is similar to but separate from the one used to time visual stimuli, whereas it is already well known that nontemporal metacognitive information is gathered during mental processing.

Conclusion

We have suggested that the processes underlying the estimation of one’s own response times may differ from those required to estimate externally defined durations. In particular, estimation of externally defined durations requires effortful processing that may not be required to estimate internally defined durations. It may be that estimation of internally defined durations does not depend on the explicit ‘temporal processing’ that is required for externally defined intervals but rather is based on nontemporal metacognitive information. This supports previous research suggesting that response time estimation may be based mainly on perceived task difficulty (Bryce & Bratzke, 2014). However, more research will be required to discriminate between this and other possibilities.

References

- Block, R. A., & Grondin, S. (2014). Timing and time perception: A selective review and commentary on recent reviews. *Frontiers in Psychology*, 5, 648. <https://doi.org/10.3389/fpsyg.2014.00648>
- 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.
- Boekaerts, M., & Rozendaal, J. S. (2010). Using multiple calibration indices in order to capture the complex picture of what affects students’ accuracy of feeling of confidence. *Learning and Instruction*, 20(5), 372–382.
- Brown, S. W. (1985). Time perception and attention: The effects of prospective versus retrospective paradigms and task demands on perceived duration. *Attention, Perception, & Psychophysics*, 38(2), 115–124.
- Brown, S. W. (1995). Time, change, and motion: The effects of stimulus movement on temporal perception. *Perception & Psychophysics*, 57, 105–116.
- Brown, S. W. (1997). Attentional resources in timing: Interference effects in concurrent temporal and nontemporal working memory tasks. *Perception & Psychophysics*, 59(7), 1118–1140.
- Brown, S. W. (2006). Timing and executive function: Bidirectional interference between concurrent temporal production and randomization tasks. *Memory & Cognition*, 34(7), 1464–1471.
- Brown, S. W. (2014). Involvement of shared resources in time judgment and sequence reasoning tasks. *Acta Psychologica*, 147, 92–96.
- Brown, S. W., & Merchant, S. M. (2007). Processing resources in timing and sequencing tasks. *Perception & Psychophysics*, 69(3), 439–449.
- 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.
- Bryce, D., & Bratzke, D. (2015a). Are introspective reaction times affected by the method of time estimation? A comparison of visual analogue scales and reproduction. *Attention, Perception, & Psychophysics*, 77(3), 978–984.
- Bryce, D., & Bratzke, D. (2015b). Multiple timing of nested intervals: Further evidence for a weighted sum of segments account. *Psychonomic Bulletin & Review*. <https://doi.org/10.3758/s13423-015-0877-5>
- Bryce, D., Seifried-Dübon, T., & Bratzke, D. (2015). How are overlapping time intervals perceived? Evidence for a weighted sum of segments model. *Acta Psychologica*, 156, 83–95.
- Buffardi, L. (1971). Factors affecting the filled-duration illusion in the auditory, tactual, and visual modalities. *Perception & Psychophysics*, 10(4), 292–294.
- Casini, L., & Macar, F. (1997). Effects of attention manipulation on judgments of duration and of intensity in the visual modality. *Memory & Cognition*, 25(6), 812–818.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2013). Applied multiple regression/correlation analysis for the behavioral sciences. New York: Routledge.
- 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.
- Cousineau, D. (2005). Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson’s method. *Tutorials in Quantitative Methods for Psychology*, 1(1), 42–45.
- Droit-Volet, S. (2010). Stop using time reproduction tasks in a comparative perspective without further analyses of the role of the motor response: The example of children. *European Journal of Cognitive Psychology*, 22(1), 130–148.
- Fisk, J. E., & Sharp, C. A. (2004). Age-related impairment in executive functioning: Updating, inhibition, shifting, and access. *Journal of Clinical and Experimental Neuropsychology*, 26(7), 874–890.
- Fortin, C., & Breton, R. (1995). Temporal interval production and processing in working memory. *Perception & Psychophysics*, 57(2), 203–215.
- Grondin, S. (1993). Duration discrimination of empty and filled intervals marked by auditory and visual signals. *Attention, Perception, & Psychophysics*, 54(3), 383–394.
- Han, H. J., Viau-Quesnel, C., Xi, Z., Schweickert, R., & Fortin, C. (2010). Self-timing in memory and visual search tasks. *Proceedings of Fechner Day*, 26(1), 357–362.
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in psychology*, 4, 863.
- Lejeune, H., & Wearden, J. H. (2009). Vierordt’s *The Experimental Study of the Time Sense* (1868) and its legacy. *European Journal of Cognitive Psychology*, 21(6), 941–960.
- Macar, F. (2002). Expectancy, controlled attention and automatic attention in prospective temporal judgments. *Acta Psychologica*, 111(2), 243–262.
- 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.
- Matthews, W. J. (2013). How does sequence structure affect the judgment of time? Exploring a weighted sum of segments model. *Cognitive Psychology*, 66(3), 259–282.
- Meck, W. H. (1984). Attentional Bias between Modalities: Effect on the internal clock, memory, and decision stages used in animal time discrimination. *Annals of the New York Academy of Sciences*, 423(1), 528–541.
- Miller, G. W., Hicks, R. E., & Willette, M. (1978). Effects of concurrent verbal rehearsal and temporal set upon judgements of temporal duration. *Acta Psychologica*, 42(3), 173–179.
- Mioni, G., Stablum, F., McClintock, S. M., & Grondin, S. (2014). Different methods for reproducing time, different results. *Attention, Perception, & Psychophysics*, 76(3), 675–681.
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41(1), 49–100.
- 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.
- Ogden, R. S., Wearden, J. H., & Montgomery, C. (2014). The differential contribution of executive functions to temporal generalisation, reproduction and verbal estimation. *Acta Psychologica*, 152, 84–94.
- Rattat, A. C. (2010). Bidirectional interference between timing and concurrent memory processing in children. *Journal of Experimental Child Psychology*, 106(2), 145–162.
- Redding, G. M., & Wallace, B. (1985). Cognitive interference in prism adaptation. *Perception & Psychophysics*, 37(3), 225–230.
- Reisberg, D. (1983). General mental resources and perceptual judgments. *Journal of Experimental Psychology: Human Perception and Performance*, 9(6), 966–979.
- Rousseau, R., Poirier, J., & Lemyre, L. (1983). Duration discrimination of empty time intervals marked by intermodal pulses. *Attention, Perception, & Psychophysics*, 34(6), 541–548.
- Thomas, E. A. C., & Brown, I. (1974). Time perception and the filled-duration illusion. *Perception & Psychophysics*, 16(3), 449–458.
- Thomas, E. A. C., & Weaver, W. B. (1975). Cognitive processing and time perception. *Perception & Psychophysics*, 17(4), 363–367.
- Tse, P. U., Intriligator, J., Rivest, J., & Cavanagh, P. (2004). Attention and the subjective expansion of time. *Perception & Psychophysics*, 66(7), 1171–1189.
- van Rijn, H., & Taatgen, N. A. (2008). Timing of multiple overlapping intervals: How many clocks do we have?. *Acta Psychologica*, 129(3), 365–375.
- Wearden, J. H., Edwards, H., Fakhri, M., & Percival, A. (1998). Why “sounds are judged longer than lights”: Application of a model of the internal clock in humans. *The Quarterly Journal of Experimental Psychology: Section B*, 51(2), 97–120.
- Wessel, J. R. (2012). Error awareness and the error-related negativity: Evaluating the first decade of evidence. *Frontiers in Human Neuroscience*. <https://doi.org/10.3389/fnhum.2012.00088>
- Wing, A. M., & Kristofferson, A. B. (1973). Response delays and the timing of discrete motor responses. *Attention, Perception, & Psychophysics*, 14(1), 5–12.
- Zakay, D. (1993). Relative and absolute duration judgments under prospective and retrospective paradigms. *Perception & Psychophysics*, 54(5), 656–664.