Determinants of cognitive offloading:

Toward a metacognitive approach

by

Timothy L. Dunn

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, 2014

© Timothy L. Dunn 2014

Author's Declaration

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

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

Timothy L. Dunn

Abstract

Individuals frequently make use of the body and environment when engaged in a cognitive task. For example, individuals will often spontaneously physically rotate when faced with rotated objects, such as an array of words, putatively to offload the costs associated with stimulus rotation. We examined this idea further by independently manipulating the costs associated with both word rotation and array frame rotation. Surprisingly, we found that individuals' patterns of spontaneous physical rotations did not follow patterns of rotation costs or benefits associated with being physically rotated, findings difficult to reconcile with existing theories of strategy selection involving external resources. Individuals' subjective ratings of perceived benefits, rather, provided an excellent match to the patterns of physical rotations, suggesting that theory-based metacognitive judgments are used when deciding on-the-fly whether to incorporate an external resource such as the body. Implications for metacognition's future in theories of cognitive offloading are discussed.

Acknowledgments

I would like to thank Dr. Evan Risko for all of his effort, guidance, and time invested into this project. I would also like to thank Dr. Colin MacLeod and Dr. Daniel Smilek for serving as readers on this thesis.

Table of Contents

List of Tables

List of Figures

Introduction

Much recent work within cognitive science has been aimed at providing a deeper understanding of the embodied and embedded nature of cognition (e.g., Barsalou, 2008; Chemero, 2009; Clark, 2010; Glenberg, 2010; Hutchins, 1995; Kirsh, 1995; Proffitt, 2006; Rupert, 2004; Shapiro, 2011; Wilson, 2002). An important part of this aim involves understanding how we use our body (e.g., physical actions like gesture; Beilock & Goldin-Meadow, 2010) and physical environment (e.g., manipulation of external artifacts like calculators; Walsh & Anderson, 2009) in the course of a given cognitive act, as opposed to focusing solely on the internal processes in isolation. A prominent means of interaction between brain, body, and physical environment is the use of external processing (i.e., manipulation of the body, physical environment) in an attempt to avoid internal processing – a behavior referred to as cognitive offloading (Kirsh & Maglio, 1994; Martin & Schwartz, 2005; Risko et al., 2014; Scaife & Rodgers, 1996; Wilson, 2002). Given its prevalence, cognitive offloading represents an ideal model behavior for investigating how individuals couple internal and external processes. In the present investigation we explore one such instance of cognitive offloading, spontaneous external normalization in the context of reading disoriented text, in an effort to better understand the factors that influence the decision to incorporate an external process.

From Viewpoint Costs to Normalization

There exist numerous demonstrations that performance, across a range of tasks, can be impaired when stimuli are not presented in their canonical orientation (e.g., Corballis, 1988; Diwadkar & McNamara, 1997; Graf, 2006; Kung & Hamm, 2010; Jolicoeur, 1985; Jolicoeur & Landau, 1984; Jolicoeur, Snow, & Murray, 1987; Kolers & Perkins, 1969a; 1969b; Koriat & Norman, 1984) although this is not always the case (Hamm & McMullen, 1998; Murray, 1998; Wells & Hamm, 2009). These viewpoint costs (or lack thereof) have attracted a great deal of attention from researchers. The typical explanation involves the need, when a stimulus is presented in a non-canonical orientation, to engage in some form of internal transformation of that object to match a stored (canonical) representation of the object in memory (e.g., Graf, 2006; Jolicoeur, 1990). For example, individuals might mentally rotate an internal representation of the stimulus (as in mental rotation tasks; Shepard & Metzler, 1971) to match the corresponding upright representation stored in memory (Jolicoeur, 1990; Tarr, 1995). Alternatively, an individual may rotate an internal frame of reference to match the orientation of the stimulus (i.e., coordinate transformation; Graf, 2006; Graf, Kaping, & Bülthoff, 2005). Critically, both mental rotation and coordinate transformation represent forms of internal normalization. Thus, in these cases, the perceptual difficulties caused by stimulus rotation are dealt with via a kind of internal solution.

While individuals clearly have internal solutions at their disposal, it also seems clear that individuals often (at least) attempt external solutions as well. For example, Risko et al. (2014) demonstrated in both letter identification and reading tasks, that individuals, if free to do so, often spontaneously physically rotate their body (i.e., tilt their head) when presented with a rotated stimulus and asked to respond in some manner (e.g., to read). This physical rotation can be viewed as a form of external normalization (i.e., bringing the stimulus closer to its canonical orientation via the physical movement of the body) paralleling the internal forms of normalization discussed above. Indeed, physical rotation of the body to better match the orientation of a stimulus can be thought of as analogous to an internal frame rotation, and physical rotation of the actual stimulus back to upright (see Kirsh & Maglio, 1994) as analogous to mental rotation. The latter idea has been forwarded in embodied accounts of mental rotation where the internal process is hypothesized to reflect a simulated version of the external process

(Wexler, Kosslyn, & Berthoz, 1998; Wohlschläger & Wohlschläger, 1998). Risko et al. (2014) suggested that a similar more explicit relation might hold between physical rotation of the body and internal frame rotation.

Cognitive Offloading

In the present context, we use external normalization as a means to investigate cognitive offloading. Specifically, the behavior can be seen as an attempt to trade-off internal processing (e.g., internal normalization) for external processing (e.g., external normalization; Kirsh & Maglio, 1994; Martin & Schwartz, 2005; Risko et al., 2014; Scaife & Rodgers, 1996; Wilson, 2002). We choose this task because of the long history of cognitive research investigating internal normalization and stimulus rotation effects in general (Corballis, 1988; Graf, 2006; Jolicoeur, 1990; Kolers & Perkins, 1969a; 1969b; Koriat & Norman, 1984), the ability to control stimulus parameters and, most importantly, because individuals will readily spontaneously engage in this behaviour in a controlled setting.

One of the major theoretical tasks in understanding cognitive offloading is to determine how individuals decide on-the-fly whether to incorporate an external process into an ongoing cognitive act. In the context of external normalization, when presented with a rotated display on a given trial, what determines whether an individual physically rotates or deals with the disorientation of the stimulus internally? One general idea is that the decision is based on the relative effort that the different combinations of resources (i.e., internal versus internal plus external) will require (Clark, 2010; Gray & Fu, 2004; Gray et al., 2006; Kirsh, 2010) the assumption being that individuals will expend the least amount of effort possible (e.g., Clark, 2010; Gray et al., 2006). While intuitive, this general account is limited by the inherent difficulty in defining the construct of “effort” (i.e., what determines the least effortful combination of resources?). Gray et al. (2006) avoid this issue by directly tying the notion of effort to time (i.e., the least effortful strategy is the one that takes the least time), a notion that has a long history in cognitive psychology (e.g., Payne, Bettman, & Johnson, 1988; Reder, 1987; Reder & Ritter, 1992) and one with a solid empirical backing in investigations involving the integration of internal and external processes (e.g., Gray & Fu, 2004; Kirsh, 1995; Kirsh & Maglio, 1994; Maglio, Wenger, & Copeland, 2008; Risko et al., 2014; Walsh & Anderson, 2009). For example, Walsh and Anderson (2009) presented participants with multiplication problems across three conditions: (1) they forced participants to use an internal solution (i.e., multiplying in their head), (2) they forced participants to use an internal plus external strategy (i.e., making use of an on-screen calculator) and (3) they allowed participants to freely choose between the strategies. Results demonstrated that individuals adaptively selected between strategies based on time and maximizing pay from correct answers. That is, selection followed the strategy that resulted in the fastest solutions based on time estimates derived from the two forced conditions. In a similar vein, Risko et al. (2014) demonstrated that manipulations that increased the costs (in terms of time) associated with stimulus rotation in a condition where individuals could not physically rotate, increased the frequency of participants' physical head rotations in a condition in which they were free to engage in this behavior. This result is consistent with the time/effort-based criterion for deciding whether to incorporate an external process into an ongoing cognitive act (i.e., as the time it took to complete the task internally increased, the likelihood that an external process was incorporated increased).

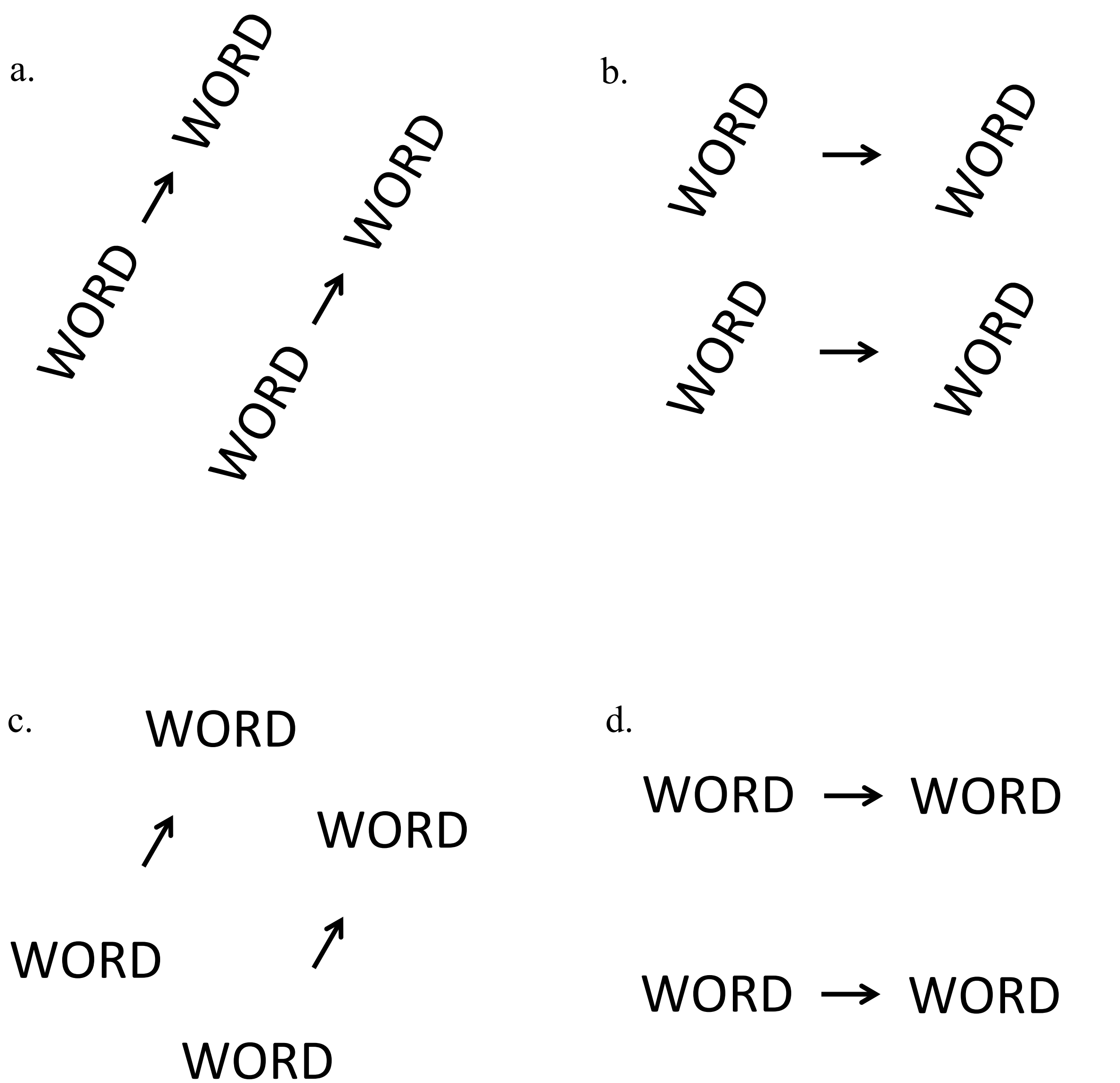
In Experiment 1, we provide a further test of a time/effort-based account of cognitive offloading as indexed by external normalization. Interestingly, results from this experiment challenged the intuitive idea that the frequency of physical rotations should follow the time costs associated with stimulus rotation. This led to a series of experiments that more closely examined the relation between "objective" effort (as indexed by time) and subjective effort, and toward a deeper consideration of the metacognitive basis of cognitive offloading.

Experiment 1

The majority of research on the effects of viewpoint costs has used individual items (e.g., a letter, a line drawing of an object; Jolicoeur, 1985; Koriat & Norman, 1984). Following in this tradition, Risko et al. (2014) suggested that the internal performance costs driving the decision to physically rotate were generated by the rotation of the individual items in the display. As such, the physical rotation would represent an attempt to normalize the orientation of the individual items (i.e., words) in the display. However, the rotation of a multi-item array (e.g., a page of text) typically involves (and did in Risko et al., 2014) not only the rotation of the individual items but also the rotation of the frame of the array (see Figure 1). This rotation of the frame could, in and of itself, produce performance costs, for example, by disrupting the natural direction (e.g., left-to-right) with which individuals navigate the text. Thus, rather than representing an attempt to normalize word orientation, physical rotation could be an attempt to normalize the frame. This would allow the individual to navigate the paragraph in a more natural left-to-right manner and to offload any costs that may be associated with the unfamiliar reading direction. Thus, physical rotation could reflect an attempt to normalize the individual items (e.g., to facilitate retrieval of words from memory), the objects' frame (e.g., to navigate the paragraph in a more natural left-to-right manner), or some combination of the two. In Experiment 1, we discriminated between the above accounts by manipulating individual word and word frame rotation independently in a reading task in which individuals were free to physically rotate. As outlined above, following a time/effort based account, we would expect to observe the highest frequencies of physical rotations in conditions associated with the highest performance costs associated with stimulus rotation.

Figure 1.

Examples of Array Conditions



Note: (a) Rotated Word-Rotated Frame (RW-RF); (b) Rotated Word-Upright Frame (RW-UF); (c) Upright Word-Rotated Frame (UW-RF); (d) Upright Word-Upright Frame (UW-UF). Each stimulus (not including UW-UF) was presented using 25 words, 4 to 5 letters per words, in 5 x 5 displays rotated at both ±60° from upright.

In Experiment 1, participants were presented with arrays of 25 words and asked to read the words aloud in a standard left-to-right, top-to-bottom manner. Participants completed the reading task in a "free-to-rotate" condition, where no instructions were given pertaining to their physical movement while reading. The word display on each trial could be presented in one of four different formats representing the factorial crossing of word and frame rotation: (1) Rotated Word-Rotated Frame: RW-RF, (2) Rotated Word-Upright Frame: RW-UF, (3) Upright Word-Rotated Frame: UW-RF, and (4) Upright Word-Upright Frame: UW-UF (see Figure 1). If the decision to physically rotate is driven solely by the costs associated with normalizing the individual words, then we would expect to observe similar frequencies of physical rotation for the RW-RF and RW-UF array types (i.e., displays in which the words are rotated) and both array types should elicit a higher frequency of physical rotation than the UW-RF and UW-UF array types (i.e., displays in which the words are not rotated). Alternatively, if the decision to rotate is driven solely by the costs associated with normalizing the frame (e.g., reading direction), then we would expect to observe similar frequencies of physical rotations for the RW-RF and UW-RF array types (i.e., displays in which the frames are rotated) and both array types should elicit a higher percentage of rotations than the RW-UF and UW-UF array types (i.e., displays in which the frames are not rotated). Lastly, if the decision to rotate is driven by some combination of the costs associated with normalizing the words and the frame, then we would expect to observe a higher frequency of rotation for the RW-RF array type (i.e., displays in which the words and frames are rotated) than either of the mixed array types (i.e., RW-UF and UW-RF). This latter prediction is based on the idea that physical rotation in either the RW-UF or UW-RF conditions would result in disorienting the non-rotated dimension (the words in the case of the UW-RF condition or the frame in the RW-UF condition) a result that (seemingly) would be incompatible with the desire to normalize both dimensions of the stimulus (hence the prediction that rotation would be less frequent in these conditions in relation to the RW-RF condition). Under the latter hypothesis, where the mixed conditions rank with respect to each other and the upright condition (i.e., UW-UF) can provide some insight into whether word rotation costs or frame rotation costs are stronger determinants of physical rotation. Specifically, a pattern in which rotations are more frequent in one condition than the other would provide evidence for the precedence of that dimension in terms of driving external normalization. Again, the most straightforward prediction based on a time/effort account of cognitive offloading is that individuals will spontaneously physically rotate most often in the conditions associated with the largest stimulus rotation costs.

Method

Participants. Thirty-two (26 females) Arizona State University undergraduate students participated in the study in exchange for either research credit or ten dollars.

Design. A one-factor (Array Type: RW-RF, RW-UF, UW-RF, UW-UF) factor within-subject design was employed.

Apparatus. The presentation of the stimuli and the recording of participants' manual responses were handled by Experiment Builder software (SR Research). Two cameras were used: One camera was placed on top of the monitor to record the participants' behavior during the study, and one camera recorded an additional monitor in an area separated from the participant in order to record the stimulus being presented. Stimuli were presented on a 24" LCD monitor and participants sat approximately 75 cm away from the monitor. Participants used a standard QWERTY keyboard to enter manual responses.

Stimuli. Stimuli consisted of 25 words arranged into 5 x 5 arrays in black on a white background in 18 point Courier New font. Words consisted of four letter nouns and verbs, one to two syllables each, with an average written word frequency of 66.3 per million. The RW-RF and UW-RF array types subtended approximately 15° x 14° (H x W), while the RW-UF and UW-UF array types subtended approximately 9.5° x 11.5°. The first word in each paragraph was colored red and arrows were positioned between words to ensure that participants read in a left-to-right and top-to-bottom manner[1]. Participants completed one block of 32 trials, with each array type occurring eight times. Each disoriented array type (RW-RF, RW-UF, and UW-RF) was presented four times to the right of gravitational upright (0°) and four times to the left of gravitational upright each, while the UW-UF array type was presented eight total times at 0°. Words were counterbalanced such that each 25-word set appeared an equal number of times in each condition.

Procedure. Participants sat in front of the screen with the keyboard in their lap for the duration of the study. Participants first read instructions on the screen stating that they were to read each word in the presented array aloud as quickly and accurately as possible, and to press the spacebar once they had finished reading all of the words. Instructions stated that there were no restrictions on bodily movement except to stay seated. No effort was made to hide the camera on top of the monitor recording the participant.

Results

Results are reported first for the frequency of physical rotation data followed by performance data (i.e., response time and accuracy; see Table 1a). Generalized (GLMM) and Linear (LMM) Mixed Models were constructed using the lme4 package (Bates, Maechler, & Dai, 2011) in R (R Development Core Team, 2010). GLMMs were used for rotation (binomial)

[1] Early pilot studies showed that participants would sometimes adopt irregular reading strategies when presented with disoriented arrays and we wanted to ensure that reading strategy remained constant across participants.

and error (counts) data and LMMs for response time data. An iterative model comparison approach was adopted. Model comparison for both generalized and linear models employed parametric bootstrapping (PB) methods in the pbkrtest package (Halekoh & Højsgaard, in press). Baseline models (outlined for each section), consisting only of main effects, were constructed using all experimental variables and were tested against more complex models that included interactions between variables. If a more complex model (i.e., higher df) significantly increased the model fit when compared to a less complex model based on PB tests (i.e., $p < .05$ for the test), then the more complex model was adopted and compared to the next most complex model. This process continued until the most complex model was tested (i.e., the highest order interaction model) or the more complex model did not significantly increase the model fit when compared to the less complex model. As a result, the simplest model producing the best model fit is reported for all analyses. If any main effects within this final model were not significant, then an additional model comparison was conducted to examine whether removal of that effect was justified.

The main effect model for all analyses consisted of the fixed effects of array type and trial (centered), with subject and item as random effects. The first three trials were treated as practice and removed from all analyses.

Physical Head Rotation

Physical head rotation was determined by the video recording of the experimental session. An individual coder (TD) who was blind to the condition (i.e., coder could not see the stimulus being presented on each trial) coded physical head rotation dichotomously. Trials in which the participant physically rotated their head were determined by the following criteria: (1)

the participant must have started the trial in an upright position (within ±10° from 0°); (2) the participant must have rotated their head a total of ±10° from their starting position; and (3) the onset of rotation must have occurred within 1000 ms of stimulus onset. To calculate intra- and inter-rater reliability 20% of the data files (8 participants) were recoded by the initial coder and by a different coder; again both raters were blind to the conditions. Both inter- and intra-rater reliability were high (K = .83 and K = .88, respectively). A $|Z| > 2$ criterion was used to evaluate the significance of each fixed factor to stay consistent with the significance criteria for analyses employing LMMs (see below). In addition, all trials in which the participant did not start upright (i.e, within ±10° from 0°) were removed, resulting in the removal of an additional 8% of all trials.

The model comparison process produced a final model consisting of array type (i.e., neither the inclusion of trial nor the inclusion of any interactions improved model fit sufficiently). Pairwise comparison across the different array types was achieved by manipulating the reference category in the model. Compared to the UWUF condition, participants rotated more in the RW-RF, b = 6.63, SE = .83, Z =8.02, RW-UF, b = 4.06, SE = .80, Z = 5.12, and UW-RF, b = 2.68, SE = .80, Z = 3.34, conditions respectively. Individuals were less likely to rotate in the RW-UF and UW-RF conditions when compared to the RW-RF condition, b = -2.57, SE = .35, Z = -7.40; b = -3.94, SE = .40, Z =- 9.93, respectively, and individuals were more likely to rotate in the RW-UF condition when compared to the UW-RF condition, b = 1.38, SE = .33, Z = 4.15. In sum, individuals physically rotated the most in the RW-RF condition, followed by the RW-UF, UW-RF, and UW-UF conditions (see Table 1b; Figure 2).

Response Time and Accuracy

Response times (RT) were calculated as the amount of time between the onset of the paragraph and the participant pressing the spacebar to indicate that they had finished reading all of the words. RTs (secs) were transformed using a reciprocal transformation (-1/RT; Masson & Kliegal, 2012) and a $|t| > 2$ criterion was used to evaluate the significance of each fixed factor (Baayen, Davidson, & Bates, 2008). Although we report RTs, it is important to mention that participants' movements were not restricted, thus complicating the interpretation of those RTs in terms of the duration of some internal process. Accuracy was determined by analyzing the vocal responses from each video file. Errors included the incorrect reading of a word, skipping a word, repeating a word, or reading the words aloud out of order. Thus, values reported in performance tables are errors per trial (i.e., the count of errors per 25 word array). Trials on which one or more errors occurred were not removed from the RT analysis as this would have resulted in the elimination of a large proportion of trials given that each trial consists of 25 different words. Accuracy models used a $|Z| > 2$ criterion to evaluate the significance of each fixed factor.

Total Response Time. Model comparison demonstrated that the model of best fit consisted of the main effect of array type and trial. Response time decreased as trial increased, b = -.0015, SE = .0002, t = -7.33. Response times were longer in the RW-RF, RW-UF and UW-RF conditions when compared to the UW-UF condition, b = .0019, SE = .0006, t = 3.16; b = .0029, SE = .0006, t = 4.75; b = .0016, SE = .0006, t = 2.71, respectively. In addition, RTs were longer in the RW-UF condition than in the UW-RF condition, b = .0012, SE = .0006, t = 2.02. No other pairwise comparisons were significant. Therefore, the RW-UF condition produced the slowest RTs followed by the RW-RF condition, and then the UW-RF condition, and all were slower than responses in the UW-UF conditions (see Table 1c; Figure 3).

Accuracy. The model comparison process produced a final model consisting of array type only (i.e., neither the inclusion of trial nor the inclusion of any interactions improved model fit sufficiently). Pairwise comparisons between conditions demonstrated that there were more errors in the RW-UF condition than in the UW-UF and UW-RF conditions, $b = .34$, $SE = .11$, $Z = 3.18$; $b = .298$, $SE = .108$, $Z = 2.74$, respectively. All other comparisons between conditions were not significant. Thus the RW-UF condition produced the greatest number of errors, followed by the RW-RF, UW-RF, and UW-UF conditions (see Table 1d).

Table 1a.

Mean Performance Results for Experiment 1

	Array Type							
Variable	UW-UF		UW-RF		RW-UF		RW-RF	
Percentage of Trials Rotated	1%	(10%)	10%	(31%)	23%	(42%)	52%	(50%)
RT (ms)	15199	(2960)	15757	(3181)	15920	(3123)	15653	(2995)
Errors	.64	(1.02)	.65	(1.08)	.89	(1.16)	.78	(1.05)

Note: UW-UF= Upright Word-Upright Frame; UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame; RT = response times. Results reported in the table are overall means. Standard deviations are presented in parentheses.

Table 1b.

Generalized Linear Mixed Model Results for Physical Head Rotations (log Odds) in Experiment 1

Random Effects	Variance	SD	
Subject	7.95	2.82	
Item	-	-	
Fixed Effects	**Estimate**	***SE***	***Z***
Intercept	-6.73	.94	-7.19
UW-RF	2.68	.80	3.33
RW-UF	4.06	.80	5.12
RW-RF	6.63	.83	8.02

Note: UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. A |*Z*| > 2 criterion was used to evaluate the significance of each fixed factor.

Table 1c.

Linear Mixed Model Results for Response Times (-1/RT; sec) in Experiment 1

Random Effects	Variance	SD	
Subject	1.45e-04	.012	
Item	1.40e-07	.001	
Residual	3.81e-05	.001	
Fixed Effects	Estimate	*SE*	*t*
Intercept	-.0681	.0021	-31.41
UW-RF	.0016	.0006	2.71
RW-UF	0.0029	.0006	4.75
RW-RF	0.0019	.0006	3.16
Trial	-0.0015	.0002	-7.33

Note: UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. A $|t| > 2$ criterion was used to evaluate the significance of each fixed factor.

Table 1d.

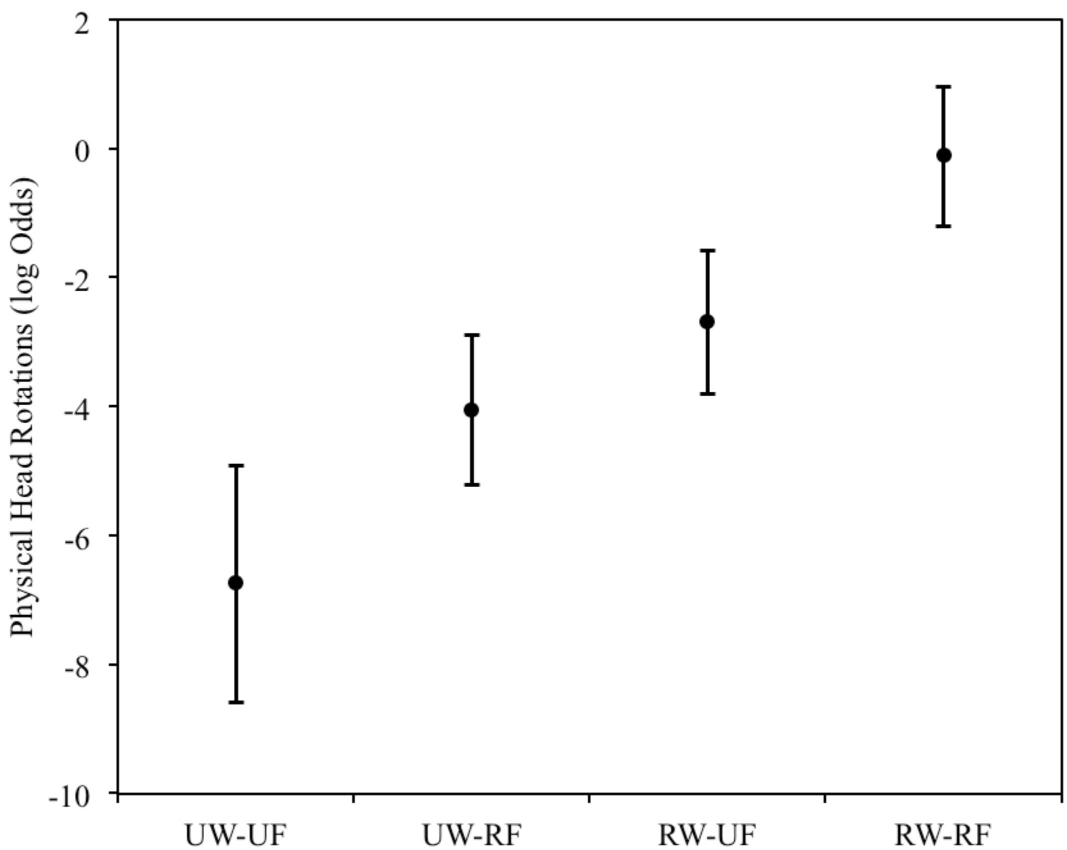
Generalized Linear Mixed Model Results for Errors (log Counts) in Experiment 1

Random Effects	Variance	SD	
Subject	.39	.62	
Item	.001	.02	
Fixed Effects	Estimate	*SE*	*Z*
Intercept	-.65	.14	-4.66
UW-RF	.06	.12	.48
RW-UF	.34	.11	3.18
RW-RF	.21	.11	1.88

Note: UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. A |*Z*| > 2 criterion was used to evaluate the significance of each fixed factor.

Figure 2.

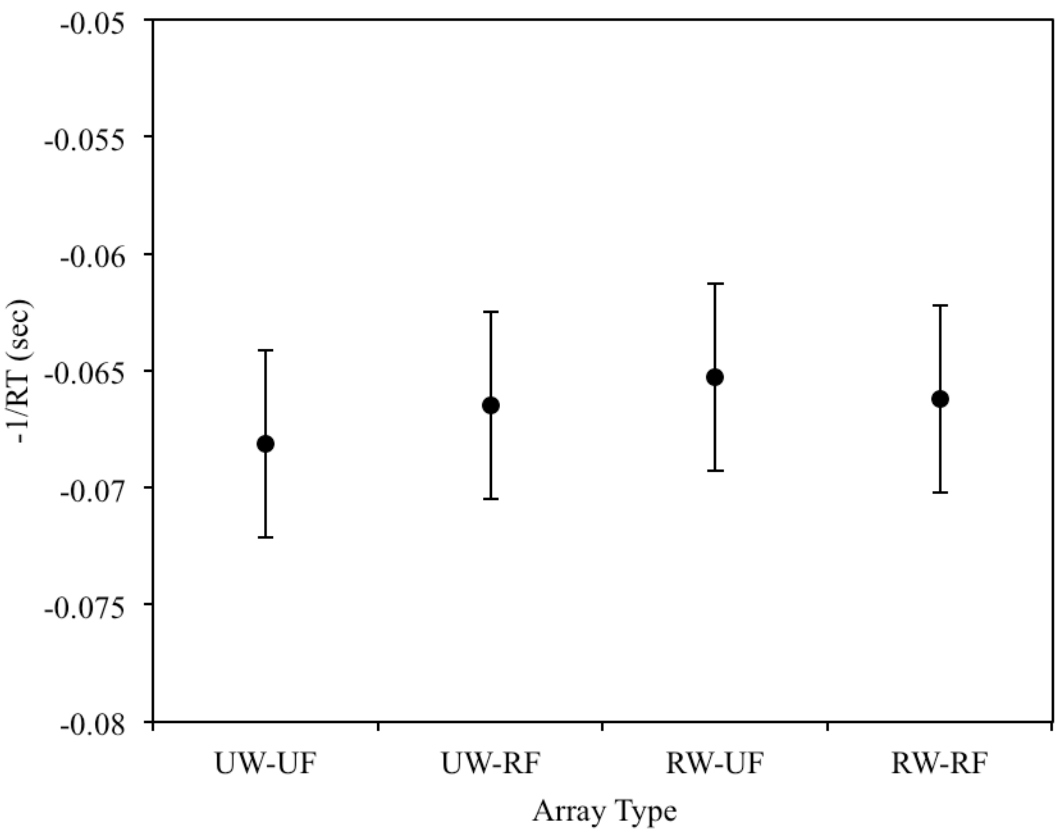
Spontaneous Physical Head Rotation Results for Experiment 1



Note: UW-UF = Upright Word-Upright Frame; UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. Error bars denote 95% confidence intervals.

Figure 3.

Response Time Results for Experiment 1



Note: UW-UF = Upright Word-Upright Frame; UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. Error bars denote 95% confidence intervals.

Discussion

The results of Experiment 1 are clear-cut. Individuals spontaneously rotated more when both the word and frame were rotated (i.e., RW-RF condition) than when only the word (i.e., RW-UF) or only the frame (i.e., UW-RF) was rotated. Therefore, spontaneous physical rotation seems to be driven by the desire to normalize both the individual items and the object's frame. Interestingly, there was also an increased frequency of physical head rotation in the RW-UF array type when compared to the UW-RF array type. Given that rotating in the RW-UF array type disorients the frame of the array and that the two array types that contain item-level manipulations (i.e., RW-RF and RW-UF) produced the highest frequencies of physical rotations, it appears that normalizing individual items in a multi-element array takes precedence over normalizing the frame of the array.

The pattern of spontaneous physical rotation observed in Experiment 1 is particularly interesting in light of the counterintuitive discovery that the costs of stimulus rotation were similar in the RW-RF and RW-UF conditions and the costs associated with the RW-RF and RW-UF conditions ware larger than that associated with the UW-RF condition. Thus, the pattern of results is underadditive in the sense that while frame rotation has a cost relative to when the words and frame are upright, there are no associated costs when the words are rotated. This pattern is interesting given that participants were far more likely to spontaneously physically rotate in the RW-RF condition than in the RW-UF condition (i.e., the pattern in terms of physical rotation does not take the same underadditive form). Thus, there appears to be a dissociation between the likelihood of physical rotation and the time costs of stimulus rotation. This is inconsistent with the idea that spontaneous physical rotation would be most frequent in the condition with the greatest performance costs (i.e., RW-RF and RW-UF take about the same amount of time but individuals are more than twice as likely to physically rotate in the former condition). We explore the basis of this dissociation further in Experiment 2.

Experiment 2

One obvious issue with the interpretation of the performance data in Experiment 1 is that individuals were allowed to physically rotate. This is particularly problematic given that there were marked differences in physical rotation across conditions. Thus, the equivalent performance in the RW-RF and RW-UF conditions might be due to individuals physically rotating more often in the latter condition, possibly speeding up performance. We addressed this issue in Experiment 2 by adding a fixed condition wherein participants were restricted from physically rotating. This condition permits a relatively direct assessment of the costs associated with the different types of rotation used in Experiment 1 and, of course, mirrors how rotation costs would have traditionally been indexed in the deep literature investigating viewpoint costs (e.g., Corballis, 1988; Graf, 2006; Jolicoeur, 1990; Kolers & Perkins, 1969a; 1969b; Koriat & Norman, 1984). Thus, if individuals are spontaneously physically rotating to subvert time based rotation costs, then when this not permitted, we should expect to find the largest costs in the RW-RF condition, followed by the RW-UF condition, and then the UW-RF condition.

In addition to adding a fixed condition, we also included a task following the primary task in which individuals provided subjective effort ratings associated with reading the texts across the various conditions in Experiment 2. As noted above, the effort construct can be difficult to operationalize and to date effort has been considered equivalent to time on task (e.g., Gray et al., 2006; Risko et al., 2014). While the existence of a relation between time and effort is certainly intuitive, the possibility that subjective estimates of effort might, at least in some conditions, deviate from estimates of effort based on time (or performance in general) remains a distinct possibility. If we consider subjective estimates of effort to reflect a kind of inference-based metacognitive judgment (Efklides, 2008; Koriat, 2007; Metcalfe, 2009; Nelson & Narens, 1990), then we might expect them to be influenced by factors other than objective task demands (e.g., preconceived biases, intuitive theories; Koriat, 2007). For example, numerous studies in the metamemory literature demonstrate dissociations between the influence of a manipulation on objective memory performance and its influence on metacognitive judgments of memory performance (e.g., Castel, Rhodes, & Friedman, 2012; Koriat, 1995; Metcalfe, Schwartz, & Joaquin, 1993; Schwartz & Metcalfe, 1992). Thus, subjective measures of effort, heretofore unexplored in this context, could aid significantly in the development of a theory of the decision processes underlying cognitive offloading. Experiment 2 represents the first attempt to do so. The straightforward prediction is that individuals will spontaneously physically rotate in conditions that are rated subjectively to be the most demanding (independent of whether this is actually objectively the case).

Method

Participants. Thirty-two University of Memphis undergraduate students (14 females) participated in the study in exchange for research credit.

Design. A two (Instruction: fixed-to-upright, free-to-rotate) x four (Array Type: RW-RF, RW-UF, UW-RF, UW-UF) within-subject design was employed.

Apparatus. Presentation of stimuli and recording of vocal responses were handled by DMDX software (Forster & Forster, 2003). One camera, which was placed on top of the monitor facing the participant, was used to record head movement. An additional monitor was placed behind the participant to record the stimulus being presented. All other apparatus was the same as Experiment 1.

Stimuli. Stimuli were similar to those in Experiment 1 except that words consisted of four and five letter nouns and verbs, one to two syllables each, with average word frequencies of 64.5

and 47.8 per million for four-letter and five-letter words, respectively. Participants completed two blocks of 32 trials with each array type occurring eight times.

Procedure. The procedure for Experiment 2 matched that of Experiment 1. However, instructions in the fixed-to-upright block stated that participants needed to keep their heads in an upright position and to try to remain as still as possible while reading aloud; the free-to-rotate block instructions remained the same as in Experiment 1. The order of instruction was counterbalanced across participants and a short break was given to participants between blocks[2]. Participants then completed two blocks of a paper and pencil version of the NASA Task Load Index measure (TLX: Hart & Staveland, 1988) for each array type. Scales ranged from 0 to 100. Instruction (i.e., fixed-to-upright vs. free-to-rotate) was blocked. For example, "Please rate how mentally demanding reading each array would be based on if you were freely able to physically rotate (but not required to)". Each disoriented array type was rated a total of eight times, once at each direction and word length combination, while the UW-UF condition was rated twice at each word-length for a total of 28 trials.

Results

Results are reported first for the frequency of physical rotation data followed by performance data[3] (see Table 2a) and subjective effort ratings. GLMM and LMM procedures matched those of Experiment 1 (e.g., model comparison, categorical variable dummy coding, releveling process, and significance criterions). The subjective effort ratings analysis used LMMs. The first three trials were removed as practice for performance analyses.

---

[2] No order effects of block were found for Experiment 1 or any of the experiments further reported.

[3] Error analyses revealed no significant effects and are thus not reported (see Table 2a for overall means).

Physical Rotation

Physical head rotation was determined by the same criteria as in Experiment 1. To calculate intra- and inter-rater reliability 25% of the data files (8 participants) were recoded by the initial coder and by a different coder; again both raters were blind to the presented array types. Both inter- and intra-rater reliability were high, K = .82 and K = .93, respectively. The main effect model consisted of the fixed effects of array type and trial, with subject and item as random effects.

Similar to Experiment 1, model comparison produced a final model consisting of only array type. The effect of each disoriented array type was significant when compared to the UW-UF condition, b = 5.04, SE = .68, Z = 7.46; b = 3.12, SE = 0.67, Z = 4.68; b = 1.89, SE = 0.69, Z = 2.76, for the RW-RF, RW-UF, and UW-RF conditions, respectively. In addition, individuals were more likely to rotate in the RW-RF condition when compared to the RW-UF condition, b = 1.93, SE = .32, Z = 6.10, and the UW-RF condition, b = 3.15, SE = .38, Z = 8.23, as well as more likely to rotate in the RW-UF condition when compared to the UW-RF condition, b = 1.22, SE = .37, Z = 3.29. Therefore the results replicated the exact pattern found in Experiment 1 (see Table 2b; Figure 2).

Response Time and Accuracy

RTs were calculated using CheckVocal software (Protopapas, 2007). Total Response Time was calculated as the time from stimulus onset to the vocal onset of the last word of the array. All RTs (secs) were transformed using the reciprocal transformation and accuracy was determined using the same criteria as in Experiment 1. One trial was removed as an extreme outlier based on visual inspection of residual plots for the fixed-to-upright analysis. An additional 5% of trials were removed due to participants hitting the spacebar to move to the next trial before they had finished reading the words in the array. The main effect model for the fixed-to-upright condition consisted of the fixed effects of array type and trial, and subject and item as random effects. The main effect model for total response time, errors, and subjective effort consisted of the fixed effects of instruction, array type and trial, with subject and item as random effects.

Fixed-to-Upright. To examine the costs associated with both word rotation and frame rotation, a model was fit using RTs from the fixed-to-upright condition only. The model comparison process demonstrated that the model of best fit consisted of array type. Response times were significantly longer for all disoriented array types when compared to the UW-UF condition, b = .0043, SE = .0007, t = 6.23; b = .0047, SE = .0007, t = 6.61; b = .0024, SE = .0007, t = 3.52, for the RW-RF, RW-UF, and UW-RF conditions, respectively. In addition, RTs were significantly slower in the RW-RF and RW-UF conditions when compared to the UW-RF condition, b = .0019, SE = .0007, t = 2.70; b = .0022, SE = .0007, t = 3.09. Critically, no significant difference was found between the RW-RF and RW-UF conditions. Thus, the RW-RF and RW-UF conditions showed similar internal costs, with both being greater than the UW-RF condition and all three being greater than the UW-UF conditions (see Table 2c; Figure 4).

Combined Fixed-to-Upright and Free-to-Rotate. If the option to physically rotate is facilitating performance for the free-to-rotate instruction, then we should observe faster RTs in those conditions with the highest frequencies of spontaneous physical rotations (i.e., RW-RF, and RW-UF to some extent). To test this hypothesis, the fixed effect of instruction was added to the model outlined in the previous section. The model comparison process demonstrated that the model of best fit included instruction, array type, and trial (i.e., there was no interaction between instruction and array type). Participants produced a similar pattern of results for array type and trial as outlined above. As for instruction, participants produced faster RTs in the free-to-rotate instruction condition when compared to the fixed-to-upright instruction condition across all array types, b = -.0012, SE = .0004, t = -2.03 (see Table 2c). Importantly, the lack of an array type by instruction interaction implies that the benefit observed as a function of instruction is not specific to an array type. To further explore this pattern, a model was constructed using data from only the free-to-rotate condition examining whether trials where individuals spontaneously physically rotated led to faster reading compared to trials where there were no rotations. Results demonstrated a significant array type x physical rotation interaction. The RW-RF and RW-UF conditions did not show significant effects of physical rotation, $t < -.63$. The UW-RF condition, however, showed a significant cost when individuals physically rotated compared to all other array types, $t > 2.33$ for all comparisons.

Subjective Effort

Subjective effort was determined by the raw scores from the “Mental Demand” subscale (i.e., “How mentally demanding was the task above?”) of the NASA-TLX. Furthermore, a $|t| > 2$ criterion was used to evaluate the significance of each fixed factor (see Table 5).

The model comparison process demonstrated that the model including array type produced the best fit to the data. The effect of each disoriented array type was significant when compared to the UW-UF condition, b = 13.73, SE = 2.18, t = 6.30; b = 10.45, SE = 2.18, t = 4.80; b = 4.84, SE = 2.18, t = 2.22, for RW-RF, RW-UF, and UW-RF, respectively. The RW-RF and RW-UF conditions produced higher effort scores when compared to the UW-RF condition, b = 8.89, SE = 1.78, t = 4.99; b = 5.61, SE = 1.78, t = 3.15, respectively. The difference between the RW-RF and RW-UF conditions closely approached our significance criterion, t = 1.84, such that individuals rated the RW-RF condition as more effortful than the RW-UF condition. Thus, the RW-RF condition produced the highest effort rating followed by the RW-UF condition, and then UW-RF condition, and all were greater than the UW-UF condition (see Table 2e; Figure 5). The lack of an effect of instruction (i.e., it was not included in the final model) means that individuals' ratings did not differ depending on whether they were asked to provide ratings based on reading fixed-to-upright or free-to-rotate.

Table 2a.

Mean Performance and Subjective Effort Results for Experiment 2

Variable	Instruction	Array Type							
		UW-UF		UW-RF		RW-UF		RW-RF	
Percentage of Trials Rotated	*Fixed-to-Upright*	-	-	-	-	-	-	-	-
	Free-to-Rotate	.01%	(12%)	7%	(26%)	15%	(36%)	39%	(49%)
RT (ms)	*Fixed-to-Upright*	16455	(3429)	17100	(4033)	17653	(3501)	17656	(3573)
	Free-to-Rotate	16316	(3592)	16691	(3720)	17087	(3771)	16976	(3530)
Errors	*Fixed-to-Upright*	.53	(1.01)	.47	(.95)	.45	(.83)	.58	(.97)
	Free-to-Rotate	.45	(.84)	.51	(.98)	.60	(.95)	.55	(.84)
Subjective Effort	*Fixed-to-Upright*	23.50	(24.74)	31.32	(24.64)	34.71	(26.27)	36.21	(27.15)
	Free-to-Rotate	23.81	(23.99)	25.55	(24.48)	33.23	(25.25)	38.28	(29.14)

Note: UW-UF= Upright Word-Upright Frame; UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame; RT = response times. Results reported in the table are overall means. Standard deviations are presented in parentheses.

Table 2b.

Generalized Linear Mixed Model Results for Physical Head Rotations (log Odds) in Experiment 2

Random Effects	Variance	SD	
Subject	4.83	2.20	
Item	-	-	
Fixed Effects	**Estimate**	***SE***	***Z***
Intercept	-5.91	.76	-7.77
UW-RF	1.89	.69	2.76
RW-UF	3.12	.67	4.68
RW-RF	5.04	.68	7.46

Note: UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. A |*Z*| > 2 criterion was used to evaluate the significance of each fixed factor.

Table 2c.

Linear Mixed Model Results for Fixed-to-Upright Condition (-1/RT; sec) in Experiment 2

Random Effects	Variance	SD	
Subject	8.55e-05	9.25e-03	
Item	5.62e-15	7.50e-08	
Residual	5.31e-05	7.29e-03	
Fixed Effects	Estimate	*SE*	*t*
Intercept	-.0635	.0017	-37.23
UW-RF	.0024	.0007	3.52
RW-UF	.0046	.0007	6.61
RW-RF	.0043	.0007	6.23

Note: UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. A $|t| > 2$ criterion was used to evaluate the significance of each fixed factor.

Table 2d.

Linear Mixed Model Results for combined Fixed-to-Upright and Free-to-Rotate conditions (-1/RT; sec) in Experiment 2

Random Effects	Variance	SD	
Subject	9.10e-05	9.53e-03	
Item	1.92e-07	4.38e-04	
Residual	5.81e-05	7.61e-03	
Fixed Effects	**Estimate**	***SE***	***t***
Intercept	-.0630	.0017	-36.35
Free-to-Rotate	-.0012	.0004	-3.38
UW-RF	.0020	.0005	3.98
RW-UF	.0039	.0005	7.72
RW-RF	.0035	.0005	6.75
Trial	-.0004	.0001	-2.13

Note: UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. A $|t| > 2$ criterion was used to evaluate the significance of each fixed factor.

Table 2e.

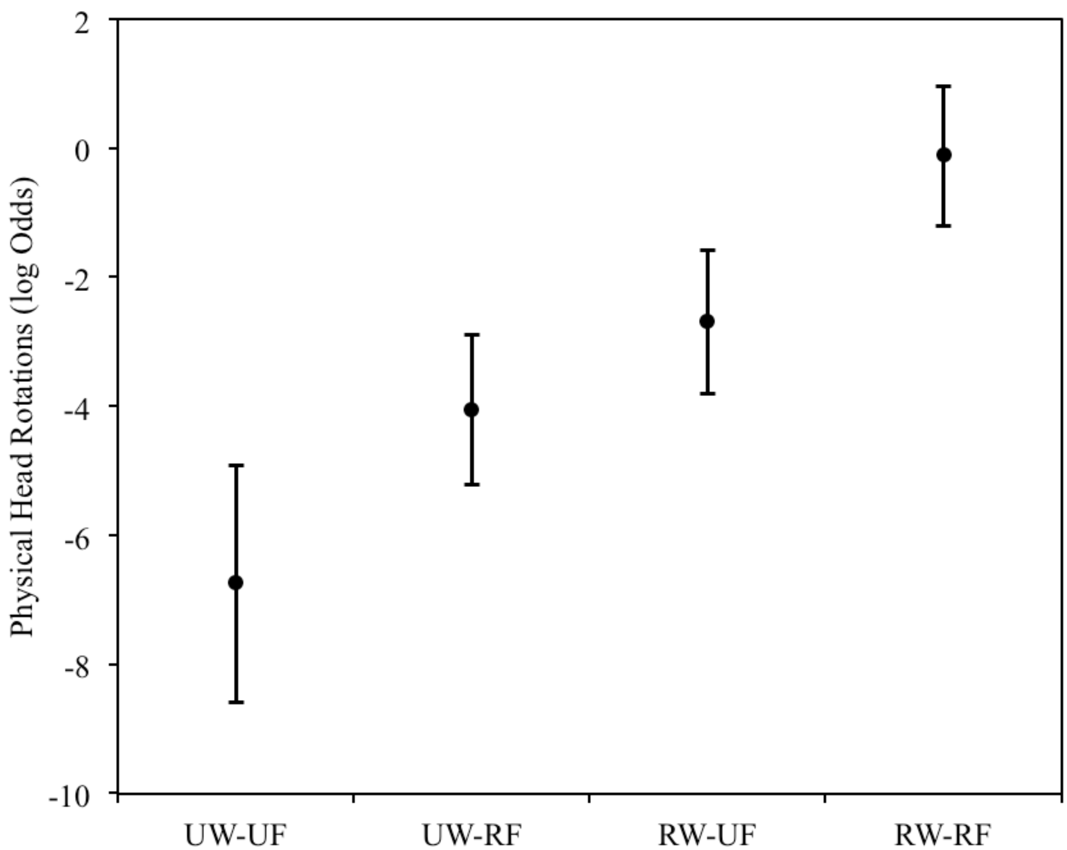
Linear Mixed Model Results for Subjective Effort Ratings in Experiment 2

Random Effects	Variance	SD	
Subject	285.66	16.90	
Item	-	-	
Residual	405.16	20.13	
Fixed Effects	Estimate	*SE*	*t*
Intercept	23.40	3.48	6.73
UW-RF	4.84	2.18	2.22
RW-UF	10.45	2.18	4.80
RW-RF	13.73	2.18	6.30

Note: UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. A $|t| > 2$ criterion was used to evaluate the significance of each fixed factor.

Figure 4.

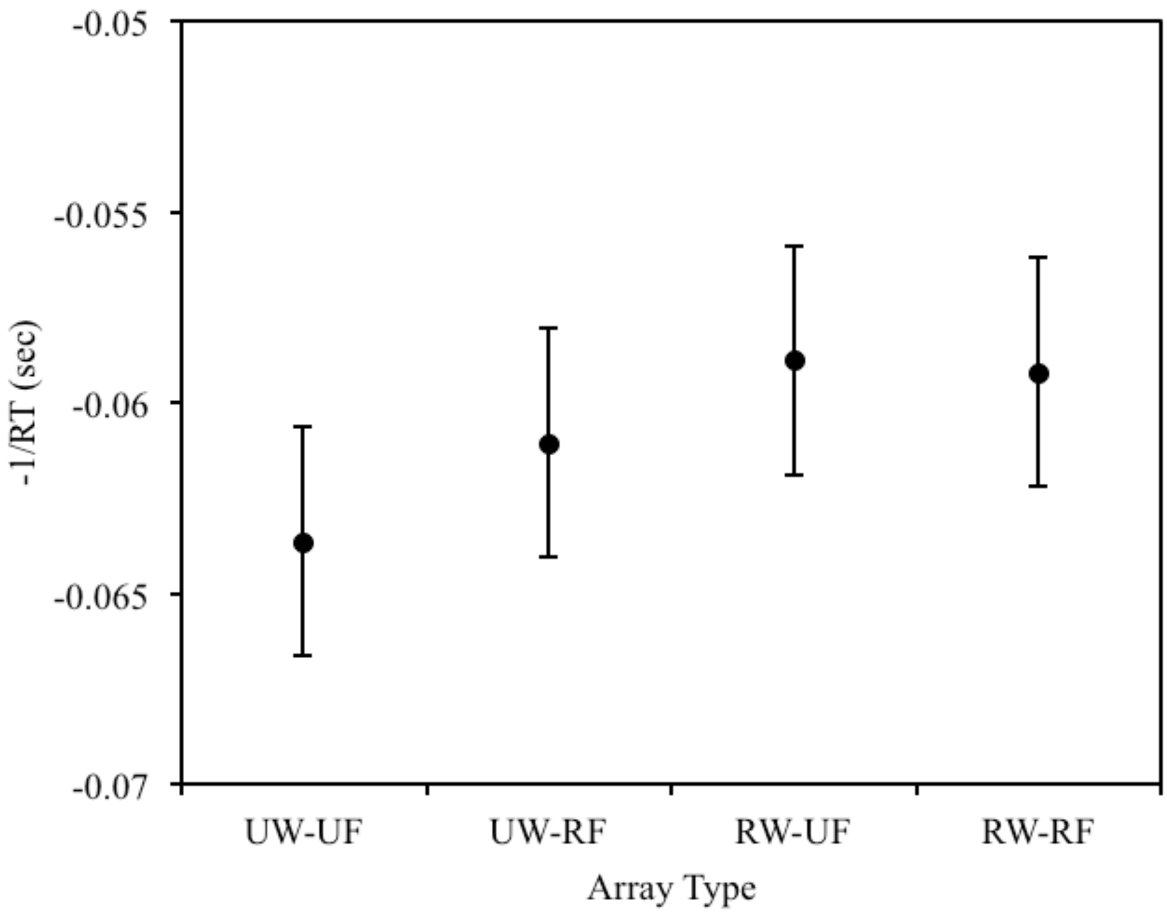
Spontaneous Physical Head Rotation Results for Experiment 2



Note: UW-UF = Upright Word-Upright Frame; UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. Error bars denote 95% confidence intervals.

Figure 5.

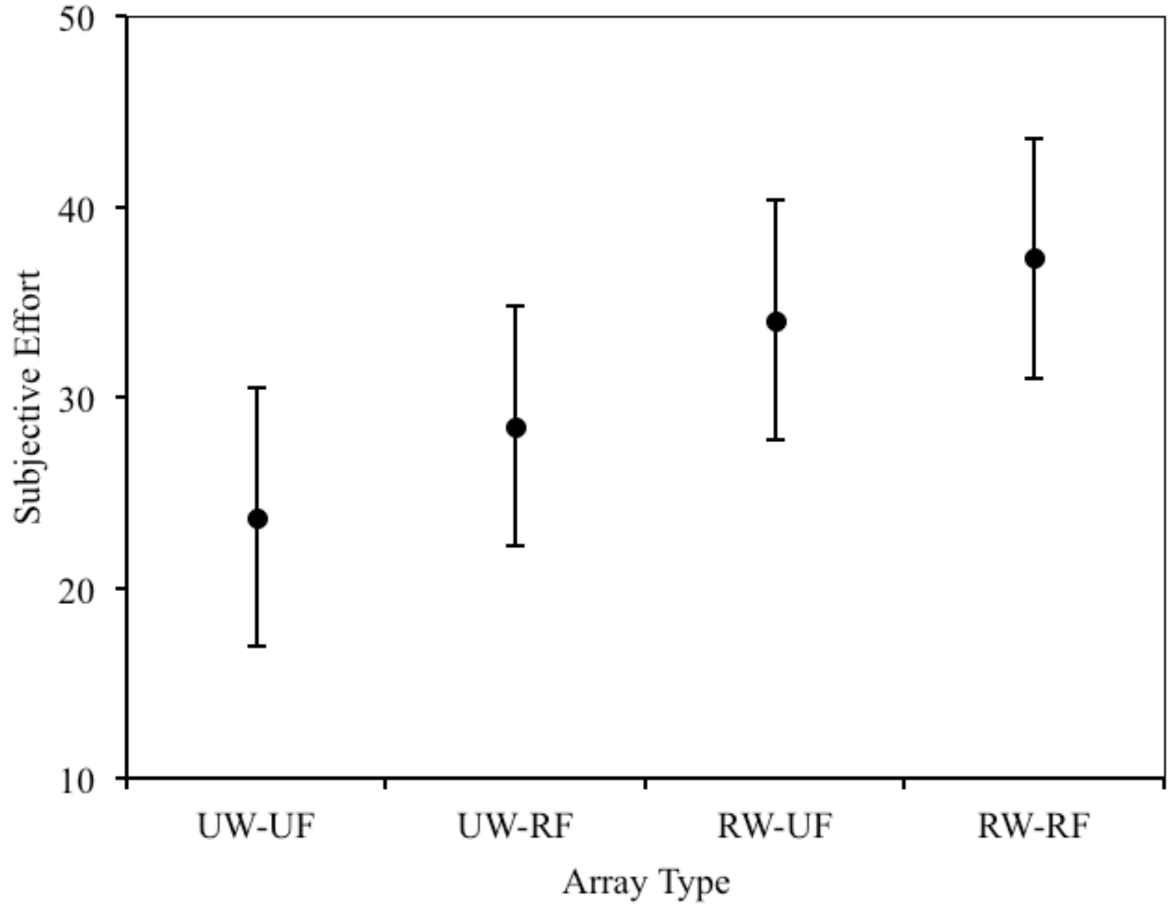
Results for Fixed-to-Upright Condition for Experiment 2



Note: UW-UF = Upright Word-Upright Frame; UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. Error bars denote 95% confidence intervals.

Figure 6.

Subjective Effort Results for Experiment 2



Note: UW-UF = Upright Word-Upright Frame; UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. Error bars denote 95% confidence intervals.

Discussion

Experiment 2 replicated the major findings from Experiment 1 for both spontaneous physical head rotations and stimulus rotation costs. Specifically, individuals more frequently rotated in the RW-RF condition, followed by the RW-UF condition, and then the UW-RF condition. This pattern did not follow the pattern in terms of stimulus rotation costs in the fixed-to-upright condition. Rather, stimulus rotation costs for the RW-RF and RW-UF conditions were similar and larger than in the UW-RF condition. Thus the dissociation between likelihood of physical rotation and stimulus rotation costs remained despite the addition of the fixed condition, which makes the interpretation of stimulus rotations costs in terms of some internal process much more clear.

One unexpected result from Experiment 2 was the lack of any benefit of physically rotating with respect to the stimulus rotation costs (i.e., physically rotating in the direction of the stimulus did not reduce the costs of stimulus rotation). Risko et al. (2013) reported a similar finding in relatively simple displays. Critically, Risko et al. (2013) also demonstrated reductions in stimulus rotation costs when individuals spontaneously physically rotated or were forced to rotate when reading larger displays (i.e., paragraphs of text). Thus, the behavior can benefit performance. That said, it is nonetheless interesting from an effort minimization (defined in terms of time) perspective that individuals spontaneously physically rotate as often as they do in Experiments 1 and 2 (at least in the RW-RF condition) despite the fact that there are no performance benefits relative to remaining upright. It is important to note that although there were no observable performance benefits, this does not mean that physically rotating has no impact on processing per se during the trial. Specifically, physically rotating requires time in the sense of both the decision making process (i.e., “should I rotate”, “which way”) and the physical act itself. Thus, physically rotating could “save” or offload cognitive processing, but these savings may be washed out by the cost associated with physical rotation. Experiment 3 will more closely examine the benefits of physical rotation across the display types used in Experiments 1 and 2.

The results from the subjective effort ratings were interesting but unclear. While mental demand appeared to increase in a manner akin to the pattern found in the likelihood of physical rotation (i.e., UW-UF < UW-RF < RW-UF < RW-RF), the paired comparison between RW-RF and RW-UF was only marginal. Nonetheless, this result raises the interesting possibility that subjective effort ratings do not follow objective performance, but rather mirror individuals’ spontaneous physical rotations. Thus, individuals appear to be deciding to physical rotate in the condition that they perceive to be the most difficult (i.e., RW-RF) despite the fact that their performance (when remaining upright) does not bear this out. However, individuals did not perceive a difference in effort between reading arrays when fixed-to-upright and free-to-rotate, a finding difficult to reconcile with any effort-based account of the behavior rendering the data difficult to interpret. In Experiment 3, we modify the design (and increase the sample size) in an attempt to clarify some of these ambiguous results.

Experiment 3

To better evaluate the potential benefit of physical rotation across the various display types used in Experiments 1 and 2, in Experiment 3 we used the fixed-to-upright condition, as in Experiment 2, and added a condition in which individuals were forced to physically rotate prior to stimulus onset. This ensures a physical orientation (roughly) consistent with the stimulus for every trial rather than relying on spontaneous rotations as in Experiments 1 and 2. Thus, any time associated with the act of physically rotating within the trial would be neutralized and the greatest benefit (if any) could be observed. Furthermore, the addition of the forced-to-rotate condition completes all of the relevant conditions of the choice/no-choice paradigm devised by Siegler and LeMaire (1997) to investigate strategy selection. Utilizing this approach provides estimates of the performance associated with the execution of each of the strategies available to an individual (i.e., forced to remain upright, forced to physically rotate). The inclusion of the forced-to-rotate condition is also important because it remains possible that despite the fact that the rotation costs are equivalent in the RW-RF and RW-UF conditions, the benefits of physical rotation might be different. In particular, the performance (in theory) benefit associated with physically rotating might be larger in the RW-RF than the RW-UF condition, possibly because physically rotating in this condition normalizes both the words and the frame of the array. Such a result would provide a clear explanation of the pattern of spontaneous physical rotations in terms of objective performance (i.e., time savings).

We also used both the fixed-to-upright and forced-to-rotate conditions in the assessment of subjective effort. The removal of the "free" condition from the subjective rating task should serve to make clearer participant's judgments of effort (i.e., in the free condition it is unclear whether participants are rating from the perspective of being rotated or upright). In addition to this change, participants provided subjective effort ratings for the different array types and the different instructions (i.e., fixed-to-upright versus forced-to-rotate) mixed within block (in Experiment 2 instruction was blocked). The prediction relating to subjective effort is similar to the hypothesis outlined for performance in this experiment.

Method

Participants. Forty-eight University of Waterloo undergraduate students (44 females) participated in the study in exchange for research credit.

Design. A two (Instruction: fixed-to-upright, forced-to-rotate) x four (Array Type: RW-RF, RW-UF, UW-RF, UW-UF) within-subject design was employed.

Apparatus and Stimuli. All apparatus and stimuli matched that of Experiment 2.

Procedure. The procedure for Experiment 3 was similar to that of Experiment 2. However, an angle was presented for seven seconds congruent to the orientation of the stimulus prior to onset. Instructions in the forced-to-rotate condition stated that participants must rotate their head to match the angle as best as possible and to keep their head rotated at that position for the entire duration of their response. Instructions for the fixed-to-upright condition remained the same as in Experiment 2 (although angles were presented prior to stimulus onset in this condition as well). The order of blocks was counterbalanced across participants and a short break was given to participants between blocks. Participants then completed mental demand ratings on a 1 to 10 Likert-type scale for each instruction and array. Unlike Experiment 2, instruction was randomized rather than blocked. This design provides the opportunity for individuals to compare both instruction and array type in a context (potentially) less influenced by the actual performance of the task. All array types were rated a total of eight times, once at each direction and number of letter per word combination for a total of 32 trials.

Results

Response Time and Accuracy

Total Response Time was calculated and transformed in the same manner as in Experiment 2. Accuracy was determined using the same criteria as Experiments 1 and 2. One trial was removed as an extreme outlier based on visual inspection of residual plots. An additional 2% of trials were removed due to participants hitting the spacebar to move to the next trial before they had finished reading the words in the array.

Total Response Time. Model comparisons demonstrated that a model consisting of the 2-way interaction between instruction x array type and the main effect of trial was the model of best fit. Participants got faster to name the array aloud as trial progressed, b = -.0005, SE = .0001, t = -4.02. The effect of each array type in the fixed-to-upright condition was similar to the pattern observed in Experiment 2 where the RW-RF and RW-UF conditions showed similar internal costs, $t = 1.30$, with both arrays showing larger costs than the UW-RF condition, $t > 5.61$, for both comparisons. Participants were faster in the forced-to-rotate condition than in the fixed-to-upright condition for the RW-RF and RW-UF conditions, b = -.0029, SE = .0005, t = -6.42; b = -.0025, SE = .0005, t = -5.49, respectively. However, the RW-RF and RW-UF conditions did not differ in the magnitude of this RT reduction in the forced-to-rotate condition, t = .72 (i.e., the benefit of physically rotating was equivalent). In the UW-RF condition, individuals were slower in the forced-to-rotate condition than in the fixed-to-upright condition, b = .0023, SE = .0005, t = 4.87. That is, individuals showed a significant cost in terms of RT when being forced to rotate. No effect of instruction was found for the UW-UF condition, $t = .07$ (see Table 3b; Figure 6).

Accuracy. Model comparisons demonstrated that the 2-way interaction model including the instruction x array type interaction and the main effect of trial produced the best fit to the data. Individuals made more errors as trials progressed, b = .06, SE = .03, Z = 2.13. The effect of each array type in the fixed-to-upright condition was similar to the pattern of RTs above where the RW-RF and RW-UF conditions showed similar errors, *Z* = -.17, with both arrays showing more errors than the UW-RF and UW-UF conditions, *Z* > 2.66, respectively. Individuals showed reduced errors in the forced-to-rotate condition when compared to the fixed-to-upright condition in the RW-RF and RW-UF conditions, b = -.28, SE = .12, Z = -2.41; b = -.29, SE = .11, Z = -2.62, respectively. Similar to RT, the RW-RF and RW-UF conditions did not differ in the magnitude of reduced errors in the forced-to-rotate condition, Z = .06. Thus, the pattern of errors across the RW-RF and RW-UF conditions matched that found in RTs. In addition, individuals made more errors in the forced-to-rotate condition than in the fixed-to-upright condition for the UW-RF condition, b = .26, SE = .12, Z = 2.09, again a pattern similar to that found in RT. No effect of instruction was found for the UW-UF condition, *Z* = -.36 (see Table 3c).

Subjective Effort

Model comparison and significance criteria matched those of the subjective effort analysis in Experiment 2. One participant could not be included in the analysis as they were unable to complete the ratings in the time allotted for the experiment (see Table 5).

Model comparisons demonstrated that the 2-way interaction model including the instruction x array type interaction produced the best fit to the data. The effect of each disoriented array type was significant when compared to the UW-UF condition in the fixed-to-upright condition, b = 2.81, SE = .10, t = 27.21; b = 2.45, SE = .10, t = 23.66; b = 1.23, SE = .10, t = 11.87, for the RW-RF, RW-UF, and UW-RF conditions, respectively. Participants rated the RW-RF condition as more effortful when compared to the RW-UF condition, b = .35, SE = .10, t = 3.45, and the UW-RF condition, b = 1.59, SE = .10, t = 15.39. In addition, the RW-UF condition was rated as more effortful when compared to the UW-RF condition, b = 1.23, SE = .10, t = 11.87. Effort ratings decreased in the forced-to-rotate instruction when compared to the fixed-to-upright condition for the RW-RF and RW-UF conditions, b = -1.42, SE = .10, t = -13.76; b = -.80, SE = .10, t = -7.74, respectively. Critically, the reduction in the effort ratings in the forced-to-rotate condition was larger for the RW-RF condition than for the RW-UF condition, b = -.62, SE = .15, t = -4.23. An increase in ratings for the forced-to-rotate instruction was found for the UW-RF condition, b = .55, SE = .10, t = 5.35. Thus, the forced-to-rotate instruction decreased subjective effort ratings when compared to the fixed-to-upright condition for the RW-RF and RW-UF conditions and this effect was larger for the RW-RF condition than for the RW-UF condition. The UW-RF condition showed an increase in subjective effort ratings in the forced-to-rotate condition when compared to the fixed-to-upright condition, and no effect of instruction was found for the UW-UF condition (see Table 3d; Figure 7).

Table 3a.

Mean Performance and Subjective Effort Results for Experiment 3

Variable	Instruction	Array Type							
		UW-UF		UW-RF		RW-UF		RW-RF	
RT (ms)	*Fixed-to-Upright*	15257	(3624)	15743	(3620)	16473	(4358)	16544	(3868)
	Forced-to-Rotate	15315	(3717)	16304	(3806)	15941	(3918)	15875	(3848)
Errors	*Fixed-to-Upright*	.32	(.84)	.31	(.85)	.50	(.92)	.47	(.91)
	Forced-to-Rotate	.37	(.90)	.39	(.86)	.36	(.86)	.35	(.84)
Subjective Effort	*Fixed-to-Upright*	1.14	(.92)	2.70	(1.29)	3.94	(1.42)	4.28	(1.40)
	Forced-to-Rotate	1.49	(.94)	3.25	(1.25)	3.13	(1.35)	2.87	(1.35)

Note: UW-UF= Upright Word-Upright Frame; UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame; RT = response times. Results reported in the table are overall means. Standard deviations are presented in parentheses.

Table 3b.

Linear Mixed Model Results for Total Response Time (-1/RT; sec) in Experiment 3

Random Effects	Variance	SD	
Subject	1.56e-04	.012	
Item	8.82e-06	.003	
Residual	3.67e-05	.006	
Fixed Effects	Estimate	*SE*	*t*
Intercept	.0685	.0019	-36.69
Forced-to-Rotate	.00003	.0005	.07
UW-RF	.0019	.0005	4.27
RW-UF	.0046	.0005	9.94
RW-RF	.0052	.0005	11.12
Force-to-Rotate:UW-RF	.0022	.0007	3.38
Forced-to-Rotate:RW-UF	-.0026	.0007	-3.90
Forced-to-Rotate:RW-RF	-.0030	.0007	-4.59
Trial	-.0005	.0001	-4.02

Note: UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. A $|t| > 2$ criterion was used to evaluate the significance of each fixed factor.

Table 3c.

Generalized Linear Mixed Model Results for Errors (log Counts) in Experiment 3

Random Effects	Variance	SD	
Subject	.58	.76	
Item	.03	.18	
Fixed Effects	Estimate	*SE*	*Z*
Intercept	-1.29	.15	-8.73
Forced-to-Rotate	.04	.12	.36
UW-RF	-.11	.13	-.84
RW-UF	.33	.12	2.86
RW-RF	.31	.12	2.66
Forced-to-Rotate:UW-RF	.21	.17	1.24
Forced-to-Rotate:RW-UF	-.34	.16	-2.05
Forced-to-Rotate:RW-RF	-.33	.17	-1.95
Trial	.06	.03	2.16

Note: UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. A |*Z*| > 2 criterion was used to evaluate the significance of each fixed factor.

Table 3d.

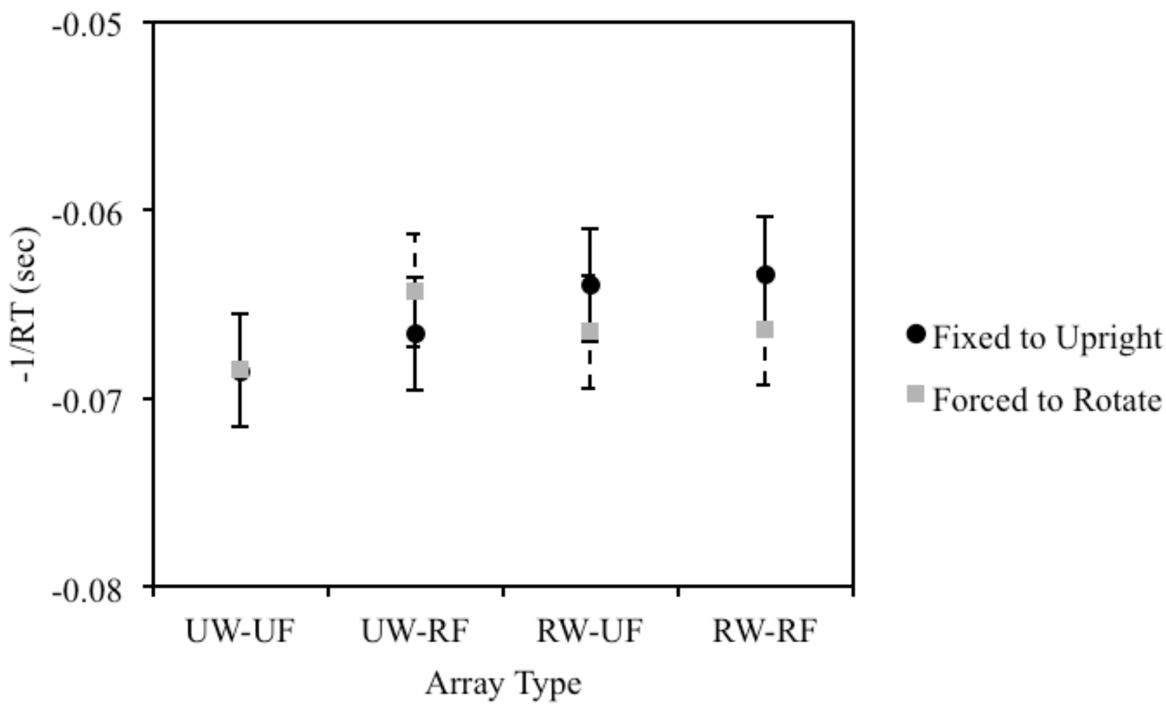
Linear Mixed Model Results for Subjective Effort Ratings in Experiment 3

Random Effects	Variance	SD	
Subject	.58	.76	
Item	-	-	
Residual	1.01	1.00	
Fixed Effects	Estimate	*SE*	*t*
Intercept	1.48	.13	11.10
Forced-to-Rotate	.02	.10	.19
UW-RF	1.23	.10	11.87
RW-UF	2.45	.10	23.66
RW-RF	2.81	.10	27.21
Forced-to-Rotate:UW-RF	.53	.15	3.64
Forced-to-Rotate:RW-UF	-.82	.15	-5.61
Forced-to-Rotate:RW-RF	-1.44	.15	-9.85

Note: UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. A $|t| > 2$ criterion was used to evaluate the significance of each fixed factor.

Figure 7.

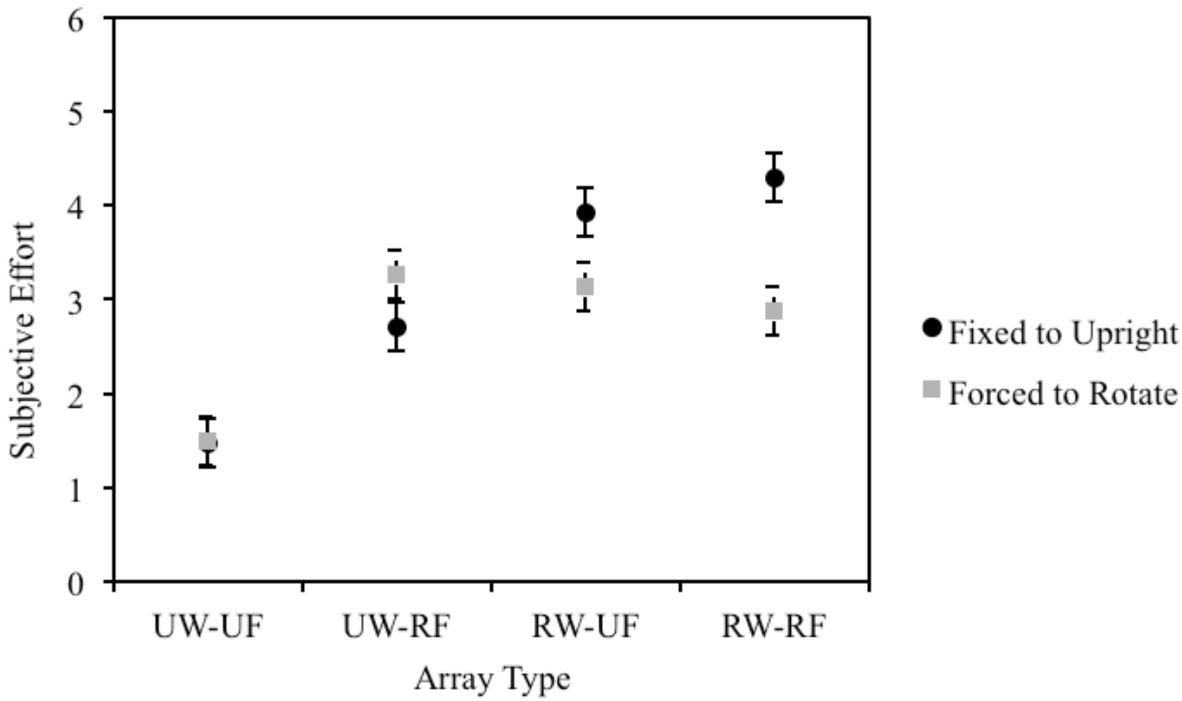
Response Time Results for Experiment 3



Note: UW-UF = Upright Word-Upright Frame; UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. Error bars denote 95% confidence intervals.

Figure 8.

Subjective Effort Results for Experiment 3



Note: UW-UF = Upright Word-Upright Frame; UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. Error bars denote 95% confidence intervals.

Discussion

Results from Experiment 3 provided several key insights. There was a benefit of being physically rotated when compared to being fixed-to-upright for the RW-RF and RW-UF conditions. A similar pattern of results was found in accuracy. Critically, the benefits in the RW-RF and RW-UF conditions were similar in magnitude. Thus, differential potential benefit in terms of performance (i.e., time savings, accuracy savings) cannot account for the fact (observed in both Experiments 1 and 2) that individuals rotate far more often in the RW-RF condition than they do in the RW-UF condition.

Although performance results were unable to explain the patterns of spontaneous physical rotation, in Experiment 3, individuals' subjective ratings of effort provided an excellent qualitative fit to this pattern. Specifically, the RW-RF condition was rated as more difficult than the RW-UF condition when participants imagined themselves fixed-to-upright, and physically rotating was perceived as being more beneficial in the RW-RF condition than in the RW-UF condition. In addition, a perceived cost was observed for the UW-RF condition. These patterns of results effectively mirror the pattern of spontaneous physical rotations in Experiments 1 and 2. Thus, while the stimulus rotation costs and the potential benefit of physically rotating are equivalent across the RW-RF and RW-UF conditions, participants in Experiment 3 perceived the stimulus rotation costs and the benefit of physically rotating in the RW-RF condition to be greater than in the RW-UF condition. This is critical because participant's behavior matches the latter (i.e., the pattern of perceived effort) and not the former (i.e., the pattern based on performance or an "objective" measure of effort). This pattern is consistent with the idea that individuals are making a kind of inference-based metacognitive judgment of the effort associated with the different strategies (i.e., remain upright, physically rotate) and using that judgment in deciding whether to adopt an external strategy (Arango-Muñoz, 2013). The fact that individuals' metacognitive judgments do not mirror actual performance could reflect a kind of metacognitive error. Experiment 4 looks to replicate and extend this pattern.

Experiment 4

In Experiment 3, we observed another dissociation, this time between perceived effort and performance (i.e., time, accuracy). The match between perceived effort judgments and spontaneous physical rotation suggests that individuals might be relying on such judgments to make their decision whether to physically rotate when free to do so. In this light, there are (at least) two ways to think about the dissociation between perceived effort ratings and performance (i.e., time, accuracy). Participants' perception of effort might differ from their perception of their performance, assuming the latter is accurate, and rely on the former when making the decision to physically rotate (i.e., participants might be trying to minimize perceived effort rather than to maximize performance). Alternatively, participant's perception of effort and performance might both be incorrect (assuming for the moment that "correct" means consistent with objective performance) thus leaving open the possibility that individuals are trying to maximize performance but are simply incorrectly judging how the various stimulus conditions and physically rotating will influence their performance. In Experiment 4, we address these two alternatives by asking participants to rate how subjectively effortful, time demanding, or accurate they would be when reading the various displays used in the previous experiments.

Method

Participants. One hundred and sixteen individuals participated using Amazon Mechanical Turk. Four individuals were removed due to providing the same ratings on every trial (e.g., simply marking "5" for every condition) as well as completing the entire study in less than half of the time allotted. Participants were all over the age of 18 years old, native English speakers, and were currently situated in the United States. Participants were compensated $2.50 upon completion of the study.

Design. A three (Dimension: mental demand, accuracy, time) x two (Instruction: fixed-to-upright, forced-to-rotate) x four (Array Type: RW-RF, RW-UF, UW-RF, UW-UF) mixed design was employed. Dimension was manipulated between subjects whereas Instruction and Array Type were manipulated within subject.

Stimuli. One list of 32 trials was randomly chosen from the set of stimuli used in Experiments 2 and 3. Each disoriented array type (RW-RF, RW-UF, and UW-RF) was presented four times to the right of gravitational upright (0°) and four times to the left of gravitational upright each, while the UW-UF condition was presented eight total times at 0°. Array types contained both four and five letter words.

Procedure. Participants chose and accepted the "human intelligence task" on Amazon Mechanical Turk. Participants then electronically signed the informed consent protocol on the first page of the survey and completed demographic questions. Instructions stated that participants were to rate either how mentally demanding, how time demanding, or how accurate they would be in performing the task (e.g., reading the words) under certain conditions (e.g., not being able to rotate their head vs. having to rotate their head). For example, "Imagine you were presented with the above display. If you had to keep your head upright, then how mentally demanding would it be to read the stimulus presented aloud?" Participants provided ratings on a 7-point Likert-type scale ranging from "Not At All…" to "Very…" for the dimension they were randomly assigned. Each array type was rated four times for each instruction. Array type and instruction were randomly presented to participants. After finishing the rating task, participants received a short debriefing and were given a coded value to enter into Amazon Mechanical Turk to indicate completion and confirmation of the data.

Results

Results are reported first for mental demand ratings, followed by accuracy and time. Accuracy and time ratings were reverse scored to match the pattern of responses of mental demand ratings. One trial was removed as an extreme outlier based on visual inspection of residual plots for the time analysis. All ratings were Z-scored at the subject level to allow comparison across the three different dimensions of ratings (see Figure 8; Table 5). LMM procedures followed those of the previous experiments.

Mental Demand. The model comparison procedure demonstrated that the model of best fit consisted of the array x instruction interaction. Individuals reduced their subjective ratings in the forced-to-rotate condition when compared to the fixed-to-upright condition in only the RW-RF condition, b = -.60, SE = .12, t = -5.07. The opposite pattern was found in the UW-RF condition, where individuals increased their subjective ratings in the forced-to-rotate condition, when compared to the fixed-to-upright condition, b = .81, SE = .12, t = 6.91. Thus, individuals showed a perceived benefit in the RW-RF condition and a perceived cost in the UW-RF condition when being forced to rotate (see Table 4.1a).

Accuracy. The model comparison procedure demonstrated that the model of best fit consisted of the array x instruction interaction and the main effect of trial. Participants' accuracy ratings increased as trials progressed, b = -.06, SE = .02, t = 2.41. Individuals reduced their subjective ratings in the forced-to-rotate condition when compared to the fixed-to-upright condition for the RW-RF and RW-UF conditions, b = -.28, SE = .09, t = -3.02; b = -.73, SE = .09, t = -8.01, respectively. The RW-RF array type showed a larger decrease in ratings in the forced-to-rotate condition when compared to the RW-UF condition, b = -.45, SE = .13, t = -3.52. The opposite pattern was found for the UW-RF condition, where individuals increased their subjective ratings in the forced-to-rotate condition when compared to the fixed-to-upright condition, b = .73, SE = .09, t = 8.03. Thus, the RW-RF condition showed the largest perceived, benefit followed by the RW-UF condition, while the UW-RF condition showed a perceived cost of being forced to physically rotate (see Table 4.2b).

Time. The model comparison procedure demonstrated that the model of best fit consisted of the array x instruction interaction. Again, individuals reduced their subjective ratings in the forced-to-rotate condition when compared to the fixed-to-upright condition for the RW-RF and RW-UF conditions, b = -.69, SE = .08, t = -9.24; b = -.36, SE = .08, t = -4.82, respectively. The RW-RF condition showed the larger perceived benefit in the forced-to-rotate condition when compared to the RW-UF condition, b = -.33, SE = .11, t = -3.13. The opposite pattern was found for the UW-RF condition, where individuals increased their subjective ratings in the forced-to-rotate condition, b = .62, SE = .08, t = 8.27. Thus, similar to mental demand and accuracy, the RW-RF condition produced the largest perceived benefit followed by the RW-UF condition, while the UW-RF condition showed a perceived cost when individuals were asked to imagine that they must physically rotate toward the direction of the array while reading (see Table 4.3c).

Table 4a.

Mean Subjective Effort Results for Experiment 4

		Array Type							
Variable	Instruction	UW-UF		UW-RF		RW-UF		RW-RF	
Mental Demand	*Fixed-to-Upright*	1.25	(.57)	2.22	(1.37)	3.28	(1.58)	3.50	(1.48)
	Forced-to-Rotate	1.23	(.54)	3.14	(1.31)	3.27	(1.57)	2.89	(1.49)
Accuracy	*Fixed-to-Upright*	6.81	(.58)	6.15	(1.12)	5.14	(1.47)	4.98	(1.47)
	Forced-to-Rotate	6.83	(.58)	5.42	(1.38)	5.49	(1.33)	5.78	(1.32)
Time	*Fixed-to-Upright*	6.66	(.54)	5.49	(1.08)	4.28	(1.57)	4.13	(1.47)
	Forced-to-Rotate	6.58	(.62)	4.69	(1.44)	4.66	(1.57)	5.01	(1.39)

Note: UW-UF= Upright Word-Upright Frame; UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame; RT = response times. Results reported in the table are overall means. Standard deviations are presented in parentheses. Accuracy and Time were rated as "Very Accurate" and "Very Fast" whereas Mental Demand was rated as "Very Mentally Demanding" at the high end of the rating scale provided.

Table 4b.

Linear Mixed Model Results for Mental Demand Ratings (Z) in Experiment 4

Random Effects	Variance	SD	
Subject	-	-	
Item	.01	.12	
Residual	.51	.71	
Fixed Effects	Estimate	*SE*	*t*
Intercept	-1.03	.07	-14.42
Forced-to-Rotate	-.01	.08	-.13
UW-RF	.70	.11	6.37
RW-UF	1.52	.11	13.77
RW-RF	1.81	.11	16.43
Forced-to-Rotate:UW-RF	.83	.14	5.79
Forced-to-Rotate:RW-UF	.01	.14	.04
Forced-to-Rotate:RW-RF	-.59	.14	-4.12

Note: UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. A $|t| > 2$ criterion was used to evaluate the significance of each fixed factor.

Table 4c.

Linear Mixed Model Results for Accuracy Ratings (Z) in Experiment 4

Random Effects	Variance	SD	
Subject	-	-	
Item	-	-	
Residual	.59	.77	
Fixed Effects	Estimate	*SE*	*t*
Intercept	-.88	.06	-13.60
Forced-to-Rotate	-.05	.09	-.51
UW-RF	.52	.09	5.64
RW-UF	1.49	.09	16.32
RW-RF	1.66	.09	18.19
Forced-to-Rotate:UW-RF	.78	.13	6.05
Forced-to-Rotate:RW-UF	-.23	.13	-1.77
Forced-to-Rotate:RW-RF	-.68	.13	-5.30
Trial	.06	.02	2.41

Note: UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. A $|t| > 2$ criterion was used to evaluate the significance of each fixed factor.

Table 4d.

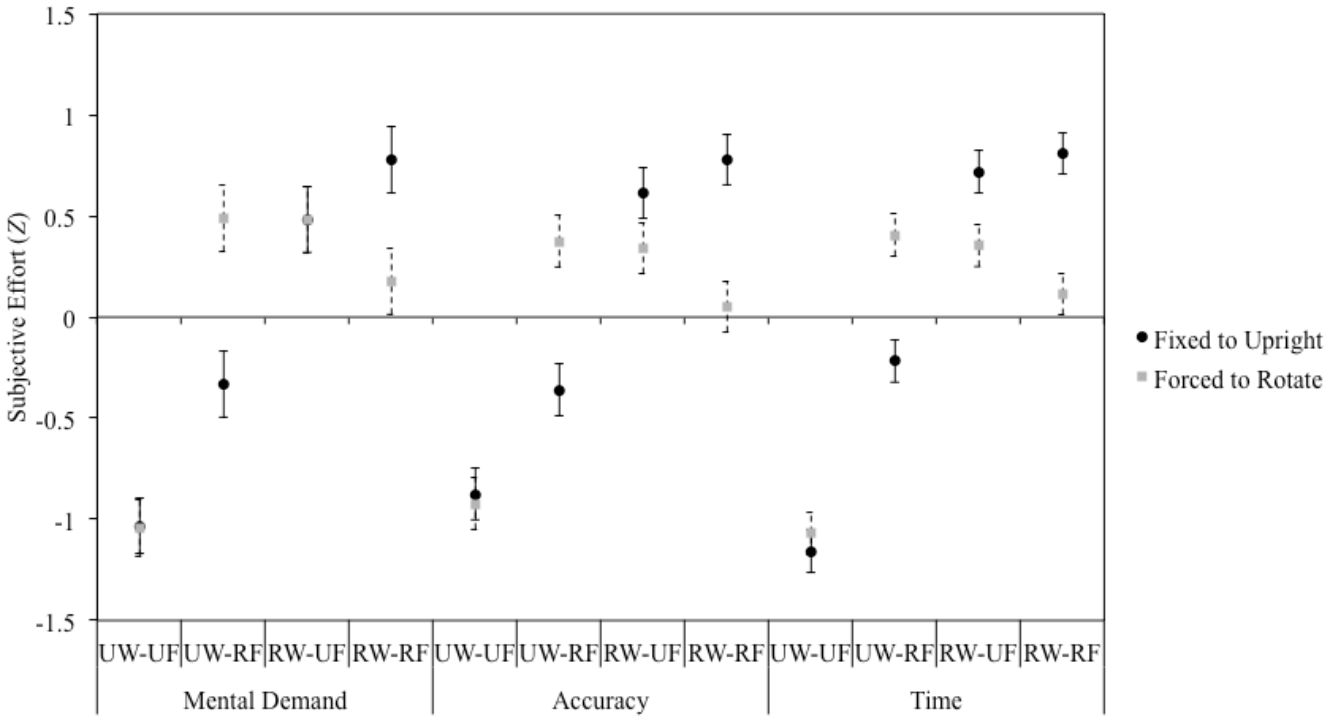
Linear Mixed Model Results for Time Ratings (Z) in Experiment 4

Random Effects	Variance	SD	
Subject	-	-	
Item	-	-	
Residual	.44	.66	
Fixed Effects	Estimate	*SE*	*t*
Intercept	-1.16	.05	-21.83
Forced-to-Rotate	.09	.09	1.18
UW-RF	.94	.08	12.52
RW-UF	1.88	.08	24.94
RW-RF	1.97	.08	26.21
Forced-to-Rotate:UW-RF	.53	.11	5.02
Forced-to-Rotate:RW-UF	-.45	.11	-4.25
Forced-to-Rotate:RW-RF	-.78	.11	-7.37

Note: UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. A $|t| > 2$ criterion was used to evaluate the significance of each fixed factor.

Figure 9.

Subjective Effort Results for Mental Demand, Accuracy, and Time for Experiment 4



Note: UW-UF = Upright Word-Upright Frame; UW-RF = Upright Word-Rotated Frame; RW-UF = Rotated Word-Upright Frame; RW-RF = Rotated Word-Rotated Frame. Error bars denote 95% confidence intervals.

Discussion

We did not find differences in patterns of subjective ratings across the three dimensions used to evaluate effort (i.e., mental demand, time, and accuracy). Nonetheless, the pattern of ratings within these dimensions stayed consistent across each array and instruction condition (although we did not find a perceived benefit of being forced-to-rotate for the RW-UF array type within mental demand ratings). Critically, the RW-RF array showed a larger perceived benefit of being forced-to-rotate when compared to the RW-UF array for all dimensions, replicating the pattern of subjective ratings in Experiment 3 and mirroring the pattern of physical rotations in Experiments 1 and 2. Given that the same pattern of ratings emerged in all dimensions, these judgments may reflect a general metacognitive evaluation of task difficulty or perceived fluency of processing. In addition, given the consistent pattern across the different judgments, it would be hard to conclude that any took precedence in the decision to physically rotate.

General Discussion

In the present investigation we used external normalization in the context of reading rotated text in an effort to better understanding the decision to attempt to offload cognition in a perceptual task. The critical manipulation involved rotating the items only, the frame only, or both in multi-element displays. The patterns observed across these displays with respect to performance, the likelihood of spontaneous physical rotation, and subjective evaluations of effort and performance have provided a number of novel contributions to understanding cognitive offloading. In particular, participants were much more likely to spontaneously physically rotate when both the word and the frame were rotated, followed by when the words were rotated, then by when only the frame was rotated. Based on previous research (e.g., soft constraints hypothesis; Gray et al., 2006; Kirsh, 1995; Maglio et al., 2008; Walsh & Anderson, 2009; Risko et al., 2014), these results suggest that performance (in terms of rotation costs) would follow the pattern of spontaneous physical rotations (i.e., RW-RF > RW-UF > UW-RF).

Interestingly, this is not what we observed. Rotation costs were similar in the conditions featuring word rotation (i.e., RWRF, RWUF) and greater than rotation costs when only the frame was rotated. This dissociation was observed in two separate experiments and also held when the potential benefits of physical rotation were measured across these conditions. That is, Experiment 3 demonstrated that the potential benefits of physically rotating were equivalent across displays featuring word rotation (i.e., RW-RF and RW-UF) and larger than in displays featuring only frame rotation (where there was actually a cost). Nevertheless, Experiments 1 and 2 demonstrated that participants freely choose to physically rotate much more often when the word and frame were rotated. Experiments 3 and 4 provided some insight into this pattern. Namely, when participants were asked to make subjective estimates of the effort (Experiment 3 and 4), time (Experiment 4) and accuracy (Experiment 4) that would be associated with reading each of the display types, participants judged the displays with word and frame rotated as more effortful, time consuming, and error prone than displays with only the word rotated. In addition, participants thought that it would be more beneficial, in terms of effort, time, and accuracy, to physically rotate when presented with displays featuring word and frame rotation than in displays featuring only word rotation. Thus, the dissociation observed between performance and the likelihood of physical rotation appears to be present also in individual's metacognitive judgments about relative effort and performance. Together these results provide a new perspective with regard to the decision processes governing cognitive offloading. We expand on this latter issue and note the implications of these results for related areas (e.g., investigations of viewpoint costs).

Toward a Metacognitive Account of Cognitive Offloading

Based on previous research (e.g., soft constraints hypothesis; Gray et al., 2006; Kirsh, 1995; Maglio et al., 2008; Walsh & Anderson, 2009; Risko et al., 2014) the critical variable in determining whether individuals decide to rely on internal processes versus integrate an external process is the time required for each approach (i.e., the faster strategy will be the one selected). In other words, individuals will attempt to offload (i.e., integrate an external process into a cognitive act in order to subvert internal processing) to the extent that it reduces the time to complete the task. Similar theoretical proposals have been made to account for strategy selection between two internal strategies (e.g., Payne et al., 1988; Reder, 1987; Reder & Ritter, 1992). The large difference in spontaneous physical rotation in the RW-RF condition relative to the RW-UF condition despite equal performance and potential benefits of rotation across those conditions are seemingly at odds with this account.

Participants' metacognitive judgments in Experiments 3 and 4, suggest a potential extension that can capture the present results. Specifically, the critical variable in determining whether individuals decide to rely on internal processes versus integrate an external process is the participant's metacognitive beliefs regarding expected performance or the effort required for each approach (i.e., internal vs. internal + external). Thus, participants may well be trying to optimize performance or minimize effort, as suggested, for example, by the soft constraints hypothesis (Gray et al., 2006), but at least in this case, participants seemingly get it wrong. From the perspective that individual's metacognitive judgments are largely inferential (Koriat, 1997), the existence of such a "metacognitive error" is not surprising, but nevertheless, it provides a critical clue with regard to the metacognitive basis of the decision to attempt to offload cognitive work.

One important question from a metacognitive perspective is why participants perceive reading the RW-RF displays as more effortful, slower, and more error prone than the RW-UF displays? Metacognitive judgments take two distinctive forms: experience-based and theory-based (e.g., Koriat, 2007). Experienced-based judgments rely on cues resulting from online cognitive processing, and often incorporate heuristics and immediate subjective experience in influencing judgments. Conversely, theory-based judgments are based on specific *a priori* beliefs and knowledge about performance and goals. The equivalent performance across the RW-RF and RW-UF conditions suggests that our observed subjective ratings are unlikely to be associated with online experienced-based judgments (i.e., there is no reason within the experiment for their experience to have differed). Also consistent with this idea is the fact that individuals in Experiment 4, who did not complete the reading task, still produced the same pattern of ratings as participants in Experiment 3, who did complete the reading task.

The alternative, that is, the idea that participants might be making a theory-based metacognitive judgment, provides an interesting perspective in the present context, given that participants are more likely to have pre-experimental exposure to displays like those encountered in the RW-RF condition (e.g., a rotated page of text). Physically rotating when reading complex displays can be beneficial (see Risko at al., 2014). Thus, participants' prior experience with such benefits could form an *a priori* theory of the most efficient way to process displays similar to the RW-RF condition that would be reliable in many cases (see Heersmink, 2012; Michaelian, 2012, for recent discussions on information selection within the context of embedded and distributed systems). For example, participants could have an intuitive theory (likely correct most of the time) that "matching" the orientation of a to-be-processed visual display leads to more fluent processing than not "matching" its orientation. From this perspective, remaining upright when the display is upright and rotating in the RW-RF displays unambiguously achieve this putative "orientation matching" goal. Rotating in the RW-UF or UW-RF conditions do not in the sense that physically rotating leads to an "orientation match" on one dimension but not the other. If we assume that matching words is perceived as conferring more fluency than matching the frame (i.e., neutralizing word rotation costs), then the existence of such an "orientation matching" theory could also explain the higher likelihood of physical rotation in the RW-UF condition. That said, given that the RW-UF condition is actually more difficult than the UW-RF condition and that it is more beneficial to physically rotate in the former condition, this decision could also be based on experience acquired during performance of the task. Thus, according to this account, the present results reflect participants making a theory-based metacognitive judgment, where the theory may reflect an intuitive "orientation matching" theory about how to best physically position their body to facilitate perceptual processing. While the participant's theory might be right in many circumstances, it is not in the present context and experience performing the task does not appear to modify that theory. Thus, participants' overt behavior becomes dissociated from performance.

What does the above mean for previous research demonstrating a close association between time and the decision to integrate an external process (Gray et al., 2006; Kirsh, 1995; Maglio et al., 2008; Walsh & Anderson, 2009; Risko et al., 2014)? In other words, how do individuals "get it right"? One potential explanation is that individuals' a priori metacognitive theory, in a given experiment, just so happens to be correct. Alternatively, the conditions within the experiment might be such that individuals can form, based on their experience within the task, an "accurate" metacognitive theory regarding the influence of integrating an external process. For example, in Risko et al. (2014), the likelihood of physical rotation increased as the number of elements in the display increased. This manipulation has a large effect on performance and the influence of increasing the number of elements in the display on performance would be relatively transparent to an individual performing the task (i.e., it takes longer to name more letters than less letters).

In a similar vein, some previous experiments that found a strong association between time and the choice to integrate an external process provided feedback during the task (e.g., Gray & Boehm-Davis, 2000; Gray & Fu, 2004; Walsh & Anderson, 2009). For example, in the Walsh and Anderson (2009) study, where individuals' decisions to use a calculator rather than multiply in their heads closely matched the strategy that would yield the best performance, participants were provided explicit feedback and paid bonuses based on their performance. In this case, feedback is providing reliable information that individuals can readily exploit during the course of the task to match their offloading decisions. The current study did not provide individuals with such information, which may have hindered the ability to adjust their strategy to be more in tune with their performance within the task. Nonetheless, feedback is not always explicitly available in real-world settings, and information related to on-line performance may be unreliable (see Schwartz, Benjamin, & Bjork, 1997). In such cases, individuals may choose to rely on theory-based metacognitive judgments to guide their strategy selection.

From the foregoing perspective, the beginnings of a metacognitive account of cognitive offloading can be forwarded. Specifically, individuals decide whether to rely on internal processes versus integrate an external process based on a kind of metacognitive judgment regarding the expected performance/effort associated with each approach (i.e., internal vs. internal + external). The extent to which the decision to integrate an external process maps onto those situations where it is the most prudent, from a performance (i.e., time) perspective, will depend on the amount of environmental support for forming an accurate metacognitive judgment. For example, where there is transparency with respect to the influence of different stimulus conditions on performance or other external supports (e.g., feedback is provided about performance), a close match might be expected. Where there is ambiguity (as in the present experiments), such a match is unlikely and participants may turn to an a priori or intuitive theory. Importantly, in the present context, simply performing the task is insufficient to alter this theory. Future work aimed at delineating these situations will provide critical insight into the metacognitive basis of our interactions with non-cognitive resources (e.g., our bodies, objects in our physical environment; Kirsh, 2004) or how we think about thinking with our body and the world.

Alternative Accounts

The metacognitive account presented here places individual's metacognitive beliefs in a causal position with regard to the decision to try to offload via external normalization while reading rotated text. One alternative account is that some unidentified variable is driving both individuals' subjective ratings and their spontaneous physical rotations, or driving just the latter with the metacognitive judgments simply reflecting a rationalization of the felt desire to rotate (e.g., "I feel like I want to rotate so it must be beneficial for me to do so"). While difficult to rule out on the basis of the present experiments, it is important to note that, if true, this would also necessitate a change in how we think about how individuals are deciding to offload cognition or integrate an external process into a cognitive act because this unidentified variable would presumably not be performance, which at present represents the dominant theoretical perspective. In other words, the empirical demonstration of a dissociation between the likelihood of spontaneous physical rotation and performance stands as a challenge to any account that claims that the primary basis of the decision to integrate an external process is based on online performance savings; this stands independent of whether the metacognitive account we have forwarded turns out to be the correct explanation.

Another alternative account would be to suggest that the RW-RF and RW-UF conditions do differ somehow in difficulty or relative benefits of physical rotation (that we could not detect) so individuals are physically rotating in response to this difference. While logically plausible, this account seems unlikely given the nearly two-fold increase in spontaneous physical rotations across the RW-RF and RW-UF conditions. In other words, if there were only a subtle increase in the likelihood of physical rotations in the RW-RF condition relative to the RW-UF condition, an explanation in terms of an undetected increase in difficulty or benefit of rotation in the RW-RF relative to the RW-UF condition would be more plausible. Rather, there is a large difference in the likelihood of spontaneous physical rotation but no detectable difference in difficulty or benefit of physical rotation (in terms of performance). In addition, this would also seemingly make a performance maximization (or time minimization) type account exceedingly difficult to falsify (i.e., RW-RF must be harder or it must be more beneficial to rotate in that condition given that they do). Thus, at present, we consider it more plausible to suggest that individuals are making an error with respect to the relative difficulty or benefit of physical rotation (as we have demonstrated), rather than accurately estimating such dimensions.

Implications for Understanding Stimulus Rotation Costs in Reading

Despite the focus of the present investigation being on the determinants of cognitive offloading, the pattern of rotation costs across our display conditions (i.e., UW-RF < RW-UF=RW-RF) in the fixed-to-upright condition also has implications for understanding the internal mechanisms underlying viewpoint costs in object identification. There exist two major theories for understanding viewpoint costs in object identification: (1) the image transformation account where internal representations of the disoriented stimulus are mentally rotated to upright (Shepard & Metzler, 1971), and (2) the frame rotation account where an internal frame of reference is rotated to match the stimulus orientation (Graf, 2006). The challenge posed by the present pattern of rotation costs, for both theories, is that frame rotation incurs a cost relative to upright (i.e., UW-UF > UW-UF) but that cost seemingly disappears when the words are also rotated (i.e., an underadditive interaction).

From an additive factors perspective (e.g., Sternberg, 1969; 1998) and assuming frame rotation and word rotation influence different stages of processing, an image transformation account would predict that the costs associated with the individual items being normalized and the disruption of reading direction by rotating the frame would simply add (i.e., an additive pattern). The fact that frame rotation costs are seemingly neutralized when both the words and the frame of the array are rotated falsifies such an account. An alternative account from an image transformation perspective would need to suggest that the costs associated with frame rotation and word rotation can be dealt with in parallel and thus the cost of frame rotation is being absorbed into the time associated with normalizing the individual items.

The frame rotation account can (to some extent naturally) accommodate the underadditive pattern. Specifically, as noted above, a frame rotation account predicts that individuals rotate an internal frame of reference when identifying a rotated object. Thus, the process by which the items are normalized (i.e., rotating an internal frame of reference) also "normalizes" the frame rotation. That is, re-orienting an internal frame of reference to match the orientation of the word in the RW-RF condition brings the internal frame of reference and the frame of the array into alignment, and as a result there is no additional cost of the frame rotation above and beyond the word rotation. Critically, this only works because the RW-RF display is "congruent" in the sense that both the words and the frame are disoriented in the same direction and to the same degree. Thus, the underadditive pattern could reflect a new type of orientation congruency effect (see Graf et al., 2005; Jolicoeur, 1990), wherein the orientation of the global (frame) and local (items) dimensions in multi-element displays facilitate word identification. This makes the straightforward prediction that an RW-RF condition, where the words are rotated in the direction opposite the frame rotation, would nullify the underadditive pattern observed here. Future work investigating rotation costs in multi-element displays will address this prediction.

Conclusion

The present investigation looked to disentangle the factors driving the decision to attempt to offload cognition, as indexed by external normalization. Results demonstrated that individuals were much more likely to spontaneously physically rotate in a rotated multi-word display when both the words and the frame were rotated than when only the words or only the frame were rotated. Critically, neither performance costs nor benefits of physically rotating could explain the patterns of spontaneous physical rotations observed, a result seemingly at odds with time-based theories of the integration of some external process. However, individuals' metacognitive ratings of effort and performance closely matched patterns of spontaneous physical rotation, leading to the notion that individuals base their decision to cognitively offload on theory-based judgments concerning the reduction of effort or performance maximization. Future work aimed at better understanding how we think about thinking with such external resources will likely yield fundamental insights into the metacognition of the embodied mind.

References

Arango-Muñoz, S. (2013). Scaffolded memory and metacognitive feelings. *Review of Philosophy and Psychology, 4*(1), 135-152.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*(4), 390-412.

Barsalou, L. W. (2008). Grounded cognition. *Annual Review of Psychology, 59*, 617-645.

Bates, D., Maechler, M., & Dai, B. (2011). lme4: Linear Mixed-Effects Models Using S4 Classes. (R package Ver. 0.999375-28) [Software]. Vienna, Austria: R Foundation for Statistical Computing.

Beilock, S. L., & Goldin-Meadow, S. (2010). Gesture changes thought by grounding it in action. *Psychological Science, 21*(11), 1605-1610.

Castel, A. D., Rhodes, M. G., & Friedman, M. C. (2013). Predicting memory benefits in the production effect: The use and misuse of self-generated distinctive cues when making judgments of learning. *Memory & Cognition, 41*(1)*,* 28-35.

Clark, A. (2010). *Supersizing the mind*. Oxford: Oxford University Press.

Corballis, M. C. (1988). Recognition of disoriented shapes. *Psychological Review, 95*(1), 115-123.

Diwadkar, V. A., & McNamara, T. P. (1997). Viewpoint dependence in scene recognition. *Psychological Science, 8*(4), 302-307.

Efklides, A. (2008). Metacognition: Defining its facets and levels of functioning in relation to self-regulation and co-regulation. *European Psychologist, 13*(4), 277-287.

Forster, K. I., & Forster, J. C. (2003). DMDX: A windows display program with millisecond accuracy. *Behavior Research Methods, Instruments and Computers, 35*(1), 116-124.

Glenberg, A. M. (2010). Embodiment as a unifying perspective for psychology. *Wiley Interdisciplinary Reviews: Cognitive Science, 1*(4), 586-596.

Goldin-Meadow, S. (2005). *Hearing gesture: How our hands help us think*. Cambridge: Harvard University Press.

Graf, M. (2006). Coordinate transformations in object recognition. *Psychological Bulletin, 132*(6), 920-945.

Graf, M., Kaping, D., & Bülthoff, H. H. (2005). Orientation congruency effects for familiar objects. *Psychological Science, 16*(3), 214-221.

Gray, W. D., & Boehm-Davis, D. A. (2000). Milliseconds matter: An introduction to microstrategies and to their use in describing and predicting interactive behavior. *Journal of Experimental Psychology: Applied, 6*(4), 322-335.

Gray, W. D., & Fu, W-T. (2004). Soft constraints in interactive behavior: The case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science, 28*(3), 359-382.

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 behavior. *Psychological Review, 113*(3), 461-482.

Halekoh, U., & Højsgaard, S. (in press). A Kenward-Roger approximation and parametric bootstrap methods for tests in Linear Mixed Models–the R package pbkrtest. *Journal of Statistical Software.*

Hamm, J. P., & McMullen, P.A. (1998). Effects of orientation on the identification of rotated objects depend on the level of identity. *Journal of Experimental Psychology: Human Perception and Performance, 24*(2), 413-426.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology, 52,* 139-183.

Heersmink, R. (2012). Mind and artifact: A multidimensional matrix for exploring cognition artifact relations. In J. M. Bishop & Y. J. Erden (Eds.), *Proceedings of the 5th AISB Symposium on Computing and Philosophy* (pp. 54–61).

Hutchins, E. (1995). *Cognition in the wild*. Cambridge: MIT Press.

Jolicoeur, P. (1985). The time to name disoriented natural objects. *Memory & Cognition, 13*(4), 289-303.

Jolicoeur, P. (1990). Orientation congruency effects on the identification of disoriented shapes. *Journal of Experimental Psychology: Human Perception and Performance, 16*(2), 351-364.

Jolicoeur, P., & Landau, M. J. (1984). Effects of orientation on the identification of simple visual patterns. *Canadian Journal of Psychology, 38*(1), 80-93.

Jolicoeur, P., Snow, D., & Murray, J. (1987). The time to identify disoriented letters: Effects of practice and font. *Canadian Journal of Psychology, 41*(3), 303-316.

Kirsh, D. (1995). The intelligent use of space. *Artificial Intelligence, 73*(1), 31-68.

Kirsh, D. (2004). Metacognition, distributed cognition and visual design. In P. Gärdenfors & P. Johansson (Eds.), *Cognition, education and communication technology* (pp.147-179). Hillsdale, NJ: Lawrence Erlbaum.

Kirsh, D. (2010). Thinking with external representations. *AI and Society, 25*, 441-454.

Kirsh, D., & Maglio, P. (1994). On distinguishing epistemic from pragmatic action. *Cognitive Science, 18*(4), 513-549.

Kliegl, R., Masson, M. E. J., & Richter, E. M. (2010). A linear mixed model analysis of masked repetition priming. *Visual Cognition, 18*(5), 655-681.

Kolers, P. A., & Perkins, D. N. (1969a). Orientation of letters and errors in their recognition. *Attention, Perception, & Psychophysics, 5*(5), 265-269.

Kolers, P. A., & Perkins, D. N. (1969b). Orientation of letters and their speed of recognition. *Attention, Perception, & Psychophysics, 5*(5), 275-280.

Koriat, A. (1995). Dissociating knowing and the feeling of knowing: Further evidence for the accessibility model. *Journal of Experimental Psychology: General, 124*(3)*,* 311-333.

Koriat, A. (1997). Monitoring one's own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General, 126*(4), 349-370.

Koriat, A. (2000). The feeling of knowing: Some metatheoretical implications for consciousness and control. *Consciousness and Cognition, 9*(2)*,* 149-171.

Koriat, A. (2007). Metacognition and consciousness. In P.D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The Cambridge handbook of consciousness* (pp. 289-325). Cambridge, England: Cambridge University Press.

Koriat, A., & Norman, J. (1984). What is rotated in mental rotation? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*(3), 421-434.

Kung, E., & Hamm, J. P. (2010). A model of rotated mirror/normal letter discriminations. *Memory & Cognition, 38*(2), 206-220.

Maglio, P. P., Wenger, M. J., & Copeland, A. M. (2008). Evidence for the role of self-priming in epistemic action: Expertise and the effective use of memory. *Acta Psychologica, 127*(1), 72-88.

Martin, T., & Schwartz, D. L. (2005). Physically distributed learning: Adapting and reinterpreting physical environments in the development of fraction concepts. *Cognitive Science, 29*(4), 587-625.

Metcalfe, J. (2009). Metacognitive judgments and control of study. *Current Directions in Psychological Science, 18*(3), 159-163.

Metcalfe, J., Schwartz, B. L., & Joaquin, S.G. (1993). The cue familiarity heuristic in metacognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(4), 851-861.

Michaelian, K. (2012). Metacognition and endorsement. *Mind & Language, 27*(3), 284-307.

Murray, J. E. (1998). Is entry-level recognition viewpoint invariant or viewpoint dependent? *Psychonomic Bulletin & Review, 5*(2), 300-304.

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and some new findings. In G. Bower (Ed.), *The psychology of learning and memory* (pp. 125-173). San Diego, CA: Academic Press.

Payne, J. W., Bettman, J. R., & Johnson, E. J. (1988). Adaptive strategy selection in decision making. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 534-552.

Proffitt, D. (2006). Embodied perception and the economy of action. *Perspectives on Psychological Science, 1*(2), 110-122.

Protopapas, A. (2007). CheckVocal: A program to facilitate checking the accuracy and response time of vocal responses from DMDX. *Behavior Research Methods, 39*(4), 859-862.

R Development Core Team. (2010). R: A language and environment for statistical computing. Vienna, Austria: R foundation for statistical computing. Retrieved from http://www.Rproject.org

Reder, L. M. (1987). Strategy selection in question answering. *Cognitive Psychology, 19*(1)*,* 90 -138.

Reder, L. M., & Ritter, F. E. (1992). What determines initial feeling of knowing? Familiarity with question terms, not with the answer. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(3)*,* 435-451.

Risko, E., F., Medimorec, S., Chisholm, J., & Kingstone, A. (2014). Rotating with rotatedtext: A natural behavior approach to investigating cognitive offloading. *Cognitive Science, 38*(3), 537-564.

Rupert, R. D. (2004). Challenges to the hypothesis of extended cognition. *The Journal of Philosophy, 101*(8), 389-428.

Scaife, M., & Rogers, Y. (1996). External cognition: How do graphical representations work? *International Journal of Human-Computer Studies, 45*(2), 185-213.

Schwartz, B. L., & Metcalfe, J. (1992). Cue familiarity but not target retrievability enhances feeling-of-knowing judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(5), 1074-1083.

Schwartz, B. L., Benjamin, A. S., & Bjork, R. A. (1997). The inferential and experiential bases of metamemory. *Current Directions in Psychological Science, 6*(5), 132-137.

Siegler, R. S., & Lamaire, P. (1997). Older and younger adults' strategy choices in multiplication: Testing predictions of ASCM using the choice/no-choice method. *Journal of Experiment Psychology: General, 126*(1), 71-92.

Shapiro, L. A. (2011). *Embodied cognition (New problems of philosophy).* New York: Routledge.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171*, 701-703.

Sternberg, S. (1969). The discovery of processing stages: Extensions of Donders' method. *Acta Psychologica, 30,* 276-315.

Sternberg, S. (1998). Discovering mental processing stages: The method of additive factors. In D. Scarborough & S. Sternberg (Eds.), *An invitation to cognitive science: Vol 4. Methods, models, and conceptual issues* (pp. 703-863). Cambridge, MA: MIT Press.

Tarr, M. J. (1995). Rotating objects to recognize them: A case study of the role of viewpoint dependency in the recognition of three-dimensional objects. *Psychonomic Bulletin and Review, 2*(1), 55-82.

Walsh, M. M., & Anderson, J. R. (2009). The strategic nature of changing your mind. *Cognitive Psychology, 58*(3)*,* 416-440.

Wells, I. C., & Hamm, J. P. (2009). The effects of inverting stimuli in a flanker task. *Canadian Journal of Experimental Psychology, 63*(1), 33-39.

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition, 68,* 77-94.

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review, 9*(4), 625-636.

Wohlschläger, A., & Wohlschläger, A. (1998). Mental and manual rotation. *Journal of Experimental Psychology: Human Perception and Performance, 24*(2), 397-412.