Judgments of effort depend on the temporal proximity to the task

by

Michelle Roshan Marie Ashburner

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Arts

in

Psychology

Waterloo, Ontario, Canada, 2019

Declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the 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.

Statement of Contributions

This thesis was co-authored with Dr. Evan Risko.

Abstract

Cognitive effort is a central construct in our lives, yet our understanding of the processes underlying our perception of effort are limited. Performance is typically used as one way to assess effort in cognitive tasks (e.g., tasks that take longer are generally thought to be more effortful); however, Dunn and Risko (2016) reported a recent case where such "objective" measures of effort were dissociated from judgments of effort (i.e., subjective effort). This dissociation occurred when participants either made their judgments of effort after the task (i.e., reading stimuli composed of rotated words) or without ever performing the task. This leaves open the possibility that if participants made their judgments of effort closer in time to the actual experience of performing the task (e.g., right after a given trial) that these judgments might better correspond to putatively "objective" measures of effort. To address this question, we conducted two experiments replicating Dunn and Risko (2016) with additional probes for immediate judgments of effort (i.e., a judgment of effort made right after each trial). Results provided some support for the notion that judgments of effort more closely follow reading times when made immediately after reading. Implications of the present work for our understanding of judgments of effort are discussed.

Acknowledgements

I would like to thank my research supervisors Dr. Jonathan Fugelsang and Dr. Evan Risko for their dedication to my academic training. Many thanks also to my thesis reader, Dr. Derek Besner. Additionally, I am grateful for the unconditional support offered by my family, friends, and my student colleagues of the Department of Psychology.

Dedication

This thesis is dedicated to the memory of Lindsay Bell.

Table of Contents

List of Figures

List of Tables

1. Introduction

While there exists a long tradition of research investigating subjective workload (Eggemeier & Stadler, 1984; Jex, 1988; Moray 1982; Yeh & Wickens, 1988), the psychological basis of cognitive effort has received increased attention from researchers over the last five years (Dunn, Gaspar, & Risko, 2018; Dunn, Inzlicht, & Risko, 2019; Dunn, Koehler, & Risko, 2017; Inzlicht, Shenhav, & Olivola, 2018; Kool & Botvinick, 2018; Kurzban, 2016; Potts, Pastel, & Rosenbaum, 2018; Yildirim et al., in press). An important dimension of this line of research is attempting to address *how* individuals judge the effort anticipated or experienced on a given task or trial of a given task (Dunn, Koehler, & Risko, 2017; Foo, Uy, & Baron, 2009; Gweon, Asaba, & Bennett-Pierre, 2017; Marshall, 2002; Song & Schwarz, 2008; Westbrook, Kester, & Braver, 2013). That is, when someone is asked how effortful they found a task to be (or will be in the case of a prospective judgment), what factors determine their judgment of effort (i.e., subjective effort)? In the present investigation we examine whether the temporal proximity between task performance and effort judgment modulates the information brought to bear on those judgments.

One way to think about judgments of effort, at least in the context of the types of cognitive tasks discussed herein, is as a type of metacognitive judgment (a judgment about one's cognitive processes; Dunn & Risko, 2016; Dunn et al., 2019). Metacognitive judgments, according to an influential framework, are viewed as inferential in nature (Koriat, 1993). That is, individuals rely on available cues to infer, for example, the likelihood that they will recall some information in the future (a judgment of learning); or, as is the focus here, the effortfulness of a given task (a judgment of effort). In general, such available cues can be roughly divided into experiential cues, for example, fluency; and belief-based cues, for example, the belief that task A is more inherently effortful than is task B. This perspective differs from the position that the effortfulness of cognitive acts can be directly accessed via an internal signal.

Research investigating metacognitive judgments has demonstrated that they can be influenced by their proximity to critical cognitive events. For example, there exists evidence that prospective judgments of confidence (i.e., how well one thinks they will perform on an upcoming trial) are less related to task performance than are retrospective judgments (i.e., how well one thinks they performed on the previous trial; Boldt & Gilbert, 2019; Fleming et al., 2016; Gilbert, 2015; Siedlecka, Paulewicz, & Wierzchon, 2016). The benefit for retrospective judgments likely derives from the act of task performance, which itself may provide cues to inform confidence judgments (e.g., experiential cues like fluency; Fleming et al., 2016). In a similar vein, Nelson and Dunlosky (1991) demonstrated that participants' judgments of learning (JOLs) on a paired-associates task were more strongly correlated with performance when made after a delay than when made immediately after learning an item. Nelson and Dunlosky (1991; see also Dunlosky & Nelson, 1997) suggested that delay impacted JOL accuracy because, immediately after study, individuals access both short-term and long-term memory to inform their judgments, where the contribution of the former is misleading. When the item is no longer in short-term memory (i.e., after a delay), a more accurate judgment can be made based on the item's current retrievability from long-term memory (Nelson & Dunlosky, 1991; see also Metcalfe & Finn, 2008; Scheck, Meeter, & Nelson, 2004; Weaver, III & Kelemen, 1997). Consistent with this general idea, Koriat and Ma'ayan (2005) demonstrated that JOLs for paired-associates elicited via a pre-JOL recall test immediately after study were more strongly associated with cues available at encoding (e.g., pre-JOL encoding fluency); whereas JOLs obtained after a delay between study and the pre-JOL recall test were more strongly associated with cues available at recall (e.g., pre-JOL retrieval fluency). Taken together, the research outlined above demonstrates that, by manipulating the proximity of metacognitive judgments to critical cognitive events, the cues brought to bear on metacognitive judgments can differ. Here we examine the influence of temporal proximity to the task on participants' use of time as a cue in making judgments of effort.

Several recent studies (Dunn et al., 2019; Dunn, Lutes, & Risko, 2016; Dunn & Risko, 2016) have examined various cues (e.g., time, errors, intrinsic properties of the stimuli) that are used to inform retrospective judgments of effort. In particular, Dunn and colleagues focused on the extent to which individuals use time, because researchers often use the time to complete a task as an index of fluency or ease of processing (Benjamin, Bjork, & Schwartz, 1998; Koriat & Ma'ayan, 2005; Thompson et al., 2013; Undorf & Erdfelder, 2013, 2015). In addition, time costs are a central factor in making decisions about resource allocation (i.e., time as the currency in making effort-based decisions; Gray et al., 2006). Dunn and Risko (2016) had participants complete multiple trials of a reading task for which there were four stimulus types, each containing 25 words: upright words in an upright frame (UW-UF); upright words in a rotated frame (UW-RF); rotated words in an upright frame (RW-UF); and rotated words in a rotated frame (RW-RF). After the reading task, participants provided judgments of effort for each stimulus type on a 7-point scale. Dunn and colleagues' results revealed that reading times across the stimulus types generally tracked with judgments of effort; however, a consistent dissociation was observed: namely, reading times for the RW-RF stimulus type and the RW-UF stimulus type were equivalent, yet the RW-RF stimulus type was consistently judged as more effortful (e.g., Dunn & Risko, 2016). One possible explanation for this dissociation, put forward by Dunn and Risko (2016), is that individuals, rather than using time, inferred effortfulness based on the stimulus cues available (e.g., stimulus orientation) and their beliefs about their perceptual/cognitive systems (e.g., processing a disoriented stimulus is hard). For example, the RW-RF stimulus type may be believed to be more effortful to read because it has two forms of rotation, whereas the RW-UF stimulus type only has one (see Figure 1).

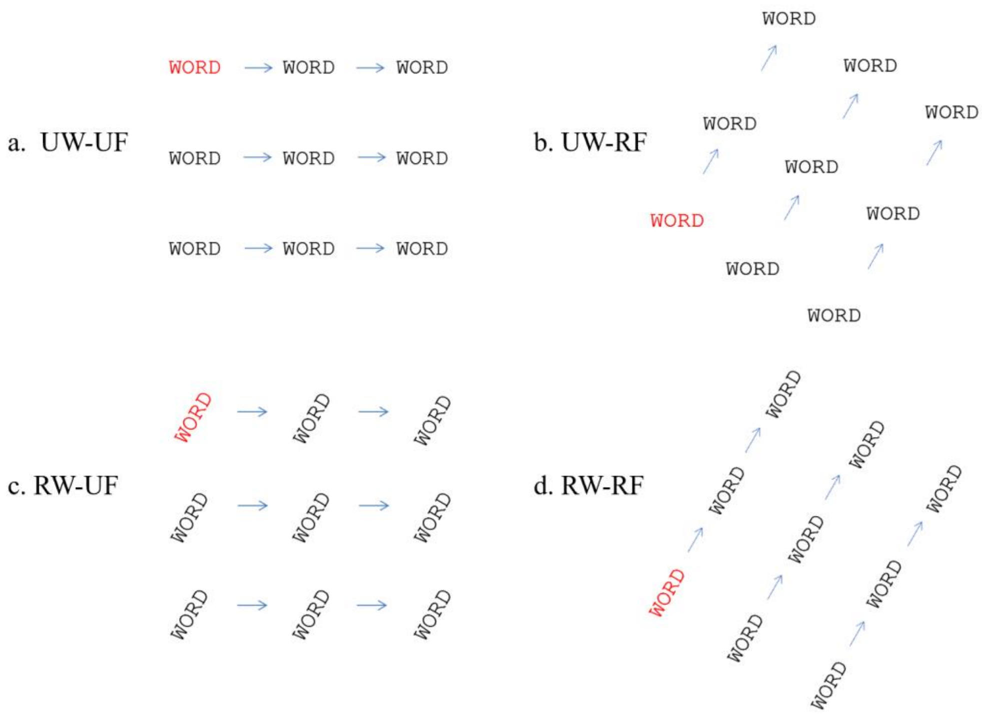

Figure 1. Examples of each stimulus type: a. UW-UF; b. UW-RF; c. RW-UF; d. RW-RF. All rotations above are 60° counterclockwise. For illustrative purposes, each stimulus display above contains nine words; however, the displays used by Dunn and Risko (2016) and here contained 25 words.

The dissociation between time and judgments of effort observed by Dunn and Risko (2016) provides a unique opportunity to investigate how the temporal proximity between a cognitive act and a judgment of effort might influence the use of a potentially important effort cue (i.e., time, fluency). As described above, Dunn and Risko (2016) used post-task judgments of effort such that participants read numerous displays of each type then completed a judgment of effort using a generic instance of each type. Thus, participants were making judgments of effort temporally separated from the experience of reading each display. This raises the interesting possibility that the dissociation between reading time, a more experiential cue, and judgments of effort might reflect a greater reliance on beliefs when there is a delay between the experience associated with the cognitive act (i.e., reading) and the judgment. We provide a direct test of this idea here.

In the present study, we used the same stimulus types and task as Dunn and Risko (2016) and added probes for immediate (i.e., made right after each trial) judgments of effort in addition to a similar series of post-task judgments of effort. Moreover, while the design of Dunn and Risko (2016) included a block wherein participants were free to rotate their heads as they read, in the current study participants were instructed to keep their head upright while reading. If the dissociation between judgments of effort and reading times reported by Dunn and Risko (2016) is due to the delay between the experience of reading and the judgment of effort, then this dissociation should be reduced or eliminated using immediate judgments of effort. Alternatively, if this dissociation is not due to the delay between the experience and judgment of effort, similar results should be observed for immediate judgments as for post-task judgments.

2. Experiment 1

2.1. Method

2.1.1. Participants

Thirty-two University of Waterloo undergraduate students participated in this experiment in exchange for course credit. This sample size was chosen based on previous research (Dunn & Risko, 2016).

2.1.2. Apparatus

The presentation of stimuli and recording of participants' responses were handled by E-Prime 3.0 software. Participants viewed all stimuli and instructions on a widescreen 22-inch LG monitor while seated at a desk and used a QWERTY keyboard for response entry. A Logitech web camera fixed on top of the monitor was used to capture audio and video of each session; the former was used to measure reading time and error count, and the latter for coding head movement.

2.1.3. Stimuli and Design

A one-factor (stimulus type) within-subjects design was employed. Each slide consisted of a stimulus composed of 25 words: 5 rows of 5 words each. Words were typed in 18-point black Courier New font on a white background. For the practice and test trials, all words were five letters in length, contained one or two syllables, and had a Kucera-Francis written word frequency ranging from 1 to 2,724 per million. For the post-test slides, all words were "WORD". There were four stimulus types: stimuli with upright words and an upright frame (UW-UF), stimuli with upright words and a rotated frame (UW-RF), stimuli with rotated words and an upright frame (RW-UF), and stimuli with rotated words and a rotated frame (RW-RF). Each stimulus type was presented eight times during test; for each of the disoriented stimulus types (i.e., UW-RF, RW-UF, and RW-RF), four of the eight were rotated counterclockwise by 60° and four were rotated clockwise by 60°.

The starting word on each slide was coloured red and blue single-headed arrows marked the direction in which participants were instructed to read. See Figure 1 above for an illustrative example of stimulus types containing 3 rows of 3 words each. A pre-ordered list of trials was made for each participant consisting of 32 randomly selected and ordered stimuli (i.e., 25-word displays). There were 33 stimuli in total from which to draw and, in lieu of randomly removing one stimulus from this set, each participant's list was formed by selecting 32 of these 33 arrays. A list of five additional stimuli that did not vary across participants comprised the practice trials. Finally, seven slides of post-test stimuli were composed entirely of the word "WORD," one slide per stimulus type and, for disoriented stimuli types, one slide per direction of rotation.

2.1.4. Procedure

Participants entered the testing room and were seated facing the centre of the monitor at about eye-level, and the keyboard was on the desk directly in front of them. After providing consent, video and audio recording began. Instructions were displayed on the screen as a research assistant (who knew about the nature of the experiment) read them and answered any questions. Participants were instructed to read each word on a slide aloud as quickly and as accurately as possible, while keeping their head upright and limiting movement. Participants were not instructed to correct errors, or to ignore them; only to proceed in the way they felt would best follow the given instructions. Moreover, errors were coded offline using the audio capture, so as not to distract participants. Once finished, participants were instructed to press the spacebar to advance the slide to a 7-point effort scale (*1 = not at all effortful*, *4 = somewhat effortful*, *7 = very effortful*) where they were asked to consider the stimulus on the previous slide and, using the keyboard, make a judgement of effort regarding their experience of reading. Upon entering their judgment, the next trial began. The five practice stimuli were first presented, in the same order for every participant: UW-UF, RW-UF, UW-RF, RW-UF, and RW-RF (only four practice trials were included in Experiment 2 so that each stimulus type appeared exactly once). After the practice trials were complete, the research assistant left the room. These practice trials were followed by 32 test trials and corresponding judgments of effort. Following this were post-task judgments of effort, wherein participants were asked to view a generic stimulus type (i.e., each word was "WORD") of each orientation and direction, for a total of seven slides (one UW-UF display and two of each disoriented display), presented in random order to each participant. Participants were provided written instructions to view the displays, but not to read them, and provide an overall judgment of effort for each stimulus type on a 7-point scale (*1 = not at all effortful*, *7 = very effortful*) using the keyboard. The experiment then concluded and participants were debriefed.

2.2. Results

Three participants were replaced. One participant experienced difficulty pronouncing many of the words; another read each word at a pace of over two seconds; and another did not wish to be video recorded. An operational error was detected at the time of coding due to instructions having been given incorrectly. Particularly, participants were given information about the nature of the experiment that could have biased their judgments. This error resulted in the need to remove these data and recruit 12 additional participants. Due to a programming error, lists for two participants were not equally composed of each stimulus type; these data were retained. The final sample had 32 participants. Any trials during which an obvious head movement was made to facilitate reading were removed from analyses. Thirty trials in all were removed for this reason; 0.8% (of total) UW-UF, 0.8% UW-RF, 2.0% RW-UF, and 8.2% RW-RF.[1] In an additional 7.3% of trials, participants prematurely pressed the spacebar before reading the last word on the slide; these trials were also removed. Finally, one trial was removed because the participant paused for an extended time (2,422 ms) after a mispronunciation. After these exclusions, a within-subject, within stimulus type search for outliers at the trial level found no reading times, error count scores, or immediate judgments of effort with $|z| > 3$. Therefore, no observations were trimmed from these data. In total, 10.2% of observations were removed. Provided the uneven exclusions for head rotations (see above), we conducted a second set of analyses excluding participants who had a disproportionate rate of head rotation for the RW-RF stimulus type, in order to have an approximately equal distribution. We excluded three participants to form a *head-tilt control subset* and the resulting proportion of trials in these data for each stimulus type was: 3.9% UW-UF, 1.3% UW-RF, 0.0% RW-UF, and 0.4% RW-RF. Overall, the results were similar. When an important deviation from the reported results (i.e., the complete sample) was found we note it in the appropriate section. All analyses were run using the open-source statistical analysis software R, version 3.4.4. The code and data are available on the Open Science Framework project webpage: https://osf.io/tgx85/. See Table 1 for mean reading time, error count, immediate judgments of effort, and post-task judgments of effort by stimulus type.

[1] This pattern of spontaneous head rotation is consistent with the results of Dunn and Risko (2016), where participants were free to rotate their heads while reading.

Table 1

Experiment 1 mean reading time (ms), mean error count, and mean judgments of effort

	Stimulus Type			
	UW-UF	UW-RF	RW-UF	RW-RF
Reading Time	15,981 (3,620)	16,337 (3,741)	17,111 (3,947)	17,338 (4,239)
Error count	1.32 (1.27)	1.14 (1.17)	1.28 (1.31)	1.33 (1.46)
Immediate judgments	2.14 (1.05)	2.63 (1.22)	3.17 (1.36)	3.57 (1.46)
Post-test judgments	1.47 (0.80)	2.33 (1.31)	2.78 (1.28)	3.52 (1.53)

Note: Standard deviations in parentheses. Immediate and post-task judgments of effort are on 7-point scales

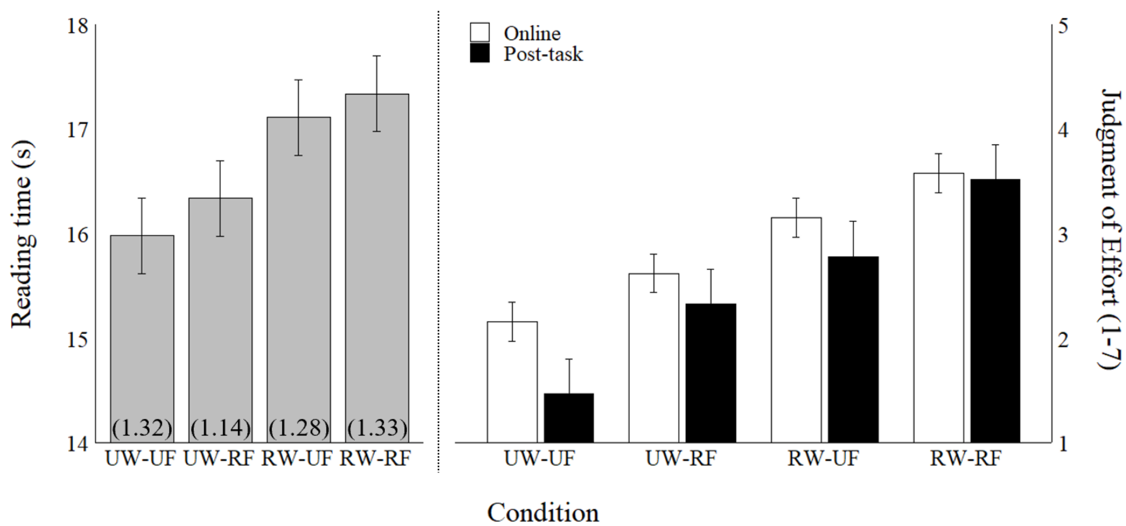

Figure 2. Reading times by stimulus type (left panel) and immediate vs. post-task judgments of effort by stimulus type (right panel) for Experiment 1. Average error count per stimulus type reported in parentheses (left panel). Error bars are Masson-Loftus 95% CI (Loftus & Masson, 1994).

2.2.1. Reading Times

Reading times for each trial were collected by the E-Prime software, measured in milliseconds from stimulus onset to when the spacebar was pressed. A one-way repeated measures ANOVA with stimulus type as the factor, corrected for sphericity violations, revealed a significant effect of stimulus type on reading time, $F(2.11, 65.52) = 16.92$, $p < .001$, $\eta_g^2 = .03$. Pairwise *t*-tests were conducted across stimulus types. Compared to UW-UF trials, individuals were no slower on UW-RF trials, $t(31) = 1.74$, $p = .092$, $d = 0.31$; however, individuals were slower on RW-UF trials, $t(31) = 5.19$, $p < .001$, $d = 0.92$, and on RW-RF trials, $t(31) = 4.87$, $p < .001$, $d = 0.86$. Compared to UW-RF trials, individuals were slower on RW-UF trials, $t(31) = 4.26$, $p < .001$, $d = 0.75$, and RW-RF trials, $t(31) = 3.82$, $p = .001$, $d = 0.68$. Individuals were not significantly slower on RW-RF trials than on RW-UF trials, $t(31) = 1.07$, $p = .292$, $d = 0.19$; $BF_{01} = 3.13$. Provided the importance of the contrast between the RW-UF and the RW-RF stimulus types, the Bayes factors are presented exclusively for this last contrast. When conducting this analysis on the head-tilt control subset, the results were qualitatively similar except that individuals were significantly slower on UW-RF trials than on UW-UF trials, $t(28) = 2.53$, $p = .017$, $d = 0.47$, which was marginal in the complete analysis. Moreover, while individuals were not significantly slower on RW-RF trials than on RW-UF trials, the Bayes Factor in support of the null for this contrast was 1.85.

2.2.2. Error Count

The error count for each trial was coded as the number of errors made while reading. An error was added when a sound or syllable in a word was repeated more than once; when a word was repeated more than once; when a word was missed; or other serious mispronunciations including pluralizing a singular word or reading the singular version of a word presented in plural form. Pauses within words were not counted as errors, nor were utterances of “um” or similar filler words. A one-way repeated measures ANOVA with stimulus type as the factor revealed that there was no effect of stimulus type on error count, $F(2.29, 71.12) = 1.95$, $p = .144$, $\eta_g^2 = .01$.

2.2.3. Effort – Immediate Judgment

Immediate judgments of effort were assessed through the report of participants’ perceived effort immediately after each trial. Each immediate judgment of effort was between 1 (*not at all effortful*) and 7 (*very effortful*). A one-way repeated measures ANOVA with stimulus type as the factor revealed a significant effect of stimulus type on immediate judgments of effort, $F(3, 93) = 44.23$, $p < .001$, $\eta_g^2 = .26$. Pairwise *t*-tests were conducted across stimulus types. Compared to UW-UF trials, individuals reported more effort on UW-RF trials, $t(31) = 3.85$, $p = .001$, $d = 0.68$; on RW-UF trials, $t(31) = 6.63$, $p < .001$, $d = 1.17$; and on RW-RF trials, $t(31) = 9.77$, $p < .001$, $d = 1.73$. Compared to UW-RF trials, individuals reported more effort on RW-UF trials, $t(31) = 4.09$, $p < .001$, $d = 0.72$; and on RW-RF trials, $t(31) = 7.37$, $p < .001$, $d = 1.30$. Critically, individuals reported significantly more effort on RW-RF trials than on RW-UF trials, $t(31) = 3.86$, $p = .001$, $d = 0.68$; $BF_{10} = 56.21$.

2.2.4. Effort – Post-task Judgment

A one-way repeated measures ANOVA with stimulus type as the factor revealed a significant effect of stimulus type on post-task judgments of effort, $F(3, 93) = 25.87$, $p < .001$, $\eta_g^2 = .30$. Pairwise *t*-tests were conducted across stimulus types. Compared to the UW-UF condition, individuals reported greater effort for the UW-RF condition, $t(31) = 3.70$, $p = .001$, $d = 0.65$; for the RW-UF condition, $t(31) = 6.24$, $p < .001$, $d = 1.10$; and for the RW-RF condition,

$t(31) = 7.05$, $p < .001$, $d = 1.25$. Compared to the UW-RF condition, individuals reported more effort for the RW-UF condition, $t(31) = 2.05$, $p = .049$, $d = 0.36$; and the RW-RF condition, $t(31) = 5.24$, $p < .001$, $d = 0.93$. Finally, individuals reported significantly more effort for the RW-RF condition than for the RW-UF condition, $t(31) = 3.06$, $p = .004$, $d = 0.54$; $BF_{10} = 8.74$.

2.2.5. Direction of Rotation

While the UW-UF stimulus type appeared upright, the other three were either rotated in a counterclockwise or a clockwise direction. To test for main effects of direction, and the stimulus type by direction interaction, a 3 (stimulus type: UW-RF, RW-UF, RW-RF) x 2 (direction of rotation: counterclockwise, clockwise) repeated measures ANOVA was conducted with each of reading time, error count, immediate judgments of effort, and post-task judgments of effort as the dependent variable. One participant was excluded from this analysis because they did not have any counterclockwise RW-RF trials.

No main effect of direction on reading time was observed, $F < 1$. There was a significant stimulus type by direction interaction on reading time, $F(1.91, 57.41) = 3.41$, $p = .039$, $\eta_g^2 < .01$. This interaction appears to reflect the fact that, on UW-RF and RW-RF trials, individuals were faster when reading counterclockwise-rotated stimuli ($M_{\text{UW-RF}} = 16{,}216$; $M_{\text{RW-RF}} = 17{,}235$) than clockwise-rotated stimuli ($M_{\text{UW-RF}} = 16{,}425$; $M_{\text{RW-RF}} = 17{,}528$); while on RW-RF trials, they were slower when reading counterclockwise-rotated stimuli ($M = 17{,}596$) than clockwise-rotated stimuli ($M = 16{,}903$).

No main effect of direction on error count was observed, $F < 1$. There was a marginally significant stimulus type by direction interaction on error count, $F(1.62, 48.96) = 3.29$, $p = .056$, $\eta_g^2 = .02$. While more errors were made reading counterclockwise-rotated stimuli ($M = 1.32$, *SD*

= 0.91) than when reading clockwise-rotated stimuli (*M* = 1.15, *SD* = 0.91) on RW-UF trials, less errors were made reading counterclockwise-rotated stimuli (*M* = 1.14, *SD* = 0.90) than when reading clockwise-rotated stimuli (*M* = 1.51, *SD* = 1.05) on RW-RF trials.

No main effect of direction on immediate judgments of effort was observed, $F(1, 30) = 2.07$, $p = .160$, $\eta_g^2 < .01$. Moreover, the stimulus type by direction interaction was not significant for immediate judgments of effort, $F < 1$.

Finally, no main effect of direction on post-task judgments of effort was observed, $F < 1$; and the stimulus type by direction interaction was not significant for post-task judgments of effort, $F(2, 62) = 1.14$, $p = .326$, $\eta_g^2 < .01$.

2.2.6. Exploratory Analysis

A two-way repeated measures ANOVA with stimulus type and judgment type (i.e., immediate or post-task judgment of effort) as factors revealed a significant interaction between stimulus type and judgment type, $F(3, 93) = 2.72$, $p = .049$, $\eta_g^2 = .01$. This interaction reflected a large difference in reported effort for the UW-UF stimulus type, $t(31) = 5.02$, $p < .001$, $d = 0.89$, which reduced in magnitude for the UW-RF stimulus type, $t(31) = 1.58$, $p = .125$, $d = 0.28$, and the RW-UF stimulus type, $t(31) = 1.84$, $p = .075$, $d = 0.33$; and was absent for the RW-RF stimulus type, $t(31) = 0.25$, $p = .806$, $d = 0.04$ (see Figure 2). To further explore the nature of the interaction between stimulus type and judgment type, we conducted three 2 (stimulus type) x 2 (judgment type) ANOVAs, comparing UW-UF to UW-RF; UW-RF to RW-UF; and RW-UF to RW-RF. This revealed a significant interaction between stimulus type and judgment type when comparing UW-UF to UW-RF stimulus types, $F(1, 31) = 4.35$, $p = .045$, $\eta_g^2 = .01$. As seen in Figure 2, the effect of judgment type was more pronounced on judgments of effort for the UW-

UF stimulus type then for the UW-RF stimulus type. There was no significant interaction between stimulus type and judgment type when comparing UW-RF to RW-UF stimulus types, $F < 1$; or when comparing RW-UF to RW-RF stimulus types, $F(1, 31) = 1.47$, $p = .234$, $\eta_g^2 < .01$. When the analysis was conducted on the head-tilt control subset, the interaction between stimulus type and judgment type when comparing UW-UF to UW-RF stimulus types did not reach significance, $F(1, 28) = 2.59$, $p = .119$, $\eta_g^2 = .01$.

2.3. Discussion

The results of Experiment 1, with respect to reading times, were consistent with previous research (Dunn & Risko, 2016). Specifically, participants were fastest when reading UW-UF and UW-RF stimulus types and slowest when reading the RW-UF and RW-RF stimulus types. There was no difference in reading times between the UW-UF and UW-RF stimulus types, although it approached significance (and was significant when analyzing the head-tilt control subset). Critically, there was no difference in reading times between the RW-UF and RW-RF stimulus types. In a similar vein, the post-task judgments of effort followed the same pattern observed by Dunn and Risko (2016), namely: UW-UF < UW-RF < RW-UF < RW-RF. Thus, the dissociation between reading times (i.e., no difference between RW-UF and RW-RF) and judgments of effort (i.e., a significant difference between RW-UF and RW-RF) was again observed. With respect to this dissociation, the pattern of immediate judgments of effort followed a similar pattern to that of the post-task judgments. Specifically, the RW-RF stimulus type was judged as significantly more effortful than the RW-UF stimulus type. Thus, the difference in judgments of effort between the RW-UF and RW-RF stimulus types reported by Dunn and Risko (2016) does not appear to be due to there having been a delay between participants' reading experience and their judgments of effort. That is, even when the judgment of effort follows immediately after reading, there was still a marked dissociation between judgments of effort and reading time. While the pattern across the RW-UF and RW-RF stimulus types was similar across immediate and post-task judgments, there was clearly an impact of when individuals made their judgments of effort. Namely, when making judgments post-task, relative to immediate, judgments of effort were much lower for the UW-UF stimulus type, slightly lower for the UW-RF and RW-UF stimulus types, and the RW-RF stimulus type was unaffected. This interaction might reflect immediate judgments better aligning with reading time. For example, the reduced difference between the UW-UF and UW-RF stimulus types in immediate judgments relative to post-task judgments could be construed as closer to the modest difference in reading times between those stimulus types. Thus, Experiment 1 provides some modest support for the idea that immediate judgments of effort more closely follow an important experiential cue (i.e., time). We replicate and extend Experiment 1 to further examine these effects.

3. Experiment 2

We decided to replicate Experiment 1 using a larger sample and alter our instructions to encourage participants to increase their reading speed (i.e., not favour accuracy over speed when reading). Moreover, a replication allowed us to address the issue of participants pressing the spacebar before they had read the 25th word, enabling us to measure reading times from stimulus onset to onset of the vocalization of the 25th word (as opposed to when the spacebar was pressed).

3.1. Method

3.1.1. Participants

Forty-eight University of Waterloo undergraduate students participated in this experiment in exchange for course credit. As we wanted to replicate the results of Experiment 1 with a larger sample, we increased our previous sample size by fifty percent. This experiment was preregistered: https://osf.io/zn4mr/.[2]

3.1.2. Apparatus

The presentation of stimuli and audio/video recording of participants' responses were identical to those in Experiment 1.

3.1.3. Stimuli and Design

As in Experiment 1, a one-factor (stimulus type) within-subjects design was employed. We found that 15 words caused regular occurrences of errors due to difficulty with pronunciation

[2] The pre-registration for Experiment 2 declares that no data had yet been collected at the time of registration; however, data from five participants had already been collected. These data had not been examined or analyzed.

in Experiment 1. These words were replaced with words of similar frequency thought to elicit fewer errors across participants. (e.g., “COCOA” replaced with “CAMEL”). In contrast to Experiment 1, wherein 33 stimuli were available to populate the randomly generated lists, only 32 stimuli were used in total (the stimulus removed was that associated with the most errors). Eight unique lists were created using the 32 stimuli such that across each of these lists, each of the 32 stimuli was presented in each orientation on exactly two trials. For the disoriented stimulus types (UW-RF, RW-UF, and RW-RF), one trial was rotated counterclockwise and the other was rotated clockwise. Furthermore, the number of practice trials was reduced from five to four (with each stimulus type presented once), and their order was randomized across participants. The stimuli were otherwise identical to those used in Experiment 1.

3.1.4. Procedure

Participants entered the testing room and were seated facing the centre of the monitor at about eye-level, and the keyboard was on the desk directly in front of them. In a change from Experiment 1 designed to decrease occurrences of head rotation, a researcher sat to the right and behind the participant throughout the experiment. Instructions were displayed on the screen as the researcher read them and answered any questions. Participants were told that, upon reading all the words on a slide, the researcher would press a key to advance them to the 7-point effort scale (*1 = not at all effortful*, *4 = somewhat effortful*, *7 = very effortful*). To ensure that participants were not favouring accuracy over speed when reading, reminder slides were presented after trial 6, 12, 18, and 24, with instructions to read as quickly as possible and to keep their head upright. The main trial procedure was otherwise identical to that for Experiment 1. The generic stimulus types for the post-task judgments were presented from a randomized list of 16 slides composed of two of each stimulus type and direction (the UW-UF generic stimulus type was presented a total of four times to keep the frequency of each stimulus type constant). The experiment then concluded and participants were debriefed.

3.2. Results

Two participants were excluded. One participant experienced difficulty pronouncing many of the words; and another had difficulty keeping their head upright for seven of the eight RW-RF trials and four of the eight RW-UF trials. Their data were removed from the analyses and were replaced by additional participants. Due to a technical error, two participants were not video recorded, but audio recordings were captured, and three participants wished to participate in the experiment without being video recorded. While absence of video was a criterion for exclusion in Experiment 1, the researcher's presence during test for the current experiment allowed for sufficient monitoring of head movement so that video was a helpful addition but not required. Therefore, the data from these five participants were retained. The final sample had 48 participants. As with Experiment 1, any trials during which the participant moved their head were removed from analyses. Forty-one trials in all were removed for this reason; 0.5% (of total) UW-UF, 0.5% UW-RF, 1.8% RW-UF, and 7.8% RW-RF. Finally, eight trials were removed due to procedural irregularities. After these exclusions, a within-subject, within stimulus type search for outliers at the trial level found no reading times, error count scores, or immediate judgments of effort with $|z| > 3$. Therefore, no observations were trimmed from these data. In total, 3.2% of observations were removed. As in Experiment 1, provided the uneven exclusion for head rotations (see above), we conducted a second set of analyses excluding participants who had a disproportionate rate of head rotation for the RW-RF stimulus type in order to have an approximately equal distribution. We excluded eight participants to form a *head-tilt control subset* and the resulting proportion of trials in these data for each stimulus type was: 0.0% UW-

UF, 0.4% UW-RF, 1.1% RW-UF, and 2.5% RW-RF. Overall, the results were similar. When an important deviation from the reported results (i.e., the complete sample) was found we note it in the appropriate section. All analyses were run using the open-source statistical analysis software R, version 3.4.4. The code and data are available on the Open Science Framework: https://osf.io/tgx85/. See Table 2 for mean reading times, error counts, immediate, and post-task judgments of effort by stimulus type.

Table 2

Experiment 2 mean reading time (ms), mean error count, and mean judgments of effort

	Stimulus Type			
	UW-UF	UW-RF	RW-UF	RW-RF
Reading Time (ms)	13,698 (2,592)	13,943 (2,455)	14,802 (2,957)	14,778 (2,924)
Error count	1.10 (0.73)	1.14 (0.72)	1.33 (0.87)	1.21 (0.81)
Immediate judgments	2.80 (1.35)	3.40 (1.27)	3.90 (1.42)	4.09 (1.46)
Post-test judgments	1.61 (0.91)	2.93 (1.18)	3.40 (1.57)	4.05 (1.47)

Note: Standard deviations in parentheses. Immediate and post-task judgments of effort are on 7-point scales

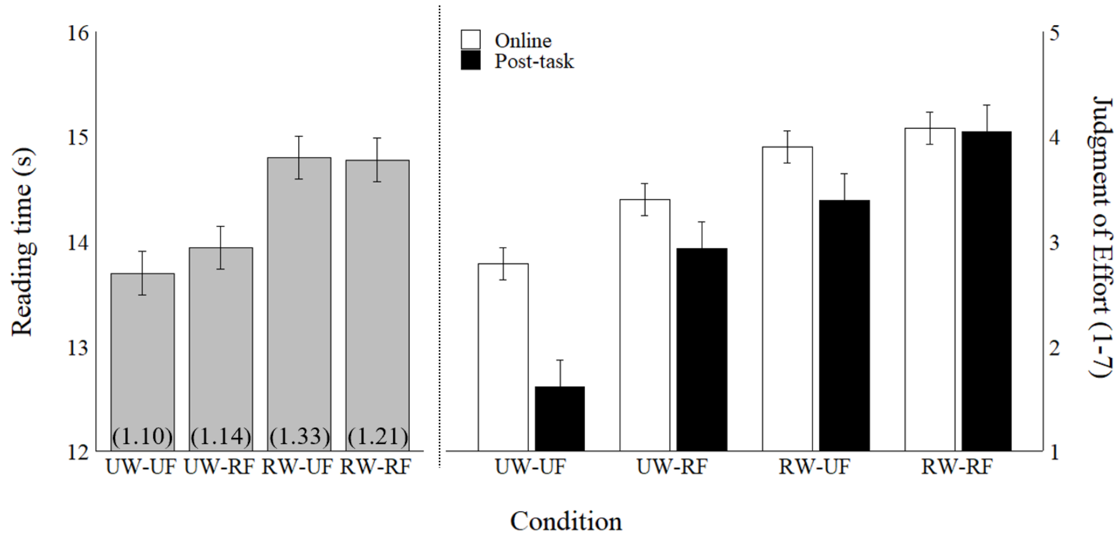

Figure 3. Reading times by stimulus type (left panel) and immediate vs. post-task judgments of effort by stimulus type (right panel) for Experiment 2. Average error count per stimulus type reported in parentheses (left panel). Error bars are Masson-Loftus 95% CI (Loftus & Masson, 1994).

3.2.1. Reading Times

Reading times for each trial were coded by a researcher using the Audacity audio file editing software, measured in milliseconds from stimulus onset to onset of reading the 25th word. A one-way repeated measures ANOVA with stimulus type as the factor, corrected for sphericity violations, revealed a significant effect of stimulus type on reading time, $F(2.28, 107.12) = 33.86$, $p < .001$, $\eta_g^2 = .05$. Pairwise t-tests were conducted across stimulus types. Compared to UW-UF trials, individuals were no slower on UW-RF trials, $t(47) = 1.91$, $p = .062$, $d = 0.28$; however, individuals were slower on RW-UF trials, $t(47) = 6.62$, $p < .001$, $d = 0.95$, and on RW-RF trials, $t(47) = 7.05$, $p < .001$, $d = 1.02$. Compared to UW-RF trials, individuals were slower on RW-UF trials, $t(47) = 5.93$, $p < .001$, $d = 0.86$, and on RW-RF trials, $t(47) = 6.16$, $p < .001$, d

= 0.89. Individuals were not significantly slower on RW-RF trials than on RW-UF trials, *t*(47) = 0.28, *p* = .782, *d* = 0.04; BF_{01} = 6.15. When conducting this analysis on the head-tilt control subset the results were qualitatively similar except that individuals were significantly slower on UW-RF trials than on UW-UF trials, *t*(39) = 2.40, *p* = .021, *d* = 0.38, which was marginal in the complete analysis.

3.2.2. Error Count

The error count for each stimulus type was coded as per Experiment 1. A one-way repeated measures ANOVA with stimulus type as the factor revealed that there was a main effect of stimulus type on error count, $F(2.93, 137.56) = 3.65$, $p = .014$, η_g^2 = 01. Pairwise *t*-tests were conducted across stimulus types. Compared to UW-UF trials, individuals made no more errors on UW-RF trials, $t < 1$; however, more errors were made on RW-UF trials, *t*(47) = 3.21, *p* = .002, *d* = 0.46. Error counts did not differ between UW-UF and RW-RF trials, *t*(47) = 1.45, *p* = .153, *d* = 0.21. Compared to UW-RF trials, individuals made more errors on RW-UF trials, *t*(47) = 2.61, *p* = .012, *d* = 0.38. Error counts did not differ between UW-RF and RW-RF trials, *t*(47) = 1.07, *p* = .289, *d* = 0.15. Finally, no more errors were made on RW-RF trials than on RW-UF trials, *t*(47) = 1.50, *p* = .140, *d* = 0.22; BF_{01} = 2.23. When conducting the analysis on the head-tilt control subset, individuals made more errors on RW-RF trials than on UW-UF trials, *t*(39) = 2.16, *p* = .037, *d* = 0.34.

3.2.3. Effort – Immediate Judgments

As in Experiment 1, immediate judgments of effort were assessed through the report of participants' perceived effort immediately after each trial. Each immediate judgment of effort was between 1 (*not at all effortful*) and 7 (*very effortful*). A one-way repeated measures ANOVA with stimulus type as the factor, corrected for sphericity violations, revealed that there was a significant effect of stimulus type on immediate judgments of effort, $F(2.51, 117.74) = 56.27$, $p < .001$, $\eta_g^2 = .22$. Pairwise t-tests were conducted across stimulus types. Compared to UW-UF trials, individuals reported more effort on UW-RF trials, $t(47) = 5.55$, $p < .001$, $d = 0.80$; on RW-UF trials, $t(47) = 9.68$, $p < .001$, $d = 1.40$; and on RW-RF trials, $t(47) = 9.42$, $p < .001$, $d = 1.36$. Compared to UW-RF trials, individuals reported more effort on RW-UF trials, $t(47) = 5.32$, $p < .001$, $d = 0.77$; and on RW-RF trials, $t(47) = 6.93$, $p < .001$, $d = 1.00$. There was a marginally significant increase in reported effort on RW-RF trials as compared to RW-UF trials, $t(47) = 1.87$, $p = .067$, $d = 0.27$; $BF_{10} = 0.78$.

3.2.4. Effort – Post-task Judgment

A one-way repeated measures ANOVA with stimulus type as the factor, corrected for sphericity violations, revealed that there was a significant effect of stimulus type on post-task judgments of effort, $F(2.29, 107.63) = 63.81$, $p < .001$, $\eta_g^2 = .38$. Pairwise t-tests were conducted across stimulus types. Compared to the UW-UF condition, individuals reported more effort for the UW-RF condition, $t(47) = 9.47$, $p < .001$, $d = 1.37$; the RW-UF condition, $t(47) = 8.21$, $p < .001$, $d = 1.18$; and the RW-RF condition, $t(47) = 12.62$, $p < .001$, $d = 1.82$. Compared to the UW-RF condition, individuals reported more effort for the RW-UF condition, $t(47) = 2.32$, $p = .025$, $d = 0.34$; and the RW-RF condition, $t(47) = 6.09$, $p < .001$, $d = 0.88$. Finally, there was a significant increase in reported effort for the RW-RF condition as compared to the RW-UF condition, $t(47) = 4.36$, $p < .001$, $d = 0.63$; $BF_{10} = 323.22$.

3.2.5. Direction of Rotation

To test for main effects of direction as well as a stimulus type by direction interaction, a 3 (stimulus type: UW-RF, RW-UF, RW-RF) x 2 (direction of rotation: counterclockwise, clockwise) repeated measures ANOVA was conducted for each of reading time, error count, immediate judgments of effort, and post-task judgments of effort as the dependent variable.

No main effect of direction on reading time was observed, $F(1, 47) = 2.47$, $p = .123$, $\eta_g^2 < .01$; nor was there a significant stimulus type by direction interaction on reading time, $F < 1$. When conducting the analysis on the head-tilt control subset, a main effect of direction on reading time was observed, $F(1, 39) = 4.74$, $p = .036$, $\eta_g^2 < .01$, whereby clockwise-rotated stimulus types were read more slowly than were counterclockwise-rotated stimulus types.

No main effect of direction and no stimulus type by direction interaction on error count was observed, *F*s < 1.

A main effect of direction on immediate judgments of effort was observed, $F(1, 47) = 4.07$, $p = .049$, $\eta_g^2 = .01$. Specifically, immediate judgments of effort were higher for clockwise-rotated stimuli ($M = 3.88$, $SD = 1.11$) than for counterclockwise-rotated stimuli ($M = 3.71$, $SD = 1.02$). No stimulus type by direction interaction on immediate judgments of effort was observed, $F(2, 94) = 2.05$, $p = .135$, $\eta_g^2 < .01$.

A main effect of direction on post-task judgments of effort was observed, $F(1, 47) = 14.37$, $p < .001$, $\eta_g^2 = .01$; such that post-task judgments of effort were higher for clockwise-rotated stimuli ($M = 3.61$, $SD = 1.41$) than for counterclockwise-rotated stimuli ($M = 3.31$, $SD = 1.37$). No stimulus type by direction interaction on post-task judgments of effort was observed, $F(2, 94) = 1.01$, $p = .369$, $\eta_g^2 < .01$.

3.2.6. Exploratory Analysis

The following results are not from pre-registered analyses; however, for a more complete picture of the data, we provide them here. A two-way repeated measures ANOVA with stimulus type and judgment type (i.e., immediate or post-task judgments of effort) as factors revealed an interaction between stimulus type and judgment type, $F(3, 141) = 14.43$, $p < .001$, $\eta_g^2 = .04$. As in Experiment 1, this interaction appears driven by a large difference in judgments of effort for the UW-UF stimulus type, with smaller differences observed for the UW-RF and RW-UF stimulus types, and no difference for the RW-RF stimulus type. Specifically, participants provided higher immediate judgments of effort than post-task judgments for the UW-UF stimulus type, $t(47) = 7.45$, $p < .001$, $d = 1.07$; and to a lesser extent, the UW-RF stimulus type, $t(47) = 3.24$, $p = .002$, $d = 0.47$, and the RW-UF stimulus type, $t(47) = 2.67$, $p = .010$, $d = 0.38$. Finally, there was no significant difference between immediate ($M = 4.08$, $SD = 0.98$) and post-task ($M = 4.05$, $SD = 1.30$) judgments of effort for the RW-RF stimulus type, $t < 1$. To further explore the nature of the interaction between stimulus type and judgment type, we conducted three 2 (stimulus type) x 2 (judgment type) ANOVAs, comparing UW-UF to UW-RF; UW-RF to RW-UF; and RW-UF to RW-RF. This revealed a significant interaction between stimulus type and judgment type when comparing UW-UF to UW-RF stimulus types, $F(1, 47) = 22.09$, $p < .001$, $\eta_g^2 = .04$. As seen in Figure 3, the effect of judgment type was more pronounced on judgments of effort for the UW-UF stimulus type than for the UW-RF stimulus type. There was no significant interaction between stimulus type and judgment type when comparing UW-RF to RW-UF stimulus types, $F < 1$. Unlike in Experiment 1, there was a significant interaction when comparing RW-UF to RW-RF stimulus types, $F(1, 47) = 9.01$, $p = .004$, $\eta_g^2 = .01$. Specifically, judgments of effort were significantly greater for the RW-RF stimulus type when judgment was made post-task, $t(47) = 4.36$, $p < .001$, $d = 0.35$, as compared to when judgment was made immediately, $t(47) = 1.87$, $p = .067$, $d = 0.13$.

To test the effect of the instructions to read as quickly as possible, three two-way mixed measures ANOVA with stimulus type as the within-subject factor and experiment as the between-subject factor were conducted. The outcome variables were reading times, immediate judgments of effort, and post-task judgments of effort, respectively. This analysis revealed a significant main effect of experiment on reading time, $F(1, 78) = 13.05$, $p = .001$, $\eta_g^2 = .14$. Participants in Experiment 2 read faster (M = 14,358, SD = 2,450) than did participants in Experiment 1 (M = 16,686, SD = 3,569). Furthermore, this analysis revealed a significant main effect of experiment on immediate judgments of effort, $F(1, 78) = 13.00$, $p = .001$, $\eta_g^2 = .11$; as well as on post-task judgments, $F(1, 78) = 6.03$, $p = .016$, $\eta_g^2 = .04$, such that participants reported higher judgments of effort in Experiment 2 ($M_{\text{immediate}} = 3.54$; $M_{\text{post-task}} = 3.00$) than did participants in Experiment 1 ($M_{\text{immediate}} = 2.88$; $M_{\text{post-task}} = 2.52$). There was no interaction between experiment and stimulus type from any of the above models, Fs < 1.10.

3.3. Discussion

The results of Experiment 2, with respect to reading times, were generally consistent with those of Experiment 1, as well as with previous research by Dunn and Risko (2016). Specifically, participants were fastest when reading the UW-UF and UW-RF stimulus types and slowest when reading the RW-UF and RW-RF stimulus types. As in Experiment 1, there was a marginal difference in reading times between the UW-UF and UW-RF stimulus types (which again was significant when analyzing the head-tilt control subset). As in Experiment 1, there was no difference in reading times between the RW-UF and RW-RF stimulus types; and as in Experiment 1, the post-task judgments of effort followed the pattern observed by Dunn and

Risko (2016), that being UW-UF < UW-RF < RW-UF < RW-RF. Thus, the dissociation between reading times (i.e., no difference between RW-UF and RW-RF) and post-task judgments of effort was again observed. With respect to this dissociation, the immediate judgments of effort followed a similar pattern to the post-task judgments; with the exception that, while the RW-RF stimulus type was judged as more effortful than the RW-UF stimulus type, this difference was only marginally significant. Also notable was that, in Experiment 2, there was a significant interaction between stimulus type and judgment type such that the difference between the RW-UF and RW-RF stimulus types was smaller when making immediate judgments than when making post-task judgments. The latter result is consistent with the idea that immediate judgments of effort more closely followed reading times. Further to this conclusion, when judgments of effort were made post-task, relative to immediate, they were markedly lower for the UW-UF stimulus type, lower for the UW-RF and RW-UF stimulus types, and the RW-RF stimulus type was unaffected. This general pattern was observed in Experiment 1 though was more pronounced here and will be examined further in the General Discussion.

In Experiment 2, a greater emphasis was placed on reading speed. This change in instructions appeared to have its intended effect. That is, a post-hoc analysis with experiment as a between-subject factor revealed that, across stimulus types, individuals read significantly more quickly in Experiment 2 than in Experiment 1. Interestingly, individuals also perceived trials as significantly more effortful, whether judged immediately or post-task. The influence of this instruction appeared to have the same effect across stimulus types, as there were no interactions between experiment and stimulus type. Thus, individuals appeared to invest more effort in Experiment 2 but this additional investment produced only a main effect.

4. General Discussion

Across two experiments, we demonstrated that judgments of effort can be influenced by their proximity to the cognitive event in question. We began this investigation focused on a previously reported dissociation between reading times and judgments of effort. One potential explanation of this dissociation was that it was due to the temporal separation of judgments of effort from the experience of reading. Overall the present results provide some support for this idea. On the one hand, individuals judged the RW-RF stimulus type as more effortful to read than the RW-UF stimulus type when providing immediate, as well as post-task, judgments in Experiment 1; and in Experiment 2, this difference was in the same direction while only marginally significant. From this perspective the immediate and post-task judgments look qualitatively similar. Importantly, however, the RW-UF versus RW-RF difference in Experiment 2 was significantly smaller for immediate judgments than post-task. Thus, while one would be hard-pressed to say the dissociation was not observed for immediate judgments, it does seem to have at least reduced in Experiment 2, a result consistent with the hypothesis that judgments made in closer proximity to the task more closely approximate an experiential cue, namely, reading times. All that said, the effect of when the judgment of effort was made also appeared to have a broader influence. That is, in both experiments there was a robust interaction between stimulus type and judgment type (i.e., immediate versus post-task judgments of effort), whereby the UW-UF stimulus type was judged as significantly less effortful post-task as compared to when judged immediately, with the UW-RF and RW-UF stimulus types exhibiting this pattern but with a notably smaller magnitude, and – as noted above – no difference was observed for the RW-RF stimulus type. We examine this interaction further below along with the broader implications of the present work for our understanding of judgments of effort.

4.1. The Stimulus Type by Judgment Type Interaction

As noted above, the most robust result from Experiments 1 and 2 was the interaction between stimulus type and judgment type. As articulated in the Introduction, one possible explanation for such an effect starts with the notion that individuals rely more on beliefs when making post-task judgments, and more on experiential cues when making immediate judgments. It seems reasonable to suggest that individuals believe that it takes less effort to read an upright display than a disoriented display; thus, relying primarily on beliefs might lead to a large separation between the only upright stimulus type (i.e., UW-UF) and the three disoriented stimulus types (i.e., UW-RF, RW-UF, RW-RF). That is, the qualitative change between “not rotated” and “rotated” might weigh heavily when individuals are inferring effort based on beliefs and are relatively separated from the experience of processing the stimulus. The idea that, in certain situations, incremental differences are less salient than are categorical differences has been suggested previously (Dunn et al., 2017; Hsee & Zhang, 2010). Critically, reading times reveal relatively modest effects for certain stimulus rotations (e.g., UW-UF vs. UW-RF). If we assume that immediate judgments are more closely tied to experiential cues (e.g., reading time), then this provides a plausible explanation for why the UW-UF stimulus type lies much closer to the rotated stimulus types when individuals are making immediate judgments (see Figures 2 and 3)

Further evidence that the stimulus type by judgment type interaction might be due to experiential cues having a greater influence on immediate judgments was present in the RW-RF versus RW-UF comparison. As described above, there was modest support for a reduction from post-task to immediate judgments in this difference. Critically, this change brings effort judgments more in line with reading times, for which there is no difference between these stimulus types. As previously suggested (Dunn & Risko, 2016), when making judgments post-task, individuals may rely on the belief that the RW-RF stimulus type is more effortful to read than is the RW-UF stimulus type, as the former comprises rotation of the frame as well as the words (i.e., it is "more" rotated).

While aspects of the stimulus type by judgment type interaction seems compatible with immediate judgments being more influenced by experiential cues, this was not universally the case. In particular, when considering the difference between the UW-RF and RW-RF stimulus types, there was no change across immediate and post-task judgments. But, in reading time, this difference is consistently one of the largest. Indeed, reading times seem largely a product of word rotation (see Figures 2 and 3). If individuals were relying more heavily on reading time when making immediate judgments, then one could reasonably expect the UW-RF vs. RW-UF difference to increase in magnitude across the judgment types. Thus, there might be an alternate explanation for the interaction between stimulus type and judgment type. It is worth noting that there was a difference across judgment types with respect to the individual words that made up the stimulus displays (i.e., "WORD" for each word for post-task displays vs. unique words for the main trial displays). It is unlikely that this difference is responsible for the stimulus type by judgment type interaction. If this was the case, then we would expect to observe a main effect of judgment type (e.g., judgments might be lower overall post-task because reading "WORD" repeatedly would be easier) but no stimulus type by judgment type interaction.

The notion that individuals use reading time more so when making immediate judgements could be further examined using a regression approach wherein reading time is viewed as a predictor of effort judgements. To this end, in a further exploratory analysis, which was not pre-registered, we employed two multi-level regression models, both with reading time as the predictor, one with immediate judgments as the dependent variable, and the other with post-task judgments as the dependent variable. Each model included random slopes for each subject and featured four observations per subject, one for each condition (i.e., mean reading time, mean immediate judgment, mean post-task judgment). In Experiment 1, this analysis revealed that reading time was a stronger predictor of immediate judgments, $B = 0.21$, $SE = 0.04$, $p < .001$, than it was of post-task judgments, $B = 0.00$, $SE = 0.04$, $p = .976$. In Experiment 2, reading time was also a stronger predictor of immediate judgments, $B = 0.34$, $SE = 0.04$, $p < .001$, than it was of post-task judgments, $B = 0.17$, $SE = 0.04$, $p = .002$. Moreover, these results do not differ qualitatively when controlling for stimulus type. This analysis provides further evidence that individuals rely more on reading time when making immediate, rather than post-task, judgments of effort.

4.2. Metacognitive Framework for Judgments of Effort

As discussed in the Introduction, a judgment of effort can be viewed as a type of metacognitive judgment. From this perspective, deciding the effortfulness of a given cognitive act involves making an inference based on available cues. The nature of such cues can be divided roughly into more experiential cues, for example, time or errors made; and more belief-based cues, for example, the belief that reading rotated text is effortful. This perspective differs from the more intuitive position that we directly read off the effortfulness of cognitive acts from some internal signal. While the present work was not a direct test of the metacognitive approach, the judgment type by stimulus type interaction reported in both experiments clearly supports it. That is, judgments of effort were demonstrated to be a function of the context in which the judgment was made (i.e. whether they were solicited immediately after reading or post-task), a result captured naturally in this framework as a shift in the particular cues relied on across judgment contexts. Future research focusing on whether the temporal proximity to other cognitive tasks plays a role in individuals' perceptions of effort would be valuable.

5. Conclusion

The present investigation aimed to determine whether the proximity to a cognitive event influences judgments of effort. The critical contribution was the discovery of a stimulus type by judgment type interaction in the context of judgments of effort, which could be interpreted as a shift in the cues used to inform judgments of effort as a function of the temporal proximity between the cognitive act and the associated judgment. Future work focusing on the effects of various contexts (e.g., single vs. joint evaluation; Dunn et al., 2017) in which judgments of effort are made will provide a deeper understanding of decisions about our expenditures of cognitive effort.

6. References

Benjamin, A. S., Bjork, R. A., & Schwartz, B. L. (1998). The mismeasure of memory: When retrieval fluency is misleading as a metamnemonic index. *Journal of Experimental Psychology: General, 127*(1), 55-68. doi: 10.1037/0096-3445.127.1.55

Boldt, A., & Gilbert, S. J. (2019). Confidence guides spontaneous cognitive offloading. doi: 10.31234/osf.io/ct52k

Dunlosky, J. & Nelson, T. O. (1997). Similarity between the cue for Judgments of Learning (JOL) and the cue for test is not the primary determinant of JOL accuracy. *Journal of Memory and Language, 36*(1), 34-49. doi: 10.1006/jmla.1996.2476

Dunn, T. L., & Risko, E. F. (2016). Toward a metacognitive account of cognitive offloading. *Cognitive Science, 40*(5), 1080-1127. doi: 10.1111/cogs.12273

Dunn, T. L., Gaspar, C., & Risko, E. F. (2019). Cue awareness in avoiding effortful control. *Neuropsychologia, 123*, 77-91. doi: 10.1016/j.neuropsychologia.2018.05.011

Dunn, T. L., Inzlicht, M., & Risko, E. F. (2019). Anticipating cognitive effort: Roles of perceived error-likelihood and time demands. *Psychological Research, 83*(5), 1033-1056. doi: 10.1007/s00426-017-0943-x

Dunn, T. L., Koehler, D. J., & Risko, E. F. (2017). Evaluating effort: Influences of evaluation mode on judgments of task-specific efforts. *Journal of Behavioural Decision Making, 30*(4), 869-888. doi: 10.1002/bdm.2018

Dunn, T. L., Lutes, D. J., & Risko, E. F. (2016). Metacognitive evaluation in the avoidance of demand. *Journal of Experimental Psychology: Human Perception and Performance, 42*(9), 1372-1387. doi: 10.1037/xhp0000236

Eggemeier, F. T., & Stadler, M. A. (1984, October). Subjective workload assessment in a spatial memory task. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 28, No. 8, pp. 680-684). Sage CA: Los Angeles, CA: SAGE Publications. doi: 10.1177/154193128402800808

Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2016). Metacognition about the past and future: quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neuroscience of Consciousness, 2016*(1). doi: 10.1093/nc/niw018

Foo, M.-D., Uy, M. A., Baron, R. A. (2009). How do feelings influence effort? An empirical study of entrepreneurs' affect and venture effort. *Journal of Applied Psychology, 94*(4), 1086-1094. doi: 10.1037/a0015599

Gilbert, S. J. (2015). Strategic use of reminders: Influence of both domain-general and task-specific metacognitive confidence, independent of objective memory ability. *Consciousness and Cognition, 33*, 245-260. doi: 10.1016/j.concog.2015.01.006

Gray, W. D., Sims, C. R., Fu, W.-T., & Schoelles, M. J. (2006). The soft constraints hypothesis: A rational analysis approach to resource allocation for interactive behaviour. *Psychological Review, 113*(3), 461-482.

Gweon, H., Asaba, M., & Bennett-Pierre, G. (2017). Reverse-engineering the process: Adults and preschoolers' ability to infer the difficulty of novel tasks. In Gunzelmann, G., Howes, A., Tenbrink, T., & Davelaar, E. (Eds.), Proceedings of the 39th Annual Conference of the Cognitive Science Society (pp. 458-463). Austin, TX: Cognitive Science Society.

Hsee, C. K., & Zhang, J. (2010). General evaluability theory. *Perspectives on Psychological Science, 5*(4), 343-355.

Inzlicht, M., Shenhav, A., & Olivola, C. Y. (2018). The effort paradox: effort is both costly and valued. *Trends in Cognitive Science, 22*(4), 337-349. doi: 10.1016/j.tics.2018.01.007

Jex, H. R. (1988). Measuring mental workload: Problems, progress, and promises. *Advances in Psychology, 52,* 5-39. doi: 10.1016/S0166-4115(08)62381-X

Kool, W. & Botvinick, M. (2018). Mental labour. *Nature Human Behaviour,* 1.

Koriat, A. (1993). How do we know that we know? The accessibility model of the feeling of knowing. *Psychological Review, 100*(4), 609-639. doi: 10.1037/0033-295X.100.4.609

Koriat, A. & Ma'ayan, H. (2005). The effects of encoding fluency and retrieval fluency on judgments of learning. *Journal of Memory and Language*, *52*(4), 478-492. doi: 10.1016/j.jml.2005.01.001

Kurzban, R. (2016). The sense of effort. *Current Opinion in Psychology, 7*, 67-70. doi: 10.1016/j.copsyc.2015.08.003

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review, 1*(4), 476-490. doi: 10.3758/BF03210951

Metcalfe, J., & Finn, B. (2008). Familiarity and retrieval processes in delayed judgments of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1084. doi: 10.1037/a0012580

Marshall, S. P. (2002, September). The index of cognitive activity: Measuring cognitive workload. In *Proceedings of the IEEE 7th conference on Human Factors and Power Plants* (pp. 7-7). IEEE. doi: 10.1109/HFPP.2002.1042860

Moray, N. (1982). Subjective mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 24*(1), 25-40. doi: 10.1177/001872088202400104

Nelson, T. O., & Dunlosky, J. (1991). When people's Judgments of Learning (JOLs) are extremely accurate at predicting subsequent recall: The "delayed-JOL effect." *Psychological Science, 2*(4), 267–271. doi: 10.1111/j.1467-9280.1991.tb00147.x

Potts, C. A., Pastel, S., & Rosenbaum, D. A. (2018). How are cognitive and physical difficulty compared?. *Attention, Perception, & Psychophysics*, *80*(2), 500-511. doi: 10.3758/s13414-017-1434-2

Scheck, P., Meeter, M., & Nelson, T. O. (2004). Anchoring effects in the absolute accuracy of immediate versus delayed judgments of learning. *Journal of Memory and Language*, *51*(1), 71-79. doi: 10.1016/j.jml.2004.03.004

Siedlecka, M., Paulewicz, B., & Wierzchon, M. (2016). But I was so sure! Metacognitive judgments are less accurate given prospectively than retrospectively. *Frontiers in Psychology, 7,* 218. doi: 10.3389/fpsyg.2016.00218

Song, H., & Schwarz, N. (2008). If It's Hard to Read, It's Hard to Do: Processing Fluency Affects Effort Prediction and Motivation. *Psychological Science, 19*(10), 986–988. doi: 10.1111/j.1467-9280.2008.02189.x

Thompson, V. A., Prowse Turner, J. A., Pennycook, G., Ball, L. J., Brack, H, Ophir, Y., & Ackerman, R. (2013). The role of answer fluency and perceptual fluency as metacognitive cues for initiating analytic thinking. *Cognition, 128*(2), 237-251. doi: 10.1016/j.cognition.2012.90.012

Undorf, M., & Erdfelder, E. (2013). Separation of encoding fluency and item difficulty effects on judgments of learning. *Quarterly Journal of Experimental Psychology, 66*(10), 2060-2072. doi: 10.1080/17470218.2013.777751

Undorf, M., & Erdfelder, E. (2015). The relatedness effect on judgments of learning: A closer look at the contribution of processing fluency. *Memory & Cognition, 43*(4), 647-658. doi: 10.3758/s1342

Weaver, C. A., & Kelemen, W. L. (1997). Judgments of learning at delays: Shifts in response patterns or increased Metamemory Accuracy? *Psychological Science, 8*(4), 318–321. doi: 10.1111/j.1467-9280.1997.tb00445.x

Westbrook, A., Kester, D., Braver, T. S. (2013). What is the Subjective Cost of Cognitive Effort? Load, trait, and aging effects revealed by economic preference. *PLoS ONE, 8*(7), e68210. doi: 10.1371/journal.pone.0068210

Yeh, Y. Y. & Wickens, C. D. (1988). Dissociation of performance and subjective measures of workload. *The Journal of the Human Factors and Ergonomics Society, 30*(1), 111-120. doi: 10.1177/001872088803000110

Yildirim, I., Saeed, B., Bennett-Pierre, G., Gerstenberg, T., Tenenbaum, J., & Gweon, H. (in press). Explaining intuitive difficulty judgments by modeling physical effort and risk. *CogSci Proceedings 2019.*