Comparing the Estimation of Internally and Externally Defined Interval Durations

by

Michael David Klein

A thesis

presented to the University Of Waterloo

in fulfilment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Psychology

Waterloo, Ontario, Canada, 2019

© Michael David Klein 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Pierre Jolicoeur PhD
Supervisor(s)	Jennifer Stolz PhD
Internal Member	Jonathan Fugelsang PhD
Internal-external Member	Catherine Burns PhD
Other Member(s)	Evan Risko PhD

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 this thesis I distinguish between two types of temporal intervals: internally-defined and externally-defined. Prior research on how humans estimate the durations of temporal intervals has been focussed almost entirely on externally-defined intervals. Because internally-defined intervals have been largely ignored, our level of understanding of how people estimate the durations of these intervals rests on whether they do so using roughly the same set of mental processes as for externally-defined intervals. I sought to collect some initial evidence regarding whether people do indeed estimate the two types of durations using similar processes. A key finding was that estimating the duration of an externally-defined interval (a stimulus on a computer screen) interfered with performance on a concurrent task whereas estimating the duration of an internally-defined interval (a response time) did not (Chapter 2). This finding extended to several different temporal and non-temporal tasks (Chapter 3). The results indicate that processes underlying estimation of internally-defined intervals may differ in meaningful ways from those underlying estimation of externally-defined intervals.

Acknowledgements

I could not have completed this without help and support from Melissa, my family, and Jenn.

Table of Contents

List of Figures

Chapter 1

Imagine you are on a fishing trip. The fish are not particularly active today so you pull out a copy of your preferred psychology journal and begin reading. Suddenly, you notice the line has gone taught – there is a fish on the other end! At this point, you have a decision to make: should you immediately put down the journal and reel in the fish, or can you first finish reading an engrossing paragraph about *Time Perception*? Your choice will essentially come down to a comparison between two predicted durations. Which duration do you expect to be longer: the amount of time between when a fish is hooked on your line and when it escapes, or the amount of time required for you to read half a paragraph? These two intervals may seem to be comparable and, indeed, in most of the literature on duration estimation, they might be treated interchangeably. In this dissertation, however, I will outline a key difference between them and explore whether this difference may lead people toward fundamentally different strategies for estimating their durations.

Consider the first interval – the amount of time that passes from when a fish is hooked on your line to when it escapes. If you are not actively interacting with the line, then you have no control over the duration of this interval[1]. You will probably perceive the onset of the interval either visually (e.g. seeing the rod move or seeing a disruption in the water) or auditorily (e.g. hearing the rod or the water). Then, after some time, the end of the interval will be demarcated by the offset of the visual or auditory signal. I will refer to this type of interval as *Externally-Defined*. Externally-Defined intervals have the property that, once they have begun, our thoughts and actions do not influence the timing of the end of the interval. To know when the interval has ended,

[1] Although the way you set up your line may affect the distribution of durations, you do not influence the individual events that occur once you have set up.

we must wait for an external stimulus to occur (or end). There are myriad examples of Externally-Defined intervals in everyday life. When we count the number of seconds between a lightning strike and its corresponding thunderclap, we are estimating the duration of an Externally-Defined interval that is defined by the amount of time between two separate stimuli. When we sit at a red light waiting for it to turn green, we might estimate the duration of the interval defined by the onset and offset of the red light.

The second interval – the time required for you to read half a paragraph – is an *Internally-Defined* interval, henceforth referred to simply as an *internal* (vs. *external*) interval. The duration of an internal interval *does* depend on our thoughts and/or actions. For example, the time from when you notice the fish on the line to when you finish reading the paragraph depends on how quickly you read. The durations of internal intervals are just as important to us in day-to-day life as those of external intervals. Examples might include paying attention to how long one takes to solve a certain type of problem while preparing for an upcoming test, or attending to how long it takes you to run across University Ave., which you must jaywalk across almost every day during your graduate studies. In general, although it is extremely important to have a sense of the duration of intervals in the outside world, this knowledge would be nearly useless if we did not also know the durations of our own thoughts and actions.

Whereas the estimation of external interval duration has been studied in great depth (see Block & Grondin, 2014), the estimation of internal interval duration has been largely ignored. This does not seem to be because researchers are not interested in internal intervals, but rather because they have not realized that there may be a difference in how we estimate durations of the two types of intervals. When internal intervals *have* been studied, it has generally been assumed that they do not differ from external intervals (e.g. Corallo, Sackur, Dehaene, & Sigman, 2008; Marti, Sackur, Sigman, & Dehaene, 2010; with one exception being Bryce & Bratzke, 2014).

If one does *not* believe there to be a difference between internal and external intervals, then there are two main reasons why the natural choice should be to study external intervals. First, they can be controlled much more precisely. When your intervals depend on your participants' performance, it is much more difficult to control their durations, which are likely to change over the course of an experiment and from participant to participant, and are not certain to follow your desired distribution. Second, most of the common manipulations in duration estimation research (e.g. dual-task interference, stimulus intensity, etc) will also influence the durations of internal intervals[2], making it much more difficult to find clean manipulations that will not be confounded with interval duration. Both difficulties are compounded by the fact that the duration of an interval itself influences peoples' estimation performance. For example, the durations of relatively longer intervals tend to be under-estimated (Zakay & Block, 1997; Lejeune & Wearden, 2009).

It is important to note that there does not necessarily *need* to be a difference between estimation of internal and external interval durations. As I explained earlier, estimating the duration of an external interval requires one to wait for an external stimulus to demarcate the end of the interval. While waiting for this stimulus, one presumably engages in some form of 'temporal processing' to keep track of how much time has passed. There is nothing stopping someone from doing this for internal intervals. Although the end of an internal interval is typically demarcated by one's own actions (e.g. their eyes reaching the end of the paragraph, or, in the context of my experiments, their key-press or button-press response), these actions can be treated as external

[2] For example, if your internal interval is a response time to some task, then implementing a dual-task manipulation will almost certainly also increase the mean duration of your intervals. This, obviously, need not be the case for external interval durations, which can be completely controlled by the researcher.

stimuli. For example, a key-press could be treated like a tactile stimulus. In fact, in this dissertation I will not try to argue that estimation processes for the two types of interval durations definitely differ, but rather that it is not clear that they do not differ. Although I argue that my main findings are difficult to explain from a 'Same Process' standpoint, many of my smaller findings fit in well with existing literature on external intervals.

If there *is* a difference, it is likely to be driven by the combination of two factors: (a) the amount of relevant 'non-temporal' information that is available during internal intervals, and (b) the relative ease with which this information can apparently be obtained compared to 'temporal' information. Focussing first on factor (b), one of the most widely accepted ideas regarding duration estimation (of external intervals) is that it is an attention-demanding process (Hicks, Miller, & Kinsbourne, 1976; Macar, Grondin, & Casini, 1994; Zakay & Block, 1995, 1996; Brown, 1997). Brown (1997) found that the performance of a concurrent non-temporal task interfered with duration 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. In a later review of 33 experiments, Brown (2006) found that duration estimation interfered with non-temporal task performance in 52% of cases. He argued that time estimation generally only interferes with performance on concurrent tasks that require executive resources, revealing that time estimation itself is heavily based on executive resources (see also Brown, 1997)[3]. Subsequent research has demonstrated that timing does interfere with

[3] Counting the proportion of significant results is not ideal meta-analytic practice, and a more basic interpretation may be that Brown was mistaking noise for signal, and that there is a consistent small interference effect (irrespective of non-temporal task type) that most studies simply had low power to detect. If one were to operationalize 'executive control', they could perform a proper meta-analysis to test the claim that duration estimation interferes specifically with executive control tasks. Unfortunately, I am not aware of any such work.

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 Miyake, Friedman, Emerson, Witzki, Howerter, & Wager, 2000; Fisk & Sharp, 2004).

Thomas and Brown (1974; see also Thomas & Weaver, 1975) argued that the outcome of mental timing is not only an encoding of the output from temporal processing, but also an encoding of the non-temporal information that was processed during the interval (see factor (a) above). The actual perceived duration is then a weighted average of these two encodings. Like temporal processing, encoding of non-temporal information likely requires attention in order to be successful (Fernandes & Moscovitch, 2000). The existence of some temporal illusions provides some evidence that non-temporal information does indeed influence duration judgements (e.g. Buffardi, 1971; Brown, 1995; Tse, Intriligator, Rivest, & Cavanagh, 2004; though these manipulations could arguably influence temporal processing in addition to non-temporal processes). The most successful and widely-accepted models of external duration estimation, however, a class of models called pacemaker-accumulator models (e.g. Zakay & Block, 1995), treat non-temporal processes as minor contributors at best. In these models, duration estimation depends mainly on a type of temporal processing wherein attention is directed toward some cyclic signal (i.e. a mental "timer" or "pacemaker", see Buhusi & Meck, 2005), with the cumulative count being stored and updated in memory (Casini & Macar, 1997; Fortin & Massé, 2000).

I propose that, whereas estimation of external interval durations may depend more on temporal processing, estimation of internal interval durations may rely mostly or exclusively on non-temporal processing. Returning to the two factors I listed earlier: (a) the non-temporal information available for internal (compared to external) intervals is much more relevant to duration estimation because this information (e.g. confidence level, Boekaerts & Rozendaal, 2010, or feeling of error, Wessel, 2012) is directly related to performance, which is in turn directly related to the duration of the interval, and (b) because attention is presumably already directed toward the task at hand, it is possible that little or no additional effort is required to encode this information. In other words, duration estimation for internal intervals may bypass the need for explicit temporal processing, instead relying on information that is already encoded anyway. Why bother invoking effortful temporal processing if you can piggy-back on processing that is already occurring? The prediction that arises most naturally from this hypothesis is that, whereas estimation of external interval durations should interfere with performance on concurrent non-temporal tasks (especially tasks that require executive resources), estimation of internal interval durations should interfere either to a lesser degree or not at all. In Chapter 2 I will present three experiments that form an initial attempt to test this prediction.[4]

[4] One previous study has also compared performance on a non-temporal task under single task and concurrent response time estimation conditions, finding that response time estimation did not interfere with non-temporal task performance (Marti, Sackur, Sigman, & Dehaene, 2010). However, because their goal was not to contrast timing of internal and external intervals, they (a) did not compare their results to a condition in which estimation of an external interval duration was required, and (b) used non-temporal 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 external intervals.

Chapter 2

The experiments presented in this chapter were designed to test the prediction that estimation of the duration of external, but not internal, intervals should interfere with performance on a concurrent non-temporal task. I will begin by explaining the tasks and measures that I used in these experiments. Next, I will outline the method used in the three experiments, and finally, I will present and discuss the results.

Tasks

My main prediction is that, if internal interval duration estimation does not require temporal processing, then it should not interfere with a concurrent task (whereas external interval duration estimation should interfere). Two tasks are required to test this prediction, and I will refer to these as the *temporal task*, which is the task that requires participants to make a duration estimate, and the *non-temporal task*, which is the concurrent task that is unrelated to duration estimation.

Duration estimation researchers have developed many different temporal tasks, each of which has both pros and cons. The most basic temporal task is probably Verbal Estimation, wherein participants simply verbally state their estimate of a duration, usually as a number of seconds (Zakay & Block, 1997). A very similar task, usually referred to as a Visual Analogue Scale (VAS) task (e.g. Bryce & Bratzke, 2014), has participants select an estimated duration by clicking or moving a marker somewhere within an analogue scale presented on a computer screen. The scale is generally presented as a range of seconds (e.g. ranging from 0 to 4 s). Although these tasks are very simple to implement and perform, they include a potentially unnecessary and unwanted source of variance: the participants are forced to convert their mental representation of a duration into specific units (e.g. seconds). This may have a relatively benign effect (i.e. simply increasing the variance of participant's estimates, or the variance between participants), but also can make results more difficult to interpret: does your manipulation influence your participants' perception of duration itself, or their conversion from that percept into explicit units?

One way to deal with this problem is to include a training phase to allow participants to calibrate themselves before the experiment begins. Another way is to avoid the problem altogether by using a different temporal task. Reproduction tasks, for example, require participants to, upon completion of an interval, produce a second interval that has the same duration as the first. This task comes with its own complications, however, since a participant's ability to reproduce a duration might depend on several potentially irrelevant factors such as their working memory capacity or their physical ability (Ogden, Wearden, & Montgomery, 2014; Ivry, 1996). Finally, a task that avoids both the need to convert into specific units and the requirement to physically reproduce an interval is a Comparison task in which participants are presented with two intervals and must choose the one that had the longer (or shorter) duration. This task, of course, introduces the complication of a second interval, and researchers must be very careful that any experimental manipulations influence only one of the two intervals, or else they might have no effect on participants' responses.[5]

The experiments presented in this chapter used the VAS task. I made this decision for two main reasons. First, because participants were required to complete two tasks on each trial, I felt it would be best to use a relatively simple temporal task in order to keep things as uncomplicated as possible. Second, previous duration estimation studies that have used internal intervals have generally used the VAS task (e.g. Corallo et al., 2008; Marti et al., 2010; Bryce & Bratzke, 2014).

[5] This is not a complete survey of commonly used temporal tasks, however it is a relatively complete list of those tasks that I strongly considered using.

Using this task therefore made my work more relevant to the published literature. The more similar my work is to the past literature, the closer it can come to answering the question of whether the estimation of internal interval duration, *as studied so far in the literature*, depends on different processes than for external intervals. Furthermore, since my key independent variable is not the duration estimates themselves, but rather the performance on the non-temporal task, it is not particularly important how well participants can translate durations into explicit units.

As the non-temporal task, I used an Equation Verification task adapted from Brown (1997). In this task, a simple mathematical equation was presented on the computer screen and participants were asked to make a keypress to indicate whether the equation was true (i.e. whether the answer that was presented was the correct answer). I selected this task because mental arithmetic is cognitively demanding, has been argued to require executive functioning (e.g. Reisberg, 1983; Redding & Wallace, 1985) and, most importantly, has been shown to be sensitive to interference from (external) duration estimation tasks (Brown, 1997). Roughly, one trial of the experiments proceeded as follows: the equation appeared on screen for some (predetermined) amount of time, participants made their Equation Verification keypress response as soon as they believed that they knew the answer, then they made an estimate of either the screen-time of the equation (i.e. the external interval) or their own response time (RT) to the verification task (i.e. the internal interval) by clicking within the VAS.

The most serious drawback in the use of these tasks is that participants might easily be able to find a way to 'cheat'[6]. Specifically, it would probably not be very difficult to repurpose the VAS for some other response that is presumably correlated with RT. For example, Bryce and Bratzke (2014) found that reported difficulty was more strongly correlated with actual RT to a simple

[6] Although I refer to this as 'cheating', I technically did not ask the participants not to do this. They were only instructed to estimate their RT and were allowed to do so however they saw fit.

perceptual task than were estimated RTs. Based on this finding, one strategy that would be relatively successful would be to simply pretend that the VAS is asking for a difficulty rating instead of a duration estimate (ignoring the labels on the scale). Note that this would only be an issue if participants were not honestly attempting to estimate their RTs. If participants were truly attempting to estimate their RTs but did so based in part on their feelings of difficulty, this would be in line with my hypothesis that estimation of internal interval durations is heavily based on non-temporal information.

Participants who realize that math problems with larger numbers take longer to solve might also ‘cheat’ by simply making their duration estimates based on the magnitude of the numbers in the math problem. In fact, participants using this strategy could come up with their estimate before even attempting to solve the problem. In this case the estimate would be completely unrelated to the actual duration except through a third variable (i.e. the number magnitude). This strategy is detectable, at least, in that the estimates from participants who use this strategy should no longer predict RTs once number magnitude has been controlled for. Even so, it would not be prudent to make any general claims from a set of experiments using only one temporal and one non-temporal task. For this reason, in Chapter 3 I will explore whether the results in this chapter generalize to a range of temporal and non-temporal tasks.

Measures

Duration estimation performance is generally measured in one (or both) of two ways: First, by calculating the mean of a set of duration estimates and looking at how it compares either to the mean of the actual durations or to the mean of estimates from another condition. Second, it can be measured by computing some measure of the degree to which participants were sensitive to trial-by-trial variations in duration.

Generally, dual-task manipulations influence both measures, with a concurrent load resulting in duration estimates that are both shorter (e.g. Hicks, Miller, & Kinsbourne, 1976; Brown, 1997) and less sensitive to trial-by-trial variation (Brown, 1997). The reason for this is that duration estimation requires attention. In pacemaker-accumulator models of duration estimation, less attention to the passage of time results in fewer 'time units' being accumulated (e.g. Casini & Macar, 1997; Fortin & Massé, 2000). As an example, imagine trying to track the passage of time by carefully listening to and counting the ticks of a clock on the wall behind you. If something else were to capture your attention for a moment, you might miss a few of the ticks, thereby underestimating the number of ticks that had occurred and therefore also the amount of time that had passed. A potential explanation for the reduced sensitivity to trial-by-trial variation in duration is that, in dual-task situations, the amount of attention directed toward the passage of time will not necessarily be consistent from trial to trial. Therefore, even if the duration on one trial is longer than on another, it may be perceived as being shorter if less attention happened to be allocated to the duration estimation task during that trial (Brown, 1997).

Throughout this dissertation I will use mean estimated duration to measure participants' overall impressions of the durations of the two types of intervals, and Pearson's correlation coefficients to measure their sensitivity to trial-by-trial variation. Unfortunately, both of these measures are influenced by the distribution of to-be-timed durations (i.e. relative underestimation for longer vs. shorter durations, Lejeune & Wearden, 2009, and higher correlations for sets of durations with a wider vs. narrower spread, Cohen, Cohen, Aiken, & West, 2013), which cannot be exactly controlled for internal intervals. Therefore, rather than analyzing data from *all* trials of each type, I will employ a trial-matching procedure in which similar internal and external intervals are matched up with one another based on their durations, and only trials that have a match are included in analysis. I will outline this procedure in more detail in the Experiment 1 Results section.

Although I will present and discuss the results of these measures, I will generally try to avoid reading too much into the results of the duration estimation tasks. The reason for this is because if durations of the two interval types are indeed estimated using completely different processes, then there is no reason to think they would be comparable at all and there is little reason to have any a priori hypothesis regarding performance on the estimation task. Therefore, the main focus throughout this dissertation will be the performance on non-temporal tasks.

Performance on the Equation Verification will also be measured in two ways: RT and error rate. I will define performance on this task as being relatively poorer if responses are slower (i.e. longer RT) and/or more error prone (i.e. higher error rate). Brown (1997) found that a concurrent temporal task (repeated interval production) increased error rate but did not influence RT (as measured by the number of responses after a set amount of time) on an Equation Verification task. However, because his procedure differed substantially from what I will present here, it would not be surprising to instead see dual-task interference manifest as either slower RTs only, or as both slower RTs and higher error rate.

General Method for Experiments 1-3

Participants. All participants were recruited through the University of Waterloo Sona system, which is an online site where students can sign up to participate in psychological studies either for course credit or for pay. All experiments presented here were posted for course credit only, with a remuneration of 1% towards the students' grade for any eligible course[7]. The experiments all required between 30 to 60 minutes to complete. Restrictions were set up so that the study only appeared on Sona for students who (a) reported on a pre-screen questionnaire that they had normal or corrected-to-normal visual acuity, and (b) had not participated in any of the previous experiments within this line of work.

Except where reported, sample sizes were determined before data collection began. Although the choice of sample size was often at least partly arbitrary, I will outline any reasoning underlying each choice in the specific method section for each experiment. The data from any participants achieving under 70% accuracy on the non-temporal task were excluded from all analyses and were replaced with data from new participants in order to maintain the desired sample size and counterbalancing.

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, while seated at the computer desk in a small room (approximately 3 × 3 m).

Stimuli. Stimuli consisted of subtraction problems displayed on the computer screen. The problems were presented 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

[7] This is treated as a bonus mark in most (but not all) courses, such that students can still achieve a grade of 100% without participating in any studies.

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 VAS presented on screen. The scale was labelled with seconds, with five explicitly marked durations, as well as one unlabeled tick mark mid-way between each of these values. The range of the scale, and therefore the values of the explicitly marked durations, differed between experiments.

Design and Procedure. The experiments each 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. There were therefore two blocked manipulations: Concurrent Timing (i.e. Estimation vs. No Estimation) and Session Type (i.e. Equation Duration vs. Response Time)[8].

Depending on whether it was part of an Estimation or a No Estimation 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

[8] Although this is the design that seems most natural to me, some find it confusing that the 'No Estimation' trials are divided across the two sessions even though they do not differ from one another (except that they were interleaved with different types of duration estimation blocks). The design that I believe is more natural to most researchers is to have three conditions: internal estimation, external estimation, and no estimation. Unfortunately, this did not become apparent to me until I had already completed five experiments. Therefore, Experiments 1 through 5 use my 2 × 2 design in which the question of interest is whether there is an interaction between Concurrent Timing and Session Type, and Experiment 6 uses (roughly) a 3 × 1 design in which the question of interest is whether the external estimation condition differs from the internal and no estimation conditions, and whether the internal and no estimation conditions differ from one another.

cross which remained on screen for 1000 ms. This was followed by the equation. The value of the minuend was randomly generated on each trial. All possible subtrahends appeared twice within every experimental block, except for 3, 7, 13, and 17, which all appeared only once per block. On each trial the answer had an 80% chance of being correct, a 10% change of being too high, and a 10% chance of being too low. The distributions of equation durations varied across experiments, but in all cases the duration depended on the magnitude of the subtrahend and the correctness of the answer because these variables were expected to also influence response times.

Participants responded by pressing 'N' on the keyboard 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 VAS 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 the 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.

Experiment 1

Method. The method of Experiment 1 was as described in the General Method section except where noted.

Participants. The sample size was set to 20 participants. It was required that the sample size be a multiple of four due to the design of the counterbalancing (i.e. the order of the two sessions, plus the order of Estimation vs. No Estimation blocks). Otherwise, there was no particular reason for the decision. The 20 participants had a mean age of 21.1 years (SD = 1.4 years), and 11 were female (9 male).

Stimuli. The VAS ranged from 0 to 4 seconds, with 0, 1, 2, 3, and 4 s explicitly marked.

Design and Procedure. 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 Estimation and No Estimation 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 durations were set to match average RTs from a pilot experiment, with the average duration of correct equations set to 650 ms for a subtrahend of 1 and increasing fairly linearly with increasing subtrahend up to 1767 ms for a subtrahend of 19. The average durations for incorrect equations ranged from 783 to 2083 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 participants' RTs but were about 35% shorter on average.

Results.

Data Preprocessing. All participants achieved at least 70% accuracy on the Equation Verification task, therefore none had to be removed and replaced. Trials were removed from analyses (except for the analysis of Equation Verification task accuracy) 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%) a participant's mean for a given condition.

Equation Verification performance. Equation verification results are shown in Figure 1. I 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, $F(1,19) = 2.82$, $MSE = 63934.1$, $p = .109$, $\eta_p^2 = .129$. Of key importance was the fact that the two factors interacted with one another, $F(1,19) = 7.78$, $MSE = 30682.8$, $p = .012$, $\eta_p^2 = .291$.

To investigate the cause of the interaction, I tested the effect of Concurrent Timing separately for each Session Type, finding that participants responded more slowly on Estimation trials than on No Estimation trials during the Equation Duration session, $t(19) = 2.38$, $SE = 85.9$, $p = .028$, $d_{rm} = 0.241$[9] 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 $Fs \leq 0.24$ (all $ps \geq .627$).

[9] 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).

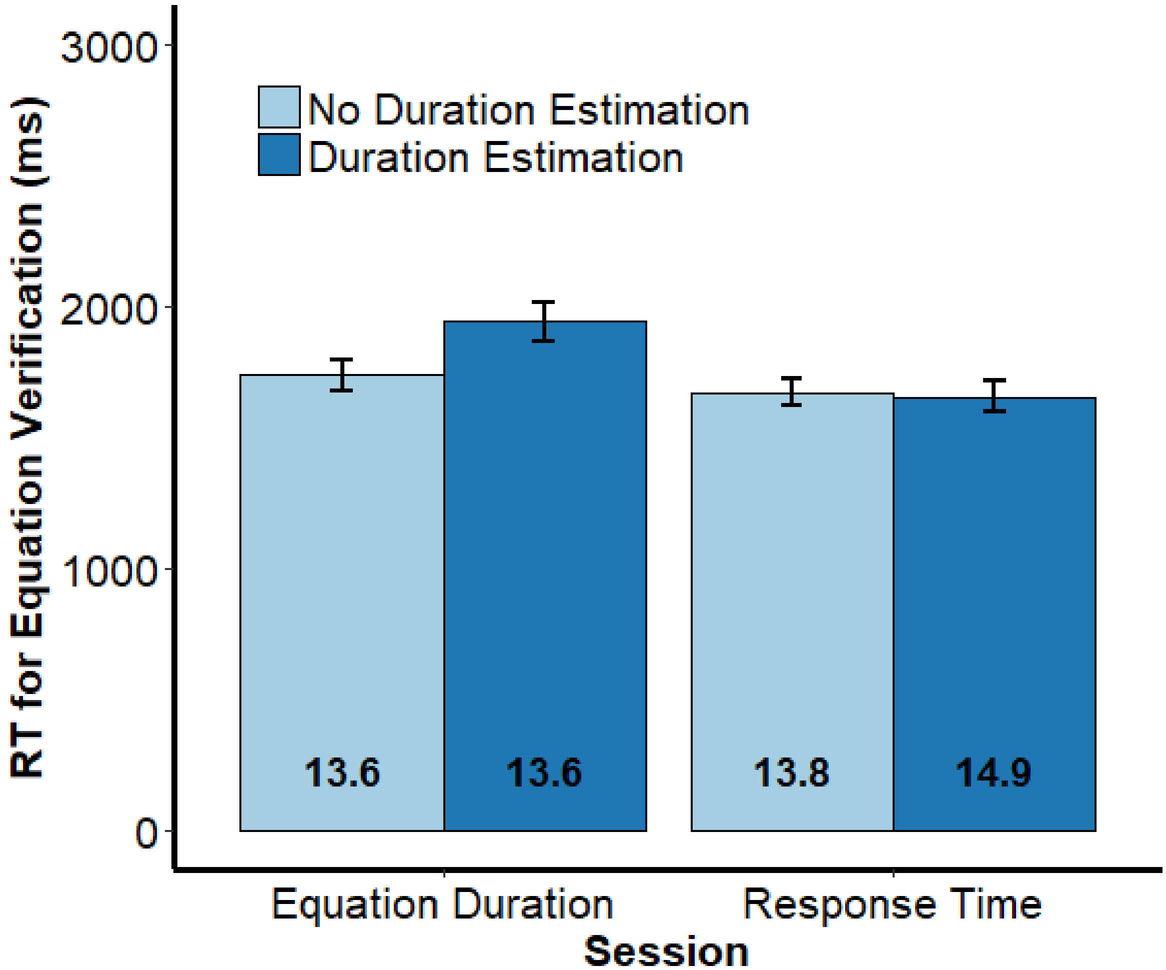

Figure 1. Response time and error percentage to the Equation Verification task in Experiment 1. The numbers within each bar show the error percentage for each condition. Error bars represent ±1 standard error, adjusted to reflect only within-subject variance (Cousineau, 2005).

Duration Estimation performance. I matched individual trials from the Equation Duration session with those from the RT 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. Ties were broken by random selection. 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, I repeated the analyses with all combinations of 5%, 10%, and 20% similarity, and 5, 10, and 20 matched pairs. I 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.

Mean of Estimates. Based on the raw data, RTs (1657 ms, SE = 162 ms), were longer than Equation Durations (1133 ms, SE = 6 ms), $t(19) = 3.21$, $SE = 163.5$, $p = .005$, $d_{rm} = 1.10$. 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 = 16.2), and amongst this set of trials, RTs (1205 ms, SE = 41 ms) and Equation Durations (1203 ms, SE = 41 ms) had about the same average duration. Participants' estimates of the two duration types differed significantly, $t(19) = 2.52$, $SE = 95.8$, $p = .021$ (.020 to .041), $d_{rm} = 0.56$ (0.46 to 0.57), with shorter estimates for RTs (1350 ms, SE = 95 ms) than for Equation Durations (1591 ms, SE = 97 ms).

Correlations. I 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 RTs (0.45, SE = 0.04) did not differ significantly from the correlation between actual and estimated Equation Durations (0.50, SE = 0.04), $t(19) = 1.19$, $SE = 0.05$, $p = .250$ (.203 to .302), $d_{rm} = 0.31$ (0.29 to 0.37).

Discussion. This experiment was designed to test whether estimation of internal vs. external interval durations would differ in level of interference with a concurrently-performed non-temporal task. Specifically, I predicted that if the durations of internal intervals (i.e. RTs) are *not* estimated in the same way as those of external intervals (i.e. Equation Durations), then the amount of interference, as measured by RT (and/or error rate) to the non-temporal task, would likely be reduced for internal intervals. Given the design of this experiment, such a reduction would manifest as an interaction between the Concurrent Timing (Estimation vs. No Estimation) and Session Type (Response Time vs. Equation Duration) factors. The interaction is indeed present in the resulting data. Specifically, in line with previous work examining interference between duration estimation and concurrent non-temporal tasks (e.g. Brown, 1997), participants performed the Equation Verification task more slowly when required to provide an estimate of the Equation Duration compared to when no estimation was required, whereas RT estimation had no effect on Equation Verification performance. This finding is consistent with the idea that estimates of one's own RTs can be accurately generated based mainly on non-temporal information, without the need for invoking attention-demanding temporal processing. I will refer to this interpretation as the 'Different Processes' interpretation. There are, of course, some reasonable 'Same Process' interpretations that do not require that duration estimation processes differ between the interval types. I will outline below what I believe to be the two strongest Same Process interpretations of Experiment 1 and will tackle one of these two interpretations in Experiment 2. The first Same Process interpretation is based on a challenge to the belief that temporal processing is consistently effortful throughout the entirety of an interval. Specifically, it is often argued that temporal processing is effortful because one must continuously monitor the pacemaker (e.g. Zakay & Block, 1995) and effortfully maintain the contents of the accumulator (e.g. Casini & Macar, 1997; Fortin & Massé, 2000) throughout the interval. However, attentional processes have also been related to a decisional stage, at which point the outcome of this temporal information accumulation process is compared to representations in memory to produce an actual estimate (e.g. Meck, 1984). It could be this decisional stage, and not the temporal information accumulation process, that interferes with Equation Verification. In fact, an executive function called 'access', which has been related to several duration estimation tasks including verbal estimation (Ogden et al., 2014), refers to the access of information from semantic memory (Ogden et al., 2014; Fisk & Sharp, 2004), which is likely to be required during a decisional stage. In Experiment 1, which involved a VAS input at the end of each trial, the decisional stage may have occurred near or after the end of the to-be-timed interval. The results of Experiment 1 are therefore 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 RTs, such that on Equation Duration trials, the end of the interval generally occurred while participants were still performing the Equation Verification task, whereas on RT 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 RT trials. According to this interpretation, the Concurrent Timing effect in the Equation Duration session should disappear if the Equation Durations were set to be longer than RTs, because the decisional stage would then occur after Equation Verification had already been completed. The second Same Process interpretation is that the findings of Experiment 1 reflect differences in how participants allocated their attention between the temporal and non-temporal tasks. Specifically, it is possible that participants allocated more attention to the performance of the Equation Verification task when they were instructed to estimate their own RTs, relative to when they estimated the Equation Durations. This would result in better duration estimation performance for Equation Duration estimation trials than for RT estimation trials, at the expense of poorer Equation Verification performance. The observed results arguably do match this pattern. Specifically, Equation Durations were estimated as being longer than their RT counterparts. Although this difference does not necessarily imply better performance, longer duration estimates under concurrent load are often interpreted as resulting from increased attentional allocation toward the passage of time (Zakay & Block, 1995). Additionally, this difference is not predicted by (but also does not contradict) the Different Processes interpretation, which makes no specific predictions regarding how internal and external durations should be perceived relative to one another. A significant shortcoming of this interpretation is that although Equation Verification RTs were longer in the Equation Duration session than in the RT session, RTs were not longer for Estimation than for No Estimation trials within the RT session. This is not easy to explain with an attentional allocation account given that all attention should be allocated to the Equation Verification task during No Estimation blocks. It presumably follows that all attention is also allocated to this task during Estimation blocks in the RT session (as predicted by the Different Processes account). Another shortcoming is that, although Equation Durations were estimated as being longer than equivalent RTs, correlations between actual and estimated durations were not significantly higher for Equation Durations than for RTs. One would expect *both* longer estimates *and* higher correlations if more attention were allocated to the passage of time during the Equation Duration session (Brown, 1997). However, although the correlations did not differ significantly, they were nominally higher for the Equation Duration session. It is possible that the lack of significant difference reflects a lack of power, so it will be worthwhile to keep an eye on the correlations in later experiments. One final note regarding the results of Experiment 1 is that care should be taken in the interpretation of the duration estimation performance. Although the trials that were analyzed were matched based on their actual duration, the matched trials were performed under different temporal contexts: the RT session involved intervals that were on average 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 two most likely consequences of this difference would have opposite effects. One consequence 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 RTs) 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, this cannot explain the difference. The other potential consequence would arise if participants tended to click near the centre of the VAS rather than at the extremes (which would occur, for example, if participants trusted that the experimenter had selected a reasonable range of durations for the VAS). In this case, intervals that were presented alongside other longer intervals (i.e. the RTs) would be relatively *under*estimated compared to those that were presented alongside shorter intervals (i.e. Equation Durations). This cannot be ruled out as a potential explanation for the difference in mean duration estimates between the interval types.

Experiment 2

In Experiment 2 I pitted the Different Processes account and the Late Locus of Interference account against one another. To do so, I simply doubled the Equation Durations from Experiment 1, leaving the rest of the procedure unchanged. I expected that this change would have the effect of making Equation Durations generally longer than RTs. If the Late Locus of Interference interpretation of Experiment 1 is correct, then there would be no interference in either the Equation Duration or the RT session, due to both interval types ending at or after the completion of non-temporal processing. If it is not correct, however, the pattern of results from Experiment 1 should be observed again here.

Method. The method of Experiment 2 was as described in the General Method section except where noted.

***Participants*.** The sample size was set to 24 participants. I had attempted to keep the sample size consistent across studies, but mistakenly thought I had tested 24 participants in Experiment 1.[10] The 24 participants had a mean age of 20.0 years (SD = 1.8 years), and 17 were female (7 male).

***Stimuli*.** Because the durations were longer in Experiment 2, I increased the maximum value of the VAS such that it ranged from 0 to 6 seconds, with 0, 1.5, 3, 4.5, and 6 s explicitly marked. Otherwise, the VAS was presented exactly as in Experiment 1.

[10] It turns out that I mixed up the sample size from Experiment 1 with that of an unrelated experiment that had actually had a sample size of 24. A careful reader will probably note that the p-value for the main test of interest in Experiment 2 (i.e. the interaction between Concurrent Timing and Session Type) is right on the border of significance. Combined with the fact that I slightly increased the sample size from Experiment 1 to 2, it would be fair if this were to raise some eyebrows. For this reason, I would like to emphasize that the change in sample size truly was an honest mistake and was not done in an attempt to get a significant interaction. (If I had really wanted to be deceitful, I would have retroactively increased the sample size of Experiment 1 as well to avoid suspicion!)

***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. This change was implemented by generating the durations in the same way as in Experiment 1, then simply multiplying each one by 2.5. Second, because each trial took longer in Experiment 2, I reduced the number of blocks for each duration type from six to four, following an ABBA pattern of Estimation and No Estimation blocks.

Results.

Data Preprocessing. One participant achieved below 70% accuracy on the Equation Verification task and was therefore removed from analysis and replaced. Thus 25 participants were tested in total to reach the desired sample size of 24. As in Experiment 1, trials were removed from analyses (except for the analysis of Equation Verification task accuracy) if they had either incorrect Equation Verification responses (12.9% of trials) or correct responses that were more than 2.5 standard deviations below (0.0%) or above (2.0%) a participant's mean for a given condition.

Equation Verification performance. Equation verification results are shown in Figure 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, $F(1,23) = 2.02$, $MSE = 38529.6$, $p = .169$, $\eta_p^2 = .081$, nor of Session Type, $F(1,23) = 0.03$, $MSE = 107999.0$, $p = .855$, $\eta_p^2 = .001$. The Session Type × Concurrent interaction was significant, $F(1,23) = 4.61$, $MSE = 35953.4$, $p = .043$, $\eta_p^2 = .167$, 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, $t(23) = 0.52$, $SE = 50.0$, $p = .605$, $d_{rm} = 0.03$, but there was during the Response Time session, $t(23) = 2.30$, $SE = 60.9$, $p = .031$, $d_{rm} = 0.16$, 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 Fs ≤ 2.55 (all *ps* $\geq .123$).

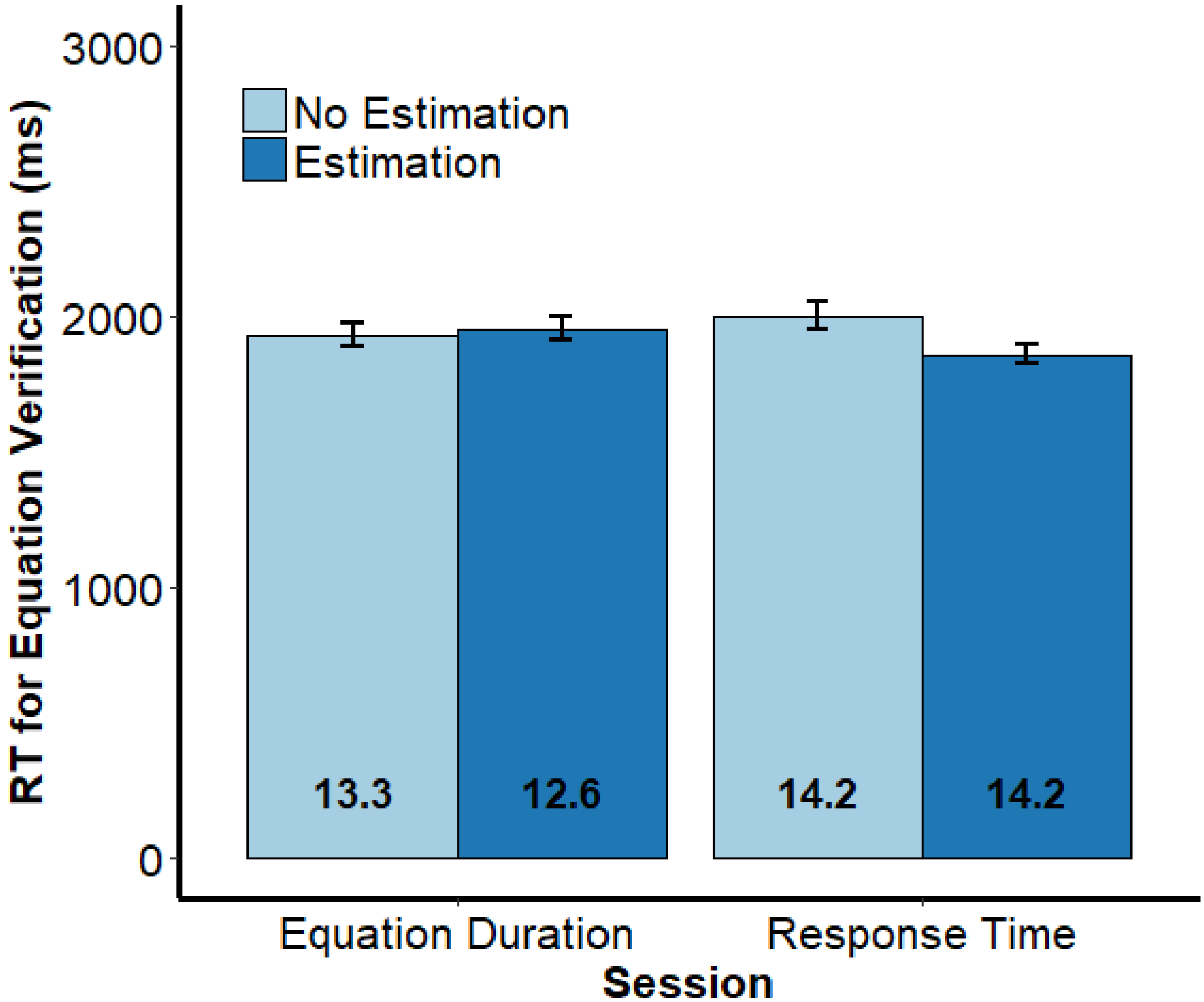

Figure 2. Response time and error percentage to the Equation Verification task in Experiment 2. The numbers within each bar show the error percentage for each condition. Error bars represent ±1 standard error, adjusted to reflect only within-subject variance (Cousineau, 2005).

Duration Estimation performance.

Mean of Estimates. Based on the raw data, Response Times (1861 ms, SE = 171 ms), were shorter than Equation Durations (2857 ms, SE = 15 ms), $t(23) = 5.60$, $SE = 177.7$, $p < .001$, $d_{rm} = 1.90$. Of the 24 participants, 19 of them met the criterion of having at least 10 matched pairs of trials (mean for those 19 participants = 27.4 pairs, SD = 10.3), and amongst this set of trials, Response Times (2457 ms, SE = 87 ms) and Equation Durations (2460 ms, SE = 85 ms) had about the same average duration. Estimates for the two duration types differed significantly, $t(18) = 3.44$, $SE = 112.8$, $p = .003$ (.001 to .021), $d_{rm} = 0.66$ (0.46 to 0.67), with shorter estimates for RTs (2026 ms, SE = 146 ms) than for Equation Durations (2414 ms, SE = 118 ms).

Correlations. The correlations between actual and estimated RTs (0.58, SE = 0.03) did not differ significantly from the correlations between actual and estimated Equation Durations (0.55, SE = 0.05), $t(18) = 0.64$, $SE = 0.05$, $p = .531$ (.153 to .959), $d_{rm} = 0.16$ (0.01 to 0.48).

Discussion. I expected that the results of Experiment 2 would either match those of Experiment 1, with an interaction and an effect of Concurrent Timing in the Equation Duration session (if the Different Processes account is correct) or would not show either of the potential main effects or the interaction (if the Late Locus of Interference account is correct). Neither of these patterns of results occurred. The Late Locus of Interference account correctly predicted that Equation Duration estimation would not slow responses on the Equation Verification task, but incorrectly predicted that there would be no effect of Concurrent Timing during the RT session as well as no interaction between Concurrent Timing and Session Type. Alternately, the Different Processes 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

RT session. This discrepancy suggests that there may have been some aspect of the design of Experiment 2 that had an unintended consequence.

One possibility is that when Equation Durations were reliably longer than RTs, 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 non-temporal information to estimate the duration of the first segment 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 (Matthews, 2013; van Rijn & Taatgen, 2008; Bryce, Seifried-Dübon, & Bratzke, 2015; Bryce & Bratzke, 2015). This strategy would explain the lack of Concurrent Timing effect in the Equation Duration session but cannot explain the effect in the RT 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 in 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 RT trials.

Finally, it is possible that the discrepancy between my predictions and the results was simply due to sampling error. However, if one of the two predictions for Experiment 2 was indeed correct, then the Concurrent Timing effect in the RT session must have been a false positive, which should occur relatively rarely. In Experiment 3, I further tweaked the procedure in an attempt to address these potential issues.

In terms of duration estimation performance, the correlations between actual and estimated durations did not differ between RTs and Equation Durations. This suggests that participants are equally sensitive to changes in the two types of intervals, and conflicts with the Attentional Allocation account of Experiment 1. However, as in Experiment 1, estimated durations were shorter for RTs 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 RTs compared to Equation Durations was not caused by contextual differences and may be instead be a true difference in how the two types of intervals are perceived.

Experiment 3

In Experiment 3, I 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 (or sampling error), 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. The method of Experiment 3 was as described in the General Method section except where noted.

***Participants*.** I originally set the sample size to be 24 participants to match Experiment 2. However, after data were collected, the effect of Concurrent Timing for Short duration trials in the External Duration session was not significant. I believed this to be a Type II error because (a) there is strong reason to believe a priori that this effect exists (i.e. the results of Experiment 1 as well as prior research discussed in the Introduction), and (b) the sample size of 24 participants is quite small, especially to detect an interaction. I therefore collected another sample of 24 participants, resulting in a total sample of 48 participants. I will present the analysis of the full sample in the main text and will present in a footnote the Equation Verification task analysis for each subsample. The 48 participants had a mean age of 20.9 years (SD = 3.3 years), and 28 were female (20 male).

***Stimuli*.** As in Experiment 2, the VAS ranged from 0 to 6 seconds, with 0, 1.5, 3, 4.5, and 6 s explicitly marked.

***Design and Procedure*.** The sequence of experimental blocks was the same as in Experiment 1 (6 blocks of 32 trials each). 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.

Results.

Data Preprocessing. Three participants achieved below 70% accuracy on the Equation Verification task, meaning that 51 participants in total were tested in order to reach the desired sample size of 48. As in the previous experiments, trials were removed from analyses (except for the analysis of Equation Verification task accuracy) if they had either incorrect Equation

Verification responses (14.9% of trials) or correct responses that were more than 2.5 standard deviations below (0.1%) or above (2.1%) a participant's mean for a given condition.

Equation Verification performance. Equation Verification results are shown in Figure 3. I conducted a 2×2×2 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, $F(1,47) = 28.55$, $MSE = 325788.3$, $p < .001$, $\eta_p^2 = .378$, but there was no main effect of Concurrent Timing, $F(1,47) = 2.48$, $MSE = 114638.1$, $p = .122$, $\eta_p^2 = .050$, or of Session Type, $F(1,47) = 1.94$, $MSE = 588046.9$, $p = .170$, $\eta_p^2 = .040$. There was a Session Type × Concurrent Timing interaction, $F(1,47) = 10.14$, $MSE = 62152.4$, $p = .003$, $\eta_p^2 = .177$, and no other significant interactions, all $Fs \leq 2.75$ ($ps \geq .104$). The interaction was driven by an effect of Concurrent Timing during the Equation Duration session, $t(47) = 2.69$, $SE = 49.0$, $p = .010$, $d_{rm} = 0.11$, with slower responses when duration estimation was required, and no effect of Concurrent Timing during the Response Time session, $t(47) = 0.93$, $SE = 34.9$, $p = .357$, $d_{rm} = 0.04$.

Repeating the same analysis using the error data instead of RTs showed that more errors were committed on Short trials than on Long trials, $F(1,47) = 4.69$, $MSE = 0.004$, $p = .035$, $\eta_p^2 = .091$. 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, $Fs \geq 2.44$ ($ps \geq .125$).[11]

[11] In the initial sample, the Session Type × Concurrent Timing interaction, $F(1,23) = 5.77$, $p = .025$, and the Session Type × Concurrent Timing × Equation Length interaction, $F(1,23) = 5.55$, $p = .027$, were both significant. As predicted by the Different Processes account, the Session Type × 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 RT 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, I conducted an exact replication to determine whether the lack of Session Type × Concurrent Timing interaction on Short trials under these conditions was replicable or was simply a Type II error. In the replication, the Session Type × 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 × Concurrent Timing interaction on Long trials) was consistent (and significant) across both samples.

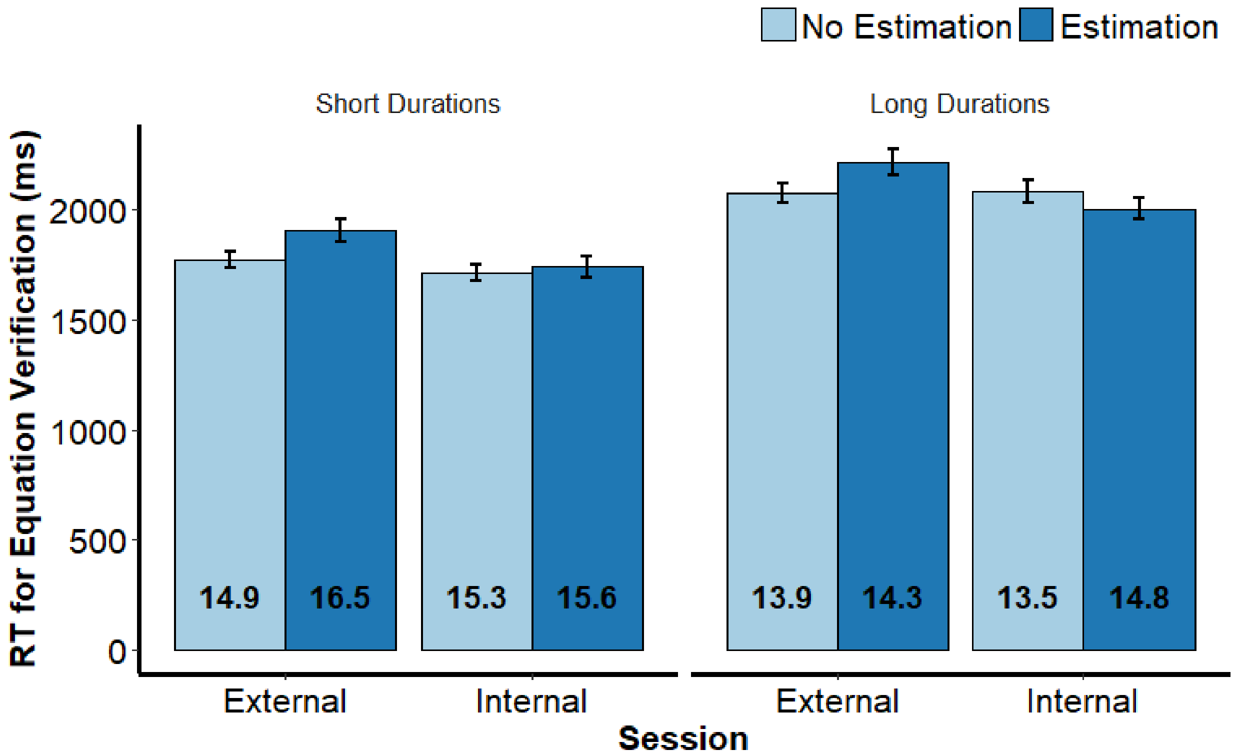

Figure 3. Response time and error percentage to the Equation Verification task in Experiment 3. The data are divided into trials on which the equation duration was drawn from either the pool of short durations or of long durations. The numbers within each bar show the error percentage for each condition. Error bars represent ±1 standard error, adjusted to reflect only within-subject variance (Cousineau, 2005).

***Duration Estimation performance*.**

Mean of Estimates. Based on the raw data, RTs (1879 ms, SE = 121 ms), were not significantly different from than Equation Durations (2023 ms, SE = 15 ms), $t(47) = 1.20$, SE = 120.0, $p = .235$, $d_{rm} = 0.23$. Of the 48 participants, all 48 of them met the criterion of having at least 10 matched pairs of trials (mean = 49.6 pairs, SD = 10.5), and amongst this set of trials, Response Times (1808 ms, SE = 73 ms) and Equation Durations (1808 ms, SE = 73 ms) had the same average duration. Estimates for the two duration types differed significantly, $t(47) = 4.39$, SE = 58.0, $p < .001$ (<.001 to <.001), $d_{rm} = 0.48$ (0.46 to 0.48), with shorter estimates for RTs (2004 ms, SE = 79 ms) than for Equation Durations (2258 ms, SE = 75 ms).

Correlations. The correlations between actual and estimated RTs (0.56, SE = 0.03) did not differ significantly from the correlations between actual and estimated Equation Durations (0.56, SE = 0.04), $t(47) = 0.03$, SE = 0.04, $p = .980$ (.804 to .980), $d_{rm} = 0.00$ (0.00 to 0.04).

Discussion. 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 is the result that was predicted from the Different Processes account, according to which estimation of interval duration relies on different types of processing depending on whether the interval is demarcated internally (e.g. RTs) or externally (e.g. a stimulus on a computer screen).

The duration 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.

Analysis of All Data

My goal for the experiments presented so far was to test the hypothesis that duration estimation would interfere with performance on a concurrent non-temporal task for external intervals but not for internal intervals. For the specific pair of tasks that were used, this hypothesis appears to have been confirmed: pooling across all RT data I have collected so far using this experimental design (111 participants across Experiments 1-3 plus a pilot experiment), there is a significant interaction between the factors of Concurrent Timing and Session Type, $F(1,110) = 25.43$, $MSE = 29083.7$, $p < .001$, $\eta_p^2 = .188$ (see Figure 4). The cause of the interaction is that participants were slower to make Equation Verification responses when they were required to also estimate the duration of the equation, $t(110) = 3.55$, $SE = 30.2$, $p = .001$, $d_{rm} = 0.11$, but not when they were required to estimate their RT for the Equation Verification task, $t(110) = 2.42$, $SE = 23.1$, $p = .017$, $d_{rm} = 0.07$, in which case they were actually faster to respond compared to a single task condition, mainly due to the surprising result of Experiment 2. Furthermore, there is no hint of an interaction in the error rate data, $F(1,110) = 0.01$, $MSE = 0.002$, $p = .925$, $\eta_p^2 = .000$, confirming that the pattern in the RT data is not the result of some speed-accuracy trade-off. In the next section I will summarize the evidence for and against a few of these interpretations.

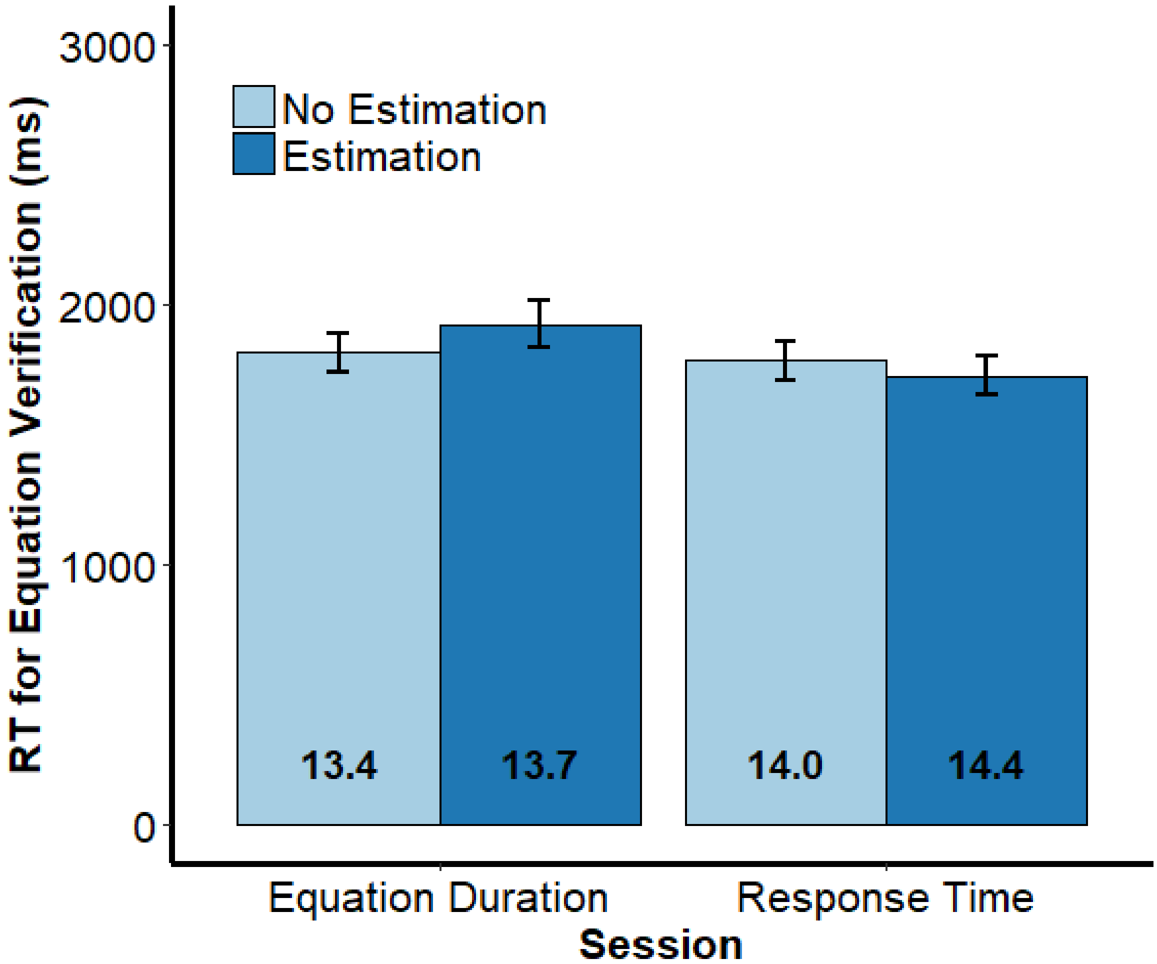

Figure 4. Response time and error percentage to the Equation Verification task for all participants in Experiments 1-3 plus a pilot experiment. The numbers within each bar show the error percentage for each condition. Error bars represent ±1 standard error, adjusted to reflect only within-subject variance (Cousineau, 2005).

Late Decisional Stage Interference. I argued earlier that a potential explanation for the results of Experiment 1 was that the cause of interference between the Equation Verification and VAS tasks could be a stage of the VAS task wherein participants must compare the temporal information they have gathered throughout the interval to some representation of duration in memory in order to come to a decision about their estimated duration. I further argued that if this were the case, the interaction between Concurrent Timing and Session Type, as well as the effect of Concurrent Timing within the Equation Duration session, should both disappear when Equation Durations were set to be longer than RTs. Examination of RT data from experiments using long Equation Durations (72 participants across Experiments 2 and 3) reveals both the interaction, $F(1,71) = 16.17$, $MSE = 45072.5$, $p < .001$, $\eta_p^2 = .185$, and the effect of Concurrent Timing, $t(71) = 2.47$, $SE = 40.7$, $p = .016$, $d_{rm} = 0.08$, to be intact. Examination of the error rate data reveals no interaction, $F(1,71) = 0.09$, $MSE = 0.001$, $p = .765$ $\eta_p^2 = .001$, arguing against a speed-accuracy trade-off. Finally, it is clear from Figure 5, which shows the distributions of the difference in Concurrent Timing effect for external versus internal durations, that the difference in Concurrent Timing effect was not reduced when long, relative to short, Equation Durations were used.

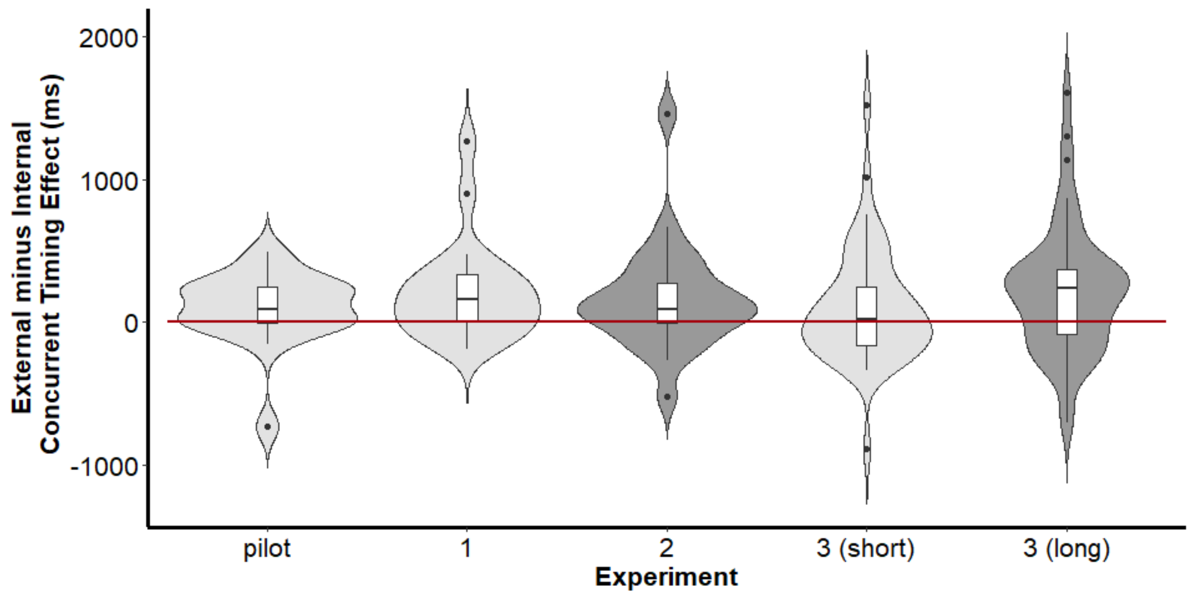

Figure 5. For each experiment, this figure shows the distribution of participants in terms of the difference between their external concurrent timing effect and their internal concurrent timing effect. The light colour represents trials with short equation durations and the dark colour represents trials with long equation durations. Note that the two distributions for Experiment 3 are based on the same set of participants, but a different set of trials.

Differences in Attentional Allocation. Another potential interpretation is that the interaction was simply caused by changes in how participants allocated their attention during the two sessions. The result that Equation Durations were estimated to be longer than RTs was consistent across experiments (see Figure 6a). These longer duration estimates combined with the longer Equation Verification task RTs in the Equation Duration session are consistent with participants allocating relatively more attention to the duration estimation task and relatively less to the Equation Verification task in the Equation Duration session compared to the RT session.

When the full picture is examined, however, it does not seem as consistent with this interpretation. First, the Attentional Allocation account predicts that the Concurrent Timing effect might be smaller for internal intervals, but there is no reason to predict that it should be non-existent, since it should not be the case that more attention is allocated to Equation Verification on internal duration estimation trials than on no duration estimation trials. Second, the second measure of duration estimation performance, the correlation between actual and estimated durations, did not differ between sessions. Although the correlation was nominally higher for Equation Durations in Experiment 1, this potential difference did not pan out in the later experiments (see Figure 6b). If participants had allocated more attention to duration estimation during the Equation Duration session, the correlations likely would have been improved relative to the RT session. That this was not the case suggests that there may instead be a different explanation for the difference in mean duration estimates. One possibility might be that participants felt pressured to respond quickly to the Equation Verification task[12] and therefore predicted that their RTs were, on average, very fast.

[12] That RTs were shorter for short compared to long Equation Duration trials in Experiment 3 seems to suggest that participants did indeed feel pressured to respond relatively quickly.

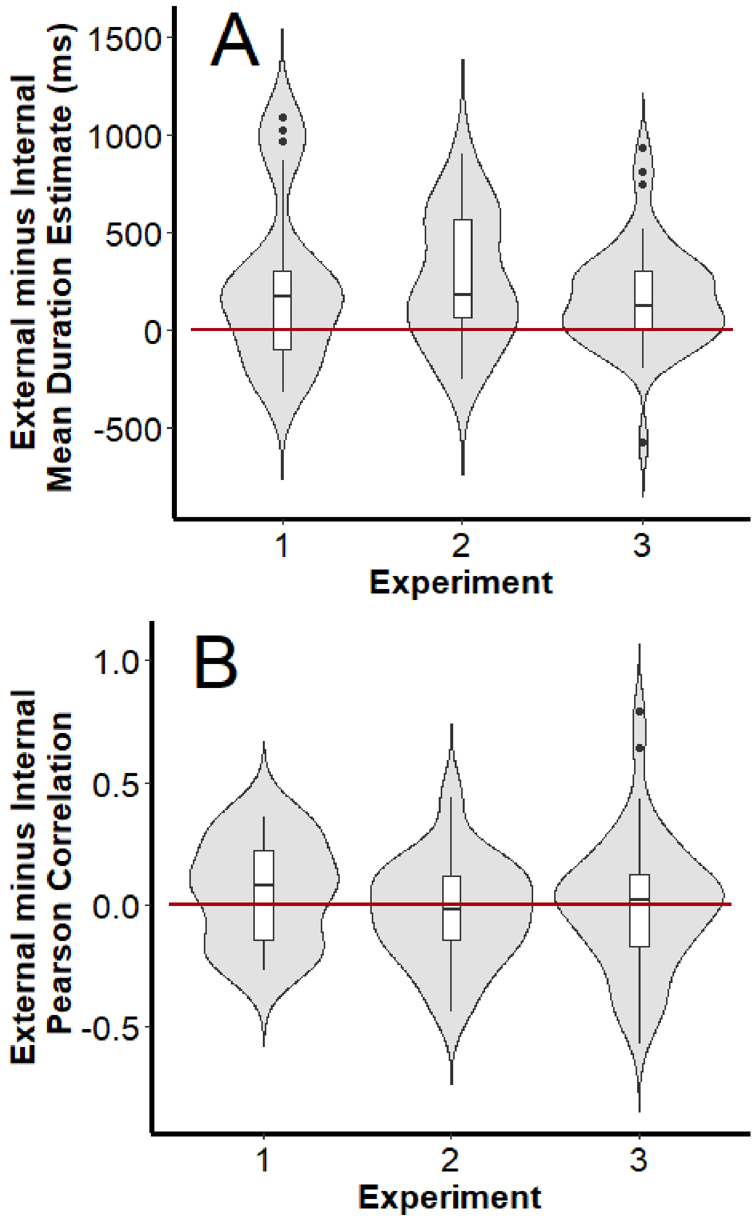

Figure 6. For each experiment, the distribution of (A) the difference in mean duration estimations between external and internal intervals, and (B) the difference in Pearson correlation coefficient for external minus internal intervals.

Subtrahend Magnitude instead of RT Estimation. One shortcoming of the Equation Verification task is that it is easy to deduce a priori which trials will have longer RTs simply by noting that it is likely to take longer to solve problems with larger subtrahends. The Pearson correlation between subtrahend and estimated RT (r = 0.24) was indeed significantly greater than zero, $t(86) = 11.23$, $p < .001$ (Figure 7a). However, when the *actual* RT was partialled out from the correlation (r = 0.02), it was no longer significantly above zero, $t(86) = 0.84$, $p = .404$, suggesting that the bivariate correlation was not caused by participants actively making their estimates based on the subtrahend magnitude, but simply because subtrahend and estimated RT were both related to actual RT. Importantly, the correlation between actual and estimated duration (r = 0.54), was only slightly reduced when subtrahend was partialled out (r = 0.50), $t(86) = 4.22$, $p < .001$.

To further confirm that participants were correctly following the instructions, I computed the partial correlations between actual and estimated durations for internal and external intervals, while controlling for the duration of the *other* interval. For example, I computed the partial correlation between actual and estimated RTs, while controlling for Equation Duration (Figure 7b). Partialling out the duration of the other interval did not significantly affect either internal, $t(86) = 0.77$, $p = .445$, or external, $t(86) = 0.05$, $p = .963$, interval correlations. Interestingly, the partial correlations between estimated and other durations, controlling for the actual durations, were negative for both internal (r = -0.06), $t(86) = 2.09$, $p = .040$, and external (r = -0.06), $t(86) = 2.42$, $p = .018$, trials. This suggests that although participants were indeed estimating the correct duration for both interval types, they either chose not to or were unable to ignore the duration of the other interval. One possibility is that participants adjusted their estimates based on the relative order of the ends of the intervals. For example, if a participant responded slowly on one trial but the Equation Duration for that trial was also quite long, the participant may have reasoned that they had not actually responded slowly because they still made their response before the offset of the equation.

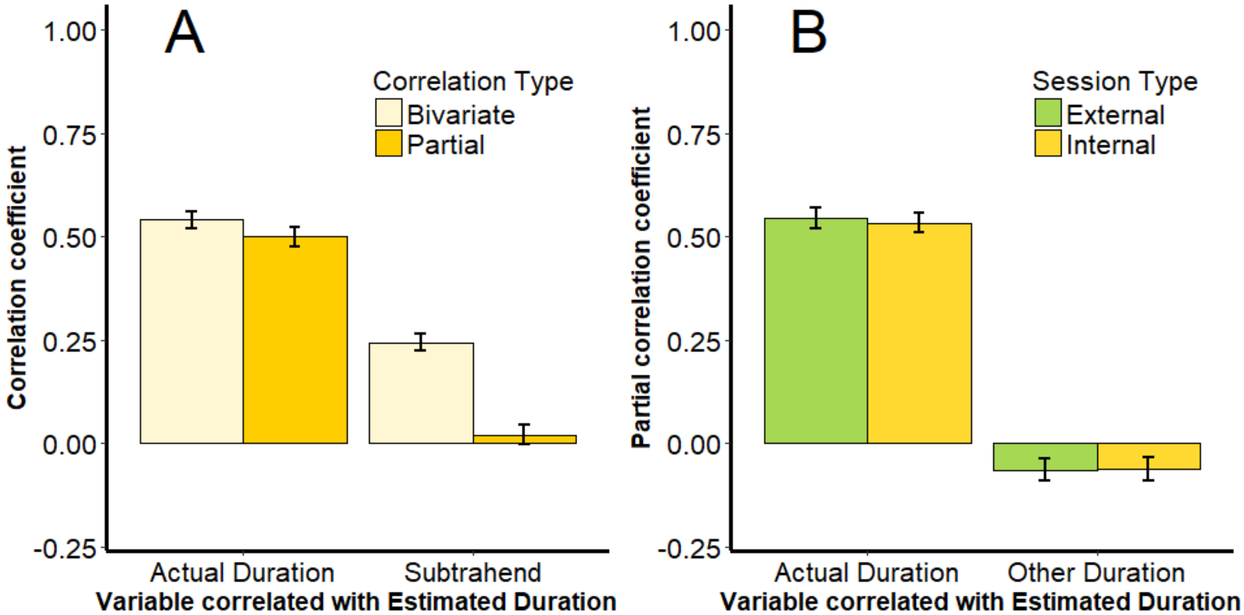

Figure 7. (A) Bivariate and partial correlations between estimated duration and either actual duration or subtrahend magnitude. (B) Partial correlations between estimated duration and either actual or other duration.

Summary

The vast majority of research on duration estimation has focused on ‘external’ intervals (Block & Grondin, 2014) – that is, intervals with durations that are not influenced by our thoughts and actions. When ‘internal’ intervals have been studied, they have generally been treated as being equivalent to external intervals (Corallo et al., 2008; Marti et al., 2010). I argue that the processes involved in the estimation of the two types of durations might differ from one another. Specifically, whereas the estimation of external interval durations largely depends on a type of processing referred to as “temporal processing” (Zakay & Block, 1997), estimation of internal interval durations may instead rely more heavily on non-temporal processing – that is, processing based on non-temporal information. This led me to predict that, whereas estimation of external interval durations should interfere with performance on concurrent non-temporal tasks (because temporal processing is effortful, Brown, 1997), estimation of internal interval durations should interfere either to a lesser degree or not at all (because it could be based on non-temporal information that is collected anyway).

Across three experiments, I tested this prediction using Equation Verification as the non-temporal task and a VAS task for duration estimates. The results of the three experiments confirm the prediction: performance on the Equation Verification task was consistently slowed when participants were required to estimate the screen time of the equation (an external interval), but not when they were required to estimate their own RT (an internal interval). This provides evidence that, at least for the specific tasks and intervals used here, estimation of internal interval duration does not rely on effortful temporal processing. Although there are some relatively reasonable alternative interpretations of the results, they all suffer from moderate to severe shortcomings, as discussed in the preceding sections.

Chapter 3

The three experiments presented in Chapter 2 all used Equation Verification as the non-temporal task and a VAS input as the temporal task. This pair of tasks matches very closely to those used in previous work with internal intervals, most of which have used the VAS (e.g. Corallo et al., 2008; Marti et al., 2010; Bryce & Bratzke, 2014), and some of which have used numerical tasks such as a number-comparison task (e.g. Corallo et al., 2008). It therefore seems likely that the findings from Chapter 2 do apply to this previous work: in duration estimation studies that have used internal intervals, participants probably did not estimate those durations in the same way as they would for external intervals. It is possible that participants in these studies estimated duration using only non-temporal information, such as their feelings of difficulty.

Nonetheless, it is important to determine whether the findings generalize beyond this very specific situation. The VAS estimation task in particular may lend itself to estimation of duration using non-temporal information because it only requires participants to make relative judgements by selecting a relative position within a scale, whereas other tasks force participants to form more explicit representations of duration. It therefore may be easier for participants to make responses based on something like feelings of difficulty (or any other factor that correlates with RT) using a VAS rather than another method such as reproduction. The goal of this chapter will be to test whether the findings from Chapter 2 generalize to (a) non-temporal tasks other than Equation Verification where it is not as easy to deduce a priori how long it should take to perform the task, and (b) temporal tasks other than the VAS where it is not as easy to 'cheat' by inputting an estimate of something other than duration.

Experiment 4

In Experiment 4 I retained the VAS temporal task but changed the non-temporal task to 'Sequence Reasoning' (see Brown & Merchant, 2007; Brown, 2014). In the Sequence Reasoning task, participants are presented with two temporally related phrases and must determine whether they are in the correct order. For example, the pair "receive a PhD" – "write a dissertation" is not in the correct order, because one generally does not receive their PhD until after they have written their dissertation. Brown and Merchant (2007) have argued that the Sequence Reasoning task relies on a similar set of executive processes as duration estimation. More importantly, they (Brown & Merchant, 2007) have shown that at least one type of temporal task, repeated interval production, interferes with performance on a concurrent Sequence Reasoning task. Additionally, unlike Equation Verification, the easiest way to determine how long a particular Sequence Reasoning problem should take is probably just to attempt to solve the problem rather than using some heuristic such as the magnitude of a subtrahend.

Method. The method of Experiment 4 was as described in the General Method section from Chapter 2 except where noted.

Participants. The sample size was set to 24 participants to match that of the previous experiments. The 24 participants had a mean age of 20.0 years (SD = 1.5 years), and 14 were female (10 male).

Stimuli. 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 VAS was the same as the one presented in Experiment 1, ranging from 0 to 4 seconds, with 0, 1, 2, 3, and 4 s explicitly marked.

Design and Procedure. The phrases remained on screen for a random amount of time between 800 and 4200 ms. The sequence of experimental blocks was the same as in Experiment 1, however each block consisted of 18 instead of 32 trials.

Results.

Data Preprocessing. Four participants achieved below 70% accuracy on the Sequence Reasoning task and were therefore removed from analysis and replaced. Thus 28 participants were tested in total to reach the desired sample size of 24. As in the previous experiments, trials were removed from analyses (except for the analysis of Sequence Reasoning task accuracy) if they had either incorrect Sequence Reasoning responses (13.5% of trials) or correct responses that were more than 2.5 standard deviations below (0.0%) or above (2.4%) a participant's mean for a given condition.

Sequence Reasoning performance. Sequence Reasoning results are shown in Figure 8. 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. RT) as factors. Overall, RTs were longer in the Phrase Duration session than in the RT session, $F(1,23) = 6.04$, $MSE = 64097.0$, $p = .022$, $\eta_p^2 = .208$. There was no main effect of Concurrent Timing, $F(1,23) = 2.56$, $MSE = 43822.7$, $p = .123$, $\eta_p^2 = .100$, however Concurrent Timing interacted with Session Type, $F(1,23) = 13.03$, $MSE = 29393.6$, $p = .001$, $\eta_p^2 = .362$. The interaction was caused by participants performing the Sequence Reasoning task more slowly when they were required to concurrently estimate the Phrase Durations, $t(23) = 2.83$, $SE = 68.9$, $p = .010$, $d_{rm} = 0.27$, but not when required to concurrently estimate RT, $t(23) = 1.57$, $SE =$

36.8, $p = .129$, $d_{rm} = 0.16$. Repeating the preceding analysis using the error data instead of RT data revealed no other significant effects or interactions, all $Fs \leq 2.77$ (all $ps > .110$).

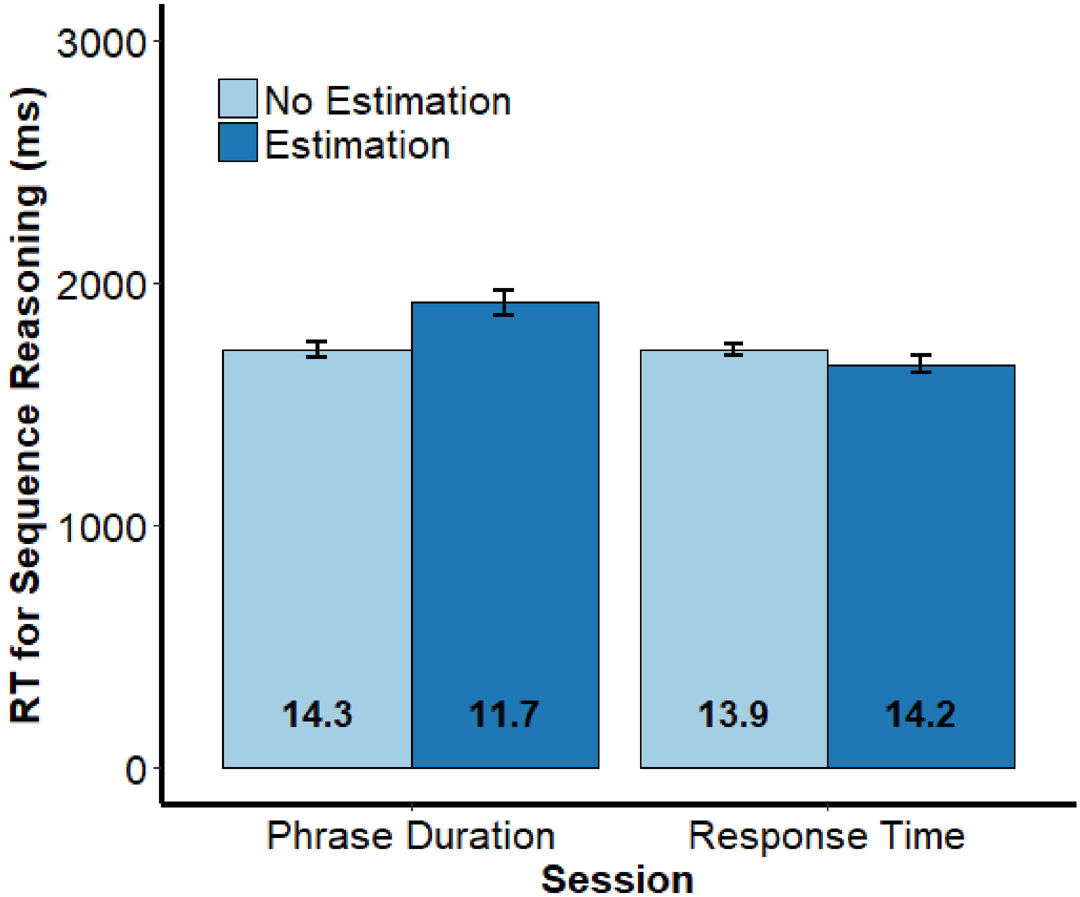

Figure 8. Response time and error percentage to the Sequence Reasoning task in Experiment 4. The numbers within each bar show the error percentage for each condition. Error bars represent ±1 standard error, adjusted to reflect only within-subject variance (Cousineau, 2005).

***Duration Estimation performance*.**

Mean of Estimates. Based on the raw data, RTs (1666 ms, SE = 77 ms), were shorter than Phrase Durations (2398 ms, SE = 58 ms), $t(23) = 7.47$, $SE = 98.0$, $p < .001$, $d_{rm} = 2.20$. 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 = 7.4), and amongst this set of trials, RTs (1815 ms, SE = 53 ms) were about the same as Phrase Durations (1818 ms, SE = 53 ms). Estimates of the two duration types did not differ significantly, $t(22) = 1.56$, $SE = 76.6$, $p = .133$ (.030 to .133), $d_{rm} = 0.42$ (0.42 to 0.67), although estimates of RTs (1660 ms, SE = 49 ms) were nominally quite a bit shorter than those of Phrase Durations (1756 ms, SE = 49 ms).

Correlations. The correlation between actual and estimated RTs (0.57, SE = 0.06) did not differ significantly from the correlation between actual and estimated Phrase Durations (0.62, SE = 0.05), $t(22) = 0.83$, $SE = 0.05$, $p = .418$ (.268 to .721), $d_{rm} = 0.20$ (-0.23 to 0.27).

Discussion. The results of Experiment 4 demonstrate that the results of Chapter 2 extend beyond Equation Verification. Whereas Phrase Duration estimation interfered with concurrent Sequence Reasoning, RT 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 RT estimations on information gathered during the trial.

Also of note is that, for the first time, RTs were not estimated as being (significantly) shorter than the external intervals. It is not clear whether this occurred because of the change in non-temporal task or simply due to chance.

Experiment 5

In Experiment 5 I tested whether the results so far would generalize from a VAS task to a Reproduction task. In the Reproduction task, instead of reporting estimates by clicking within a scale, participants attempted to produce a separate interval with the same duration as the encoded interval.

Method. The method of Experiment 5 was as described in the General Method section from Chapter 2 except where noted.

***Participants*.** The sample size was set to 24 participants to match that of the previous experiments. The 24 participants had a mean age of 19.9 years (SD = 1.5 years), and 19 were female (5 male).

***Stimuli*.** Instead of the VAS, the word “Begin” was displayed in black font at the centre of the screen to prompt participants to begin their reproduced interval. During the reproduced interval, a small fixation cross was presented at the centre of the screen.

***Design and Procedure*.** The sequence of experimental blocks matched that of Experiment 1, however each block consisted of only 28 instead of 32 trials. When participants were prompted to begin their reproduced interval, they did so by pressing the spacebar. Participants also ended their reproduced intervals by pressing the spacebar. As in the previous experiments, participants were then presented with the words "press space" and doing so started the next trial. The Equation Durations were generated using the same method as in Experiment 1.

Results.

***Data Preprocessing*.** Three participants achieved below 70% accuracy on the Equation Verification task and were therefore removed from analysis and replaced. Thus 27 participants were tested in total to reach the desired sample size of 24. As in the previous experiments, trials were removed from analyses (except for the analysis of Equation Verification task accuracy) if they had either incorrect Equation Verification responses (13.2% of trials) or correct responses that were more than 2.5 standard deviations below (0.0%) or above (2.6%) a participant's mean for a given condition.

***Equation Verification performance*.** Equation Verification results are shown in Figure 9. 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, $F(1,23) = 7.93$, $MSE = 16631.5$, $p = .010$, $\eta_p^2 = .256$, with slower responses when duration estimation was required. There was no main effect of Session Type, $F(1,23) = 1.20$, $MSE = 186212.9$, $p = .285$, $\eta_p^2 = .050$, nor was there a Session Type × Concurrent Timing interaction, $F(1,23) = 1.43$, $MSE = 43661.2$, $p = .243$, $\eta_p^2 = .059$. Repeating the analysis using error percentage instead of RTs did not reveal any significant main effects or interactions, all $Fs \leq 1.92$ (all $ps \geq .179$).

To confirm the lack of interaction, I tested an additional 20 participants, repeating the same analysis on this new sample (results shown in Figure 10). In the new sample, neither the main effects, nor the interaction, were significant, all $Fs \leq 0.95$ (all $ps \geq .341$). In the combined sample of 44 participants, there was a main effect of Concurrent Timing, $F(1,43) = 5.18$, $MSE = 35681.7$, $p = .028$, $\eta_p^2 = .107$, no effect of Session, $F < 1$, and no interaction, $F(1,43) = 2.25$, $MSE = 31576.7$, $p = .141$, $\eta_p^2 = .050$. The same analysis using error percentage revealed no significant main effects or interactions for the new sample, all $Fs \leq 0.97$ (all $ps \geq .338$), and the combined sample, $Fs \leq 0.70$ (all $ps \geq .406$).

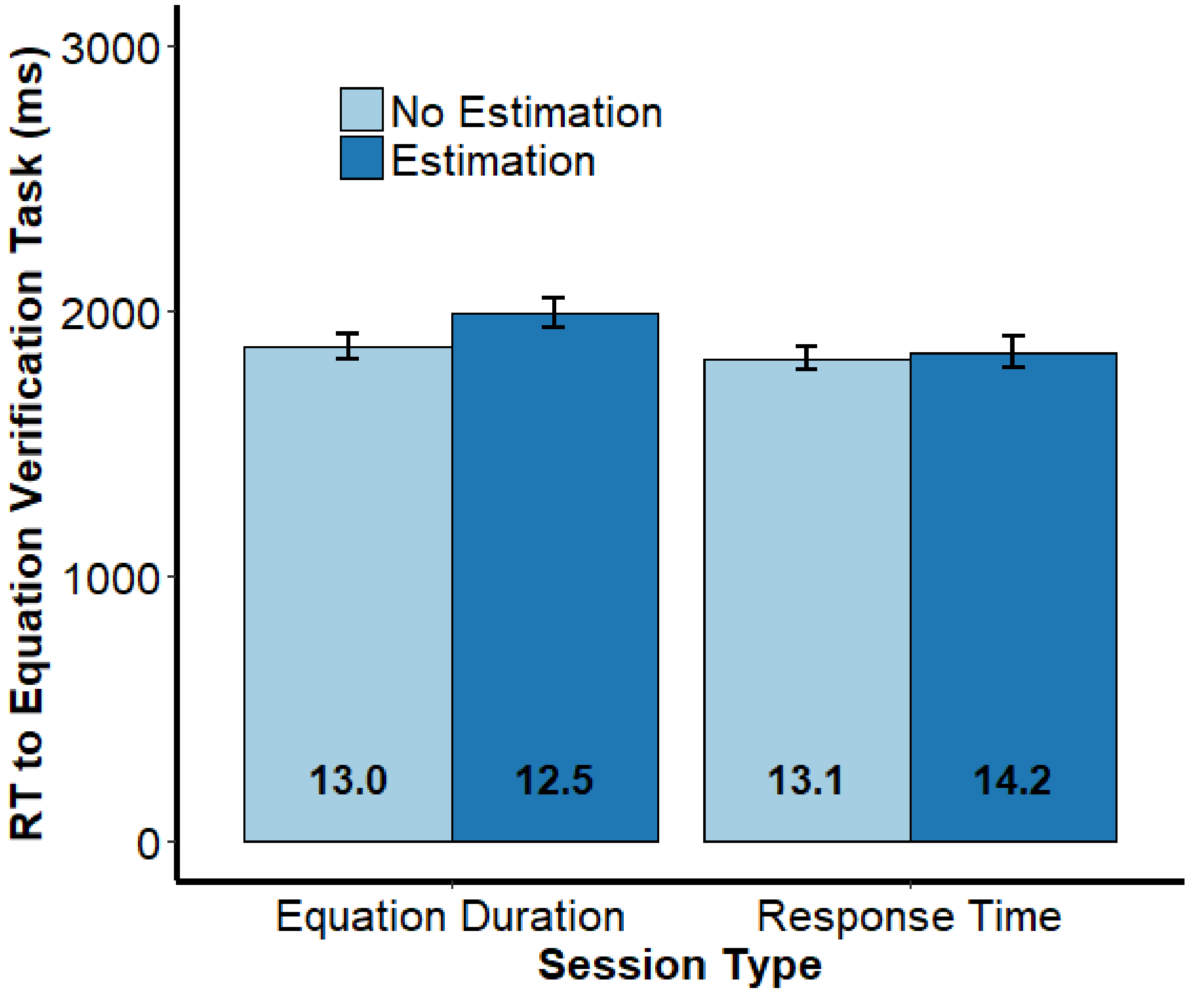

Figure 9. Response time and error percentage to the Equation Verification task in Experiment 5. The numbers within each bar show the error percentage for each condition. Error bars represent ±1 standard error, adjusted to reflect only within-subject variance (Cousineau, 2005).

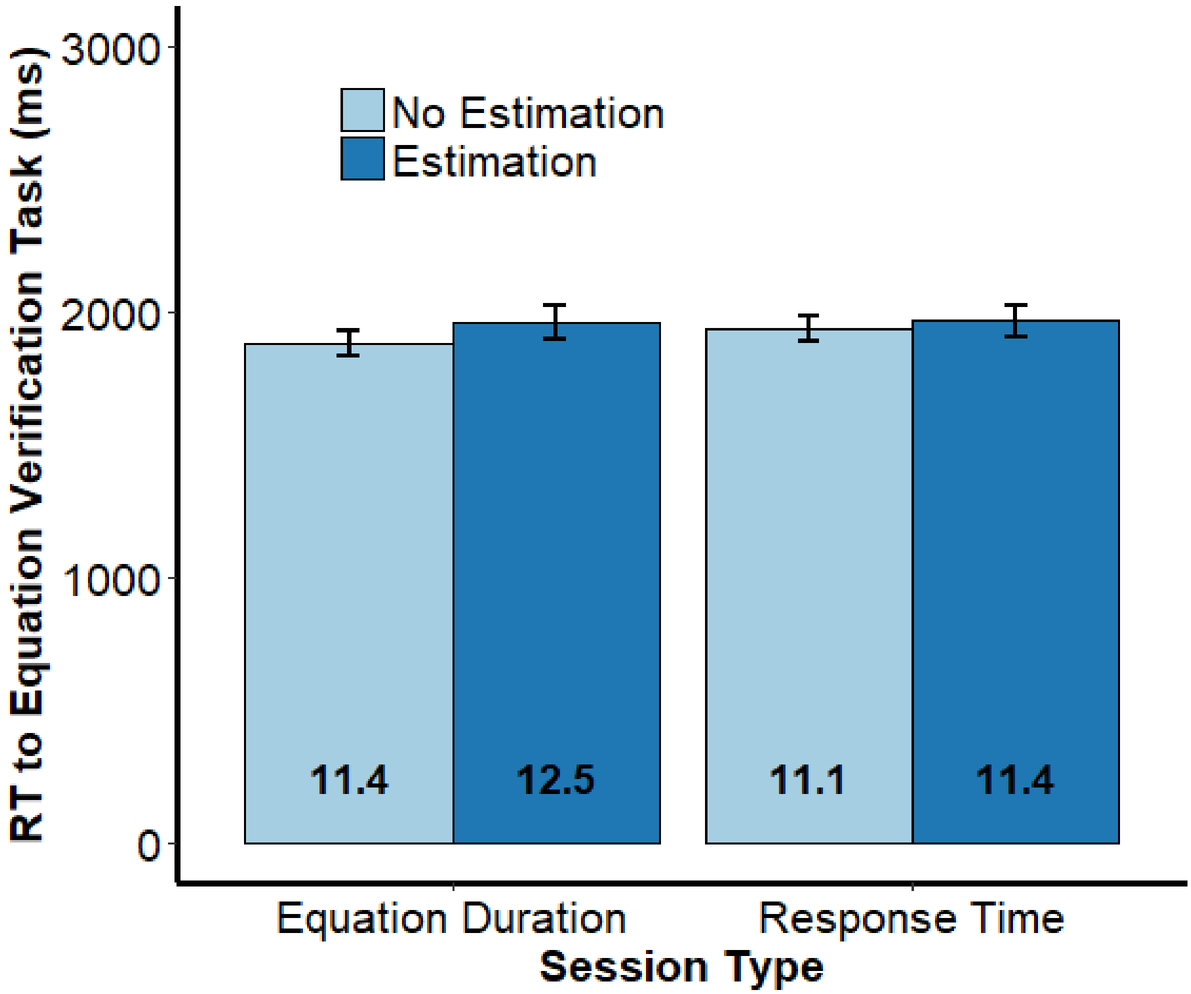

Figure 10. Response time and error percentage to the Equation Verification task for the second sample from Experiment 5. The numbers within each bar show the error percentage for each condition. Error bars represent ±1 standard error, adjusted to reflect only within-subject variance (Cousineau, 2005).

***Duration Estimation performance*.**

Mean of Estimates. Based on the raw data, RTs (1818 ms, SE = 136 ms), were longer than Equation Durations (1150 ms, SE = 4 ms), $t(23) = 4.91$, $SE = 136.1$, $p < .001$, $d_{rm} = 1.45$. All 24 of the participants met the criterion of having at least 10 matched pairs of trials (mean = 36.8 pairs, SD = 12.4), and amongst this set of trials, RTs (1280 ms, SE = 31 ms) were about the same as Equation Durations (1275 ms, SE = 31 ms). Estimates of the two duration types differed marginally, $t(23) = 1.92$, $SE = 92.7$, $p = .068$ (.065 to .242), $d_{rm} = 0.43$ (0.31 to 0.44), with shorter estimates for RTs (1347 ms, SE = 68 ms) than for Equation Durations (1524 ms, SE = 48 ms).

Correlations. The correlations between actual and estimated RTs (0.32, SE = 0.04) did not differ significantly from the correlations between actual and estimated Equation Durations (0.29, SE = 0.04), $t(23) = 0.62$, $SE = 0.04$, $p = .543$ (.543 to .889), $d_{rm} = 0.11$ (0.02 to 0.11).

Discussion. The Equation Verification results from the previous experiments did not extend to the use of a duration Reproduction task. Whereas in previous experiments there was an effect of Concurrent Timing for Equation Durations but not for RTs, here there was a main effect of Concurrent Timing that did not interact with Session Type. This was true for the original sample of 24 participants, the second sample of 20 participants, and the pooled sample of all 44 participants. As discussed in the introduction to this experiment, the lack of interaction may indicate that it is difficult to make accurate duration reproductions using non-temporal information alone, therefore making temporal processing the preferred strategy when a Reproduction response is required. However, even with the additional 20 participants, the lack of significant interaction does not seem to me to be extremely convincing evidence that estimation of internal interval duration differs between the Reproduction and VAS tasks. From examination of Figure 9 (and 10), it does not appear that there was any interference from Concurrent Timing in the RT session, and it seems very unlikely that there would be a three-way interaction if the results of Experiment 5 were compared to those of previous experiments. Because of this, it seems possible that there is a true underlying interaction that the experiment failed to detect. I believe that the best interpretation of these results would be to conclude that they do not provide strong evidence in either direction (i.e. for or against an interaction). A stronger conclusion will need to wait until after Experiment 6.

Experiment 6

In Experiment 6 I sought to directly compare multiple temporal tasks within a single experiment. Given the inconclusive results of Experiment 5, I felt it was important to directly compare results from VAS and Reproduction sessions rather than relying on the comparison of results from separate experiments. In addition to the VAS and Reproduction tasks, I included a Comparison task to further examine the generalizability of the findings of Chapter 2. A second duration was required to accommodate the Comparison task, which requires two durations to be compared to one another, so I took this opportunity to alter the non-temporal task as well, switching it to a ‘Colour Change’ task in which participants were presented with two circles and were asked to identify whether the colour of the second circle was the same as that of the first.

The use of the Colour Change task represents a large step away from the non-temporal tasks used in the previous experiments because it is much easier than the other tasks, with responses that are likely to be much quicker. Whereas the other tasks required deeper thought, potentially with multiple mental operations (e.g. the various steps in solving a math problem), answers on the Colour Change task are more likely to come to the participants without the need for any reasoning. This could mean that there is less useful non-temporal information available for participants to use for RT estimation. Previous work has found that when participants were asked to estimate their RTs to very basic tasks such as simple RT tasks (i.e. just pressing a key when a stimulus appears), they could do so accurately only on trials with outlying RTs (C. MacLeod, personal communication, October 19, 2017). It will therefore be interesting to see whether participants must revert to temporal processing for Colour Change task RTs.

Method. The Experiment 6 method differed substantially from the General Method presented in Chapter 2.

Participants. The sample size of Experiment 6 was set to be 48 participants because there were 48 separate counterbalanced versions of the experiment, as explained below in the Design and Procedure section. In addition to removing participants who achieved less than 70% accuracy on the Colour Change task, I removed those who were RT outliers (i.e. more than 2.5 standard deviations above the mean of all participants) on externally-defined trials of the Colour Change task. I made this decision because some participants actively waited for the stimulus to disappear on these trials before making their Colour Change task response, thereby inflating their RTs specifically in the externally-defined condition.

Stimuli. Throughout the experiment, the background screen colour was grey, rather than white as in the previous experiments. The stimuli for the Colour Change task were two filled circles, each presented one at a time at the centre of the screen. Each circle was either black or white, and had a diameter of approximately 2 cm.

During the VAS blocks, the VAS was the same as in previous experiments except that it ranged from 0 to 2 s, with 0, 0.5, 1, 1.5, and 2 s explicitly labelled.

***Design and Procedure*.** The experiment consisted of seven blocks of 40 trials each. This included an Internal and an External block for each of the three duration estimation tasks (VAS, Reproduction, Comparison), plus a No Estimation block. The counterbalancing was arranged as follows: the Internal and External blocks for each task were always performed back-to-back, with each of the 24 possible orders of the four tasks being performed once with the Internal block first and once with the External block first, for a total of 48 different versions of the experiment.

The experiment began with eight practice trials of the Colour Change task alone and each block, except for the No Estimation block, was preceded by eight trials of practice of the Colour Change task plus the relevant duration estimation task. Every trial began with a small (~1 cm) black fixation cross presented at the centre of the screen for 1s. The fixation cross was replaced by the first circle, which had a 50/50 chance of being either black or white. This circle remained on screen for a period of time equal to the median of the participant's previous five RTs (including practice trials). The screen remained blank for 300-500 ms (the exact duration was drawn uniformly at random) before the second circle appeared. The second circle also had a 50/50 chance of being either black or white, and also had a 50/50 chance of being the same colour as the first circle. Participants indicated whether the colour of the second circle was the same as that of the first circle by clicking either the left or the right mouse button. The duration of the second circle was equal to that of the first circle, multiplied by 0.70, 0.85, 1.15, or 1.30. If participants responded before the offset of the second circle, the duration estimation input screen appeared immediately after offset of the second circle. Otherwise, the screen went blank until the participant responded.

During the VAS phase, participants simply clicked within the VAS as they did in previous experiments. Immediately above the VAS was either the text "Estimate the duration of the last circle" or "Estimate your response time", depending on the block. Participants were instructed before the Internal VAS block that their RT was defined as the time from when the second circle appeared to when they made their mouse-click response. As soon as the participant clicked within the VAS, it disappeared and was replaced by the text "press space". Pressing space ended the trial, leading to the fixation cross for the next trial.

During the Reproduction phase, either the text "Hold down the right mouse button for a duration equal to your response time" or "Hold down the right mouse button for a duration equal to the duration of the last circle" appeared on screen instead of the VAS. Pressing the right mouse caused the text "…timing…" to appear just below the instruction, and the trial ended on release of the right mouse button.

During the Comparison phase, the text "Was your response time shorter or longer than the duration of the first circle?" or "Was the duration of the last circle shorter or longer than the duration of the first circle?" appeared on screen. Just below the text were two white boxes, one containing the word "shorter" and the other "longer". The participants made their responses by clicking within one of the two boxes. The fixation cross for the next trial appeared as soon as one of the boxes was chosen.

Finally, in the No Estimation phase, the fixation cross for the next trial appeared as soon as participants made their Colour Change response (or as soon as the second circle disappeared, if the Colour Change response was made first).

Results.

Data Preprocessing. All participants achieved at least 70% accuracy on the Colour Change task, meaning that none were removed from analyses due to poor accuracy. However, eight participants were removed as RT outliers on the Colour Change Task. That is, 56 participants had to be tested before there were at least 48 remaining after subject-level outlier removal. Trials were removed from analyses (except for the analysis of Colour Change task accuracy) if they had either incorrect Colour Change responses (6.4% of trials) or correct responses that were more than 2.5 standard deviations below (0.0%) or above (2.6%) a participant's mean for a given condition.

Colour Change performance. Colour Change task results are shown in Figure 11. To test whether there were any differences between the seven trial types, I computed a one-way ANOVA on the Colour Change task RTs. The effect of trial type was significant, $F(3.2,151.2) = 23.41$, $MSE = 66208.8$, $p < .001$, $\eta_p^2 = .332$[13], suggesting differences in performance requiring further investigation. Overall, there are nine comparisons of interest: The No Estimation condition vs. each of the six Estimation conditions, and, for each of the three types of Estimation tasks, the Internal vs. External conditions. I used repeated measures t-tests for each of the nine comparisons, using the Dunn-Sidák[14] correction for multiple comparisons.

RTs on the Colour Change task were faster in the No Estimation condition than in any of the three External Estimation conditions: VAS, $t(47) = 6.94$, $SE = 23.0$, $p < .001$, $d_{rm} = 0.88$, Reproduction, $t(47) = 5.28$, $SE = 48.2$, $p < .001$, $d_{rm} = 1.04$, and Comparison, $t(47) = 6.72$, $SE = 45.4$, $p < .001$, $d_{rm} = 1.18$. Additionally, all three Internal Estimation conditions resulted in faster RTs than their External counterparts: VAS, $t(47) = 8.45$, $SE = 23.2$, $p < .001$, $d_{rm} = 1.08$, Reproduction, $t(47) = 3.65$, $SE = 49.3$, $p < .001$, $d_{rm} = 0.67$, and Comparison, $t(47) = 5.90$, $SE = 47.8$, $p < .001$, $d_{rm} = 1.06$. Finally, RTs in the No Estimation condition were faster than in the Internal Reproduction condition, $t(47) = 3.28$, $SE = 22.7$, $p = .002$, $d_{rm} = 0.46$, but did not differ from either the Internal VAS, $t(47) = 2.54$, $SE = 14.5$, $p = .014$, $d_{rm} = 0.37$, or Internal Comparison conditions, $t(47) = 0.94$, $SE = 24.2$, $p = .350$, $d_{rm} = 0.16$.

[13] I applied the Greenhouse-Geisser correction because Mauchly's Test of Sphericity was significant, $\chi^2(20) = 136.0$, $p < .001$.
[14] The Dunn-Sidák correction adjusts the critical p-value to control the family-wise error rate. It is defined as $\alpha_{SID} = 1 - (1-\alpha)^{\frac{1}{m}}$, where α_{SID} is the adjusted critical value, α is the unadjusted critical value, and m is the number of null hypotheses being tested, Šidák, 1967. For m = 9, this results in a critical value of 0.0057.

The procedure was repeated using the Colour Change error rate instead of RTs, finding no overall effect of trial type, $F(6,282) = 0.79$, $MSE = 0.002$, $p = .581$, $\eta_p^2 = .016$.

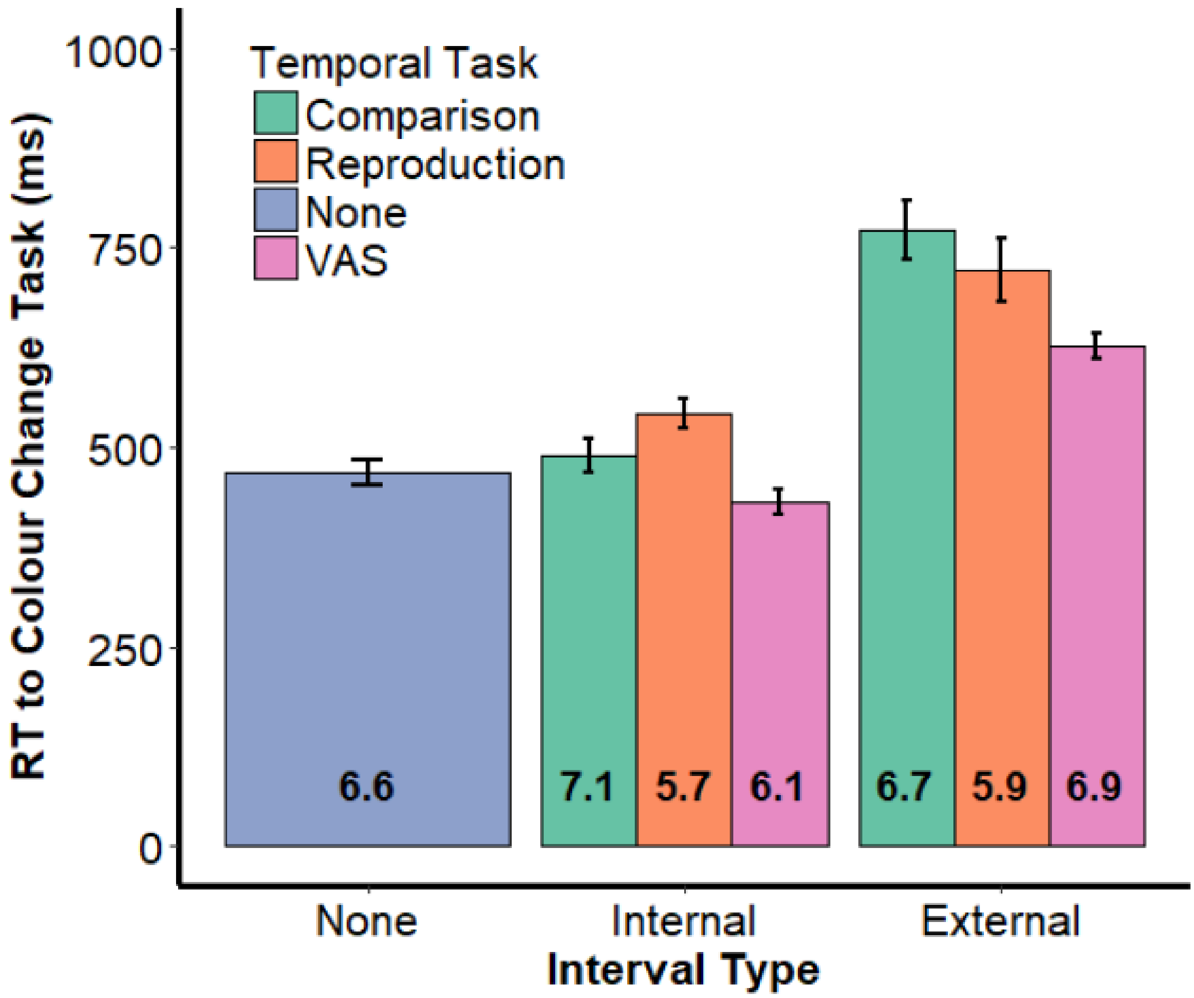

Figure 11. Response time and error percentage to the Colour Change task in Experiment 6. The numbers within each bar show the error percentage for each condition. Error bars represent ±1 standard error, adjusted to reflect only within-subject variance (Cousineau, 2005).

***VAS Task Duration Estimation performance*.**

Mean of Estimates. Based on the raw data, Response Times (431 ms, SE = 13 ms), were shorter than circle durations (621 ms, SE = 28 ms), $t(47) = 8.46$, $SE = 22.5$, $p < .001$, $d_{rm} = 1.05$. Of the 48 participants, 47 met the criterion of having at least 10 matched pairs of trials (mean = 20.9 pairs, SD = 5.1), and amongst this set of trials, Response Times (478 ms, SE = 17 ms) and Circle Durations (479 ms, SE = 17 ms) had about the same average duration. Although estimates of the two duration types did not differ significantly in the main analysis, $t(46) = 1.50$, $SE = 31.2$, $p = .141$ (.011 to .215), $d_{rm} = 0.23$ (0.19 to 0.44), they did differ in two of the nine analyses, with Response Times being estimated as longer than Circle Durations.

Correlations. The correlations between actual and estimated RTs (0.38, SE = 0.04) did not differ significantly from the correlations between actual and estimated circle durations (0.39, SE = 0.05), $t(46) = 0.19$, $SE = 0.04$, $p = .849$ (.584 to .969), $d_{rm} = 0.03$ (0.01 to 0.09).

Reproduction Task Duration Estimation performance.

Mean of Estimates. Based on the raw data, RTs (542 ms, SE = 27 ms), were shorter than circle durations (761 ms, SE = 51 ms), $t(45) = 4.12$, $SE = 53.1$, $p < .001$, $d_{rm} = 0.76$. Of the 48 participants, 46 met the criterion of having at least 10 matched pairs of trials (mean = 22.8 pairs, SD = 4.8), and amongst this set of trials, RTs (556 ms, SE = 21 ms) and circle durations (558 ms, SE = 21 ms) had about the same average duration. Estimates of the two duration types did not differ significantly, $t(45) = 1.30$, $SE = 35.9$, $p = .202$ (.202 to .386), $d_{rm} = 0.16$ (0.07 to 0.16).

Correlations. The correlations between actual and estimated RTs (0.19, SE = 0.04) were significantly lower than the correlations between actual and estimated circle durations (0.36, SE = 0.05), $t(45) = 3.29$, $SE = 0.05$, $p = .002$ (.001 to .086), $d_{rm} = 0.61$ (0.33 to 0.67).

Comparison Task Duration Estimation performance.

I modified the matching procedure slightly to accommodate the Comparison Task. Instead of matching on duration, I matched the trials based on the ratio of duration (either the 2nd circle duration or the RT) to the duration of the first circle. Unfortunately, because there were only a few unique ratios for External trials (i.e. the ones I set in the Method: 0.70, 0.85, 1.15, and 1.30), the results of the matching procedure were relatively volatile from run to run. For this reason, I repeated the procedure 100 times and calculated the mean of the summary statistics (i.e. means and correlations) across the 100 runs for each participant.

Mean of Estimates. Based on the raw data, RT ratios (1.06, SE = 0.01), were larger than circle ratios (exactly 1.00), $t(47) = 6.02$, $SE = 0.01$, $p < .001$, $d_{rm} = 1.12$. Of the 48 participants, all 48 met the criterion of having at least 10 matched pairs of trials (mean = 22.1 pairs, SD = 3.4), and amongst this set of trials, RT ratios (0.99, SE = 0.01) and circle duration ratios (0.98, SE = 0.01) were about the same on average. Overall, participants were much more likely to make 'longer' responses for circle duration ratios, $t(47) = 2.35$, $SE = 0.04$, $p = .023$ (<.001 to .067), $d_{rm} = 0.49$ (0.41 to 0.83)[15].

Correlations. The correlations between RT ratios and proportion of 'longer' responses (0.26, SE = 0.04) were significantly lower than the correlations between circle duration ratios and proportion of 'longer' responses (0.42, SE = 0.04), $t(47) = 2.78$, $SE = 0.06$, $p = .008$ (.002 to .012), $d_{rm} = 0.57$ (0.57 to 0.76)[16].

Discussion. The key questions in Experiment 6 were whether, for each of the three temporal tasks, there would be (a) an effect of Concurrent Timing for external intervals (i.e. slower Colour Change task performance on External Estimation trials than on No Estimation trials), (b)

[15] I excluded the "20 pairs within 5% analysis" because only 3 participants met these criteria.
[16] Again, I excluded the "20 pairs within 5% analysis" because only 3 participants met these criteria.

an effect of Concurrent Timing for internal intervals (i.e. slower Colour Change task performance on Internal Estimation trials than on No Estimation trials), and (c) a *larger* effect of Concurrent Timing for external compared to internal intervals (i.e. slower Colour Change task performance on External Estimation trials than on Internal Estimation trials). Finding that Colour Change performance was poorest on External Estimation trials (i.e. answering "yes" to questions a and c) would be roughly equivalent to finding the Concurrent Timing × Session Type interaction in previous experiments.

For the VAS task, the answers to these three questions are "yes", "no", and "yes". That is, the requirement to estimate the duration of the second circle using a VAS slowed performance on the Colour Change task relative to both No Estimation and Internal Estimation trials, whereas the requirement to estimate RT to the Colour Change task did not have a significant impact on performance relative to No Estimation trials. This finding is in line with the results of Experiments 1-4 and extends those results to a third non-temporal task.

Of note is that, relative to both Equation Verification and Sequence Reasoning, the Colour Change task seems to be a very basic task that requires little in the way of in-depth cognitive processing. Indeed, whereas RTs were about two seconds on average in the previous experiments, they averaged only about half a second in Experiment 6. Furthermore, based on my own experience of having performed the Colour Change task, I suspect that, had I asked participants how they came up with their responses, they would provide answers such as "it just came to me" or "the answer was obvious – it was just a different colour". Despite this, participants seem to have been able to accurately estimate their RTs while avoiding the effortful temporal processing that was required to estimate the circle duration. One key difference between the Colour Change task and the simple RT task that has been used in previous work (i.e. C. MacLeod, personal communication, October 19, 2017), is that, although both tasks are very simple compared to Equation Verification, the Colour Change task requires at least some amount of processing between stimulus perception and response execution (i.e. a response selection phase), whereas the simple RT task does not. One might speculate that the non-temporal information generated during this response selection phase, even when it is very easy to select the correct response, plays a key role in RT estimation.

The duration estimation results for the VAS task actually differed slightly from most previous experiments, with RTs and circle durations actually being estimated as being equivalent (except for on two of the nine runs of the matching procedure, for which RTs were actually estimated to be longer than circle durations). If RT estimates are based largely on non-temporal information instead of temporal processing, it would not be surprising to see these estimates vary widely between tasks. Given that RTs were reliably estimated to be shorter than equation durations in experiments that used the Equation Verification task, it seems that that task may lend itself to relatively shorter RT estimates. As I speculated in a previous section, one potential cause for this might be that, because participants were not very comfortable performing math-related tasks, they felt that they were responding more quickly than they were comfortable with and therefore estimated that their RTs were very short.

For the Comparison task, Colour Change task performance followed the same pattern of results as for the VAS task: the requirement to estimate whether the duration of the second circle was longer than that of the first circle slowed performance on the Colour Change task relative to both the No Estimation and Internal Estimation conditions, whereas the requirement to estimate whether the participants’ RT to the Colour Change task was longer than the duration of the first circle had no effect on Colour Change performance relative to the No Estimation condition. This finding demonstrates that the results of Experiments 1-4 do generalize beyond only a VAS temporal task.

That the Comparison task did not cause interference for internal intervals may be considered surprising given that it is much more difficult than the VAS task; it requires not only that participants have a rough idea of how long they took to perform the non-temporal task, but also that they can accurately compare this duration to that of the first circle. If participants estimated their RT for the Colour Change task using fundamentally different processes than they used to estimate the duration of the first circle, then one would expect Comparison performance to be poorer in the internal condition compared to the external condition in which participants presumably engaged similar processes for estimation of both circles. That is indeed what was found, with lower correlations between the duration ratio and Comparison task response for internal relative to external trials. Although this would seem to provide support for the Different Processes account, it is important to note that a proponent of a Same Processes account would probably make the same prediction. Modality plays such an important role in our perception of duration that it has been suggested that there are different internal ‘clocks’ for different modalities (Wearden, Edwards, Fakhri, & Percival, 1998). One would therefore expect that, even if circle durations and RTs were both estimated using temporal processing, it would be easier to compare the durations of two circles than to compare the duration of a circle to a RT.

The final temporal task, Reproduction, produced a slightly different pattern of results. Like the other temporal tasks, the requirement to reproduce the duration of the second circle resulted in slowed Colour Change task performance relative to both No Estimation and Internal estimation trials. Unlike the other temporal tasks, the requirement to reproduce one’s RT to the Colour Change task also slowed performance on this task relative to No Estimation trials. This may help to explain the ambiguous result of Experiment 5: although there does seem to be interference from both internal and external interval reproduction, the effect is much smaller for internal than for external intervals.

The Reproduction condition of Experiment 6 produced the first result so far that can be fully explained by the Attentional Allocation account that I introduced earlier. That is, although Colour Change performance was better on Internal Estimation trials than on External Estimation trials, it was best on No Estimation trials. Furthermore, both measures of duration estimation performance were consistent with less attention being allocated to duration estimation in the Internal condition, with both shorter reproductions and lower correlations between actual and reproduced durations for internal relative to external intervals. It therefore seems possible that the same processes are involved in reproduction of both internal and external intervals. Why this might be the case for the Reproduction task but not the Comparison task will be discussed in the next chapter.

Chapter 4

The aim of the present research was to explore the possibility that duration estimation may rely on different types of cognitive processes depending on whether the interval is externally-defined or internally-defined. This question is of key importance because despite the significance of internal durations in day-to-day life – to interact effectively with the world we must be able to estimate *both* internal and external durations – researchers have focussed almost exclusively on external intervals and have generally assumed that there is no difference in how the durations of the two types of intervals are estimated. In this chapter, I will discuss the ways in which these processes might differ, the evidence provided by the experiments presented in earlier chapters, and some alternative interpretations of this work.

The most widely-accepted type of model of (external) duration estimation is the pacemaker-accumulator model (e.g. Zakay & Block, 1995). According to this type of model, we can estimate durations by taking advantage of signals (i.e. 'pacemakers'), usually within our brains (see Buhusi & Meck, 2005), that occur at regular cycles. Because these cycles are regular, we can use the number of cycles that have passed as a proxy for the amount of time that has passed. It is generally thought that attention must be directed toward the source of the signal throughout the interval, meaning that duration estimation is an attention-demanding task (Zakay & Block, 1995; Hicks, Miller, & Kinsbourne, 1976; Brown, 1997). It has also been argued that the ongoing count of cycles (i.e. in the 'accumulator') must be effortfully maintained in working memory for the duration of the interval (Casini & Macar, 1997; Fortin & Massé, 2000).

It is possible that the durations of internal intervals are estimated in the same way. Returning to the example of reading a paragraph of a journal article, one could estimate the amount of time required to read the paragraph by dividing attention between the reading task and the passage of time while maintaining the accumulated amount of time that has passed in working memory. Unfortunately, this would probably result in both poor reading comprehension and a poor estimate of the amount of time passed. I suggest that, instead, one might estimate how long they took to read the paragraph after the fact, using non-temporal information gathered throughout the interval. How smoothly did your eyes move across the page? How many difficult words were there? Did you have to re-read any sections? All of this information, along with our basic knowledge of how long it takes, on average, to read a paragraph, should be enough to develop a relatively accurate estimate of how long it took to read this particular paragraph, without the need to engage attention-demanding temporal processing that could interfere with the reading of the paragraph itself.

The general approach I have taken within this dissertation has been to examine the effect of duration estimation in dual-task situations, using the logic that, if duration estimation for external intervals does indeed require attention, then it should interfere with concurrently performed tasks that also require attention, whereas if the duration of internal intervals can be estimated while avoiding this attention-demanding temporal processing, then it should not necessarily interfere with concurrent tasks.

Across six experiments, this prediction was generally confirmed. Experiments 1 through 3 show that estimating the duration of an equation by clicking within a VAS slowed verification of whether that equation was correct, relative to single task Equation Verification. This was not the case when participants were required to estimate their own RT to the Equation Verification task, in which case they performed just as well as on single task trials. This finding appears to hold both when equation durations are shorter than RTs and when they are longer. In Experiment 4, this finding was extended from the Equation Verification task to a Sequence Reasoning task.

Participants were slower to identify the correct ordering of two short phrases when required to estimate the screen-time of the phrases, relative to a single task condition, but estimating their own RT to the Sequence Reasoning task had no such effect. Experiment 6 demonstrated that this result further generalizes to a much easier Colour Change task in which participants judged whether the colour of a circle was the same as that of a previous circle. Estimating the duration of the circle using a VAS slowed responses to the Colour Change task, whereas estimating RT had no effect. Furthermore, the same was true when, instead of making their estimates using a VAS, participants were asked to compare the duration to that of the first circle. That is, judging whether the second circle was longer or shorter in duration than the first circle slowed Colour Change performance, but judging whether the RT to the Colour Change task was longer or shorter than the duration of the first circle did not affect performance.

These results did *not* generalize to a duration Reproduction task, in which participants were required to reproduce either the duration of the stimulus (i.e. the second circle or the equation) or their RT. In Experiment 4, the requirement to reproduce an interval slowed responses to the Equation Verification task, and this effect did not interact with Session Type. In Experiment 6, duration Reproduction also slowed performance on the Colour Change task, although the amount of slowing was greater when participants reproduced the duration of the second circle than when they reproduced their RTs.

There are many ways in which the VAS and Reproduction tasks differ. One of the most important is that Reproduction, in addition to relying on temporal processing, relies on the participant’s ability to accurately control the motor system. Differences between duration estimation by Reproduction and other means (such as clicking within a VAS) can therefore be attributable either to differences in a participant's perception of duration or to differences in their ability to produce an interval matching that perception (e.g. Droit-Volet, 2010; Wing & Kristoffersson, 1973). Another key difference is that 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). Unfortunately, neither of these differences can explain my results because both occur during the Reproduction itself whereas the difference I observed, that is, the interference (or lack thereof) with a non-temporal task, occurred during the encoding of the interval. Instead, it may be the case that it is simply too difficult or effortful to make accurate temporal reproductions on the basis of non-temporal information (e.g. feelings of difficulty). Whereas the VAS 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. That is, although it is clear how non-temporal information can be used to make accurate estimates using a VAS, it is much less clear how this non-temporal information can be compared to the information stored in an accumulator during the reproduced interval. It was presumably difficult enough to accurately compare these two types of information that participants may have resorted to temporal processing even on RT estimation trials.

It may seem to be a surprise that the results of the Comparison task matched those of the VAS rather than those of the Reproduction task. On the surface, it seems that Comparison is more similar to Reproduction than to the VAS task, given that the Reproduction task is essentially a comparison between the original and the reproduced durations. If it was difficult for participants to compare the non-temporal information from their RT to the accumulated temporal information from their reproduced interval, it follows that it should also be difficult to compare the non-temporal information from their RT to the accumulated temporal information from the first circle interval. I do not have a particularly strong explanation for this discrepancy, although there are a few possible causes. For example, it could be that participants ‘standardized’ both durations before comparing them, using reasoning such as “the duration of the first circle seemed longer than usual and my response time seemed pretty fast, so I’ll make a ‘shorter’ response”. Another possibility is that participants attempted to make their reproductions using some other strategy such as trying to ‘play back’ the events that occurred during the interval, which would presumably force them to put effort into encoding those events. Finally, it is possible that there is not actually a difference between the two temporal tasks. The difference in Colour Change task RT between external and internal interval trials actually was not significantly different between the two temporal tasks, $t(47) = 1.40$, $SE = 72.8$, $p = .167$, $d_{rm} = 0.30$. Although these may or may not be convincing explanations, it is also not clear how a ‘Same Process’ account would explain differing effects of these tasks. In the following sections, I will outline and discuss some potential alternative interpretations to the present work.

Attentional Allocation

As discussed earlier, one potential ‘Same Process’ interpretation of the results is that they were caused by differences in how participants allocated their attention between the tasks, with more attention allocated to duration estimation on external trials relative to internal trials. This account can only fully explain the Reproduction condition in Experiment 6, in which Colour Change performance was better for internal than for external interval trials (but still worse than single task trials) and duration estimation performance was worse for internal than for external intervals. In all other conditions and experiments, non-temporal task performance was just as good on internal interval trials as on single task trials, and/or duration estimation was not better for external than for internal trials. The attentional allocation account therefore does not seem to be a generally strong interpretation of the overall results presented here.

Late Decisional Stage Interference

Another interpretation that I discussed earlier is that the cause of interference between the temporal and non-temporal tasks was a late decisional stage of the temporal tasks rather than the continual need to direct attention toward the passage of time. This does not seem to be a strong interpretation of the results for two reasons. First, Concurrent Timing interfered with Equation Verification even on Experiment 3 ‘Long’ trials when the late decisional stage generally should have occurred after Equation Verification responses had been made and therefore could not have interfered with these responses. Second, to accept this interpretation, one must abandon the widely accepted view that duration estimation requires attention throughout the entirety of the to-be-timed interval.

Expectancy of Interval Offset

A different potential interpretation is that the difference between duration estimation of internal and external 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 toward the passage of time and away from concurrent non-temporal processing.

If, in my experiments, interval-offset expectancy drew attention away from the non-temporal task, it may have produced the observed pattern of results, with high expectancy (and therefore interference) early on when there was a chance that the interval might end early (all experiments except for Experiment 2), but no expectancy (and no interference) until later if the intervals were generally long (Experiment 2). There might have been no interference on internal 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 duration estimation, but also Equation Verification, Sequence Reasoning, and performance on the Colour Change task. Second, it assumes that estimation of external interval durations would not interfere with performance on the non-temporal tasks in the absence of expectancy, which, as I discussed in the previous section, would suggest that attention does *not* need to be directed toward the passage of time throughout the interval.

A Dedicated RT Estimation Mechanism

A final alternative account is based on the idea of decentralized timing mechanisms. Although most duration estimation studies are interpreted within a single pacemaker 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 depends on different timing mechanisms, my 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 I am proposing (i.e. duration estimation of internal intervals based on non-temporal information that is always collected during mental processing) in that it also proposes separate mechanisms for duration estimation of internal and external intervals. However, unlike what I am proposing, these mechanisms, despite being separate, are not necessarily fundamentally different. One advantage of this idea is that it avoids the issue of which comes first: the representation of RT duration, or the non-temporal information? I have been assuming that non-temporal information such as confidence or feeling of rightness[17] arises prior to, and influences, perceived RT duration. However, others have argued for the reverse, for example that feelings of rightness are based on response fluency, for which RT is a common measure (Thompson, 2009). In fact, given the importance of metacognitive information, if there is indeed a dedicated system devoted to timing our RTs to basic cognitive tasks, then one of its primary roles would most likely be to help to generate such information.

Summary

In summary, I have suggested that the processes underlying the estimation of one's own response times may differ from those required to estimate the durations of externally-defined intervals. I believe that the most reasonable interpretation of the six experiments that I have presented here is that duration estimation of externally-defined intervals requires some type of effortful or attention-demanding processing that is not also required to estimate the durations of internally-defined intervals. Estimation of the duration of internally-defined intervals may instead depend on non-temporal metacognitive information, or, perhaps, an ‘internal clock’ dedicated specifically to the timing of our own cognitive processes. In the future, time perception researchers

[17] While discussing the example of reading a paragraph of a journal article, I suggested some other types of information such as the number of difficult words. For cognitive tasks that are performed within a few seconds, however, that sort of information is probably less abundant.

should not assume internally- and externally-defined intervals to be equivalent and should work to uncover the mechanisms underlying our ability to accurately estimate the durations of internally-defined intervals.

References

Block, R. A., & Grondin, S. (2014). Timing and time perception: A selective review and commentary on recent reviews. *Frontiers in psychology*, 5.

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. (2015). Multiple timing of nested intervals: Further evidence for a weighted sum of segments account. *Psychonomic bulletin & review*, 1-7.

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.

Buhusi, C. V., & Meck, W. H. (2005). What makes us tick? Functional and neural mechanisms of interval timing. *Nature Reviews Neuroscience*, 6(10), 755-765.

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*. 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.

Fernandes, M. A., & Moscovitch, M. (2000). Divided attention and memory: evidence of substantial interference effects at retrieval and encoding. *Journal of Experimental Psychology: General*, 129(2), 155.

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.

Fortin, C., & Massé, N. (2000). Expecting a break in time estimation: Attentional time-sharing without concurrent processing. *Journal of Experimental Psychology: Human Perception and Performance*, 26(6), 1788- 1796.

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.

Hicks, R. E., Miller, G. W., & Kinsbourne, M. (1976). Prospective and retrospective judgments of time as a function of amount of information processed. *The American journal of psychology*, 719-730.

Ivry, R. B. (1996). The representation of temporal information in perception and motor control. *Current opinion in neurobiology*, 6(6), 851-857.

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.

Šidàk, Z. (1967). Rectangular confidence region for the means of multivariate normal distributions. *Journal of the American Statistical Association*, 62, 626–633.

Thomas, E. C., & Brown, I. (1974). Time perception and the filled-duration illusion. *Perception & Psychophysics*, 16(3), 449-458.

Thomas, E. A., & Weaver, W. B. (1975). Cognitive processing and time perception. *Perception & Psychophysics*, 17(4), 363-367.

Thompson, V. A. (2009). Dual process theories: A metacognitive perspective. *Two Minds: Dual Processes and Beyond*. Oxford University Press, Oxford.

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*, 6.

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.

Zakay, D., & Block, R. A. (1995). An attentional-gate model of prospective time estimation. *Time and the dynamic control of behavior*, 167-178.

Zakay, D., & Block, R. A. (1996). The role of attention in time estimation processes. In *Advances in psychology* (Vol. 115, pp. 143-164). North-Holland.

Zakay, D., & Block, R. A. (1997). Temporal cognition. *Current directions in psychological science*, 6(1), 12-16.