Acta Psychologica 192 (2019) 153–162

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Unpacking, summing and anchoring in retrospective time estimation

Michael M. Roy[a,b,*], Tatem Burns[a,c], Joseph R. Radzevick[d]

[a] Elizabethtown College, USA
[b] North-West University, Potchefstroom, South Africa
[c] DePaul University, USA
[d] Gettysburg College, USA

ARTICLE INFO

Keywords:
Time
Memory
Bias
Anchoring
Unpacking
Summing

ABSTRACT

We examined whether or not interventions that have been used to try to influence predictions of future task duration – unpacking, summing and anchoring – had a similar effect on retrospective estimations of duration. In three studies, participants experienced a number of short stimuli, such as watching videos, before estimating the duration for each of the stimuli and the overall duration. The first estimation given served as an anchor for all following estimates. If the first estimation was highly biased in one direction, then subsequent estimates were more likely to also be biased in the same direction. Additionally, separate estimates for a number of individual tasks differed from the estimates for all of the tasks combined. This incongruity happened even though all estimates were given in sequence. Overall, results indicated that memories of past task duration could be influenced by the manner in which they were elicited.

1. Introduction

People are often inaccurate and biased when estimating duration. This is true both for estimations of past task duration (Block & Zakay, 1997; Fraisse, 1963; Ornstein, 1969; Poynter, 1989; Wallace & Rabin, 1960) and future task duration (Buehler, Griffin, & Peetz, 2010; Halkjelsvik & Jørgensen, 2012; Roy, Christenfeld, & McKenzie, 2005). There appear to be a number of similarities in when and where estimations of past and future task duration are likely to be biased due to factors such as the actual task duration and familiarity with the task (Roy et al., 2005; Roy & Christenfeld, 2007, 2008; Roy, Christenfeld, & Jones, 2013; Thomas, Handley, & Newstead, 2004, 2007). However, there are differences in the manner in which past and future task duration typically have been investigated. The focus of research on retrospective time estimation generally involves theoretical mechanisms that are relevant to estimating duration, such as attention and memory storage (Block & Zakay, 1997; Grondin, 2010). Research on future task duration, while still theoretically grounded, tends to focus more on interventions that could improve predictive accuracy (Buehler et al., 2010; Halkjelsvik & Jørgensen, 2012; Roy et al., 2005). For example, research has examined techniques such as having participants try to think about all the subcomponents of a task before estimating how long it will take them to complete the full task (unpacking: Kruger & Evans, 2004) and supplying participants with the duration of another relevant task before estimating task duration (anchoring: König, 2005). Here we examined whether or not interventions that have been used to try to influence predictions of future task duration – unpacking, summing and anchoring – had a similar effect on retrospective estimations of duration.

1.1. Unpacking

One technique that has been used to potentially improve predictions of future task duration is unpacking, where individuals break down a task into component parts before estimating total task duration (Kruger & Evans, 2004; Tversky & Koehler, 1994). It has been proposed that unpacking could improve time estimation by forcing estimators to include certain, potentially neglected, subcomponents in their estimate and to give those subcomponents the proper weight when formulating their overall estimate (Tversky & Koehler, 1994). Results of studies using unpacking have been mixed. Studies generally found that unpacking led to longer estimates of duration (Connolly & Dean, 1997; Hadjichristidis, Summers, & Thomas, 2014; Kruger & Evans, 2004; Liu, Li, & Sun, 2014; Min & Arkes, 2012; Tsai & Zhao, 2011), which was helpful when there was a tendency to underestimate task duration (Connolly & Dean, 1997; Kruger & Evans, 2004). However, other studies found no impact of the intervention (Byram, 1997; Jorgensen, 2004). Further, though some found that unpacking led to longer

* Corresponding author at: Elizabethtown College, Department of Psychology, 1 Alpha Dr., Elizabethtown, PA 17601, USA.
E-mail address: roym@etown.edu (M.M. Roy).

https://doi.org/10.1016/j.actpsy.2018.11.012
Received 12 June 2018; Received in revised form 26 November 2018; Accepted 27 November 2018
0001-6918/

estimates, results from these studies do not indicate the exact mechanisms that led to the longer estimates – whether it was due to including neglected information (Tversky & Koehler, 1994) or some other issue such as anchoring (Hadjichristidis et al., 2014). Here we examined whether or not unpacking led to longer estimates of past task duration.

1.2. Summing

A technique for reducing bias in predicted future duration that is similar to unpacking is summing, or segmentation (Forsyth & Burt, 2008; Macan, 1994). Where unpacking involves thinking about the component parts before making an overall estimate, segmentation is estimating the duration of each of the component parts and summing them together. The sum of the component parts is used instead of the overall estimate. Like unpacking, a potential reason for why summing might lead to better estimation is that supplying estimates for each individual part can help ensure that all portions are accounted for and properly weighted (Macan, 1994). Alternatively, it might be that making estimates for the shorter portions of the task reduces the overall tendency to underestimate that is found in longer tasks (Forsyth & Burt, 2008). In support, Forsyth and Burt (2008) found that when a future task was likely to be underestimated, using the sum of estimates for component parts decreased the tendency to underestimate. However, they found that when estimation was likely to be unbiased, segmentation actually led to overestimation of task duration. In both cases, results consistently indicated that the sum of individual parts tended to be different from the overall estimate (see also Tversky & Koehler, 1994; Van Boven & Epley, 2003, for similar finding with other judgments). Here we examine the relationship between summed estimates of individual parts and estimates for the whole event in retrospective estimates of duration. However, one difference between this study and previous studies was that we compared sum and whole *within* individual participants instead of *between* participants. We were interested in examining if being forced to estimate both the individual components and the overall estimate would help bring both of these estimates in line with each other.

1.3. Anchoring

Anchors have been employed to help improve a number of different types of judgments (Bartoshuk, Duffy, Fast, Green, & Snyder, 2002; Furnham & Boo, 2011). Research indicates that anchors can influence estimation of future task duration by pushing the estimation toward the anchor (König, 2005). Anchors, either from being directly supplied or through previous experience, can influence estimated future duration both for the better and for the worse (König, 2005; König, Wirz, Thomas, & Weidmann, 2015; Thomas et al., 2007; Thomas & Handley, 2008; Thomas, Newstead, & Handley, 2003). Anchoring can occur even when individuals are not prompted to use an anchor (Thomas & Handley, 2008). This can explain why time estimation often improves with experience (Tobin & Grondin, 2012; 2015) and feedback (Roy, Mitten, & Christenfeld, 2008).

Memory for past task duration is often biased (Block & Zakay, 1997; Fraisse, 1963; Ornstein, 1969; Poynter, 1989; Wallace & Rabin, 1960). However, even though people are bad at estimating absolute duration – the exact duration of a task – they appear to be good at estimating the relative duration of a task – whether it is shorter or longer than other tasks (Roy et al., 2008). Supplying a correct anchor could therefore improve subsequent judgment of duration by giving people a correct starting point from which to adjust upward or downward from depending on the relative duration of the task.

Anchoring also supplies a potential explanation for how unpacking and summing influence estimates of duration (Hadjichristidis et al., 2014). In unpacking and summing, participants make a series of duration estimates. The first estimates might serve as an anchor for all subsequent estimates and therefore influence an overall shift in bias.

1.4. Current studies

Our research on unpacking, summing and anchoring in retrospective estimation began with a stroke of serendipity. In Study 1, we were originally interested in the impact of individual differences on estimation of duration. Fortunately, our design allowed us to also examine the impact of anchoring and summing on the retrospective estimates. In Study 1, we found similarities to results from future task duration studies in the influence of anchoring and summing. Studies 2 and 3 were designed to more directly investigate these issues as well as unpacking. In the first two studies, we found that subsequent estimations were influenced by the bias in an initial estimate. If the first estimate was highly biased in one direction (overestimation or underestimation), then subsequent estimates were likely to be biased in the same direction. For example, when the first task was very short, participants were more likely to overestimate not only the duration of that task, but also all subsequent tasks and the total duration. Further, participants' estimates for the individual components did not add up to their overall duration estimate. In Study 3, we examined whether supplying participants with the correct duration for the first in a series of short films would improve subsequent estimations of duration. Results indicate that instead of making all subsequent estimates more accurate, supplying the correct duration for the first task simply shifted the pattern of bias for subsequent estimates.

2. Study 1

The original goal of Study 1 was to examine the influence of musical training on time estimation. Previous research has found that musicians were significantly better at estimating the duration of musical stimuli than were non-musicians (e.g., Panagiotidi & Samartzi, 2013). We were interested in attempting to replicate those results, and to see if increased ability in estimating duration might also apply to non-musical stimuli such as listening to a newscast and filling out a questionnaire. Further, we attempted to determine if certain individual differences such as increased rumination and depression, which have been linked with musicians (e.g., Jones, Roy, & Verkuilen, 2014; Roy, Radzevick, & Getz, 2016; Verhaeghen, Khan, & Joormann, 2005), might mediate the relationship between musical ability and time estimation. To do so, we had participants estimate the duration of a song, a newscast, a questionnaire, and all of the stimuli combined. Because participants provided estimates for the individual tasks followed by the overall estimate of duration and we varied which task came first in the sequence, we were able to examine the relationship between the individual estimates and the overall estimate (summing) and the impact of the first estimation given on all subsequent estimations (anchoring).

2.1. Method

2.1.1. Participants

Four-hundred participants (Age $M = 36.8$, $SD = 12.3$) recruited from Amazon Mechanical Turk (see Buhrmester, Kwang, & Gosling, 2011, for more information about Mechanical Turk) completed the survey. We set this target by multiplying the number of scales to be used in the regression analysis – 4 – by 100. Nineteen participants did not follow directions or did not correctly answer questions that checked attention to the task, so we eliminated them from our sample to leave 381 remaining participants. All participants were paid $0.50 for taking part in the study. An Institutional Review Board approved the research and participants completed an online informed consent form prior to beginning the study.

2.1.2. Materials

2.1.2.1. Audio stimuli. Participants listened to the 2 min 50 s song "Staying in Love" by Raphael Saadiq and a 2 min 57 s National Public Radio news report about astronauts with a new espresso machine in space (Brumfield, 2015).

2.1.2.2. Questionnaire. Participants indicated responses for the 20-item Center for Epidemiologic Studies Depression Scale (CES-D; Radloff, 1977), the 10-item Ruminative Response Scale (RRS; Treynor, Gonzalez, & Nolen-Hoeksema, 2003), the 15-item Uses of Music Inventory (UOM; Chamorro-Premuzic & Furnham, 2003), and 10 questions about their specific musical training (see Appendix for detail on each scale).

2.1.3. Design & procedure

Participants completed this study using an online survey platform directed from Amazon Mechanical Turk. Participants were told that they would be listening to some audio clips and filling out some questionnaires and that they would be answering questions about the stimuli afterward. Importantly, participants were not told that they would be answering questions about the duration of the stimuli because prior knowledge that duration is the focus of the study can influence estimates of duration (Block & Zakay, 1997). First, participants either listened to the song or news report, with the order of the audio stimuli randomized. After the first audio piece finished, they completed the questionnaire items. Then, participants listened to the audio piece that was not presented initially, song or news report. Once the second audio piece ended, they separately estimated the time duration of the first audio piece, questionnaire section, and the second audio piece. After giving these separate time estimations for each section in the order that they were presented, participants estimated the amount of time it took to complete the entire study.

2.1.4. Dependent variable

Bias was measured by taking the log of the ratio of estimated duration divided by actual duration, called log proportional error (Roy & Christenfeld, 2007, 2008). For log proportional error, a negative score indicates underestimation, a score of zero indicates no bias and a positive score indicates overestimation. By taking the log of the ratio of estimated duration over actual duration, the data were normalized; as discussed in the next section, there was positive skew in estimates of duration for all of the experiments reported here. Use of log proportional error allows for more direct comparison of bias in estimated duration for components that differ in actual duration.

2.2. Results

2.2.1. Descriptive statistics

As is common, there was a tendency for estimates of duration to be positively skewed (and not normally distributed - Kolmogorov-Smirnov test): song skewness = 0.324 ($D(381) = 0.092$, $p < .001$); news report skewness = 1.048 ($D(381) = 0.127$, $p < .001$); questionnaire skewness = 1.644 ($D(381) = 0.178$, $p < .001$); overall estimate skewness = 0.977 ($D(381) = 0.098$, $p < .001$). In terms of median duration (median used due to skew in data), participants indicated that the song lasted 2.95 min ($IQR = 1.00$; actual duration 2.83 min), that the news report lasted 3.08 min ($IQR = 1.50$; actual duration 2.95 min), that the questionnaire lasted 5.00 min ($IQR = 6.72$; actual duration $Median = 4.01$ min, $IQR = 2.13$), and that the whole study took 10.93 min ($IQR = 5.55$; actual duration $Median = 12.10$ min, $IQR = 3.00$). As indicated by a single sample t-test on log proportional error (log of estimated/actual duration), participants significantly overestimated the duration of the news report, $M = 0.022$, $SD = 0.164$, $t(380) = 2.57$, $p = .011$, $d = 0.132$, and the questionnaire, $M = 0.106$, $SD = 0.332$, $t(380) = 6.20$, $p < .001$, $d = 0.318$, exhibited no significant bias for the song, $M = 0.002$, $SD = 0.121$, $t(380) = 0.27$, $p = .789$, $d = 0.014$, and underestimated the overall duration, $M = -0.069$, $SD = 0.265$; $t(380) = -5.06$, $p < .001$, $d = -0.259$. The individual difference measures in the questionnaire were, for the most part, not predictive of bias in time estimation (see Appendix).

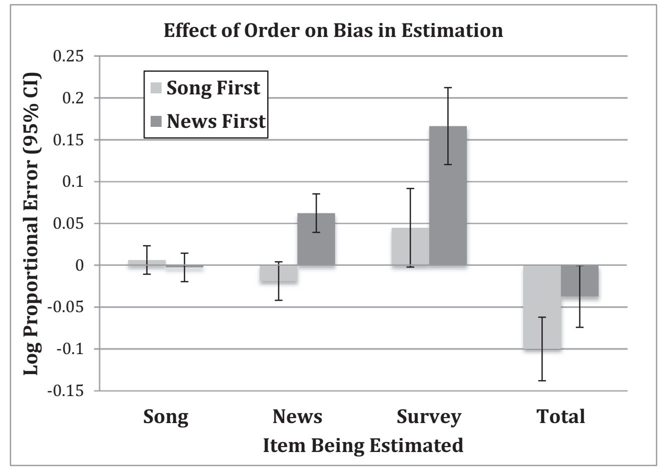

Fig. 1. Mean (± 95% Confidence Interval) bias in estimated duration (log of estimated duration/actual duration) for all tasks.

2.2.2. Anchoring

A 4 (task estimated) × 2 (order of tasks) mixed-model ANOVA on bias (log proportional error) was used to examine the effect of anchoring on estimation. There was a significant main effect of task with participants differing in bias for the 4 estimates $F(3,1137) = 43.95$, $p < .001$, $\eta_p^2 = 0.104$. As can be seen in Fig. 1, post-hoc LSD tests indicate amount and direction of bias was significantly different between all paired comparisons of estimated items ($ps < .03$). There was a significant main effect of order with a greater tendency to overestimate when participants heard the news report first ($M = 0.047$, $SE = 0.010$) than when they heard the song first ($M = -0.017$, $SE = 0.010$), $F(1,379) = 18.95$, $p < .001$, $\eta_p^2 = 0.048$. The overall main effect of order was qualified by a significant interaction between task and order $F(3,1137) = 6.32$, $p < .001$, $\eta_p^2 = 0.017$. As indicated by a simple effects test, when participants estimated the news report first there was a significant tendency to supply longer estimates for the news report itself ($p < .001$), the questionnaire ($p < .001$), and the total duration ($p = .021$), but not for the song ($p = .475$).

2.2.3. Sum versus whole

As indicated earlier, participants overestimated two of the three individual components (and showed no bias on the third) but underestimated the overall duration. Using log proportional error as the dependent variable, participants underestimated how long the whole study took ($M = -0.069$, $SE = 0.014$; $t(380) = -5.06$, $p < .001$, $d = -0.26$), but the sum of their individual estimates overestimated the duration, ($M = 0.075$, $SE = 0.009$; $t(382) = 8.63$, $p < .001$, $d = 0.44$). The difference between the sum of the estimates and the overall duration was confirmed by a 2 (Sum v Whole) × 2 (Order) mixed-model ANOVA, $F(1,379) = 121.18$, $p < .001$, $\eta_p^2 = 0.242$ (see Fig. 2). The analysis also indicated an overall significant main effect of order, $F(1,379) = 13.55$, $p < .001$, $\eta_p^2 = 0.035$, with participants providing longer estimates when hearing the news report first ($M = 0.037$, $SE = 0.013$) than when hearing the song first ($M = -0.030$, $SE = 0.013$). The interaction was not significant, $F(1,379) = 0.142$, $p = .706$, $\eta_p^2 < 0.001$. Even though the sum of participants individual estimates was related to their overall estimate of the whole task, $r(380) = 0.375$, $p < .001$, they were not equal.

2.3. Discussion

The original intent of the study, to examine in more detail the relationship between time perception and musical training, was not realized. However, other interesting results emerged from this study. The newscast elicited longer estimates of duration than did the song despite

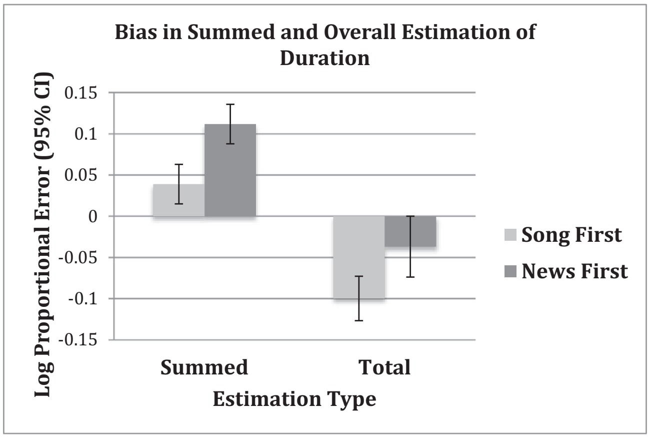

Fig. 2. Mean (± 95% Confidence Interval) bias in estimated duration (log of estimated duration/actual duration) for the total duration estimation and for the sum of the tasks.

having similar overall durations. One possible explanation for this difference has to do with experience and feedback. Participants likely had extensive exposure to songs and song duration because when they listen to songs digitally the duration is clearly displayed. People tend to be accurate in estimates for tasks where they have received frequent feedback on timing (Tobin & Grondin, 2012; 2015), but inaccurate when feedback has been absent (Bisson, Tobin, & Grondin, 2012; Tobin, Bisson, & Grondin, 2010). In support of this exposure account, the song was both the only stimuli to not show any effect of task order and to be estimated accurately. Well-learned stimuli may be immune to manipulations that shift bias in estimation. Whatever its origin, the overestimated duration for the newscast meant that the order of the tasks mattered in terms of bias in individual and overall estimates found in Study 1. When the newscast was first, subsequent estimates were more likely to be in the same direction, with bias in the first estimate serving as an anchor for subsequent estimation.

Additionally, participants' individual estimates did not add up to their overall estimate. Even though the overall estimate occurred directly after the previous three estimates, the sum of the components did not add up to the overall estimate. Participants indicate that, overall, the individual parts lasted longer than the whole. This is even more remarkable because the overall duration was longer than the individual parts because it also included time spent on instructions between each of the individual components. Neither the sum nor the whole appeared to be preferable in that both were somewhat equally biased, only in opposite directions.

Even though participants' estimate of the total was not equal to the sum of the components, the total estimate was influenced by the estimates given for the components. Bias in the first estimate transferred to subsequent estimates, including the total. These results indicate that any shift in bias due to unpacking may not have been due to processes such as thinking more deeply about all the components of the task (Tversky & Koehler, 1994), but simply because estimating duration for one of the smaller components first leads to a shift in bias that was transferred to the total. Shorter tasks tend to be overestimated while longer tasks tend to be underestimated (Lejeune & Wearden, 2009; Roy et al., 2013; Roy & Christenfeld, 2008; Tobin & Grondin, 2009; Yarmey, 2000). Starting by estimating the duration of one of the smaller components of the task could lessen an overall tendency to underestimate the total duration by shifting bias upward.

The results of Study 1 need to be interpreted with caution because the study was not set up specifically to study anchoring, summing or unpacking. As stated previously, the sum and the whole were of different durations because there was instruction time between each of the components. Further, the whole estimate was given only after the components. To truly examine the full effects of unpacking, the total duration estimate given after unpacking would need to be compared to the total duration estimate given without unpacking.

3. Study 2

We conducted Study 2 to examine the phenomena found in Study 1 more directly and in greater detail, making three major changes to our procedure. First, we limited stimuli to three science videos that varied in duration. Previous research has found that shorter tasks tend to be overestimated while longer tasks tend to be underestimated (Lejeune & Wearden, 2009; Roy et al., 2013; Roy & Christenfeld, 2008; Tobin & Grondin, 2009; Yarmey, 2000). Here, duration of the first video was varied between short and long to measure the impact of bias of the initial estimate on all subsequent estimates. Second, participants viewed the three videos sequentially without interruption so that the sum of the individual components was the same as the overall duration. Finally, we varied the placement of the overall estimate so that it occurred either before or after the estimates of the individual components, allowing us to examine the effect of unpacking on the overall estimate of duration.

3.1. Method

3.1.1. Participants

Fifty-four college students at a small liberal arts college in the United States completed the study. Demographic information was not collected. The Institutional Review Board approved the research and participants provided informed consent prior to beginning the study. Sample size was determined by the number of participants that signed up for the study in return for course credit over the course of one semester.

3.1.2. Procedure

Participants were told that they would be watching videos and then answering questions about those videos after they were completed, but were not told the nature of the questions that would be asked. Participants watched three short science videos from the YouTube channel SciShow (n.d.). The videos, on the influence of grapefruit juice on medications (2 min 2 s; short), on how fingerprints form (2 min 52 s; medium), and on ancient grain wheat (3 min 55 s; long), varied in duration from each other by approximately 1 min. Videos were presented either in the order of short, medium, long or long, medium, short, with assignment to condition determined randomly.

Participants estimated the duration for each of the three videos and the overall duration of all the videos combined (8 min 49 s). Estimates for the individual videos were supplied in the order that they were presented. Estimations for the total duration were given either before the individual estimates or after the individual estimates, also determined randomly. Overall, the study was a 2×2 design varying the order of the videos and the placement of the overall estimate.

For each of the individual videos, participants also indicated how much they liked the video (5-point scale), how familiar they were with the video (3-point scale), and how much attention they paid to the video (5-point scale).

3.2. Results

3.2.1. Descriptive statistics

Again, there was a tendency for estimates of duration to be positively skewed (and not normally distributed - Kolmogorov-Smirnov test): grapefruit video skewness = 0.528 ($D(53) = 0.131$, $p = .023$); fingerprint video skewness = 0.668 ($D(54) = 0.119$, $p = .056$); ancient grain video skewness = 0.093 ($D(54) = 0.094$, $p = .020$); overall estimate skewness = 0.786 ($D(54) = 0.159$, $p = .002$). In terms of median duration, participants indicated that the grapefruit video lasted

2.77 min (IQR = 1.13; actual duration 1.97 min), that the fingerprint video lasted 3.17 min (IQR = 1.02; actual duration 2.87 min), that the ancient grain video lasted 4.00 min (IQR = 1.70; actual duration 3.92 min), and that the whole task took 9.93 min (IQR = 2.63; actual duration 8.82 min). As indicated by a single sample t-test on log proportional error (log of estimated/actual duration), participants significantly overestimated the duration of the grapefruit video, $M = 0.129$, $SD = 0.137$, $t(52) = 6.86$, $p < .001$, $d = 0.941$, and the fingerprint video, $M = 0.033$, $SD = 0.120$, $t(53) = 2.03$, $p = .047$, $d = 0.275$, but exhibited no significant bias for the grain video, $M = -0.018$, $SD = 0.130$, $t(53) = -1.01$, $p = .317$, $d = -0.137$, or for the overall estimate, $M = 0.022$, $SD = 0.106$, $t(53) = 1.56$, $p = .125$, $d = 0.212$.

3.2.2. Anchoring

To examine the overall effect of video and estimation order, bias (log proportional error) was examined in a 4 (estimate: short, medium, long, overall) × 2 (order of videos: long first or short first) × 2 (overall estimate placement: first or last) mixed-model ANOVA. Estimates were influenced by the duration of the video, with shorter videos more likely to be overestimated, $F(3,147) = 37.89$, $p < .001$, $\eta_p^2 = 0.436$. As can be seen in Fig. 3, post-hoc LSD tests indicate that bias was significantly different between all individual estimates ($ps < .001$) except for between bias in the estimate for the medium (fingerprint) video and the overall estimate ($p = .420$).

There was a main effect of total estimation placement, with shorter overall estimates given when the total estimate came first ($M = 0.005$, $SE = 0.019$) than when it came last ($M = 0.079$, $SE = 0.020$), $F(1,49) = 7.40$, $p = .009$, $\eta_p^2 = 0.131$. The main effect of order – whether the short (grapefruit) video ($M = 0.062$, $SE = 0.019$) or the long (ancient grain) video was first ($M = 0.022$, $SE = 0.020$) - was not significant, $F(1,49) = 2.17$, $p = .147$, $\eta_p^2 = 0.042$. However, there was a significant interaction between the item being estimated and the order of the videos with longer estimates more likely when the shorter video came first for all estimates except for the wheat video, $F(3,147) = 6.03$, $p = .001$, $\eta_p^2 = 0.110$ (see Fig. 3). All other interactions were not significant ($ps > .32$, $\eta_p^2s < 0.03$).

To further examine the interaction between the item being estimated and the order of the videos, a series of 2 (order of videos: long first or short first) × 2 (overall estimate placement: first or last) ANOVAs were conducted examining bias for each individual estimate. For the short (grapefruit) video, estimates were significantly longer when the video was first ($M = 0.170$, $SE = 0.024$) than when the video was last ($M = 0.088$, $SE = 0.026$), $F(1,49) = 5.37$, $p = .025$, $\eta_p^2 = 0.099$. Bias for the video was significantly less when the total estimate was supplied first ($M = 0.092$, $SE = 0.024$) than when it was given last ($M = 0.166$, $SE = 0.025$), $F(1,49) = 4.35$, $p = .042$, $\eta_p^2 = 0.082$. The interaction was not significant, $F(1,49) = 0.99$, $p = .324$, $\eta_p^2 = 0.020$.

For the medium (fingerprint) video, estimates were significantly longer when the short video was first ($M = 0.066$, $SE = 0.021$) than when the long video was first ($M = -0.002$, $SE = 0.023$), $F(1,49) = 4.65$, $p = .036$, $\eta_p^2 = 0.085$. When the total estimate was given, first ($M = 0.009$, $SE = 0.022$) or last ($M = 0.055$, $SE = 0.023$), did not significantly influence estimates for this video, $F(1,49) = 2.03$, $p = .160$, $\eta_p^2 = 0.039$. The interaction was not significant, $F(1,49) = 0.12$, $p = .676$, $\eta_p^2 = 0.004$.

For the long (ancient grain) video order did not matter, with no significant difference in estimates when it was viewed first ($M = -0.001$, $SE = 0.025$) than when it was viewed last ($M = -0.028$, $SE = 0.023$), $F(1,50) = 0.64$, $p = .428$, $\eta_p^2 = 0.013$. The lack of difference due to order found here contrasts with the significant effect of order found in the previous two videos and explains the significant video type by order interaction found earlier. However, bias for the video was significantly lower when the total estimate was given first ($M = -0.062$, $SE = 0.023$) than when it was given last ($M = 0.032$, $SE = 0.025$), $F(1,50) = 7.70$, $p = .008$, $\eta_p^2 = 0.133$. The interaction was not significant, $F(1,50) = 0.23$, $p = .879$, $\eta_p^2 < 0.001$.

For the overall total duration, order - short video first ($M = 0.045$, $SE = 0.018$) or long video first ($M = 0.001$, $SE = 0.020$) - did not significantly influence estimation of duration, $F(1,50) = 2.62$, $p = .112$, $\eta_p^2 = 0.050$. Indicating that unpacking influenced the overall duration, bias for the total duration estimate was significantly lower when it was given first ($M = -0.017$, $SE = 0.018$) than when it was given last ($M = 0.063$, $SE = 0.019$), $F(1,50) = 9.11$, $p = .004$, $\eta_p^2 = 0.154$. The interaction was not significant, $F(1,50) = 0.46$, $p = .501$, $\eta_p^2 = 0.009$.

3.2.3. Sum versus whole

Even though participants did not exhibit bias for their overall estimate, the sum of their estimates significantly overestimated the actual duration, $M = 0.042$, $SD = 0.108$, $t(52) = 2.81$, $p = .007$, $d = 0.386$. As found in the previous study, the sum of the individual estimates did not equal the overall estimate (see Fig. 4). As indicated by a 2 (sum v whole) × 2 (order of videos: long first or short first) × 2 (overall estimate: first or last) mixed model ANOVA, there was a significant difference in bias (log proportional error) for the sum of the estimates and the overall estimate, $F(1,49) = 5.73$, $p = .021$, $\eta_p^2 = 0.105$. Additionally, both sum and overall estimates were influenced by when the

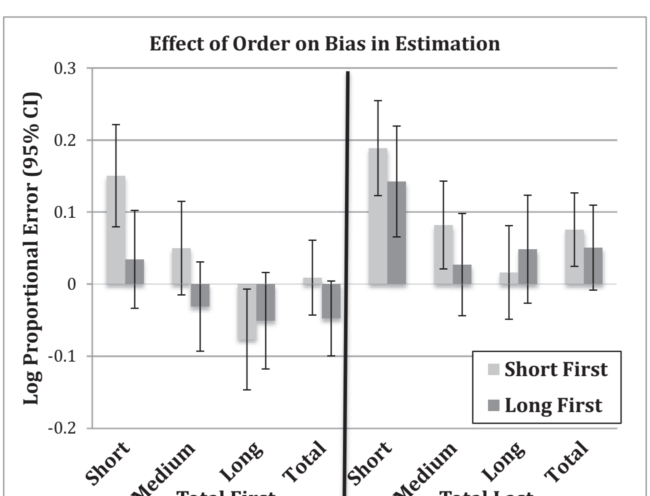

Fig. 3. Mean (± 95% Confidence Interval) bias in estimated duration (log of estimated duration/actual duration) for all film clips.

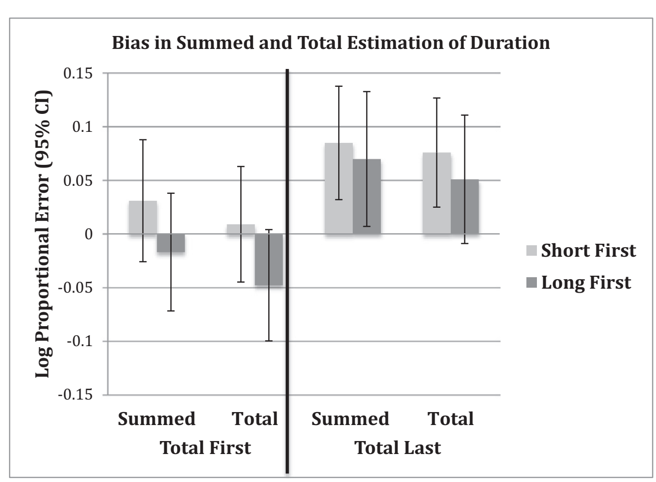

Fig. 4. Mean (± 95% Confidence Interval) bias in estimated duration (log of estimated duration/actual duration) for the total duration estimation and for the sum of the film clips.

total estimate was given, $F(1,49) = 8.407$, $p = .006$, $\eta_p^2 = 0.146$, with overestimation for both more likely when the overall estimate was given last ($M = 0.070$, $SE = 0.019$) than when it was given first ($M = -0.006$, $SE = 0.018$). The main effect of order - short video first ($M = 0.050$, $SE = 0.018$) or long video first ($M = 0.014$, $SE = 0.019$) - was not significant, $F(1,49) = 1.88$, $p = .176$, $\eta_p^2 = 0.037$. All interactions were not significant ($ps > .48$, $\eta_p^2 s < 0.01$).

3.2.4. Familiarity, liking, and attention

Participants exhibited a low level of familiarity for the videos: the percentage of participants indicating that they were unfamiliar with the video was 78% for the grapefruit video, 67% for the fingerprint video, and 80% for the ancient grain video. The differences in familiarity for the videos was significant with participants more familiar with the fingerprint (medium) video ($M = 1.47$, $SE = 0.099$) than there were with either the grapefruit (short) video ($M = 1.23$, $SE = 0.064$) or the ancient grain (long) video ($M = 1.21$, $SE = 0.056$), $F(2,104) = 8.74$, $p < .001$, $\eta_p^2 = 0.114$. Self-ratings of liking and attention for the videos revealed a different pattern with lower overall ratings for the ancient grain (long) video. Participants gave lower ratings of liking for the ancient grain (long) video ($M = 2.94$, $SE = 0.130$) than they did for either the grapefruit (short) video ($M = 3.62$, $SE = 0.121$) or the fingerprint (medium) video ($M = 3.76$, $SE = 0.100$), $F(2,104) = 17.23$, $p < .001$, $\eta_p^2 = 0.249$. In addition, participants gave lower ratings of attention spent on the video for the ancient grain video ($M = 3.26$, $SE = 0.121$) than they did for either the grapefruit video ($M = 3.87$, $SE = 0.114$) or the fingerprint video ($M = 3.87$, $SE = 0.093$), $F(2,104) = 11.73$, $p < .001$, $\eta_p^2 = 0.184$.

To examine whether differences in bias in estimation of duration could be explained by differences in familiarity, liking and attention, we performed a series of pairwise multiple regression analyses. First, differences in bias, familiarity, liking and attention were computed for each pair of videos (grapefruit and fingerprint; grapefruit and ancient grain; fingerprint and ancient grain). Next, multiple regression was used to determine if individual differences in change scores for familiarity, liking, and attention could predict individual differences in change scores for bias for estimated duration for each pair. Results indicate no significance for the fingerprint and grapefruit videos, $F(3,52) = 0.829$, $p = .484$, $R^2 = 0.048$, the ancient grain and grapefruit videos, $F(3,52) = 1.13$, $p = .345$, $R^2 = 0.065$, and the ancient grain and fingerprint videos, $F(3,52) = 2.03$, $p = .122$, $R^2 = 0.108$. Differences in ratings for the videos did not predict differences in bias for the videos. Further, within each video, bias in estimated duration was not predicted by familiarity, liking and attention for the grapefruit, $F(3,52) = 0.707$, $p = .552$, $R^2 = 0.042$, fingerprint, $F(3,53) = 1.78$, $p = .163$, $R^2 = 0.096$, and ancient grain, $F(3,53) = 2.66$, $p = .058$, $R^2 = 0.138$, videos.

3.3. Discussion

Replicating the findings of Study 1, an initial estimate of duration tended to influence subsequent estimates of duration and the sum of the estimates for individual components did not equal the overall estimated duration. For the short and medium videos, starting the series with the shorter video led to longer individual estimates. Whether the overall estimate came first or last, the sum of participants individual estimates did not add up to the overall estimates. The overall estimate placement did matter: the overall estimate was shorter when given first than when given last. In line with an anchoring account, when the overall estimate was given first it also led to shorter individual estimates for the short and the long videos. Results indicate that both order of presentation and order of what is estimated can cause changes in initial bias, which then transfers to later estimates. It should be noted that these anchors – the specific video estimated and when the overall duration was given – did not significantly influence all of the subsequent estimations. It is possible that certain tasks factors not measured here influenced the likelihood of transferring bias from a previous estimate to that particular estimate. It is also possible that the lack of significant results for some of the estimates was due to power issues. Even though not significant, a number of the non-significant estimates were in the predicted direction. For these estimates, effects sizes were fairly small indicating that a very large sample would be needed to detect differences.

Even though the pattern was similar, with longer individual estimates linked with longer overall estimates, individual estimates did not add up to the overall estimates in any of the conditions. Participants appear to be good at maintaining the relative relationship of all the estimates (including part to whole), but not the absolute relationship (the parts did not add up to the whole).

The unpacking manipulation did lead to longer estimates of total duration when compared to when the total duration estimate was supplied before unpacking. It appears that the total duration, as with the other individual estimates of duration, was influenced by bias in the first estimate. By first estimating the duration of one of the shorter components, estimation was shifted upward. In this case, there was little bias in participants' original estimates for the total duration, causing the unpacking manipulation to lead to increased bias in the form of overestimation in the estimation of total duration.

Here we replicated previous research finding that shorter tasks are more likely to be overestimated and longer tasks to be underestimated (Lejeune & Wearden, 2009; Roy et al., 2013; Roy & Christenfeld, 2008; Tobin & Grondin, 2009; Yarmey, 2000). In this case, the cutoff between overestimation and underestimation was around 3.5 min for the science videos. As with previous research, the cutoff between what is considered short and long is likely task dependent (Roy & Christenfeld, 2008; Yarmey, 2000). Because participants gave multiple estimations, regression to the mean could explain the pattern of results found here (Fraisse, 1963). Alternatively, the pattern of results could be due to participants rounding to certain whole values such as 3 min (Forsyth & Burt, 2008; Huttenlocher, Hedges, & Bradburn, 1990). In support, 22% of responses in Study 2 were whole numbers (36% in Study 1 and 23% in Study 2). Whatever the reason for the direction of the bias for the first estimation given, that bias was then likely to be passed onto the subsequent estimations.

4. Study 3

The goal for Study 3 was to examine the effects of manipulating bias in the initial estimate before having participants make subsequent estimations. Specifically, we provided feedback about the actual duration of an initial stimulus to examine whether correcting this bias would affect subsequent estimations. Previous research has indicated that receiving a correct anchor can decrease bias and increase accuracy in estimation of duration (Roy et al., 2008). Here, a portion of the participants were told the correct duration of the first video clip before making subsequent estimations.

4.1. Method

4.1.1. Participants

Fifty-nine college students (66% Female; Age $M = 19.0$, $SD = 1.0$) at a small liberal arts college in the United States completed the study. The Institutional Review Board approved the research and participants provided informed consent prior to beginning the study. Again, sample size was determined by number of participants that could be recruited for class credit during a single semester.

4.1.2. Procedure and design

Participants were told that they would be watching videos and then answering questions about those videos after they were completed, but were not told the nature of the questions that would be asked. Participants were randomly assigned to either the correct anchor supplied or the anchor self-generated condition. Participants in the correct anchor supplied condition were told the duration of the first video before estimating the duration of the next two videos whereas participants in the anchor self-generated condition estimated the duration of the first video before supplying the other two estimates. All participants first watched the medium duration video on how fingerprints form (2.87 min). Next the participants watched the short video on the influence of grapefruit juice on medications (2.03 min) and the long video on ancient grain wheat (3.92 min) with order of the two videos counterbalanced. After all videos were finished, half of the participants estimated the duration of the first video on fingerprints while the other half were supplied with the correct duration as an anchor. Participants then estimated the duration for each of the two remaining videos and, finally, the overall duration of all the videos combined (8.82 min). The overall design was a 2 (correct anchor supplied vs. anchor self-generated) × 2 (short-long vs. long-short) between-subjects experiment.

For each of the videos, participants supplied ratings of liking, familiarity and attention. Finally, participants supplied demographic information on age and gender.

4.2. Results

4.2.1. Descriptive statistics

Consistent with previous studies, there was a tendency for estimates of duration to be positively skewed (and not normally distributed - Kolmogorov-Smirnov test): fingerprint video skewness = 0.336 ($D(30) = 0.117$, $p = .200$); grapefruit video skewness = 1.194 ($D(59) = 0.139$, $p = .006$); ancient grain video skewness = 0.557 ($D(59) = 0.093$, $p = .200$); overall estimate skewness = 0.993 ($D(59) = 0.205$, $p < .001$). For participants that estimated the duration of the fingerprint video, median estimated duration was 3.68 min ($IQR = 1.80$; actual duration 2.87 min). For all participants combined, median duration was 2.50 min ($IQR = 1.75$; actual duration 1.97 min) for the grapefruit video, 3.67 min ($IQR = 1.88$; actual duration 3.92 min) for the ancient grain video, and 9.92 min ($IQR = 2.83$; actual duration 8.82 min) for the whole task. As indicated by a single sample *t*-test on log proportional error (log of estimated/actual duration), participants that estimated the duration for the fingerprint video, $M = 0.099$, $SD = 0.162$, $t(29) = 3.23$, $p = .003$, $d = 0.611$, significantly overestimated the duration. When all participants were taken together, participants tended to overestimate the grapefruit video, $M = 0.063$, $SD = 0.279$, $t(58) = 1.74$, $p = .088$, $d = 0.226$, underestimated the ancient grain video, $M = -0.043$, $SD = 0.164$, $t(58) = -2.03$, $p = .047$, $d = -0.265$, and show little bias for the overall estimate, $M = 0.023$, $SD = 0.139$, $t(58) = 1.25$, $p = .215$, $d = 0.165$.

4.2.2. Anchoring

The tendency for participants in the anchor self-generated condition to overestimate video duration transferred to the subsequent estimates as indicated by a 3 (short, long, total) × 2 (correct anchor supplied or anchor self-generated) × 2 (order of videos) mixed model ANOVA on log proportional error. There was a significant main effect of anchoring with shorter overall estimates by participants that received a correct anchor ($M = -0.033$, $SE = 0.030$) than by participants that estimated the anchor ($M = 0.056$, $SE = 0.028$; $F(1,55) = 4.54$, $p = .038$, $\eta_p^2 = 0.076$). There was no effect of order - whether or not the medium video was followed by the short or the long video, $F(1,55) = 0.33$, $p = .569$, $\eta_p^2 = 0.006$. As indicated by a main effect of video (shorter, longer, total), direction of bias varied with overestimation for the shorter video ($M = 0.063$, $SE = 0.036$) and total duration ($M = 0.023$, $SE = 0.017$), but underestimation for the longer video ($M = -0.043$, $SE = 0.021$; $F(2,110) = 7.35$, $p = .001$, $\eta_p^2 = 0.118$). Simple effects indicate there was not a significant difference in bias for estimates of the short video and the overall estimate ($p = .209$), but that both the short and overall estimates differed significantly from bias in estimation for the long video ($ps < .005$).

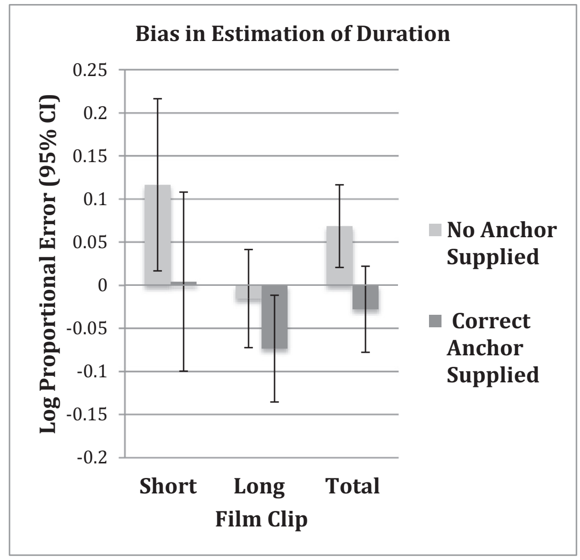

Fig. 5. Mean (± 95% Confidence Interval) bias in estimated duration (log of estimated duration/actual duration) for all film clips.

As can be seen in Fig. 5, which reports the mean and 95% CI for the three videos, supplying a correct anchor shifted the pattern of bias for the three estimates. Exposure to the correct anchor eliminated the overestimation tendency found in both the short estimate and the overall estimate when the anchor was self-generated. However, supplying a correct anchor did not eliminate bias overall. For the long video, supplying a correct anchor caused participants to underestimate the duration of the video in comparison to participants who showed no bias when the anchor was self-generated.

4.2.3. Sum versus whole

To examine the difference between the sum of the estimates and the whole estimate, estimates for all three videos were added together. For participants in the correct anchor supplied condition, the actual duration for medium (fingerprint) video was used since they did not supply an estimate. The 2 (sum v whole) × 2 (order of videos) × 2 (correct anchor supplied v anchor self-generated) mixed model ANOVA indicated that, unlike the previous two studies, there was no difference between the sum of the estimates and the whole duration, $F(1,55) = 0.299$, $p = .586$, $\eta_p^2 = 0.005$. Only the effect of supplying the correct anchoring was significant, $F(1,55) = 8.23$, $p = .006$, $\eta_p^2 = 0.130$: when the anchor was self-generated, participants tended to overestimate both in terms of the sum of their estimates ($M = 0.072$, $SE = 0.027$) and for their overall estimate ($M = 0.087$, $SE = 0.028$), but when given a correct anchor, participants slightly underestimate their sum ($M = -0.028$, $SE = 0.014$) and their overall estimate ($M = -0.028$, $SE = 0.017$).

4.2.4. Familiarity, liking, and attention

Participants did not appear to be familiar with the videos: the percentage of participants indicating that they were unfamiliar with the video was 67% for the fingerprint video, 83% for the grapefruit video, and 81% for the ancient grain video. In comparing ratings for the grapefruit (short) and ancient grain (long) videos, which were rated by all participants, there were no differences in familiarity for the grapefruit video ($M = 1.17$, $SE = 0.050$) or the ancient grain (long) video ($M = 1.20$, $SE = 0.059$), $F(1,55) = 0.612$, $p = .437$, $\eta_p^2 = 0.011$. Again, participants gave lower ratings of liking for the ancient grain (long) video ($M = 2.90$, $SE = 0.116$) than they did for the grapefruit (short) video ($M = 3.39$, $SE = 0.102$), $F(1,55) = 19.62$, $p < .001$, $\eta_p^2 = 0.263$. Similarly, participants gave lower ratings of attention spent on the video for the ancient grain video ($M = 3.34$, $SE = 0.139$)

than they did for the grapefruit video ($M = 3.85$, $SE = 0.102$) $F(1,55) = 14.64$, $p < .001$, $\eta_p^2 = 0.210$.

As with the previous study, individual differences in liking and attention for the videos did not appear to be related to individual differences in bias for estimated duration for the videos. Multiple regression analysis for the pair of videos examining if differences in familiarity, liking and attention could predict differences in bias for estimated duration was not significant for the ancient grain and grapefruit videos, $F(3,58) = 2.05$, $p = .118$, $R^2 = 0.100$. Similarly, within condition, familiarity, liking and attention were not significant predictors of bias for either the grapefruit, $F(3,58) = 0.665$, $p = .158$, $R^2 = 0.034$, or the ancient grain, $F(3,58) = 0.711$, $p = .550$, $R^2 = 0.037$, video.

4.3. Discussion

Results from this study indicate again that bias in the original estimate influenced bias in all subsequent estimates. Participants that estimated the duration of the initial video on fingerprints tended to overestimate the duration. Bias in the initial estimate was then transferred to subsequent estimates with these participants also overestimating the duration of the grapefruit video and the overall estimate. Supplying a correct anchor removed overestimation for these two estimates. However, supplying participants with the correct duration of the initial video caused them to underestimate the duration of the ancient grain video. Removing the tendency to overestimate for one video increased the tendency to underestimate for another. Instead of reducing bias for all subsequent estimations, supplying a correct anchor simply shifted that pattern of bias found when no anchor was supplied. To put another way, anchoring did remove the tendency for the total duration estimate to be biased. It did so by shifting the overall pattern of bias that followed.

Unlike the previous two studies, there was no difference found between the sum of the individual estimates and the overall estimate. Even though the sample size is fairly small (as with Study 2), this does not appear to be a power issues as the effect size was < 1%. It is possible that starting with the medium video shifted bias such that bias in the overall estimate was now more similar to bias in the sum of the individual estimates. In support, Study 2 found that bias for the medium video was not significantly different from bias in the overall estimate. It may be that starting with the medium video helped center the individual estimates so that they were more in line with the overall estimate.

5. General discussion

Results from three studies indicate that anchoring, unpacking and summing influenced estimation for remembered task duration in a manner similar to that found for predicted task duration (Roy et al., 2005). Bias in the first estimation given served as an anchor for all following estimates. Estimating individual components first (unpacking) led to longer estimates of the overall duration. Additionally, the summed separate estimates for a number of individual tasks differed from the overall estimate of duration.

5.1. Anchoring

In all three studies, findings indicated that bias in the first estimation influenced bias in all subsequent estimations. If participants greatly overestimated the first task because it was very short (Study 2), then all subsequent tasks were also more likely to be overestimated. The bias in the first estimation given served as an anchor for all following estimates. Note that in this way the effect of anchoring is qualitatively different from that found in other anchoring studies in time estimation. In the majority of studies examining anchoring in estimation of future task duration, estimates were pushed up and down by the absolute value of the anchor with short anchors leading to underestimation and long anchors leading to overestimation (König, 2005; König et al., 2015. Thomas & Handley, 2008; Thomas et al., 2007; Thomas et al., 2003). Here, in essence, the relationship was reversed. In Study 2, starting estimation with the shortest of the clips led to an increased likelihood of overestimation, likely because shorter tasks tend to be overestimated (Lejeune & Wearden, 2009; Roy et al., 2013; Roy & Christenfeld, 2008; Tobin & Grondin, 2009; Yarmey, 2000). Participants were aware of the relative duration of the stimuli – which were shorter and which were longer – and in maintaining those relative relationships applied the bias on the first estimate to subsequent estimates.

In support, the results of Study 3 indicate that supplying a correct anchor removed the tendency to overestimate the duration for the shorter video and for the overall duration but *increased* the tendency to underestimate for the longer video. Instead of correcting all following estimates (e.g. Roy et al., 2008), the anchor appeared to shift the pattern of bias, with some estimates improved but others impaired.

5.2. Unpacking and summing

Similar to studies on predicted task duration (Connolly & Dean, 1997; Hadjichristidis et al., 2014; Kruger & Evans, 2004; Liu et al., 2014; Min & Arkes, 2012; Tsai & Zhao, 2011), unpacking caused an increase in the estimated overall duration of the stimuli. In Study 2, half the participants gave the overall estimation before estimating the individual components while the other half estimated the individual components first. The tendency to overestimate was greatest when the overall estimate was given after the individual estimates (unpacking condition). Because estimates of the overall duration were fairly unbiased when given without unpacking, unpacking increased bias in overall duration with participants significantly overestimating how long the videos lasted. The influence of individual estimates on the overall estimate was reciprocal: giving the overall estimate first also influenced the individual components resulting in shorter subsequent estimates for the individual videos.

As with other research on the effect of summing (Forsyth & Burt, 2008; Tversky & Koehler, 1994; Van Boven & Epley, 2003), our results indicated that the sum of the individual estimates did not always add up to the overall estimate. The sum of the individual components was longer than the overall estimate in Study 1 and shorter in Study 2. The added individual components were no more accurate than the overall duration estimate. What is remarkable about this study is that difference in sum and whole happened in a within-subjects design, unlike previous research that was mostly between subjects. The incongruity between sum and whole happened even though all estimates were given in sequence.

These results indicate the potential mechanism that causes a shift in estimation for both unpacking and summing. Both interventions require participants to make a series of estimates. As had been speculated (Hadjichristidis et al., 2014), it was the anchoring provided by the first estimate that influenced the shift in both the individual estimates and the whole duration estimate that came after. In both summing (Forsyth & Burt, 2008) and unpacking (Connolly & Dean, 1997; Hadjichristidis et al., 2014; Kruger & Evans, 2004; Liu et al., 2014; Min & Arkes, 2012; Tsai & Zhao, 2011) the common result is for an increase in estimated duration. This is likely because the shorter subcomponent that is estimated first is more likely to be overestimated (Lejeune & Wearden, 2009; Roy et al., 2013; Roy & Christenfeld, 2008; Tobin & Grondin, 2009; Yarmey, 2000). The tendency for an initial estimate to be longer is then passed on to all subsequent estimations, including the sum and the whole.

5.3. Limitations

We found similar results using two different sets of stimuli and two different samples of participants. What was found through serendipity in Study 1 was confirmed through purposeful, targeted research in Studies 2 and 3. However, all studies employed individual tasks that lasted a few minutes and were all similar in overall duration (approximately 2 to 4 min). It is possible that the manipulations used here could have very different effects for tasks that last just a few seconds or that take hours to complete. Further, all studies employed exactly three component parts. Future research should vary both the duration and number of the subcomponents. Finally, as noted in Studies 2 and 3, larger sample sizes than those used here would allow for more detailed examination of the factors manipulated in these studies.

Further research should also investigate in more detail the specific instances where bias is likely for the stimuli that is estimated first. In Study 1 we speculated that experience with the task could explain the pattern of results, but this explanation should be directly tested in future research. It would also be worthwhile to examine the impact of regression (Fraisse, 1963) and rounding (Forsyth & Burt, 2008; Huttenlocher et al., 1990) on bias when repeated estimates are given.

5.4. Summary

Results indicate that memory for past task duration can be easily manipulated by the manner in which it is elicited and the structure of what is to be estimated. None of the techniques examined here – anchoring, unpacking and summing – led to consistently improved estimates of duration for these studies. Rather, these techniques caused a shift in bias that sometimes improved estimation but at other times made it worse. It was dependent upon the starting point of original bias in estimation to determine whether these interventions improved or exacerbated bias in remembered duration. These results could have important implications for a number of real-world settings where retrospective estimations for duration are important such as witness testimony about duration of a crime (Flowe, Mehta, & Ebbesen, 2011) and billable hours spent on a task such as software development (Jorgensen, 2004). For example, in witness testimony remembered duration might be important in establishing likelihood of events and alibis (Loftus, Schooler, Boone, & Kline, 1987). Results here indicate that how the memories are elicited could be important for juries to consider. Further, given that estimations of future task duration are based on memories of previous task duration (Roy et al., 2005, 2008; Roy & Christenfeld, 2007, 2008), how memories of past task duration are elicited could influence the amount of time allocated for similar tasks in the future.

Acknowledgements

Research for Study 1 supported by the Elizabethtown College Summer Scholarship, Creative Arts and Research Program 2015.

Appendix A

Measures

Center for Epidemiologic Studies Depression Scale (**CES-D**; Radloff, 1977). This 20-item self-report scale measures individual depression using a 4-point Likert-type scale (0 = *Rarely or None*, 3 = *Mostly or Always*). Four statements within the scale indicate positive factors, rather than negative factors, with a total score one can obtain between 0 and 60 points. The lower the score, the less symptoms of depression indicated. Participants were asked to indicate how often they had certain thought or feelings such as how often “I had crying spells” or “I thought my life had been a failure”.

Ruminative Response Scale (**RRS**; Treynor et al., 2003). The 10-item self-report subscale of the RRS identifies reflective rumination and brooding rumination on a 5-point Likert scale (1 = *almost never*, 5 = *almost always*). Statements of reflective rumination include, how often do you “go away by yourself and think about why you feel the way you do.” Statements of brooding rumination include, “how often do you think, ‘what am I doing to deserve this?’”.

Uses of Music Inventory (**UOM**; Chamorro-Premuzic & Furnham, 2003). This 15-item questionnaire identifies participant motives for listening to music on a 5-point Likert-type scale (1 = *strongly disagree*, 5 = *strongly agree*). Three subscales within the survey identify emotional uses of music (e.g.: “listening to music really affects my mood”), cognitive uses of music (e.g.: “I often enjoy analyzing complex musical compositions”), and background use of music (e.g.: “I enjoy listening to music while I work”).

Musical Training Questionnaire. This questionnaire was created to ask participants about specific musical training experiences. Years of formal and informal training, and types of instruments played, were recorded. Self-reports of how often participants listen to music, think about music, use a metronome during practice, and play to other tracks during practice were rated on a 5-point Likert-type scale (1 = *not at all*, 5 = *most of the time*). Listening to music for enjoyment, playing to music for enjoyment, and a self-report of one's musical meter was rated on a 5-point Likert-type scale (1 = *completely disagree*, 5 = *completely agree*). Lastly, a self-report rating of overall musical ability was measured on a 10-point scale (1 = *novice*, 10 = *expert*).

Results

Multiple regression equations with CES, RRS, UOM, and musical training used to predict bias in estimated duration were not significant for the song, $F(7,380) = 1.36$, $p = .222$, $R^2 = 0.025$, the news clip, $F(7,380) = 1.44$, $p = .186$, $R^2 = 0.026$, or for the questionnaire, $F(7,380) = 1.69$, $p = .111$, $R^2 = 0.031$. The overall regression equation was significant for the total estimate, $F(7,380) = 2.20$, $p = .034$, $R^2 = 0.040$, with the only significant individual predictor being that participants who preferred to listen to music for cognitive reasons (appreciating the complexity of the music) underestimated overall duration, $\beta = -0.112$, $p = .05$. It is not clear why this one aspect of music appreciation would be predictive of bias in estimation for the total duration and not for any of the individual predictions, including the song. It is possible that the relationship between bias in time estimation and a more cognitive appreciation of music was simply a type I error, but it is possible that a preference for complex music might indicate thoughtfulness that could somehow translate into differences in how duration is estimated.

References

Bartoshuk, L. M., Duffy, V. B., Fast, K., Green, B. G., & Snyder, D. J. (2002). Labeled scales (e.g., category, Likert, VAS) and invalid across-group comparisons: What we have learned from genetic variation in taste. *Food Quality and Preference, 14*, 125–138.

Bisson, N., Tobin, S., & Grondin, S. (2012). Prospective and retrospective time estimates of children: A comparison based on ecological tasks. *PLoS One, 7*(3), e33049.

Block, R. A., & Zakay, D. (1997). Prospective and retrospective durations judgments: A meta-analytic review. *Psychonomic Bulletin & Review, 4*, 184–197.

Brumfield, G. (2015, April 15). Space station gets a coffee bar [radio broadcast episode]. *Morning edition*. Washington, DC: National Public Radio.

Buehler, R., Griffin, D., & Peetz, J. (2010). The planning fallacy: Cognitive, motivational, and social origins. *Advances in experimental social psychology (Vol. 43, pp. 1-62)*. Academic Press.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*, 3–5.

Byram, S. J. (1997). Cognitive and motivational factors influencing time prediction. *Journal of Experimental Psychology: Applied, 3*, 216–239.

Chamorro-Premuzic, T., & Furnham, A. (2003). Personality predicts academic performance: Evidence from two longitudinal university samples. *Journal of Research in Personality, 37*, 319–338.

Connolly, T., & Dean, D. (1997). Decomposed versus holistic estimates of effort required for software writing tasks. *Management Science, 43*, 1029–1045.

Flowe, H. D., Mehta, A., & Ebbesen, E. B. (2011). The role of eyewitness identification evidence in felony case dispositions. *Psychology, Public Policy, and Law, 17*, 140–159. https://doi.org/10.1037/a0021311.

Forsyth, D. K., & Burt, C. D. (2008). Allocating time to future tasks: The effect of task segmentation on planning fallacy bias. *Memory & Cognition, 36*, 791–798.

Fraisse, P. (1963). *The psychology of time.* New York: Harper & Row.

Furnham, A., & Boo, H. C. (2011). A literature review of the anchoring effect. *The Journal of Socio-Economics, 40*, 35–42. https://doi.org/10.1016/j.socec.2010.10.008.

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

Hadjichristidis, C., Summers, B., & Thomas, K. E. (2014). Unpacking estimates of task duration: The role of typicality and temporality. *Journal of Experimental Social Psychology, 51*, 45–50.

Halkjelsvik, T., & Jørgensen, M. (2012). From origami to software development: A review of studies on judgment-based predictions of performance time. *Psychological Bulletin, 138*, 238–271.

Huttenlocher, J., Hedges, L. V., & Bradburn, N. M. (1990). Reporting of elapsed time: Bounding and rounding processes in estimation. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 16*, 196–213.

Jones, M., Roy, M. M., & Verkuilen, J. (2014). The relationship between reflective rumination and musical ability. *Psychology of Aesthetics, Creativity, and the Arts, 8*, 219–226. https://doi.org/10.1037/a0035634.

Jorgensen, M. (2004). Top-down and bottom-up expert estimation of software development effort. *Information and Software Technology, 46*, 3–16.

König, C. J. (2005). Anchors distort estimates of expected duration. *Psychological Reports, 96*, 253–256. https://doi.org/10.2466/pr0.96.2.253-256.

König, C. J., Wirz, A., Thomas, K. E., & Weidmann, R. Z. (2015). The effects of previous misestimation of task duration on estimating future task duration. *Current Psychology, 34*, 1–13.

Kruger, J., & Evans, M. (2004). If you don't want to be late, enumerate: Unpacking reduces the planning fallacy. *Journal of Experimental Social Psychology, 40*, 586–598. https://doi.org/10.1016/j.jesp.2003.11.001.

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*, 941–960.

Liu, Y., Li, S., & Sun, Y. (2014). Unpacking a time interval lengthens its perceived temporal distance. *Frontiers in Psychology, 5*, 1345. https://doi.org/10.3389/fpsyg.2014.01345.

Loftus, E. F., Schooler, J. W., Boone, S. M., & Kline, D. (1987). Time went by so slowly: Overestimation of event duration by males and females. *Applied Cognitive Psychology, 1*, 3–13.

Macan, T. H. (1994). Time management: Test of a process model. *Journal of Applied Psychology, 79*, 381–391.

Min, K. S., & Arkes, H. R. (2012). When is difficult planning good planning? The effects of scenario-based planning on optimistic prediction bias. *Journal of Applied Social Psychology, 42*, 2701–2729.

Ornstein, R. E. (1969). *On the experience of time.* Middlesex, England: Penguin Books.

Panagiotidi, M., & Samartzi, S. (2013). Time estimation: Musical training and emotional content of stimuli. *Psychology of Music, 41*, 620–629.

Poynter, D. (1989). Judging the duration of time intervals: A process of remembering segments of experience. In I. Levin, & D. Zakay (Eds.). *Time and human cognition: A life-span perspective* (pp. 305–322). Amsterdam: North-Holland.

Radloff, L. S. (1977). The CES-D scale: A new self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385–401.

Roy, M. M., & Christenfeld, N. J. S. (2007). Bias in memory predicts bias in estimation of future task duration. *Memory & Cognition, 35*, 557–564.

Roy, M. M., & Christenfeld, N. J. S. (2008). Effect of task length on remembered and predicted duration. *Psychonomic Bulletin & Review, 15*, 202–208.

Roy, M. M., Christenfeld, N. J. S., & Jones, M. (2013). Actors, observers and the estimation of task duration. *Quarterly Journal of Experimental Psychology, 66*, 121–137. https://doi.org/10.10080/17470218.2012.699973.

Roy, M. M., Christenfeld, N. J. S., & McKenzie, C. R. M. (2005). Underestimation of future duration: Memory incorrectly used or memory bias. *Psychological Bulletin, 131*, 738–756.

Roy, M. M., Mitten, T. S., & Christenfeld, N. J. S. (2008). Correcting memory improves accuracy of predicted duration. *Journal of Experimental Psychology: Applied, 14*, 266–275.

Roy, M. M., Radzevick, J. R., & Getz, L. (2016). The manifestation of stress and rumination in musicians. *Muziki, 13*, 100–112. https://doi.org/10.1080/18125980.2016.1182385.

SciShow (n.d.). Home [YouTube channel]. Retrieved from https://www.youtube.com/user/scishow/featured.

Thomas, K. E., & Handley, S. J. (2008). Anchoring in time estimation. *Acta Psychologica, 127*, 24–29.

Thomas, K. E., Handley, S. J., & Newstead, S. E. (2004). The effects of prior experience on estimating the duration of simple tasks. *Current Psychology of Cognition, 22*, 83–100.

Thomas, K. E., Handley, S. J., & Newstead, S. E. (2007). The role of prior task experience in temporal misestimation. *Quarterly Journal of Experimental Psychology, 60*, 230–240.

Thomas, K. E., Newstead, S. E., & Handley, S. J. (2003). Exploring the time prediction process: The effect of task experience and complexity on prediction accuracy. *Applied Cognitive Psychology, 17*, 655–673.

Tobin, S., Bisson, N., & Grondin, S. (2010). An ecological approach to prospective and retrospective timing of long durations: A study involving gamers. *PLoS One, 5*(2), e9271.

Tobin, S., & Grondin, S. (2009). Video games and the perception of very long durations by adolescents. *Computers in Human Behavior, 25*, 554–559.

Tobin, S., & Grondin, S. (2012). Time perception is enhanced by task duration knowledge: Evidence from experienced swimmers. *Memory & Cognition, 40*, 1339–1351.

Tobin, S., & Grondin, S. (2015). Prior task experience affects temporal prediction and estimation. *Frontiers in Psychology, 6*, 916.

Treynor, W., Gonzalez, R., & Nolen-Hoeksema, S. (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research, 27*, 247–259. https://doi.org/10.1023/A:1023910315561.

Tsai, C. I., & Zhao, M. (2011). Predicting consumption time: The role of event valence and unpacking. *Journal of Consumer Research, 38*, 459–473.

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review, 101*, 547–567. https://doi.org/10.1037//0033 295x.101.4.547.

Van Boven, L., & Epley, N. (2003). The unpacking effect in evaluative judgments: When the whole is less than the sum of its parts. *Journal of Experimental Social Psychology, 39*, 263–269. https://doi.org/10.1016/s0022-1031(02)00516-4.

Verhaeghen, P., Khan, R., & Joormann, J. (2005). Why we sing the blues: The relation between self-reflective rumination, mood, and creativity. *Emotion, 5*, 226–232. https://doi.org/10.1107/1528-3542.5.2.22.

Wallace, M., & Rabin, A. I. (1960). Temporal experience. *Psychological Bulletin, 57*, 213–236.

Yarmey, A. D. (2000). Retrospective duration estimations for variant and invariant events in field situations. *Applied Cognitive Psychology, 14*, 45–57.