Using Eye Tracking to Investigate the Evaluation-Performance Relationship in Visual Attention

by Adam J. Brown

B.S. in Psychology, University of Maryland
M.A. in Psychology, Northeastern University

A dissertation submitted to

The Faculty of
The College of Science of
Northeastern University
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

July 24, 2017

Dissertation directed by

Stephen G. Harkins
Professor of Psychology
Acknowledgements

The truth about this acknowledgements section is that I had every intention to write a very nice version of it, but I ran out of time. I’d like to thank my advisor, Stephen G. Harkins, for a really interesting/fun/stressful/informative five years. I’d also like to thank my cohort, Molly Sands, Isha Vicaria, Darrell Penta, Tina Gruene, Amanda Dupuis, and Jin X. Goh, for always being there to hear me complain when things were more stressful than interesting or fun. I’d like to especially thank Jin because I’m not sure how long I would have lasted in this program if you hadn’t been here, seriously. I’d also like to thank Mia Radic for listening me drone on and on and on about this stuff and supporting me through what were some of the most difficult times that I can recall. Finally, I’d like to thank anyone who said, “you’ll be fine.”
Abstract of Dissertation

Recently, two accounts have attempted to account for the effects of evaluation on performance, the mere effort account (Harkins, 2006; McFall, Jamieson, & Harkins, 2009) and the focus of attention account (Muller & Butera, 2007; Normand, Bouquet, & Croizet, 2014). The focus of attention account suggests that undergoing evaluation causes a reduction in processing capacity, which, in turn, leads individuals to prioritize the most relevant information and disregard less focal information. In support of this notion, Normand, Bouquet, and Croizet (2014) report an experiment in which participants were asked to look at a fixation cross followed by a brief abrupt onset cue that flashed in one of four locations forming a square followed by a display of four letters (three Qs and one O) in the same square. The participants’ task was to identify the location of the O as quickly and as accurately as possible. The brief onset cue appeared either in the location where the O would appear (valid trials), in a location where the Q would appear (invalid trials), or onsets flashed in all four locations (neutral trials). Normand et al., found that participants subject to evaluation exhibited smaller cueing effects (i.e., the difference between invalid and valid trial reaction times) than their non-evaluated counterparts, which they interpreted as evidence for the cue having less impact on reaction times.

Alternatively, the mere effort account argues that undergoing evaluation leads participants to be highly motivated to perform well, which potentiates (i.e., makes more likely) the prepotent (i.e., dominant) response. Task performance is then dependent on whether this response is correct, as well as whether or not participants have the knowledge, opportunity and motivation to correct it if it is incorrect. Normand et al. proposed that the spatial cue presented prior to the target array represents a dominant response, which should have produced larger, not smaller, cueing effects. However, this prediction for mere effort does not take into account the
motivation to correct this response when it is incorrect, as seen in previous work (e.g., McFall, Jamieson, & Harkins, 2009). Additionally, Normand et al. draw conclusions about this process by simply interpreting terminal reaction time data even though a more definitive measure is available. The current work aimed to replicate and expand upon this research with the addition of eye-tracking, which allows a direct test of the two accounts.

In Experiment 1, using the stimulus parameters reported by Normand et al. (2014, Exp. 1), we found that the potential for evaluation did affect terminal reaction times, but not because participants looked away from or toward the cue. In fact, our best estimate is that participants looked at the cue only 30% of the time, perhaps because the letters in Normand et al.’s display were so large that the participants could see the target without even moving their eyes.

In Experiment 2, by changing stimulus parameters, it appeared that we were successful at producing a prepotent response since participants looked at the cue on 74% of the trials. However, once again, the significant evaluation effect in terminal reaction times was not the result of evaluated participants looking away from or toward the cue more than non-evaluated participants. Instead, through exploratory analyses, we found that what appeared to be a strong tendency to look at the cue was the result of a bias to look toward the top-left of the target display, which was accentuated for evaluated participants.

In Experiment 3, we attempted to better understand this location bias and concluded that this behavior was the actual prepotent response in these designs. In both Experiments 2 and 3, undergoing experimenter evaluation led to potentiation of the location bias, which produced slower reaction times for valid trials in Experiment 2, and, interestingly, no effect on overall reaction times in Experiment 3. Each of these experiments support the mere effort account and not the focus of attention account.
Experiments 4 and 5 represent an attempt to produce a set of stimulus parameters that simultaneously eliminate the location bias seen in Experiments 2 and 3 and make eye movements to the abrupt onset the prepotent response. In Experiment 4, we found that the stimulus parameters actually produced a ceiling effect for eye movements to the abrupt onset, which precluded testing the potentiation hypothesis of mere effort. In Experiment 5 in which we reduced the potency of the cue, we found that evaluated participants were faster to respond on invalid trials, and trended towards being faster on valid trials, but the eye-tracking data revealed no evidence for potentiation of the prepotent response (i.e., orienting to the location of the abrupt onset). Additional analyses revealed that evaluated participants were fastest on trials in which they did not orient to the abrupt onset or to the target location, suggesting that being evaluated still led to motivated responding, but without potentiation of the prepotent response.

Taken together, these findings support a motivational account of performance under evaluation despite the limitations revealed in the mere effort account. In addition, the research demonstrates the pitfalls of the common practice of using single behavioral measures (e.g., reaction time) to infer mediating process.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acknowledgements</td>
<td>ii</td>
</tr>
<tr>
<td>Abstract</td>
<td>iii</td>
</tr>
<tr>
<td>Table of Contents</td>
<td>vi</td>
</tr>
<tr>
<td>List of Figures</td>
<td>vii</td>
</tr>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Baseline 1</td>
<td>9</td>
</tr>
<tr>
<td>Experimenter Evaluation Experiment 1</td>
<td>14</td>
</tr>
<tr>
<td>Baseline 2</td>
<td>22</td>
</tr>
<tr>
<td>Experimenter Evaluation Experiment 2</td>
<td>25</td>
</tr>
<tr>
<td>Baseline 3</td>
<td>35</td>
</tr>
<tr>
<td>Experimenter Evaluation Experiment 3</td>
<td>39</td>
</tr>
<tr>
<td>Baselines 4a-4c</td>
<td>46</td>
</tr>
<tr>
<td>Experimenter Evaluation Experiment 4</td>
<td>52</td>
</tr>
<tr>
<td>Experimenter Evaluation Experiment 5</td>
<td>57</td>
</tr>
<tr>
<td>General Discussion</td>
<td>66</td>
</tr>
<tr>
<td>References</td>
<td>76</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1: Sequence of events on an invalid trial 80

Figure 2: Reaction times by cue type and evaluation condition
   for Evaluation Experiment 2 81

Figure 3: Percentage of eye movements to uncued locations in Evaluation Experiment 2
   by display location 82

Figure 4: Reaction times by cue type and display location condition
   for Evaluation Experiment 2 83

Figure 5: Reaction times by cue type, display location, and evaluation condition
   for Evaluation Experiment 2 84

Figure 6: Percentage of eye movements split by location and cue type
   for Baseline 3 85

Figure 7: Percentage of uncued eye movements split by evaluation condition and location
   for Experimenter Evaluation Experiment 3 86

Figure 8: Percentage of cued eye movements split by evaluation condition and location for
   Experimenter Evaluation Experiment 3 87

Figure 9: Invalid reaction times split by evaluation condition and eye movement for
   Experimenter Evaluation Experiment 5 88

Figure 10: Valid reaction times split by evaluation condition and eye movement for
   Experimenter Evaluation Experiment 5 89
Introduction

Individuals often find themselves in important situations facing evaluative scrutiny (e.g., a job interview or an exam), and while these high-pressure situations create a desire in individuals to perform well, facing evaluation occasionally leads to underperformance. As a society that regularly deals with evaluation in social, academic and career contexts, a better understanding of the processes that underlie how we perform in these scenarios is required. More than a century of social psychology research has been aimed at understanding the effect of the potential for evaluation on task performance. Yet, there is still a debate as to what drives performance when an individual is faced with evaluation. At the core of this research is explaining how, and when, the potential for evaluation leads to either facilitated or debilitated performance.

A survey of the psychological literature reveals five different research traditions that have found that the potential for evaluation affects task performance: social loafing, social facilitation, goal-setting, intrinsic motivation/creativity, and achievement-goal theory. Although these traditions have proposed processes that could account for evaluation effects (e.g., focus of attention, working memory deficits, withdrawal of effort, drive theory), the models do not agree on the mediating process(es), nor is there any compelling evidence favoring one account over the others.

For example, drive approaches (e.g., Zajonc, 1965) suggest that participants are putting out more effort when they work in the presence of others, and this increased effort enhances performance on simple tasks (correct answer likely in the set of possible responses), but debilitates performance on complex tasks (correct answer unlikely in the set of possible responses). Bond’s (1982) self-presentation approach suggests that participants are working hard
in the presence of others, but concern over the embarrassment of potential failure causes
cognitive interference. Baron’s (1986) focus-of-attention interpretation would suggest that it is
not a motivational effect but a cognitive one. Participants working on difficult tasks are working
as hard as participants facing easy tasks, but the presence of others leads to a narrowing of
attention that debilitating performance on complex tasks because these tasks require attention to a
wide range of cues.

In contrast to these approaches, Carver and Scheier (1981) suggest that debilitating
performance on complex tasks results from the withdrawal of effort. That is, the presence of
others makes participants self-aware, leading them to be “more cognizant of both the level of
performance being manifested at the moment and the salient standard” (Geen, 1989, p. 32).
When they find that they have little chance of bringing their performance into alignment with the
standard, they stop trying.

In fact, the same set of potential explanations proposed to account for the effects of
evaluation on performance have been with us for over 100 years. For example, Triplett (1898)
suggested “brain worry” as one explanation for his findings, which reappeared as “cognitive
interference” in the 1980s (e.g., Bond, 1982), and is now termed “working memory deficit” (e.g.,
DeCaro, Thomas, Albert & Beilock, 2011). Another explanation, range of cue utilization
(Easterbrook, 1959), that was used by Geen (1976) and Baron (1986) to account for performance
effects remains with us, but is now termed focus of attention (e.g., Muller & Butera, 2007).

More recently, two accounts have attempted to account for these effects, the mere effort
account (Harkins, 2006; McFall, Jamieson, & Harkins, 2009) and the focus of attention account
(Muller & Butera, 2007; Normand, Autin, & Croizet, 2015; Normand, Bouquet, & Croizet,
2014), with the former suggesting that changes in motivation cause performance effects, and the
latter presenting a cognitive explanation. Specifying the mediating process is key to the theoretical development of each of these research traditions as well as to any effort to integrate them. Understanding the mediating process is also key to the implementation of effective behavioral training programs. For example, the intervention that would be designed to counter the effect of evaluation on working memory would be completely different from one designed to counter the effect of evaluation on focus of attention.

**The Mere Effort Account**

Harkins (2006) argued that the field's failure to resolve this issue may stem from the fact that our efforts have been focused broadly on theory construction, rather than on the tedious analysis required to learn how performance unfolds on a given task. Although it would appear that a molecular analysis of task performance would be an integral part of theory development, this type of analysis has not been undertaken. Instead, researchers have chosen tasks simply because they were convenient or because they have been used before. To this end, Harkins undertook a molecular analysis of the effects of evaluation on the performance of a specific task, the Remote Associates Test (RAT).

The RAT requires participants to look at a set of three words (e.g., *sandwich*, *house*, and *golf*) and generate a fourth word that is related to each word in the given triad (in this case *club*). Harkins (2001) has shown that the potential for evaluation produces the typical pattern of performance on this task: Participants who anticipate evaluation by the experimenter solve more triads shown by a pretest to be simple than do no-evaluation participants, whereas participants who anticipate experimenter evaluation solve fewer triads shown by a pretest to be difficult than do no-evaluation participants.
Harkins’s (2006) analysis suggested the mere effort account. This explanation argues that the potential for evaluation motivates participants to want to do well, which potentiates whatever response is prepotent (i.e., dominant) on the given task. If the prepotent response is incorrect and participants do not know, or lack the knowledge or time required for correction, performance is debilitated. However, if the prepotent response is correct, or if participants are able to recognize that their prepotent tendencies are incorrect and are given the opportunity to correct, performance will be facilitated.

Support for the mere effort account has also been found using the antisaccade task (Jamieson & Harkins, 2007; McFall et al., 2009, Exp. 4). The antisaccade task is a visual attention task, which asks participants to report the orientation of a target arrow randomly presented on either the left or right side of a display. However, prior to the target appearing on the screen, a visual cue is presented for 400 ms on the opposite side of the screen from where the target will appear. Participants are instructed to look away from this initial cue to the opposite side of the screen to where the eventual target will appear. On the antisaccade task, the prepotent response is to orient the eyes to the cue that abruptly appears, despite the explicit instructions to look away, meaning that participants know that the prepotent response is incorrect. Thus, participants’ performance on the antisaccade depends on their motivation and opportunity to correct the prepotent response. McFall et al. (2009) found that when participants had a long enough response window on the antisaccade (i.e., the opportunity to correct), the more motivated evaluation participants outperformed (i.e., responded more quickly) non-evaluated controls. McFall et al. also collected eye-tracking data to aid in the interpretation of the reaction time data. They found that evaluated participants made more reflexive saccades (i.e., the prepotent response was potentiated), but these participants were faster to launch corrective saccades faster than non-
evaluated participants. Evaluated participants were also faster at launching correct saccades (i.e., saccades away from the spatial cue) than non-evaluated participants. For both types of saccades, evaluated participants were also faster to press the response key once their eyes arrived at the target location than the non-evaluated participants.

Additional support for the mere effort account has been shown on the Stroop task (McFall et al., 2009) and anagrams (McFall et al., 2009). It has also been extended to explain stereotype threat effects on the antisaccade task (Jamieson & Harkins, 2007), Graduate Record Exam quantitative problems (Jamieson & Harkins, 2009; Seitchik, Jamieson & Harkins, 2014), Modular Arithmetic (Seitchik & Harkins, 2015), Stroop Color-Word Task (Jamieson & Harkins, 2011), and sensorimotor tasks (Huber, Brown, & Sternad, 2016; Huber, Seitchik, Brown, Sternad & Harkins, 2015). In this line of research, undergoing threat is suggested to lead to an increase in motivation to disconfirm the negative stereotype about oneself or one’s group, which leads to potentiation of the prepotent response. Thus, motivation is afforded a central role in producing stereotype threat effects, which is in contrast to the prevailing view that stereotype threat effects are produced by decreases in working memory capacity (for a review, see, Schmader, Hall, & Croft, 2015). Critically, for any paradigm to test the predictions of the mere effort account, a prepotent response must be identified for the task.

The Focus of Attention Account

On the other hand, the focus of attention account suggests that the potential for evaluation triggers concerns about one’s performance and can hijack attention from the task at hand, reducing the amount of information that can be processed. Due to this decrease in processing capacity, individuals are suggested to prioritize the most relevant information and disregard less focal information. In essence, this account argues that something akin to “tunnel vision” occurs.
As a result, an irrelevant stimulus will be more distracting when it shares a feature that is critical to performance on the task than when it does not. Thus, a distractor that does not match the attentional set produced by the task will lead to better performance by participants subject to evaluative pressure than those that are not, whereas a distractor that does match the attentional set will lead to poorer performance by these participants.

Recently, Normand, Bouquet, and Croizet (2014) provided evidence in support of the focus of attention account in a series of experiments. In their first experiment, using a modified version of the paradigm presented in Muller and Butera (Exp. 5, 2007), Normand et al. sought to extend focus of attention effects from self-evaluative pressures (e.g., social comparison) to experimenter evaluation. Participants were asked to look at a fixation cross that was displayed for one of four durations ranging from 850 to 1500 ms. A brief abrupt onset then flashed for 30 ms approximately 7° from the fixation cross in one of four locations forming a square followed by a 50 ms blank, and a display of four letters (three Qs and one O) in the same square. The participants’ task was to identify the location of the O as quickly and as accurately as possible. On one third of the trials, the dot flashed in the location where the O would appear (i.e., the onset cue was valid). On one third of the trials, the dot flashed in a location where the Q would appear (i.e., the onset was invalid). For the remaining trials, dots flashed in all four locations (i.e., the onsets were neutral). Ordinarily, one would expect that a cue with a sudden onset would capture participants’ attention (e.g., Yantis & Jonides, 1990), but according to Normand et al., the dot does not match the task demands (identifying the location of the letter O) and is, therefore, irrelevant. As a result, participants subject to evaluative pressure should be able to filter out its effects, making them faster on the trials when the cue is invalid but slower on the trials when the cue is valid than participants not subject to this pressure. Consistent with this prediction,
Normand et al. (2014) reported that participants subject to evaluation displayed a smaller cueing effect (i.e., the difference between the reaction times for invalid and valid cues) than those who were not. Specifically, evaluated participants had faster reaction times on invalid trials, but evaluation condition did not affect valid trial reaction times.

The results of Normand et al. (2014, Exp. 1) show that experimenter evaluation can affect performance on this visual attention task, and can be interpreted to support the focus of attention account. Normand et al. also note that their findings are inconsistent with the mere effort account, which suggests that being subject to experimenter evaluation makes participants reflexively orient to the location of the abrupt onset more often than non-evaluated participants, leading to poorer, not better, performance by evaluated participants.

However, as suggested by the results on the antisaccade task, evaluated participants may have made more reflexive saccades to the cue, but were also more motivated to correct this error than participants who were not evaluated, which could lead to faster reaction times for invalid trials. Thus, the smaller cueing effects found in Normand et al. may have been the outcome of the same process described in McFall et al. (2009), rather than the result of focusing of attention.

It is worth noting that the valid reaction times reported in Normand et al. (2014) appear inconsistent with the predictions drawn from both accounts. For focus of attention, if evaluative pressure causes an individual to filter out the cue, then reaction times should become slower on valid trials, but this does not occur. On the other hand, mere effort suggests that making more reflexive responses to the cued location should make evaluated participants faster on these trials, but this does not appear to be the case either.

In the current work, we plan to revisit the paradigm presented in the first experiment of Normand et al. (2014) with the addition of eye-tracking in order to provide a better
understanding of the processes that produce these effects. For example, eye-tracking data afford more a straightforward means of testing the claims of the focus of attention and the mere effort accounts. Prior to manipulating evaluative pressure, baseline (i.e., control) experiments will be run in order to identify the prepotent response, if any, in the specific experimental paradigm. This prepotent response will be used to make predictions from the perspective of the mere effort account and will allow tests of mere effort against the focus of attention account.
Baseline 1

As described above, Normand et al.’s (2014) first experiment presents individuals with abrupt onsets that are suggested to be irrelevant to finding the target. The purpose of this baseline experiment was to better understand the role eye movements play in performing this task without an evaluation manipulation.

Method

Participants. 21 undergraduate students from Northeastern University participated in the experiment for partial fulfillment of a course requirement. None had any prior experience with the task.

Task. In the visual attention task, each trial began with the presentation of a fixation cross in the center of the screen for a randomly determined interval ranging from 850 to 1500 ms. A spatial cue, a round dot subtending 0.44° x 0.44° visual angle, was then presented for 30 ms. For one third of the trials, the cue appeared in the same spatial location as the eventual target (i.e., it was a valid cue). For one third of the trials, the cue appeared in a different location than the eventual target (i.e., it was an invalid cue). For the remaining trials, a cue was presented in all four locations (i.e., the cue was neutral). Following the presentation of the cue, there was a 50 ms blank and then the target display was presented. The target display consists of 3 Qs and 1 O each subtending 2.16° x 2.41° and displayed in a square 6.78° from the fixation cross (see Figure 1). Participants were instructed to locate the letter O as quickly and as accurately as possible by pressing a key on the number pad; they were instructed to press 1 for bottom-left, 3 for bottom-right, 4 for top-left, and 6 for top-right. The target display remained on the display until a response was made. The next trial began after a 1750 ms intertrial interval. Participants completed 144 trials in two blocks of 72 trials separated by a 1-minute rest period. Angular
distances were controlled through the use of a chinrest positioned 56 cm from the monitor. Eye movements were collected at a rate of 60 Hz with a resolution of 0.25°.

**Procedure.** Participants came into the lab individually and were told that they would be completing a visual attention task. Prior to the instructions of the task being given, a male researcher calibrated the eye tracker prior to leaving the lab. Then, a second male researcher delivered the instructions for the task and participants completed 15 practice trials. Participants were instructed to locate the O as quickly and as accurately as possible. Following the instructions, the second researcher left the lab and participants completed the visual attention task.

**Data preparation.** A number of criteria were put in place to ensure that the eye movements recorded were due to the stimuli presented and not to failures to follow instructions. Participants were required to fixate on the fixation cross. If a participant’s eye position strayed more than 2.82° from the central position during a 100 ms pretrial window, the trial was considered as having a bad baseline and was excluded from the analysis. Additionally, saccades away from the cross that occurred faster than 80 ms were considered anticipatory and were also excluded. Participants were excluded from the analysis if less than 60% of their eye movement data remained after these criteria were used.

Eye movements were categorized as saccades if participants shifted their gaze outside of the rectangle defined for looking at the fixation cross. The target display was divided into four equally sized areas that corresponded to the locations of the target letters. If a participant’s eyes moved more than 2.82° from the fixation cross toward a given quadrant of the display, the movement was scored as going to that location. These eye movements were classified as either going to the location of the cue, to the location of the target, or to the location of one of the other
locations, depending on what was presented at the screen location that the eyes moved to first. For valid trials, looking at the cue and looking at the target were scored as the same. Finally, the percentage of trials that participants looked to a location other than the cue or the target was averaged across the number of these locations that were presented. On invalid trials, an “other” location appeared in two of the four target display locations, so the overall percentage was divided by two. For valid trials, three of the target display locations matched this description, and, so, the overall percentage was divided by three.

Results

Two participants were excluded from these analyses due to missing more than 40% of their eye data, leaving a sample of 19 participants.

Reaction time. Correct reaction times were analyzed using a paired-samples t-test for invalid and valid trials\(^1\). Participants showed a cueing effect such that the reaction times for invalid trials were slower (\(M = 588\) ms, \(SD = 55\) ms) than valid trials (\(M = 544\) ms, \(SD = 61\) ms), \(t(18) = 6.88, p < .001, d = .723\).

Eye movements. We next examined the eye movements associated with these reaction times. The analyses for valid and invalid cues were conducted separately because they are not balanced. That is, for invalid cues, on any given trial the participants could have looked at the cue location, the target location, one of the other two locations, or continued to fixate the center of the screen after the cross was removed. However, for valid cues, the target and cue are in the

\(^1\) Mean error rate across all studies presented is equal to 2.15%, ranging from 0.42% to 3.38%, and does not differ by condition. Error rates were likely low because the target display remained available until participants responded.
same location, there are three other locations where the participants could have looked, and trials in which participants continue to fixate at the center of the screen.

The repeated measures ANOVA on invalid trial movements (Movement: cue, target, other, fixation) yielded a significant effect, $F (3, 54) = 4.86, p < .01, \eta_p^2 = .213$. Participants oriented to a location in which the cue or the target did not appear less ($M = 8.1\%, SD = 4.9\%$) than they oriented to the location of the target ($M = 30.1\%, SD = 14.1\%$), $F (1, 54) = 11.24, p < .01, d = 1.12$, less than they continued looking at the location of the fixation cross ($M = 29.5\%, SD = 27.9\%$), $F (1, 54) = 10.60, p < .01, d = 1.09$, and less than they looked at the cued location ($M = 24.2\%, SD = 15.3\%$), $F (1, 54) = 5.98, p = .02, d = .82$, which did not differ among themselves, $F$s < 1.

The repeated measures ANOVA for valid trial movements (Movement: cue/target, other, fixation) yielded a significant effect, $F (2, 36) = 12.43, p < .01, \eta_p^2 = .408$. Participants were more likely to look at the cue/target ($M = 49.3\%, SD = 23.5\%$) than to look to one of the other three display locations ($M = 5.8\%, SD = 4.5\%$), $F (1, 36) = 24.29, p < .01, d = 1.64$, and more than to not make an eye movement ($M = 33.2\%, SD = 30.5\%$), $F (1, 36) = 3.34, p = .07, d = .61$. Finally, participants were more likely to not make an eye movement than they were to orient to one of the other locations, $F (1, 36) = 9.64, p < .01, d = 1.03$.

Comparing invalid and valid trials, participants were more likely to look to the location of the cue/target on valid trials than the cue on invalid trials, $t (18) = 7.51, p < .001, d = 1.72$. This difference is likely due to the fact that cue and target occupy the same location for valid trials. Indeed, launch latencies to the cue for invalid trials were faster ($M = 256$ ms, $SD = 47$ ms) than those to the location of the cue/target on valid trials ($M = 315$ ms, $SD = 51$ ms), $t (16) = -
5.39, $p < .001$, $d = 1.31$, suggesting that the valid trials did indeed include a combination of faster saccades to the cue and slower saccades to the target.²

**Discussion**

In the baseline experiment, we were able to replicate Normand et al.’s cueing effect, such that participants were slower to respond when the cue was spatially invalid. However, the eye movement results suggest that looking at the abrupt onset cue is not a prepotent response on this version of the task. Participants oriented to the peripheral cue only 24.2% percent of the time on invalid trials, which is the better estimate of this behavior because the cue and target occupy the same position on valid trials.

² Two participants were missing one of these values (i.e., did not have any of the trial type and thus did not have a launch time for it) and, thus, were dropped from the analysis.
Experimenter Evaluation Experiment 1

The results of the first baseline suggest that the stimulus conditions used by Normand et al. (2014, Exp. 1), while producing an effect of evaluation on overall reaction times, do not permit us to pit the mere effort and focus of attentions explanations against each other, because looking at the cue is not a prepotent response. Nonetheless, we replicated Normand et al.’s experiment, while simultaneously measuring eye movements in an attempt to understand what produced the evaluation effect that they obtained.

The focus of attention account predicts that participants subject to evaluation produce smaller cueing effects because they ignore the cues to a greater extent than non-evaluated participants. The baseline experiment revealed that participants exhibit a number of different behaviors in their performance of this task, and it is possible the undergoing evaluation impacts some or all of these behaviors. For example, the effect of evaluation may only occur on trials on which participants make eye movements to the cued location, or only on trials on which participants continue to fixate at the location of the cross, or on both.

On the other hand, because looking at the cue is not a prepotent response, the mere effort account does not predict that looking at the cue will be potentiated, and so, the cue would not be expected to affect performance. However, if performing this version of the task comes down to simply finding the target as quickly as possible, the mere effort account could predict that evaluated participants would outperform non-evaluated participants because the evaluated participants are more motivated to perform well. This motivational effect has been demonstrated on other “simple” tasks (e.g., vertical mental subtraction, Seitchik & Harkins (2015); the prosaccade task, Jamieson & Harkins (2007), McFall et al. (2009); control stimuli on the Stroop, Jamieson & Harkins (2011), McFall et al., (2009)).
Method

Participants. 44 undergraduates were run from Northeastern University’s introductory psychology participant pool.

Procedure. The task was identical to the one described in the baseline experiment. After participants completed the block of practice trials, evaluation pressure was manipulated. The experimenter informed half of the participants that we were interested in their individual performance and that he would evaluate their performance once they had completed this task. The other half of the participants was told that we were interested in group performance and that we would not be looking at their individual scores (no evaluation condition). These participants were also instructed to press the letter “A” after completing the task to average their score with those of previous participants. Once the no evaluation participants finished the task, instructions on the display prompted them to average their scores. All participants were told to perform the task as quickly and as accurately as they could. Participants then completed the visual attention task.

Following this, participants were asked to respond to a manipulation check for experimenter evaluation with the critical item, “To what extent does the experimenter know how well you performed?” (1 = Not at all, 11 = Knows exactly).

Results

One participant was excluded from all analyses for having less than 60% usable eye movement data, leaving a sample of 43 participants.

Manipulation check for experimenter evaluation. Participants in the experimenter evaluation condition indicated that the experimenter knew how well they performed to a greater
extent \((M = 8.77, SD = 2.62)\) than participants in the no evaluation condition \((M = 3.48, SD = 2.70)\), \(t\ (41) = 6.54, p < .001\), suggesting that our manipulation of evaluation was successful.

**Reaction time.** The reaction time data for correct trials were analyzed using a 2 (Experimenter Evaluation: Yes vs. No) x 2 (cue validity: invalid vs. valid) ANOVA, with the first factor being between subjects, and the second within subjects. This analysis yielded a significant main effect for cue validity, \(F (1, 41) = 116.35, p < .001, \eta_p^2 = .74\), such that participants were slower to respond to invalid trials \((M = 602 ms, SD = 80 ms)\) than valid trials \((M = 557 ms, SD = 79 ms)\). There was also a significant main effect for experimenter evaluation, \(F (1, 41) = 4.07, p = .05, \eta_p^2 = .09\). Regardless of cue validity, participants subject to evaluation were faster to respond \((M = 557 ms, SD = 72 ms)\) than their non-evaluated counterparts \((M = 603 ms, SD = 82 ms)\). The interaction effect did not approach significance, \(F < 1\).

**Eye movements.** Eye-tracking data were once again analyzed using separate ANOVAs for the invalid and valid trials due to the unbalanced nature of the design. A 2 (Evaluation) x 4 (Movement: cue, target, other, fixation) ANOVA on invalid trials yielded a significant movement main effect, \(F (3, 123) = 9.42, p = .001, \eta_p^2 = .187\), such that participants were less likely to make an eye movement to one of the two display regions that did not contain the cue or the target \((M = 8.6\%, SD = 4.7\%)\) than to the cued location \((M = 29.6\%, SD = 19.8\%)\), \(F (1, 123) = 19.37, p < .01, d = .96\), than to the target location \((M = 31.2\%, SD = 14.9\%)\), \(F (1, 123) = 22.41, p < .01, d = 1.03\), or than to continue fixating at the screen’s central location \((M = 22.0\%, SD = 29.6\%)\), \(F (1, 123) = 7.88, p < .01, d = .61\). Eye movements to the target were marginally more likely than trials without an eye movement away from the fixation cross, \(F (1, 123) = 3.71, p = .06, d = .42\). No other differences between these movements reached significance. The interaction effect approached significance, \(F (3, 123) = 2.38, p = .07, \eta_p^2 = .055\). Evaluated
participants looked to the cue less \((M = 23.2\%, SD = 17.6\%)\) than the non-evaluated participants \((M = 36.3\%, SD = 20.0\%)\), \(F(1, 123) = 3.76, p = .05, d = .30\). Evaluated participants were marginally more likely to continue fixating at the screen’s central location \((M = 27.9\%, SD = 33.0\%)\) than non-evaluated participants \((M = 16.0\%, SD = 24.9\%)\), \(F(1, 123) = 3.12, p = .08, d = .28\). Evaluation condition did not influence the proportion of the other two eye movements, \(F < 1\).

A 2 (Evaluation) x 3 (Movement: cue/target, other, fixation) ANOVA was conducted on the valid trials. Again, the ANOVA yielded a significant effect for movement type, \(F(2, 82) = 34.59, p < .001, \eta^2_p = .458\). On valid trials, participants, regardless of evaluation condition, looked to the location of the cue/target more \((M = 53.7\%, SD = 23.1\%)\) than to one of the three uncued locations \((M = 7.8\%, SD = 5.3\%)\), \(F(1, 82) = 66.61, p < .001, d = 1.78\), and more than the percentage of trials on which they did not move their eyes from the fixation point \((M = 22.7\%, SD = 29.5\%)\), \(F(1, 82) = 30.38, p < .001, d = 1.20\). Participants were less likely to look to an uncued location than they were to not make an eye movement, \(F(1, 82) = 7.02, p < .01, d = .58\). The interaction effect did not approach significance, \(F(2, 82) = 1.98, p = .15, \eta^2_p = .046\).

Launch times for invalid and valid launches were compared in a 2 (Evaluation) x 2 (cue type: invalid vs. valid) ANOVA. This analysis yielded a main effect for cue type, \(F(1, 38) = 11.70, p < .01, \eta^2_p = .235\), such that invalid launches to the cue were faster \((M = 254 ms, SD = 92 ms)\) than valid launches \((M = 296 ms, SD = 50 ms)\), suggesting that the latter are once again composed of faster launches to the cue and slower launches to the target, replicating the finding for the baseline condition. Neither the evaluation main effect nor the interaction effect approached significance \((ps > .26)\).
In an attempt to determine how our manipulation of evaluation affected reaction time as a function of eye movement, we ran separate 2 (Evaluation) x 2 (Cue validity) ANOVAs for the three types of eye movements that occurred for both valid and invalid trials (i.e., to the cue or the cue/target, to a location other than the cue or the target, continuing at the center of the screen). An additional independent t-test was used to test the effect of evaluation on reaction times on invalid trials on which participants looked to the target.3

The ANOVA for trials that had eye movements to the cue yielded a significant effect of cue validity, $F(1, 38) = 110.89, \eta_p^2 = .745$, such that invalid trials were significantly slower ($M = 656$ ms, $SD = 94$ ms) than valid trials ($M = 555$ ms, $SD = 78$ ms), as expected. The main effect for evaluation was also significant, $F(1, 38) = 4.83, p = .03, \eta_p^2 = .113$. Regardless of cue validity, evaluated participants were faster ($M = 579$ ms, $SD = 76$ ms) than their non-evaluated counterparts ($M = 632$ ms, $SD = 89$ ms). The interaction was not significant, $F < 1$.

The ANOVA for trials that had eye movements to a location other than the cue or the target yielded a significant evaluation main effect, $F(1, 38) = 6.24, p = .02, \eta_p^2 = .141$, such that evaluated participants responded faster on these trials ($M = 598$ ms, $SD = 98$ ms) than the non-evaluated participants ($M = 668$ ms, $SD = 100$ ms). The cue validity main effect and the interaction were not significant, $Fs < 1$.

3This method serves to retain the maximum number of participants that have a RT for each type of movement, whereas other analyses aimed at understanding how different eye movements affect terminal reaction time suffer from large amounts of data loss. For example, analyzing these data across movement type separately for invalid and valid trials, similar to the approach taken for the eye movement analyses, lowers n from 43 to 26 for invalid trials and to 31 for valid trials.
The ANOVA for trials on which participants did not make an eye movement away from the screen’s central location yielded a marginal main effect for evaluation condition, $F(1, 27) = 3.01, p = .09, \eta^2_p = .10$, such that evaluated participants responded faster on these trials than non-evaluated participants. The main effect for cue validity and the interaction did not approach significant, $Fs < 2.02, ps > .16$.

Finally, the independent t-test for invalid trials on which participants looked to the location of the target location was not significant, $t(40) < 1$. However, the means were in the same direction as the previous main effects. The overall effect of experimenter evaluation was significant across these four analyses, combined $Z = 2.78, p < .01$.

**Discussion**

Using Normand et al.’s (2014, Exp. 1) paradigm, we did not replicate the interaction that they reported between evaluation and cue validity for reaction times. We did find a main effect for evaluation: evaluated participants outperformed non-evaluated participants, regardless of cue validity.

The fact that Normand et al. found an interaction between cue validity and evaluation, whereas we found a main effect for evaluation may be the result of a difference between our participants and those of Normand and her colleagues. Inspection of Normand et al.’s means shows that they found a cueing effect, not because evaluation participants were faster on invalid trials and slower on valid trials than no evaluation participants, but only because evaluated participants were faster on invalid trials. There was no difference on valid trials in their experiment. It is possible that Normand et al.’s participants were at some kind of floor for valid trials, which would make the effect of evaluation appear to interact with cue type, rather than producing a main effect, as seen in our results. Consistent with this possibility is the fact that
Normand et al.’s participants were faster than our participants on valid trials; our participants had a mean reaction time of 557 ms, whereas Normand et al.’s participants had a mean of 496 ms.

In any event, in Experiment 1, as in the baseline experiment, looking at the cue was not the prepotent response. On invalid trials, our best estimate of cue potency, participants looked at the cue approximately 30% of the time. That said, the eye movement results reveal that evaluated participants looked to the cue less and continued to fixate at the location of the cross marginally more than the no evaluation participants on invalid trials. This finding is consistent with the focus of attention prediction that the cue would be less influential under evaluative pressure, which would make reaction times for these trials faster. However evaluated participants were also faster on valid trials, which runs counter to the focus of attention prediction.

Given that looking at the cue is not the prepotent response, the mere effort account would not predict that making this response would be potentiated. But, if performance of the task under these conditions is reduced to simply finding the target as quickly as possible (i.e., serial search), the greater motivation of the evaluated participants could lead them to outperform the non-evaluated participants, regardless of eye movements. This outcome is consistent with what we found in Experiment 1, and with what has been found in previous research on “simple” tasks (e.g., McFall et al., 2009; Jamieson & Harkins, 2007; Seitchik & Harkins, 2015).

By splitting reaction time data by eye movement type, we found that evaluated participants were faster on trials on which they made eye movements to the cue, to a screen location besides where the cue or target was presented, and marginally faster for trials on which no eye movement was made, regardless of cue validity, suggesting that the overall main effect seen for terminal reaction time is not simply produced by looking to the cue less frequently. These findings show, once again, that the current design is not suitable to test the predictions of
the mere effort and focus of attention accounts because the effect of evaluation cannot be isolated to one type of eye movement.

The fact that the cue does not impact performance in this paradigm may be a result of the fact that the letters in Normand et al.’s display are so large (2.16° x 2.41°) that they can be recognized at an angular distance greater than 30° from the center of the display (Anstis, 1974). As a result, participants could see the target from the center of the display without moving their eyes. In fact, on approximately 31% of the trials, the baseline and Experiment 1 participants identified the target location without moving their eyes. Even on the trials on which the participants did move their eyes, it is possible that some of these movements were made to verify the identification of the target that they had already made without eye movement, rather than to find the target.

To pit mere effort and focus of attention against each other, the prepotent response on the task must be to look at the abrupt onset cue. In the next version of the paradigm, in an attempt to address this issue, we changed the stimulus parameters such that participants were required to move their eyes to successfully perform the task. Perhaps having to move one’s eyes to find the target will also make it more likely that participants will look at the abrupt onset cue.
Baseline 2

The goal of the second baseline was to modify the stimulus conditions used in Experiment 1 in order to produce a prepotent response that could then be used to pit the mere effort account against the focus of attention account. Specifically, the letter size in the target display was reduced so as to require eye movement to perform the task, which may also increase the potency of the abrupt onset cue. In a second change, we increased the interval between the cue display and target display by 75 ms (from 50 ms to 125 ms). This amount of time was chosen because the difference between the mean latencies of launches to the cue on invalid trials and launches to the cue/target on valid trials from the first baseline suggested that this additional time would result in separation of the two response types.

Method

Participants. 21 undergraduate students from Northeastern University participated in the experiment for partial fulfillment of a course requirement. None had any prior experience with the task.

Task and procedure. The task was exactly the same as the one used in the first baseline and Experiment 1 with two key changes. First, the letters in the target display subtended .26° x .23°, which was below the threshold of recognition with an eccentricity of 6.78° from the fixation point (Anstis, 1974), the eccentricity used by Normand et al. in their experiment and the one used in Baseline 1 and Experiment 1. Second, after the cue was presented for 30 ms, a blank screen appeared for 125 ms after which the target display was presented. After calibrating the eye-tracker, participants were given the task instructions, completed 15 practice trials and two blocks of 72 trials.

Results
Three participants were excluded from these analyses due to missing more than 40% of their eye data, leaving a sample of 18 participants.

**Reaction time.** Correct reaction times were analyzed using a paired-samples $t$-test for invalid and valid trials. Participants showed a cueing effect such that the reaction times for invalid trials were slower ($M = 1256$ ms, $SD = 110$ ms) than valid trials ($M = 976$ ms, $SD = 200$ ms), $t(17) = 6.02$, $p < .001$, $d = 1.68$.

**Eye movements.** As in the first paradigm, there are an unequal number of possible eye movements for invalid and valid trials. That is, invalid trials can lead to an eye movement to the cued location, the target location, or one of the other two display locations, whereas valid trials can only lead to an eye movement to the location of the cue/target or one of the other three locations. The new letter size was selected so that participants should be unable to perform this task without making eye movements, and, indeed, less than 1% of the correct responses were made without an eye movement. Because this rate was so low, these trials were excluded from the analyses that follow.

For invalid trials, a repeated measures ANOVA for eye movement type (movement: to the cue, to the target, to other) yielded a significant effect of movement, $F(2, 34) = 52.46$, $p < .001$, $\eta^2_p = .755$. Participants were more likely to look at the cue ($M = 63.6\%$, $SD = 22.4\%$) than they were to look at either the target ($M = 11.2\%$, $SD = 8.2\%$), $F(1, 34) = 79.72$, $p < .001$, $d = 3.16$, or the other two locations ($M = 12.6\%$, $SD = 7.8\%$), $F(1, 34) = 75.51$, $p < .001$, $d = 3.07$. Participants did not differ in the extent to which they looked at the target location and the two other locations, $F < 1$.

A dependent $t$-test was conducted on the percentages of eye movements made to the location of the cue/target and to the other three locations on valid trials. This analysis yielded a
significant effect, $t(17) = 6.58, p < .001, d = .88$, with participants being more likely to make an
eye movement to the location of the cue/target ($M = 59.8\%, SD = 22.4\%$) than to an uncued
distance ($M = 13.4\%, SD = 7.5\%)$.

Once again, the rate of looking at the cue was compared across invalid and valid trials.
The current stimulus conditions did not reveal a difference between these movements, $t(17) =
1.46, p = .16, d = .34$. Additionally, the latencies of the launch times for these eye movements
did not differ, $t(17) = 1.44, p = .17, d = .34$. Taken together, these findings suggest that valid
trials that are being scored as looking at the cue/target are likely to be mostly composed of
movements to the cue, rather than to the target. These findings also suggest that cue types can be
averaged to produce an estimate of how potent the cue is in the current paradigm. On average,
participants looked at the cue on 61% of the trials that were presented.

**Discussion.** After modifying the stimulus conditions of the basic paradigm used in
Normand et al. (2014), there was still a strong cueing effect. Furthermore, the modifications to
the task eliminated two problems in the first experiment: Performance of the task requires eye
movement; and the eye movements scored as going to the cue/target for valid trials now appear
to be the same as eye movements scored as going to the cue for invalid trials. That is, with the
new task parameters, we now have a reliable measure of the percentage of trials on which
participants look at the valid and invalid trials. Most importantly, the average rate at which
participants make an eye movement to the cue (61%) suggests that this behavior is the prepotent
response for this version of the task.
Experimenter Evaluation Experiment 2

Now that we have stimulus parameters for this task that produce a prepotent response, we can pit the mere effort account against the focus of attention account. Mere effort predicts that participants subject to evaluation will make more eye movements to the cue (i.e., the prepotent response will be potentiated), and when this response is incorrect (i.e., on invalid trials), evaluated participants will be faster to correct this error than non-evaluated participants, which will lead to faster reaction times. Making more eye movements to the cue is also predicted to make evaluated participants faster on valid trials. In tandem, these behaviors are expected to make evaluated participants faster than non-evaluated participants, regardless of trial type. On the other hand, the focus of attention account predicts that participants subject to evaluation should be better able to ignore the abrupt cue, meaning that they should make fewer eye movements to this location than non-evaluated participants. As an outcome of this, these participants should be slower on valid trials and faster on invalid trials, which will create a smaller cueing effect overall.

Method

Participants. 42 undergraduates were run from Northeastern University’s introductory psychology participant pool

Procedure. The task was identical to the one described in Baseline 2. After participants completed the block of practice trials, evaluation pressure was manipulated by using the same manipulation presented in Evaluation Experiment 1. Participants then completed 2 blocks of 72 trials, followed by a questionnaire containing the manipulation check item.

Results
Two participants were excluded from the analysis for having less than 60% usable eye data based on the criteria put in place. An additional participant was lost due to the computer program crashing in the middle of the test blocks, leaving a sample of 39 participants.

**Manipulation check for experimenter evaluation.** Participants in the experimenter evaluation condition indicated that the experimenter knew how well they performed to a greater extent ($M = 8.21$, $SD = 2.44$) than participants in the no evaluation condition ($M = 1.90$, $SD = 1.29$), $t(37) = 10.16$, $p < .001$, suggesting that our manipulation of evaluation was successful.

**Reaction time.** A 2 (Evaluation) x 2 (Cue Validity: invalid vs. valid) ANOVA with evaluation as a between subjects factor and cue validity as a within subjects factor was used to analyze reaction times for correct trials. This analysis yielded a significant main effect for cue validity, $F(1, 37) = 194.11$, $p < .001$, $\eta_p^2 = .840$, with participants responding faster to valid trials ($M = 899$ ms, $SD = 173$ ms) than invalid trials ($M = 1251$ ms, $SD = 141$).

The Evaluation x Cue Validity interaction also reached significance, $F(1, 37) = 5.52$, $p = .02$, $\eta_p^2 = .130$. On invalid trials, there was no difference between evaluated participants ($M = 1239$ ms, $SD = 143$ ms) and non-evaluated participants ($M = 1263$ ms, $SD = 142$ ms), $F < 1$. However, on valid cues, no experimenter evaluation participants identified the target more quickly ($M = 854$ ms, $SD = 116$ ms) than evaluation participants ($M = 948$ ms, $SD = 210$ ms), $F(1, 37) = 7.17$, $p < .05$, $d = .88$ (see Figure 2). The main effect for evaluation did not approach significance, $F < 1$.

**Eye Movements.** The eye movements associated with these reactions times were analyzed separately for invalid and valid trials due to the unbalanced nature of these data. A 2 (Evaluation) x 3 (movement: cue, target, other) ANOVA was conducted with evaluation as a between subjects factor and movement as within subjects. This analysis revealed a significant
movement effect, $F(2, 74) = 319.88, p < .001, \eta^2_p = .896$. Participants were more likely to make an eye movement to the cued location ($M = 74.2\%, SD = 16.6\%$) than to the location of the target ($M = 7.8\%, SD = 7.1\%$), $F(1, 74) = 477.64, p < .001, d = 5.01$, or to one of the other two display locations ($M = 9.0\%, SD = 6.0\%$), $F(1, 74) = 460.53, p < .001, d = 4.92$. Participants looked at the target as often as they looked to one of the other two display locations, $F < 1$. The Evaluation x Movement interaction was not significant, $F < 1$. A direct test of whether evaluated and non-evaluated participants differed in the extent to which they looked at the cue first was not significant, $F = 1.05, p > .30$.

A 2 (Evaluation) x 2 (Movement: cue/target, other) ANOVA was conducted on the valid trials. This analysis yielded a significant main effect for movement type, $F(1, 37) = 360.31, p < .001, \eta^2_p = .907$, such that participants looked to the cue more ($M = 74.6\%, SD = 16.6\%$) than the other three empty locations ($M = 8.46\%, SD = 5.5\%$). The interaction term did not reach, but approached, significance, $F(1, 37) = 2.62, p = .11, \eta^2_p = .066$. Once again, a direct comparison of the extent to which evaluated and non-evaluated participants looked at the cue was not significant, $F(1, 37) = 2.86, p = .10, d = .24$.

These movement findings are not consistent with either account. To replicate Normand et al.’s findings, we should have found that evaluated participants had slower reaction times than non-evaluated participants on invalid trials, but instead found a difference on valid trials. However, the pattern of reaction time data is potentially consistent with the focus of attention account (i.e., not looking at the cue would make valid trial reaction times slower). Yet, evaluated participants did not look at the cue significantly less than non-evaluated participants.

Mere effort was also not supported because it predicts that the evaluated participants would look at the cue more than non-evaluated participants, but are also motivated and should be
able to correct for this potentiated response. Evaluated participants did not look at the valid cue more than non-evaluated participants, and their reaction times were slower, not faster, than non-evaluated participants.

Taken together, these findings suggest that the results are produced by some other process in addition to or instead of the eye movements that would have been consistent with the focus of attention or mere effort accounts. One possibility is that the participants were taking some systematic approach to searching for the target. For example, they could have a tendency to start their search for the target in a particular display location. In fact, a number of participants volunteered the fact that they started their search in the top-left of the display, similar to the way one would start reading in the top-left. In the next set of analyses, we sought to determine whether there was any evidence of a specific search process that could help explain the pattern of findings.

**Location analyses.** To test this possibility, we examined the percentage of first looks from the fixation to one of the four locations when no cue was presented in that location in a 2 (Evaluation) x 4 (Location: top-left, top-right, bottom-left, bottom-right) ANOVA with evaluation as a between subjects factor and location as a within subjects factor. This analysis revealed a location main effect, $F(3, 111) = 39.16, p < .001, \eta^2_p = .514$. The most striking finding is that there was a marked tendency for participants to first look in the top-left of the display even though no cue was present ($M = 18.4\%, SD = 13.7\%$), which was stronger than the tendency to first look to any other location ($M_{\text{top-right}} = 6.1\%, SD = 5.8\%; M_{\text{bottom-left}} = 2.5\%, SD = 2.4\%; M_{\text{bottom-right}} = 2.9\%, SD = 3.4\%, ps < .05$, see Figure 3). Of course, each time the participant looked to the top-left even though no cue was present, there was a cue in some other
location to which the participant did not respond (e.g., looking to the top-left when it was uncued would be a cost for a valid trial occurring at a different location).

Next, we tested to see if this tendency to make the first eye movement to the top-left of the display was affected by our manipulation of evaluation. Indeed, participants subject to evaluation made more first looks to the top-left ($M = 22.2\%, SD = 14.8\%$) than non-evaluated participants ($M = 15.1\%, SD = 11.5\%$), $F (1, 111) = 7.07, p < .01, d = .87$.

To see whether this tendency to look to the top-left made it less likely that participants subject to experimenter evaluation would look at cued locations, we summed the proportion of looks to valid and invalid cues for each location and analyzed these data in a 2 (Evaluation) x 4 (Location) ANOVA. This analysis showed that evaluation and no evaluation participants did not differ in the proportion of looks to cued locations in either the top-left or the top-right, $Fs < 1$. In the bottom-left, participants subject to evaluation were marginally less likely to look at this location ($M = 15.9\%, SD = 7.3\%$) than no-evaluation participants ($M = 18.4\%, SD = 6.4\%$), $F (1, 111) = 3.04, p = .08, d = .57$. Finally, in the bottom-right, participants subject to evaluation looked at this location less ($M = 12.4\%, SD = 7.2\%$) than no-evaluation participants ($M = 16.3\%, SD = 7.0\%$), $F (1, 111) = 7.64, p < .01, d = .91$.

To examine how this location preference affected overall performance, we reanalyzed our reaction time data in a 2 (Evaluation) x 2 (Cue Validity: Invalid vs. Valid) x 4 (Location) ANOVA. In addition to the significant cueing effect and evaluation x cue validity interaction seen originally, this analysis revealed a significant effect of location, $F (3, 111) = 32.30, p < .001, \eta^2_p = .466$, which must be interpreted in the context of the cue validity x location interaction, $F (3, 111) = 10.20, p < .001, \eta^2_p = .216$ (see Figure 4). The reaction times show that participants exhibit the largest cueing effect in the top-left corner of the display, $F (1, 111) =$
177.07, $p < .001$, $d = 3.05$, a smaller cueing effect at the top-right, $F(1, 111) = 119.86, p < .001$, $d = 2.51$, a smaller cueing effect at the bottom-left, $F(1, 111) = 84.13, p < .001$, $d = 2.10$, and the smallest cueing effect at the bottom-right, $F(1, 111) = 36.33, p < .001$, $d = 1.38$. Analyzing by cue type suggests that reaction times to invalid trials do not change based on display location, $F(3, 111) = 1.17$, $p > .32$, whereas reaction times on valid trials vary based on display location, $F(3, 111) = 27.00, p < .001$. Specifically, as suggested by a significant linear contrast, $F(1, 37) = 55.82, p < .001$, $\eta^2_p = .601$, valid reaction times increase from the top-left to the top-right to the bottom-left to the bottom-right (see Figure 4).

Taken together with the eye movement data, these reaction time findings suggest that looking more to the top-left of the display when the cue occurs elsewhere does not directly impact invalid reaction times, but changes how quickly participants respond to valid trials. For invalid trials, a majority of the targets appear in a location other than the top-left, so looking to the top-left location first has the same effect as looking to an invalid cue positioned elsewhere. For valid trials, looking to the top-left rather than to the cued location has the same effect as an invalid trial (i.e., a slower reaction time). Furthermore, larger costs to reaction time should occur for valid trials as the position they are presented in changes from top-left to top-right to bottom-left to bottom-right, because participants are completing the task as if they are reading.

Because the evaluated participants looked to the top-left of the display more than the non-evaluated participants, we assessed whether this behavior affected their reaction times. Comparing the valid reaction times of our participants at each of the four locations revealed that evaluated participants were slower than no-evaluation participants for these trials at the bottom-left, $F(1, 111) = 7.16, p < .01$, $d = .88$, and the bottom-right, $F(1, 111) = 7.61, p < .01$, $d = .91$, but not at the two top locations, $Fs < 1.42, ps > .23$ (see Figure 5).
**Discussion.** In Baseline 2, we made changes to the paradigm used by Normand et al. (2014) and in Baseline 1 and Experiment 1 so that we could pit the focus of attention and mere effort accounts against each other. Specifically, we reduced the letter size so that participants were actually required to move their eyes to perform the task. Participants did not have to move their eyes to perform the task in Experiment 1. We also increased the time between the offset of the cue and the onset of the target by 75 ms so that we could separate eye movements to the cue and to the target on valid trials.

The results of Baseline 2 suggested that making these changes created the conditions necessary for the test of these accounts. Participants moved their eyes on 99% of the trials that they got correct, and the percentage of looks to the cue on invalid trials and to the cue/target on valid trials did not differ, nor did their launch times. And, unlike in Experiment 1, looking to the cue appeared to be the prepotent response, occurring on 61% of the trials.

We then added the evaluation manipulation to this paradigm in Experiment 2, and found that the evaluation manipulation made no difference in reaction times on invalid trials, whereas participants subject to experimenter evaluation were slower in their responses on valid trials. This finding could be seen as consistent with the focus of attention prediction that evaluated participants would be slower on valid trials, but the eye movement data do not show that evaluated participants looked at the valid cues significantly less than the no evaluation participants, which would be the focus of attention prediction. In fact, evaluation had no effect on eye movements. Of course, this finding is also inconsistent with the mere effort account, which would predict more looks to the cue, along with correction leading to better, not worse, performance by participants subject to evaluation.
Given that evaluation produced a reaction time effect, but not because of looking away from (or toward) the abrupt onset cue, we sought to determine what might be responsible for this effect. One possibility, suggested by comments by our participants, is that they were exhibiting a location preference such that they began their search for the target in the top-left quadrant, showing what might be termed a “reader’s bias.” And, in fact, in Experiment 2, when we took where the participants looked first into account, this is exactly what we found: Participants tended to look in the top-left quadrant first even when no cue was presented there. And, as a consequence, participants tended to look in the bottom quadrants less.

Of course, when participants look in the top-left quadrant when no cue is present, they miss the facilitating effect of the valid cues in the bottom quadrants, which slows their reaction times in these locations. There is no effect for invalid cues, because looking at the top-left quadrant when there is no cue produces essentially the same effect as looking at an invalid cue in the bottom quadrants. In both cases, participants are looking at a location in which the target will not be presented, and, as a result, this has no effect on the overall invalid reaction times. This accounts for the finding that reaction times on invalid trials do not change as a function of location, whereas the reaction times for valid trials in the bottom quadrants of the display are slower than those in the top of the display.

Finally, this analysis shows that this tendency to look to the top left (and look less at the bottom) is stronger in participants who are subject to evaluation than those that are not. That is, participants subject to evaluation look to the top left (exhibit the “reader’s bias”) to a greater extent than no evaluation participants. And, as a result, their reaction times on valid trials in the bottom of the display are slower than those of no evaluation participants, whereas their times do not differ on invalid trials. It is this reaction time difference on this subset of trials that produces
the Evaluation x Cue Validity interaction in overall reaction times. Thus, what appears to be a cue validity effect is actually a location effect.

Looking back at Baseline 2, which suggested that looking at the abrupt onset cue was the prepotent response, we see the same pattern of effects. Participants showed a bias towards looking to the top-left of the display when no cue was presented there ($M = 20.9\%$, SD = 18.9\%). It is also the case that, in these data, the reaction times for invalid cues do not change as a function of location, whereas the reaction times for valid cues are slower in the bottom half of the display than in the top half. Taken together, these findings suggest that, in this paradigm, looking at the abrupt onset cue is not the prepotent response. Instead, it appears that a more powerful tendency is to look to the top-left quadrant and this tendency is potentiated by the potential for evaluation.

Of course, there are number of possible explanations as to why the location effect overpowers the effect of the abrupt onset cue, but we will focus on two, each of which is a consequence of the 75 ms that were added to the blank between the offset of the cue and the onset of the target. One explanation, suggested by comments of participants after their sessions, is that they saw something flash on the screen before the target display appeared. In some cases, participants could describe the sequence of events exactly. In other cases, participants said that one of the “letters” flashed before the rest of the letters, or that all of the “letters” flashed. This was not the case in Experiment 1, in which few, if any, participants mentioned anything about the display. It is possible that participants, seeing that something was going on with the display, decided to implement a strategy focusing on location to avoid being affected by whatever was happening before the whole display was presented. The evaluated participants were more motivated to perform well, and implemented this approach to a greater extent than the non-
evaluated participants. If this were the case, removing the 75 ms would increase the potency of
the cue, and decrease the potency of the location effect.

A second possibility is that adding this blank had the effect of increasing the strength of
the cue. That is, because the cue was not masked, it could persist in the visual system, essentially
adding to the length of time that the cue was displayed. If this were the case, if we were to
eliminate this extra time, this location effect would be enhanced, and the abrupt onset cue would
produce fewer eye movements to its location. We tested these possibilities in Baseline and
Experiment 3.
Baseline 3

In Baseline 3, we eliminated the 75 ms that we had added to the time between the offset of the cue and the onset of the target. To the extent that participants look to the top left to avoid being affected by the flashing cues, in Baseline 3, because the cues are much less noticeable with the 75 ms taken out, the location effect should be minimized and performance may then be driven by the abrupt onset cue. On the other hand, if adding the 75 ms had the effect of increasing the potency of the abrupt onset cue, eliminating this time should decrease its potency, thereby increasing the strength of the location effect.

Baseline 3 will also tell us whether the 75 ms is necessary to separate looks to the cue versus look to the target on valid trials. That is, in Baseline 1 and Experiment 1, on valid trials some of the looks were to the cue and some were to the target. With the two changes, adding 75 ms and small letters, in Baseline 2 and Experiment 2, looks and launch times on valid trials were the same as on invalid trials (i.e., both produced by the cue). However, it is possible that the small letters alone are sufficient to produce this effect. Baseline 3 will show us whether this is the case.

Method

Participants. 21 undergraduate students from Northeastern University participated in the experiment for partial fulfillment of a course requirement. None had any prior experience with the task.

Task and procedure. The task is a combination of the two tasks from the previous experiments. An abrupt onset cue was presented for 30 ms, at either one of the four target display locations or at all four locations (for neutral trials), and the target display from baseline experiment 2 (i.e., the smaller target letters) was displayed following a 50 ms blank after the cue
offset. After calibrating the eye-tracker, participants were given the task instructions, completed 15 practice trials and two blocks of 72 trials.

Results

Reaction time by location. Reaction times for this task were analyzed using a 2 (cue validity) x 4 (cue location: top-left, top-right, bottom-left, bottom-right) within-subjects ANOVA. This analysis yielded the typical cueing effect, $F(1, 20) = 31.91, p < .001, \eta_p^2 = .615$. There was also a significant cue location effect, $F(3, 60) = 22.66, p < .001, \eta_p^2 = .531$. However, these effects must be interpreted in the context of the significant interaction, $F(3, 60) = 39.24, p < .001, \eta_p^2 = .662$. At the top-left and top-right locations of the display there were significant cueing effects, $F(1, 60) = 118.47, p < .001, d = 3.44$ and $F(1, 60) = 70.83, p < .001, d = 2.66$, respectively. For the remaining two locations there was no cueing effect (i.e., invalid and valid reaction times were the same), $Fs < 1.8, ps > .18$.

Location analyses. To examine the role of a location preference in this design, we tested the percentage of first looks to uncued locations in a repeated measures ANOVA for location (Top-Left, Top-Right, Bottom-Left, Bottom-Right). This analysis yielded a significant location effect, $F(3, 60) = 36.64, p < .001, \eta_p^2 = .647$ (see Figure 6, top half). Participants made more eye movements to the top-left of the display when no cue was presented ($M = 35.4\%, SD = 19.0\%$) than to any other display location ($M_{\text{top-right}} = 9.9\%, SD = 8.1\%$; $M_{\text{bottom-left}} = 4.0\%, SD = 2.8\%$; $M_{\text{bottom-right}} = 3.4\%, SD = 4.1\%$), $ps < .05$.

A second ANOVA was conducted on the summed proportions of valid and invalid trials on which participants looked to the cued location at each of the four display locations. This analysis yielded a significant location effect, $F(3, 60) = 56.34, p < .001, \eta_p^2 = .738$ (see Figure 6 bottom half). Participants were more likely to orient to the cues that appeared at the top-left ($M =$
20.9%, $SD = 4.1\%$) than when they appeared at the top-right ($M = 14.3\%, SD = 6.4\%), F (1, 60) = 22.87, p < .001. Participants were more likely to make an eye movement to the cue when it appeared in the top-right location than when it was displayed in either the bottom-left ($M = 5.7\%, SD = 5.6\%), F (1, 60) = 38.83, p < .001, or the bottom-right ($M = 6.4\%, SD = 5.3\%), F (1, 60) = 32.60, p < .001. Participants did not differ in the amount they made an eye movement to the cue when it was presented in the bottom-left or the bottom-right, $F < 1$.

**Discussion**

The results of the third baseline showed that there was an overall cueing effect, but it was produced by significant effects only in the top two locations. There was no cueing effect in either the bottom-left or the bottom-right. The eye movement results indicate that participants made a majority of their first eye movements to the top left of the display, both when a cue was presented at this location and when it was not.

It is also worth noting that when location is not taken into account, movements scored as looking to the cue would still appear to be the prepotent response ($M = 48.1\%, SD = 15.2\%). The results of this baseline show that a majority of these looks are from when the cue is presented in the top-left of the display, suggesting that this number may be inflated due to the tendency to begin the search in the top-left of the display.

Another key difference between the second and third baseline involves the potency of the cue. The results indicate that removing the additional 75 ms that were added to the Baseline 2 and Experiment 2 made the cue less potent and accentuated the location preference displayed in those experiments. Whereas the cue in the second baseline still attracted attention to the display’s lower locations on enough trials to create a cueing effect at these locations, the cue in the third baseline did not. Overall, participants looked to the cued locations about 13% less in the third
baseline than the second baseline. The only difference between these two experiments is the added time between the cue display’s offset and the target display’s onset, which suggests that the duration of this blank changes the potency of the cue. It appears that because the cue is not masked, it continues to persist in the visual system, and the longer the blank is, the more potent the cue becomes.

We next manipulated experimenter evaluation in this paradigm. If the prepotent response for this task is looking to the top-left of the display, then being evaluated should potentiate this response, according mere effort. The prediction of the focus of attention account is not affected by location; evaluated participants are still predicted to look at the cue less in this paradigm, as long as it is not central to task performance, which it is not.
Experiment 3

Method

Participants. 42 undergraduates were run from Northeastern University’s introductory psychology participant pool.

Procedure. The task was identical to the one described in baseline experiment 3. After participants completed the block of practice trials, evaluation pressure was manipulated by using the same manipulation presented in the first evaluation experiment. Participants then completed 2 blocks of 72 trials, followed by a questionnaire containing the manipulation check item.

Results

One participant was excluded from the analysis for having less than 60% usable eye data based on the criteria put in place, leaving a sample of 44 participants.

Manipulation check for experimenter evaluation. Participants in the experimenter evaluation condition indicated that the experimenter knew how well they performed to a greater extent ($M = 9.68$, $SD = 2.05$) than participants in the no evaluation condition ($M = 3.91$, $SD = 2.89$), $t (42) = 7.628, p < .001$, suggesting that our manipulation of evaluation was successful.

Reaction time by location. Reaction times were analyzed using a 2 (Evaluation) x 2 (cue validity) x 4 (location) ANOVA with the first factor as between subjects and the last two factors as within subjects. This analysis yielded a cue validity x location interaction, $F (3, 126) = 29.84$, $p < .001$, $\eta^2_p = .415$, replicating the baseline experiment. The cueing effect was significant in the top-left, $F (1, 126) = 79.47, p < .001$, $d = 1.95$, and in the top-right corner of the display, $F (1, 126) = 56.62, p < .001$, $d = 1.64$, but not in the bottom-left, $F (1, 126) = 2.16, p = .14$, $d = .32$, or the bottom-right, $F (1, 126) = 1.16, p > .25$, $d = -.23$. The main effects for cue validity, $F (1, 42)$
= 62.48, \( p < .001, \eta^2_p = .598 \), and for location, \( F(3, 126) = 40.28, p < .001, \eta^2_p = .490 \), should be interpreted in the context of this interaction. No other effects approached significance, \( ps > .32 \).

**Location analyses.** Participants’ first eye movements were analyzed using a 2 (Evaluation) \( \times \) 4 (location) ANOVAs for trials on which these movements went to uncued locations. This analysis yielded a significant interaction, \( F(3, 126) = 2.74, p = .05, \eta^2_p = .061 \) (see Figure 7). Evaluated participants made more eye movements to the top-left of the display (\( M = 34.5\%, SD = 19.5\% \)) than non-evaluated participants (\( M = 26.0\%, SD = 16.9\% \)), \( F(1, 126) = 4.97, p = .03, d = .69 \). Evaluated participants also tended to make fewer first eye movements to the top-right of the display than non-evaluated participants when no cue was presented there, \( F(1, 126) = 2.82, p = .10, d = .52 \). No differences emerged between the two groups at the other two locations, \( Fs < 1 \). The location main effect, \( F(3, 126) = 39.83, p < .001, \eta^2_p = .487 \) should be interpreted in the context of this interaction.

A second 2 (Evaluation) \( \times \) 4 (location) ANOVA was conducted for the summed invalid and valid cues. This analysis only yielded the location main effect, \( F(3, 126) = 59.36, p < .001, \eta^2_p = .586 \) (see Figure 8). Participants showed a preference for the top-left (\( M = 19.0\%, SD = 5.3\% \)), followed by the top-right (\( M = 14.4\%, SD = 5.8\% \)), bottom-left (\( M = 6.7\%, SD = 6.0\% \)), and bottom-right (\( M = 6.0\%, SD = 5.4\% \)), but evaluation did not modify this preference.

**Discussion**

The third baseline experiment identified looking to the top-left corner of the display as the prepotent response. In the experimenter evaluation experiment using these stimulus conditions, we found that evaluated participants made more of these eye movements (i.e., this behavior was potentiated). However, unlike the second experimenter evaluation experiment, making more of these eye movements did not lead to slower overall reaction times. In fact, there
were no differences between evaluated and non-evaluated participants when only assessing reaction times.

Baseline 3 showed that this design did not lead to cueing effects at the bottom-left and bottom-right of the display, which is where evaluated participants showed slower reaction times for valid trials in Experiment 2. That is, in Experiment 2, evaluated participants were making more looks to the top-left of the display when the cue appeared in the bottom-left and bottom-right, which for invalid trials did not alter reaction times, but made valid reaction times slower (e.g., if a valid trial occurred in the bottom-right and a participant looked to the top-left, this trial essentially became an invalid trial). However, in Experiment 3, removing the additional 75 ms between the cue and the target made the cue less potent, which led participants to make fewer eye movements to the bottom-left and bottom-right when the cue appeared there, eliminating the cueing effect. Thus, when evaluated participants made more eye movements to the top-left of the display when it was uncued, it did not come at the same cost as it did in Experiment 2, eliminating the evaluation effect (i.e., all participants have slow reaction times in the bottom-left and bottom-right because they are frequently looking elsewhere on valid trials).

Predictions for the mere effort account require that the prepotent response be identified. Clearly, our initial assumption that looking at the cue would be the prepotent response in the current paradigm was incorrect. Taken together, the results of the second experimenter evaluation experiment and the results from the third baseline and evaluation experiments suggest that the prepotent response for this task is to orient to the top-left of the display because the cue is not strong enough to capture attention on a majority of trials. These results also show that this behavior is potentiated by experimenter evaluation.
On the other hand, the focus of attention account does not make predictions that would be influenced by location. This account suggests that evaluated participants should ignore information that is not central to the completion of the task, which in the current set of experiments has been the cue (i.e., the location of this cue should not change how central it is to complete the task). Furthermore, this account should predict that participants would be less likely to rely on this cue at all four locations, not a specific subset of locations. Even if participants did rely only on a subset of locations, then the focus of attention account would still predict that being evaluated should lead to smaller cueing effects in both the second and third evaluation experiments because evaluation participants should have slower valid reaction times and faster invalid reaction times, despite showing a location bias (i.e., the cue should still have a smaller effect for evaluated participants at each of the locations).

Despite not conforming to our assumptions about how performance on this task is produced (i.e., the spatial cue is not the primary driver of performance), these two experiments can still provide tests of predictions drawn from the perspective of the focus of attention account and the mere effort account. In fact, even though a behavior other than orienting to the cue is producing the overall cueing effect, the results from Experiments 2 and 3 suggest that undergoing evaluation increases motivation, leading to an increase in the most likely response (i.e., eye movements to the top-left), which produces slower reaction times in the second experiment and no differences in the third experiment. These effects are congruent with the mere effort account, but do not support the predictions drawn from the focus of attention perspective.

This set of experiments demonstrates the importance of attempting to understand exactly how evaluation produces its effects on the given task. It would have been possible for us to run an experiment with the parameters used in any of these three evaluation experiments. Had we
done so and used only overall reaction time as the dependent measure, as is often done, we could have concluded that evaluation improves performance, that evaluation debilitates performance, or that evaluation has no effect on performance. And, in each case, we feel certain that an account could be developed that would provide a convincing explanation for that particular outcome. In the current line of research, the use of eye-tracking data has provided insight into exactly how evaluation produced the three different outcomes that we see in the three paradigms that we used.

Simply removing 75 ms between the offset of the cue display and the onset of the target display impacted the types of eye movements participants made to complete the task, which influenced terminal reaction times. In Baseline and Evaluation Experiment 2, participants oriented to the top-left of the display on a majority of trials, but still oriented to the cued location at the bottom-left and the bottom-right of the display often enough to create a significant cueing effect at these locations. However, participants in baseline and experiment 3 were about 50% less likely to orient to the cues presented in the bottom-left and bottom-right of the display than participants in the second baseline and evaluation experiment, which removed the cueing effect at these locations. This difference between these two paradigms suggests that the location bias and the effectiveness of the cue at capturing attention are negatively related. The prepotent response for both of these paradigms was to look to the top-left of the display first, but having a more salient cue led to fewer trials where participants first eye movement went to the top-left.

In the next set of experiments, we made additional changes to the task with the goal of finding a set of task parameters that removes the location bias seen in the first experiments and makes orienting to the abrupt onset cue the prepotent response. Our first attempt at achieving this goal was informed by a similar visual attention task described in Schreij, Los, Theeuwes, Enns,
and Olivers (2014). In their work, the target display was a diamond instead of a rectangle. We hypothesized that removing a stimulus from the top-left quadrant of the display could eliminate the reader’s bias seen in the earlier designs. In addition to this change, the cue display was presented for 50 ms instead of 30 ms and the duration of the interval between the cue’s offset and the target display’s onset was 100 ms, an intermediate value to the 125 ms used in baseline and experiment 2 and the 50 ms used in baseline and experiment 3. Schreij et al. reported that 60% of participants’ first saccades go to the location of the cue, which could have been influenced by a location bias, but may have been produced by a cue that is a prepotent response. Collectively, these changes represented an attempt to remove the location bias.

In another baseline experiment we made the three changes described above as a first attempt at removing the location bias. These changes led to the same result as our previous baselines. Participants exhibited cueing effects that differed depending on location, looking at cues appearing in the top, left, and right more than those presented at the bottom. Participants also showed a preference for the top and left, but not the right, locations even when a cue was not presented there, despite orienting to the cue about 73% of the time overall. We next strengthened the cue by doubling the length of time it was presented to 100 ms. Similar to the first baseline conducted using a diamond, participants did not orient as often to cues presented at the bottom display location, which led to smaller cueing effects for trials occurring at the bottom of the display. However, participants looked at the cue on approximately 87% of the trials in this design and showed equal cueing effects at the remaining three screen locations.

The two baselines using a diamond display suggested that cues at the bottom were not potent enough to capture attention as often as the other three screen locations. However, looking
to cues in the top three locations in the diamond was not affected by location. Thus, we removed the bottom location and used a triangle for the target display.
Baseline Experiment 4a

Method

Participants. 20 undergraduates were run from Northeastern University’s introductory psychology participant pool. None had any prior experience with the task.

Task and Procedure. The task used was a modified version of the task described in Baseline 3. As described in Schreij et al. (2014), the display locations were rotated to make a diamond, with the cues being presented for 50 ms, followed by a 100 ms blank. Due to the results of the two baselines using a diamond, the bottom location was removed from the target display, leaving a triangle. The cue and targets were still displayed 6.78° from the center of the display in the top, left, and right locations. Participants completed 15 practice trials, followed by 2 blocks of 72 actual trials, and were instructed to locate the O as quickly and as accurately as possible.

Results

One participant was excluded for having less than 60% of the data left after using the exclusion criteria stated above, leaving a sample of 19 participants.

Reaction time by location. A 2 (cue validity) x 3 (location: top, left, right) within-subjects ANOVA was used to analyze correct reaction times. This analysis revealed a cue validity x location interaction, \( F (2, 16) = 5.99, p < .01, \eta^2_p = .250 \), suggesting that the size of the cueing effect varied depending on the display location. Participants were faster to respond to valid trials when they were presented at the left of the display \((M = 670 \text{ ms}, SD = 149 \text{ ms})\) than they were when they were presented at the top \((M = 721 \text{ ms}, SD = 141 \text{ ms})\), \( F (1, 16) = 4.94, p = .04, d = .74 \). Participants were marginally faster to respond to valid trials at the top of the display than they were when these trials occurred on the right of the display \((M = 764 \text{ ms}, SD = 132 \text{ ms})\),
Evaluation and Visual Attention

\( F(1, 16) = 3.51, p = .07, d = .62 \). On the other hand, reaction times did not differ as a function of location for invalid trials. The significant cue validity effect, \( F(1, 18) = 173.81, p < .001, \eta_p^2 = .906 \), and the marginal location effect, \( F(2, 36) = 2.97, p = .06, \eta_p^2 = .142 \), should be interpreted in the context of this interaction.

**Location analyses.** The eye movements associated with these reaction times were analyzed using separate repeated-measures ANOVAs on location for trials where participants’ first looks either went to the cue or went to an uncued location. The ANOVA for eye movements to an uncued location revealed a marginally significant location effect, \( F(2, 36) = 3.04, p = .06, \eta_p^2 = .144 \). Participants were more likely to make an eye movement to the left of the display (\( M = 10.8\%, SD = 12.3\% \)) than to the right of the display when no cue was presented there (\( M = 4.3\%, SD = 3.9\% \)), \( F(1, 36) = 5.73, p = .02, d = .64 \). Participants looked to the top of the display when a cue was not presented there on 6.7\% of trials, but this did not differ from either of the other two display locations.

The ANOVA for looking to the cue revealed a significant location effect, \( F(2, 36) = 3.84, p = .03, \eta_p^2 = .176 \). Participants oriented more to the cue when it was presented on the left of the display (\( M = 29.4\%, SD = 3.1\% \)) than when it was presented on either the top (\( M = 24.9\%, SD = 8.4\% \)), \( F(1, 36) = 4.81, p = .03, d = .73 \), or the right of the display (\( M = 23.9\%, SD = 8.5\% \)), \( F(1, 36) = 7.24, p = .01, d = .90 \). Overall, participants looked at a cued location on 78.1\% of the trials.

**Discussion**

The reaction times from the current baseline still show evidence of the display location influencing the size of the cueing effect. Specifically, participants were faster on valid trials that occurred on the left of the display than they were at the right of the display, suggesting that
participants preferred making eye movements to the left of the display first. The eye movement
data confirm this; participants looked to the left of the display more than to the top or right
whether a cue appeared there or not. This finding suggests that the percentage of trials on which
participants are scored as “looking at the cue” may still be inflated by location preference. In the
next baseline, similar to the second diamond baseline, we extended the duration of the cue
display to determine whether a more potent cue would eliminate the location preference.

**Baseline Experiment 4b**

**Method**

**Participants.** 21 undergraduates were run from Northeastern University’s introductory
psychology participant pool. None had any prior experience with the task.

**Task and Procedure.** The task used was identical to the one used in the Baseline
Experiment 4a except for one change. The cue display was presented for 100 ms rather than 50
ms. Participants still completed 15 practice trials, followed by 2 blocks of 72 actual trials, and
were instructed to locate the $O$ as quickly and as accurately as possible.

**Results**

Two participants were excluded for having less than 60% of their data left after using the
exclusion criteria stated above, leaving a sample of 19 participants.

**Reaction time by location.** Reaction times for correct trials were submitted to a 2 (cue
validity) x 3 (location) ANOVA. This analysis yielded a significant cueing effect, $F(1, 18) =
198.48, p < .001, \eta_p^2 = .917$, such that participants were faster on valid trials ($M = 687$ ms, $SD =
202$ ms) than invalid trials ($M = 965$ ms, $SD = 184$ ms). The main effect for location and the
interaction term did not approach significance, $F$s < 1.61, $ps > .21$. 
**Location analyses.** Eye movements were submitted to separate ANOVAs for movements that went to cued and uncued locations. The ANOVA for eye movements to uncued locations did not yield a significant effect, $F(2, 34) = 1.62$, $p = .21$, $\eta_p^2 = .087$, suggesting that participants did not show a location bias when they did not orient to the cue. Similarly, the ANOVA for eye movements to the cue did not yield a significant location effect, $F < 1$. Participants in this task oriented to the cued location 89.4% of the time.

**Discussion**

Baseline 4b provided a set of reaction times that did not depend on location, and eye movements that were uniform at all display locations. However, participants oriented to the cue on 90% of the trials. The upper bound of the 95% CI suggests that this mean is potentially as high as 94%, which, despite leaving a small window to see a potentiated behavior, could just as easily be a ceiling for this effect. As a result, we conducted another baseline experiment (Baseline 4c) in which the cue was presented for 80 ms. In Baseline 4a, the cue was presented for 50 ms and participants looked at it on 78.1% of the trials, whereas in Baseline 4b the cue was presented for 100 ms, and participants looked at it on 89.4% of the trials. If we find that the looking percentage for the 80 ms cue display falls between these values, allowing for potentiation under evaluation, and, in addition, there is no location bias, we will have the conditions necessary to test mere effort against focus of attention.

**Baseline Experiment 4c**

**Method**

**Participants.** 22 undergraduates were run from Northeastern University’s introductory psychology participant pool. None had any prior experience with the task.
**Task and Procedure.** The task used was identical to the one used in the Baseline Experiment 4b except for one change. The cue display was presented for 80 ms rather than 100 ms. Participants still completed 15 practice trials, followed by 2 blocks of 72 actual trials, and were instructed to locate the *O* as quickly and as accurately as possible.

**Results**

One participant was excluded for having less than 60% of their data left after using the exclusion criteria stated above, leaving a sample of 22 participants.

**Reaction time by location.** Reaction times for correct trials were submitted to a 2 (cue validity) x 3 (location) ANOVA. This analysis yielded a significant cueing effect, $F(1, 21) = 111.70, p < .001, \eta^2_p = .842$, such that participants were faster on valid trials ($M = 690$ ms, $SD = 124$ ms) than invalid trials ($M = 910$ ms, $SD = 138$ ms). The main effect for location, $F(2, 42) = 2.89, p = .07, \eta^2_p = .121$, and the interaction term, $F(2, 42) = 1.37, p = .27, \eta^2_p = .061$, did not reach significance.

**Location analyses.** Eye movements were submitted to separate ANOVAs for movements that went to cued and uncued locations. The ANOVA for eye movements to uncued locations did not yield a significant effect, $F(2, 42) = 1.12, p = .34, \eta^2_p = .051$, suggesting that participants did not show a location bias when they did not orient to the cue. Similarly, the ANOVA for eye movements to the cue did not yield a significant location effect, $F(2, 42) = 1.64, p = .21, \eta^2_p = .072$. Participants in this task oriented to the cued location 80.7% of the time.

**Discussion**

Presenting the cue for 80 ms led to effects that were in between the effects seen for Baselines 4a and 4b. Participants exhibited a large cueing effect with a location effect that only approached significance, and no interaction between cue type and location. Additionally,
participants only oriented to the cued location 80.7%, which is unlikely to be a ceiling for this effect. Finally, participants did not show a preference for one location over another, regardless of whether or not their eye movements went to a cued or an uncued location.

Taken together, these effects suggest that the current stimulus conditions provide the best possible parameters to conduct the next experimenter evaluation experiment. Despite showing a possible location effect in reaction times, the current results appear to have lowered the looking rate enough for potentiation to be observed. It is also worth noting that a location preference is not seen in the eye movement analyses, suggesting that these stimulus parameters are the conditions necessary to test the mere effort account against the focus of attention account.

We next manipulated experimenter evaluation in this paradigm. Because the prepotent response is now to look to the location of the abrupt onset, mere effort predicts that evaluated participants will show more of this behavior than their non-evaluated counterparts. Mere effort also predicts that for invalid trials, evaluated participants will be more motivated to correct the incorrect prepotent response, which should make them faster for this trial type than non-evaluated participants. In addition to this, mere effort predicts that evaluated participants should be faster for valid trials, on which they are more likely to look at the cue/target location than non-evaluated participants. On the other hand, the predictions from the focus of attention account do not change; evaluated participants are predicted to filter out the abrupt onset cue because it is not thought to be central to task completion. In terms of eye movements, focus of attention predicts that evaluated participants will make fewer eye movements to the location of the abrupt onset cue than non-evaluated participants, which should make these participants slower for valid trials and faster for invalid trials.
Experimenter Evaluation Experiment 4

Method

Participants. 46 undergraduates were run from Northeastern University’s introductory psychology participant pool.

Procedure. The task was identical to the one described in baseline experiment 4c. After participants completed the block of practice trials, evaluation pressure was manipulated by using the same manipulation presented in the first evaluation experiment. Participants then completed 2 blocks of 72 trials, followed by a questionnaire containing the manipulation check item.

Results

One participant was excluded from the analysis for having less than 60% usable eye data based on the criteria put in place, leaving a sample of 45 participants.

Manipulation check for experimenter evaluation. Participants in the experimenter evaluation condition indicated that the experimenter knew how well they performed to a greater extent ($M = 8.70$, $SD = 2.44$) than participants in the no evaluation condition ($M = 2.45$, $SD = 2.09$), $t(43) = 9.20$, $p < .001$, $d = 2.75$ suggesting that our manipulation of evaluation was successful.

Reaction time. Overall reaction times were analyzed using a 2 (Evaluation) x 2 (cue validity) ANOVA with the first factor as a between subjects variable and the second as a within subjects variable. The main effect for cue validity was significant, $F(1, 43) = 430.90$, $p < .001$, $\eta^2_p = .909$, such that participants were faster to respond on valid trials ($M = 635$ ms, $SD = 112$) than they were to respond on invalid trials ($M = 902$ ms, $SD = 83$ ms). The Evaluation main effect and the interaction term did not approach significance, $Fs < 1$. 
**Eye movements.** The baseline experiment suggested that looking to the cue was the prepotent response for this task. We next tested whether experimenter evaluation led to potentiation (or a decrease) of this response. Because invalid and valid trials have a different number of possible responses, we tested these trial types separately. For invalid trials, a 2 (Evaluation) x 3 (Movement: cue, target, or other) ANOVA revealed a main effect for where participants looked, $F(2, 86) = 1046.30, p < .001, \eta_p^2 = .965$. This analysis suggests that participants looked at the cue on 88.1% of the trials ($SD = 11.1\%$), whereas they looked at the target location on 5.5% of the trials ($SD = 5.5\%$), and the other location on 6.4% of the trials ($SD = 6.5\%$). Experimenter evaluation condition did not have an effect on eye movements to the cue for invalid trials.

The analysis of the valid trials in a 2 (Evaluation) x 2 (Looking at the cue/target or looking at one of the other two locations), also revealed a main effect for looking, $F(1, 43) = 1843.87, p < .001, \eta_p^2 = .977$. Participants looked at the cue/target on 89.4% of the trials ($SD = 8.7\%$), whereas they looked at one of the other locations on 5.3% ($SD = 4.3\%$). Once again, evaluation condition did not have a significant effect on the type of eye movement made. Overall, participants looked to the cued location on 88.8% of the trials.

**Location analyses.** Despite not finding significant effects in the overall analyses, a location analysis was carried out to test whether location biases occluded effects in overall reaction time. Participants’ first eye movements were analyzed using a 2 (Evaluation) x 3 (location) ANOVAs for trials on which these movements went to uncued locations. This analysis yielded a significant location main effect, $F(2, 86) = 5.51, p < .01, \eta_p^2 = .114$, such that participants looked to the top of the display first more often when a cue was not presented there ($M = 5.5\%, SD = 6.0\%$) than they looked to the right of the display ($M = 2.0\%, SD = 3.3\%$), $F$
(1, 86) = 6.12, \( p = .02 \). No other differences emerged. The evaluation x location interaction did not reach significance \( F < 1 \).

A second 2 (Evaluation) x 3 (location) ANOVA was conducted for the summed invalid and valid cues. This analysis also yielded a location main effect, \( F (2, 86) = 4.77, p = .011, \eta_p^2 = .100 \). Participants looked less to cues presented on the right of the display (\( M = 28.2\% \), \( SD = 4.9\% \)) than they did to cues presented at the left of the display (\( M = 30.0\% \), \( SD = 4.1\% \), \( F (1, 86) = 4.06, p = .05 \), and less than cues presented at the top of the display, (\( M = 30.4\% \), \( SD = 3.9\% \), \( F (1, 86) = 5.44, p = .02 \). The interaction between location and evaluation did not reach significance, \( F < 1 \).

**Discussion**

The results from this experiment revealed no significant effects of experimenter evaluation on reaction times or on eye movements (i.e., showed no support for either the mere effort account or the focus of attention account). However, the additional location results reveal that participants looked less at cues presented at the right of the display compared to the other two locations, suggesting that the results of the baseline may have underestimated the size of this effect. Additionally, the rate at which participants looked at the cued location was underestimated. Mean percentages were used as the best estimate of the rate that participants looked to the cue, but closer examination of the distribution of this behavior in Baseline 4c reveals a negatively skewed distribution, suggesting that this measure of central tendency may not have been an appropriate descriptive. In fact, the median for Baseline 4c is 86.7\%, which is more consistent with the value found in the current evaluation experiment. This value is also similar to the rate observed for Baseline 4b, which was proposed to be potentially at a ceiling for this behavior.
Of Baselines 4a-4c, the conditions from Baseline 4c were thought to be optimal because they appeared to remove the location preference and the ceiling effect seen in Baseline 4a and Baseline 4b, respectively. However, the results of the current experiment suggest that these stimulus conditions did not remove either of these effects. For the current line of work, both of these effects are problematic, but the ceiling effect prevents the predictions from the mere effort account to be tested, making it a larger issue. This leaves Baseline 4a as the final option for testing these two accounts in this type of paradigm because it does not show evidence of a ceiling effect for looking to the cue.

Critically, the location preference seen in Baseline 4a differs from those seen in Experiments 2 and 3. In the earlier experiments, the location preference was the prepotent response (i.e., looking to the top-left), which inflated the rate of trials scored as “looking to the cue” to the point that this behavior appeared to be the prepotent response. As an example of this, consider the results of Baseline Experiment 2. On average, participants in this baseline looked to the cued location 61% of the time, with 24% of the trials overall occurring in the top-left and only 11% of trials occurring in the bottom-right. Thus, if the true potency of the cue was closer to 11% at each display location, this response would not be prepotent. However, in Baseline 4a, participants look to the cued location approximately 78% of the time, with 29% of trials occurring in the left, but only 25% and 24% when it is in the top and right, respectively. Thus, even in the location that participants look to the cued location the least, this behavior occurs frequently enough to still consider it a prepotent response (i.e., on the right of the display, participants look to the cued location about 24% of the time).

The results of Baselines 4a-4c, as well as Evaluation Experiment 4, suggest that there may not be a set of conditions that can completely remove location preferences without
approaching a ceiling for looking at the cue. However, the results of Baseline 4a appear to create a set of parameters in which looking to the cue is the prepotent response and this behavior does not approach a ceiling. Thus, we ran another experimenter evaluation experiment using the stimulus conditions from Baseline 4a. The predictions from both the mere effort and the focus of attention accounts remain the same as they were for Evaluation Experiment 4.
Experimenter Evaluation Experiment 5

Method

Participants. 44 undergraduates were run from Northeastern University’s introductory psychology participant pool.

Procedure. The task was identical to the one described in Baseline Experiment 4a. After participants completed the block of practice trials, evaluation pressure was manipulated by using the same manipulation presented in the first evaluation experiment. Participants then completed 2 blocks of 72 trials, followed by a questionnaire containing the manipulation check item.

Results

Two participants were excluded from the analysis for having less than 60% usable eye data based on the criteria put in place, leaving a sample of 42 participants.

Manipulation check for experimenter evaluation. Participants in the experimenter evaluation condition indicated that the experimenter knew how well they performed to a greater extent ($M = 8.40, SD = 2.80$) than participants in the no evaluation condition ($M = 2.68, SD = 2.85$), $t(43) = 6.55, p < .001, d = 2.02$, suggesting that our manipulation of evaluation was successful.

Reaction time. Overall reaction times were analyzed in a 2 (Evaluation) x 2 (cue validity) ANOVA with evaluation as a between subjects factor and cue validity as a within subjects factor. This analysis revealed a cue validity main effect, $F(1, 40) = 206.47, p < .001, \eta^2_p = .838$. Reaction times for valid cues were faster ($M = 726$ ms, $SD = 148$ ms) than for invalid cues ($M = 945$ ms, $SD = 114$ ms). To test directly for evaluation effects, we contrasted evaluated and non-evaluated participants’ reaction times for valid and for invalid cues, using the error term for the interaction. For invalid cues, evaluated participants were significantly faster ($M = 913$ ms,
Evaluation and Visual Attention

SD = 69 ms) than non-evaluated participants (M = 974 ms, SD = 138 ms), F (1, 40) = 7.80, p < .01. However, for valid cues, the comparison only approached conventional significance levels, F (1, 40) = 3.03, p = .09. Evaluated participants tended to be faster (M = 706 ms, SD = 79 ms) than non-evaluated participants (M = 744 ms, SD = 191 ms).

**Eye movements.** To determine the source of this pattern of differences, we examined these effects in the context of the eye movements. Once again, because there are unequal numbers of possible eye movements for invalid and valid trials, we tested these cue types separately. For invalid trials, a 2 (Evaluation) x 3 (Movement: cue, target, other) ANOVA revealed a main effect for where participants looked, F (2, 80) = 318.71, p < .001, ηp² = .888. This analysis suggests that participants looked at the cue on 77.6% of the trials (SD = 15.4%), whereas they looked at the target location on 10.8% of the trials (SD = 8.0%), and the other location on 11.6% of the trials (SD = 8.8%). Experimenter evaluation condition did not affect these rates, F < 1.

The analysis of the valid trials in a 2 (Evaluation) x 2 (Movement: cue/target, other), also revealed a main effect for looking, F (1, 40) = 291.29, p < .001, ηp² = .912. Participants looked at the cue/target on 77.4% of the trials (SD = 16.6%), whereas they looked at the other locations on 11.3% (SD = 8.3%). Once again, evaluation did not influence these rates, F < 1.

These findings suggest that neither looking toward the cue, nor away from it contributes to the effect that evaluation had on overall reaction time differences in this experiment. As the next step, we examined whether a location preference was the source of the overall reaction time differences.

**Location analyses.** To test this possibility, as in Experiment 2, we examined the percentage of first looks from the fixation to one of the three locations when no cue was
presented in that location in a 2 (Evaluation) x 3 (Location) ANOVA with evaluation as a between subjects factor and location as a within subjects factor. Unlike in Experiment 2, this analysis did not yield a significant location effect, $F(2, 80) = 2.52, p = .09$. However, there was a tendency for participants to look to the top more ($M = 10.4\%, SD = 11.9\%$) than to the right ($M = 5.3\%, SD = 10.3\%$) or to the left ($M = 6.8\%, SD = 8.2\%$). The main effect for evaluation, as well as the interaction term did not approach significance, $Fs < 1$.

A second 2 (Evaluation) x 3 (Location) ANOVA was conducted for trials on which participants did orient to the cued location. Neither main effect nor the interaction were significant, $Fs < 1$, suggesting participants did not favor one location over another in this paradigm, regardless of experimenter evaluation condition.

**Reaction times by eye movement.** Similar to Evaluation Experiment 1, the current experiment yields a significant evaluation effect in overall reaction times that does not appear to be the outcome of making more or less eye movements to the cue. Additionally, the effect on reaction times is not the outcome of a location preference. We next sought to test whether evaluation condition interacted with the reaction times for each trial type to produce the effect at the level of overall reaction times.

The current reaction time effects suggest that the effect of evaluation was different on invalid and valid trials, unlike the evaluation main effect in the first evaluation experiment. Additionally, in the first experiment, we were specifically interested in how the size of the cueing effect potentially varied depending on the type of eye movement that participants made, which is not of interest in the current results. Thus, rather than relying on multiple ANOVAs, as we did in Evaluation Experiment 1, separate linear mixed effects models were used for invalid and valid trials. Both models relied on modeling participants as a random factor, with trial type
nested within participant to predict reaction time. This method also serves to maximize the amount of data available to interpret these effects (i.e., individual ANOVAs suffer from data loss).

Experimenter evaluation condition was entered at the level of participants and was effect coded (Evaluation = -1, No Evaluation = +1). Using model comparison techniques that compare the omnibus model to reduced models that have specific terms missing (e.g., a model that removes all effects of evaluation), this model can be used to test for main effects and interaction terms akin to those found in ANOVA designs. For interpretive clarity, the $X^2$ value that is used for these comparisons has been transformed into $F$ statistics.

For invalid trials, two effect coded variables were used to examine the relationship between specific eye movements and reaction time, with the first variable coded as: Cue = -1, Target = +1, Other = 0, and the second variable coded as: Cue = -1, Target = 0, Other = +1. In addition to these within-subject effects, interactions between evaluation and these variables were also entered into the omnibus model.

On these trials, there is a significant evaluation x movement interaction, $F (2, 1664) = 2.96, p = .05$, suggesting that reaction times differ by evaluation condition and the type of eye movement made (see Figure 9). In order to parse this interaction, follow-up models were run to test for an evaluation effect for each movement type. For ease of interpretation, evaluation was recoded as a dummy variable (No Evaluation = 0, Evaluation = +1), such that beta weights can be interpreted as the difference in reaction times for these conditions in milliseconds. The evaluation conditions did not differ in reaction time when participants made an eye movement to the location of the cue, $b = -53.52, SE = 31.71, t (40) = -1.69, p = .10$. Similarly, these two conditions did not differ in reaction time when participants made an eye movement directly to
the location where the target was presented, \( b = -66.55, SE = 42.40, t (35) = -1.57, p = .13 \).

However, evaluated participants were faster to respond than the no evaluation participants when participants looked to the location that did not display the cue or the target, \( b = -144.80, SE = 64.18, t (35) = -2.26, p = .03 \). The marginally significant main effect for Evaluation, \( F (1, 40) = 3.57, p = .07 \) and the significant main effect for Eye Movement, \( F (2, 1664) = 146.46, p < .001 \), are best interpreted in the context of the interaction.

For valid trials, a parallel model was run, but because there are only two possible eye movements there is only one within-subjects variable (Cue/Target = -1, Other = +1). This contrast, in addition to the effect of evaluation (Evaluation = -1, No Evaluation = +1) and the interaction term were used to predict reaction time. This model yielded a significant interaction effect, \( F (1, 1698) = 13.15, p < .001 \). Once again, follow-up models were conducted for each movement type using a dummy coded Evaluation variable. For eye movements that went to the cue/target, the evaluation conditions did not differ in reaction times, \( b = -28.62, SE = 35.50, t (40) = -0.81, p = .42 \). Again, evaluated participants were faster than the non-evaluated participants when their first eye movements went to one of the two locations that did not display the cue/target, \( b = -129.53, SE = 41.27, t (38) = -3.14, p < .01 \). The marginal evaluation main effect, \( F (1, 40) = 3.76, p = .06 \), and the significant eye movement main effect, \( F (2, 1698) = 903.38, p < .001 \), are best interpreted in the context of this interaction (Figure 10).

**Discussion**

We found that evaluated participants were faster than non-evaluated participants on invalid trials, but not significantly faster on valid trials, although this difference did approach significance. We also found that looking at the cue was the prepotent response with participants across the cue types looking at the cue on 78.3% of the trials. Given that the level of this value is
not so high as to produce concern about ceiling effects, it should be possible to see whether
evaluation produced potentiation of this response, which would support the mere effort account,
or led to looking away, which would support focus of attention. In fact, we see neither of these
responses, suggesting that the reaction time effect is not the outcome of orienting to the cue. We
also did not find evidence that this outcome was due to a location preference similar to that seen
in Evaluation Experiment 2, supporting the notion that orienting to the cued location was the
prepotent response for this task.

We next included information about where participants looked first (e.g., to the cue) to
determine whether these behaviors produced the overall reaction time results. For invalid trials,
this analysis revealed that the manipulation of evaluation had different effects depending on the
type of eye movement that a participant made. Specifically, this analysis revealed that evaluated
participants were faster than their non-evaluated counterparts when their first eye movement
went to an uncued location without the target, but did not significantly differ for eye movements
that went to the cued location or to the target location, although these effects did show a trend
toward significance (cue: $p = .10$; target: $p = .13$).

A similar pattern was found for valid trials, such that evaluated participants were faster
than non-evaluated participants when they looked to an uncued location, but did not differ when
they looked at the cue/target location. However, in this case, there was no tendency toward a
significant evaluation effect, $p = .42$.

Thus, the effect of evaluation is the largest for trials on which participants do not orient to
the cued location or to the target location, regardless of trial type (i.e., Invalid Other and Valid
Other movements). For each of these trial types, participants make a volitional launch away from
the fixation cross to a location that did not display the eventual target, meaning that additional
eye movements must be made to find the target. This series of events provides the most opportunity for motivation to have an effect (i.e., there are multiple eye movements to be made).

On the other hand, trials on which participants look directly to the target (i.e., Valid cue/target and Invalid target movements) require no correction and, thus, provide less opportunity for the effect of motivation. In addition, launches drawn by the cue are faster \( (M = 224 \text{ ms}, SD = 39 \text{ ms}) \) than volitional launches \( (M = 314 \text{ ms}, SD = 43 \text{ ms}) \), which speeds up the search process for both evaluated and non-evaluated participants, thereby reducing the role that motivation can play.

This pattern of effects explains why valid trials on which participants looked to the cue/target show the smallest effect of evaluation: the search process is started sooner due to the cue and there is no need to correct the search process through additional movements. And it also provides an explanation as to why the reaction time effects are marginal for invalid cues when the participants look at the cue or at the target. In the case of the cue looks, all participants get the advantage of the speeded eye movement to the cue, and motivation can only affect the subsequent behavior to find the target. In the case of the target looks, all participants get the advantage of being at the target location, and motivation can only affect the initial volitional launch and the response time to the target.

Put together, these results show why, in the overall reaction time analysis, there is a significant evaluation effect for invalid trials, but only a marginal effects for valid trials. On invalid trials, the cue and target effects are only marginal, because in each case, there is less opportunity for motivation to have an effect, but there are small effects in favor of the evaluated participants. The strong evaluation effect is produced when participants look to an uncued
location, and there is no benefit from the cue or the target. When these trials are combined with the marginal cue and target effects, the overall reaction time yields a significant effect.

On the other hand, on valid trials, the great majority of the time participants look at the cue/target, which provides all participants with the benefits of the speeded launch and finding themselves at the target location with their first eye movement. These trials swamp the effect of the significant evaluation effect produced when participants look first at an uncued location, yielding the marginal evaluation effect for valid trials.

Together, the results of this experiment suggest that participants faced with experimenter evaluation were more motivated to perform the task, as evidenced by faster reaction times, but this increase in motivation did not lead to potentiation of the prepotent response, nor did it produce faster performance for both trial types. As noted above, these findings do not support the predictions drawn from either account, but are better aligned with an account focused on the motivational effects of experimenter evaluation than an account implicating possible cognitive or perceptual effects.

However, failing to find potentiation in this paradigm suggests that the mere effort account needs revision. There are two important, potentially related, distinctions to be made between the current work and the antisaccade research that provided the mere effort account predictions. The first deals with awareness about the experimental paradigm. On the antisaccade task, participants understand that they will see a cue flash on the screen and that they are to look away from this cue, but in the current work, participants are not told beforehand that a small spatial cue will be presented in one of the three display locations or, in the case of neutral trials, in all three locations, prior to the target array being presented. In fact, during debriefing approximately two-thirds of our participants reported that they did not notice anything about the
flashing cues, and, of the participants that did notice, few could actually report back the order of events displayed in the experiment, with most of these participants reporting that letters would “flash” or “blink,” rather than reporting a cue prior to the display. Thus, our participants are, for the most part, completely unaware of the cue, which may be required to see potentiation. The second difference between the current experiment and the antisaccade deals with the kind of task each represents. The antisaccade is an inhibition task, in which participants are told to “look away” from a stimulus, whereas in the paradigms used in the current work, most participants are unaware of the spatial cues and cannot be told how to interact with them (e.g., told to look away). Furthermore, on valid trials there is no reason to inhibit an eye movement to the cued location (i.e., the cue is presented at the eventual target location), which renders this possible instruction illogical for this task. While the mere effort account has utility in describing a variety of tasks, most of which are not inhibition tasks, the paradigm used in this experiment does not conform to the predictions of this account. We return to this topic in the General Discussion.
General Discussion

Two competing theories have been proposed to explain the effect of evaluation on task performance. The mere effort account argues that the potential for evaluation motivates individuals to perform well, which potentiates the prepotent response. If the prepotent response is incorrect and participants do not know, or lack the knowledge or time required for correction, performance is debilitated. However, if the prepotent response is correct, or if participants are able to recognize that their prepotent tendencies are incorrect and are given the opportunity to correct, performance will be facilitated. Using a visual attention task, Normand et al. (2014) present evidence that appears inconsistent with the mere effort account, and, instead, favors the focus of attention account. This account suggests that the potential for evaluation triggers concerns about one’s performance, reducing the amount of information that can be processed. Due to this decrease in processing capacity, individuals are suggested to prioritize the most relevant information and disregard less focal information. However, Normand et al. relied on terminal reaction times to infer the mechanisms that underlie these effects without actually measuring these processes. In the current research, we revisited the visual attention paradigm used in Normand et al. (2014) with the addition of eye-tracking in order to provide a better understanding of the processes that produce these effects.

Baseline 1 revealed that the stimulus conditions used in Normand et al (2014, Exp. 1) were not suitable to pit the mere effort account against the focus of attention account. Specifically, looking to the spatial cue was not the prepotent response, which is required to test the predictions of each account, perhaps because the letters in the target display were so large that participants did not need to make an eye movement to respond correctly. Nevertheless, we set out to replicate the reaction time results in Normand et al. (2014, Exp. 1). Experiment 1
revealed that evaluated participants were faster to respond, regardless of trial type (i.e., invalid or valid) and the type of eye movement made (i.e., to the cue, to the target, etc.). This result is consistent with a motivational account of evaluation, which suggests that increased effort would lead to faster reaction times, independent of the way a participant completed a given trial.

We then reduced the target letter size and increased the amount of time the screen was blank between the cue and the target, which we believed made looking to the cued location the prepotent response in Baseline 2, providing the necessary conditions to test the mere effort account against the focus of attention account. Instead, the results of Experiment 2 show that our evaluation manipulation did not increase or decrease the rate at which participants looked to the cue and that evaluated participants were slower than non-evaluated participants on valid trials, but there was no difference on invalid trials. These findings are inconsistent with the predictions drawn from both accounts if looking to the cue is the prepotent response. In an exploratory analysis, we found that looking to the cued location was not the prepotent response, and that participants were using a different systematic approach to completing the task (i.e., looking to the top-left of the display first). Evaluated participants also showed an increase in this behavior compared to the non-evaluated participants, which made these participants slower to respond on valid trials that occurred in the bottom-left and bottom-right display locations, accounting for the differences in overall reaction time.

In Baseline and Experiment 3, we found that decreasing the gap between the cue and target displays led to an increase in the location bias, suggesting that this was the prepotent response for this paradigm. The reaction times revealed that even though looking to the spatial cue was not the prepotent response, an overall cueing effect was still produced. Closer examination of this effect revealed that it was produced by significant cueing effects at the top-
left and top-right of the display, but significant cueing effects were not produced in the bottom display locations. We also found that evaluated participants made more eye movements to the top-left of the display than did non-evaluated participants, suggesting that this behavior was potentiated. However, this difference in behavior no longer led to differences in terminal reaction time, as seen in Evaluation Experiment 2. Specifically, because this set of stimulus conditions no longer produces cueing effects at the bottom-left and bottom-right of the display, the cost in reaction time that was seen for looking more to the top-left in the second experiment is removed in Evaluation Experiment 3 (i.e., looking to the top-left does not make evaluated participants slower in the bottom display locations).

The results of the second and third experiments are consistent with the mere effort account, albeit not in the way we had anticipated. That is, in this paradigm, looking to the top-left of the display was a much stronger behavior than looking to the spatial cue. Our initial predictions resulted from the assumption that eye movements to the cue were the main driver of performance, similar to the antisaccade task. However, the location analyses in these experiments revealed that looking to the top-left of the display first was the prepotent response and that undergoing evaluation potentiated this response.

On the other hand, once location is taken into account, the focus of attention account is no longer able to explain these results. This account suggests that participants would be less likely to rely on this cue at all four locations, not a specific subset of locations. Additionally, the focus of attention account would still predict that being evaluated should lead to smaller cueing effects in both the second and third evaluation experiments because evaluation participants should have slower valid reaction times and faster invalid reaction times because the cue should still have a smaller effect for evaluated participants at each of the locations.
We next ran a series of baselines in order to isolate a set of stimulus conditions in which looking to the spatial cue was the prepotent response. Using the parameters from Baseline 4c, we ran Evaluation Experiment 4 and found all null effects, for both reaction times and eye movements. Participants in this experiment looked to the cued location on approximately 89% of trials, which potentially represents a ceiling for this effect. If that is the case, then potentiation of this behavior is not possible. Indeed, examining the distribution for this behavior in Baseline 4c suggested that relying on the mean led us to underestimate the rate at which participants looked to the cue. Yet, the focus of attention account would still predict that evaluated participants look less to the cued location, which could still be observed despite the ceiling effect, but no evidence of this was found.

To avoid this ceiling we ran a fifth evaluation experiment using the stimulus conditions from Baseline 4a. We found that evaluated participants were faster than their non-evaluated counterparts on invalid trials, but did not differ on valid trials. Again, we found that these conditions did not differ in the rate that they oriented in the cue, which is inconsistent with the predictions drawn from both accounts, and that a location bias did not produce the difference in overall reaction times.

A final analysis revealed that the largest evaluation effect was produced on the subset of trials on which participants did not look to the cue. Specifically, evaluated participants were faster to respond than non-evaluated participants when their first eye movement went to a location in which the cue or the target was not presented. This finding is consistent with the notion that the reason that the evaluated participants are outperforming the non-evaluated participants is due to increased motivation. However, this increase in motivation does not lead to significantly faster reaction times for invalid trials on which participants move to the cue, nor for
invalid trials on which participants move to the location of the eventual target, however, both of these effects approached significance. This pattern of effects suggests that trials on which a participant arrives at the target (e.g., invalid trials on which participants orient directly to the target) there is not an opportunity for this increased motivation to show because participants are at the target location. Additionally, for invalid trials on which participants look to the cue, it is possible that the advantage provided from looking to the cue (i.e., beginning the search process sooner) is large enough to cancel out the effect of motivation on responding (see Figure 9). Finally, the smallest effect for evaluation was observed for valid trials on which participants made an eye movement to the cue, which combines the two advantages of beginning the search process sooner due to the cue and not having to correct once at the target location. Thus, the effect of motivation has the smallest opportunity to be produces for these trials.

This explanation also accounts for the lack of significant effects in Evaluation Experiment 4. If effects in this version of the paradigm were the outcome of motivated serial search on trials on which participants orient to a location other than the cue or the target, then the ceiling effect for looking at the cue observed in this experiment would have eliminated these effects. That is, there were not enough of these trials to produce an evaluation effect in overall reaction times.

The results from the final two experiments are at odds with both the mere effort account and the focus of attention account. In both experiments, the predictions that evaluation would lead to more (mere effort) or less (focus of attention) eye movements to the cued location were not supported. It is unclear what effects focus of attention would predict without a change in eye movements. On the other hand, mere effort affords a central role to motivation, which leads to the prediction that evaluation participants could outperform no evaluation participants in absence
of potentiation. As suggested above, it is unclear how any account relying upon a perceptual or cognitive explanation could explain these effects.

Across five experiments, we found that undergoing experimenter evaluation leads to effects that are consistent with an account that proposes this process leads to an increase in motivation. Of these experiments, two show potentiation of a prepotent response (Experiments 2 & 3), consistent with the mere effort account. Of the three other experiments, one paradigm did not have an identifiable prepotent response (Experiment 1) and two did not show potentiation of the prepotent response identified (Experiments 4 & 5). As introduced above, characteristics of the current paradigms may have differed from other tasks in which potentiation has been observed (e.g., the antisaccade task) in critical ways that do not lead to potentiation.

Most notably, performance in this experimental paradigm, from our initial reading of it in Normand et al. (2014) through Evaluation Experiment 5, was expected to operate similarly to other inhibition tasks, but repeatedly did not. Indeed, the paradigm used has elements that appear in the antisaccade task (i.e., a spatial cue followed by a target display), but, as suggested above, participants are unaware of this cue and the instruction to “look away” from the cue cannot be given, nor would it make sense on the large portion of trials on which the cue is presented in the location of the eventual target. In fact, all of the inhibition tasks that have been used in support of mere effort (i.e., the Stroop task, the Sustained Attention to Response Task and the antisaccade task) have an instruction regarding a behavior that is to be prevented. For example, on the Stroop task, participants are told that they are to call out the ink color that the word is printed in rather than reading out the word, meaning that the impulse to read the word must be inhibited. Our current results suggest that this type of instruction may be required in order to see effects that are congruent with all aspects of the mere effort account.
Without an instruction, as evidenced in Experiment 5, participants in both conditions look to the cue on the majority of trials and evaluation does not moderate this behavior. Furthermore, the results from Experiments 2 through 5 indicate that the potency of the spatial cue is dictated by the stimulus conditions in place (e.g., exposure length) and the manipulation of evaluation alone does not impact the rate at which participants orient to the cue. When put into the context of previous antisaccade findings that reveal that undergoing evaluation (e.g., McFall et al., 2009) or stereotype threat (Jamieson & Harkins, 2007) both lead to increased eye movements to the spatial cue, the role that the instruction to “look away” may play in potentiation becomes clearer. On the one hand, as suggested by accounts focused on working memory deficits (e.g., Schmader & Johns, 2003), individuals in the evaluation or threat condition may lack the cognitive resources necessary to inhibit this eye movement to the cue, which leads to an increase in this behavior. On the other hand, increased motivation may lead individuals to be unable to look away, whereas the less motivated conditions are better able to do so, suggesting that the response is potentiated. In both cases, this would suggest that the potency of the cue itself is not varied due to manipulations of evaluation or stereotype threat, but the way that individuals interact with these cues are influenced.

The current work only focused on the first experiment presented in Normand et al. (2014). In order to extend the focus of attention account, in a series of studies, Normand et al. manipulated the relevance of the peripheral cue. Focus of attention suggests that because evaluation leads to a decrease in the amount of information that can be processed, evaluated participants prioritize the most central information for a task and discount the rest. In this case, focus of attention predicts that evaluated participants will pay greater attention to the cues that match the eventual target (i.e., are perceptually relevant), leading to an increase in cueing effects
for these items, and disregard the cues that do not match (i.e., are perceptually irrelevant), leading to a decrease in cueing effects for these items.

In their second experiment, Normand et al. (2014) had participants fixate on a cross for 1000 ms. Cues were then flashed at all four eventual target locations, but three of the cues were black and the one remaining cue was green on half of the trials and red on the other half. Following a 50 ms blank, a target display containing two black Qs, a red O and a green O was presented for 60 ms and then a blank screen was presented for 1440 ms as a response window. Participants were instructed to locate the red O, making the red cue relevant to the task and the green cue irrelevant to the task. Normand and her colleagues found that participants subject to experimenter evaluation showed larger cueing effects than participants who were not evaluated when the cue color matched the target color, but smaller cueing effects when the cue and target color did not match.

Collectively, the effect of these changes is unknown. Most relevant to the current research is whether or not the cue display in this paradigm produces a prepotent response (i.e., do participants orient to the location of the color cue when it is presented simultaneously with black cues at the other three locations). In addition to this, the target display is no longer presented until participants respond, which, paired with stimuli letters that are once again as large as they were in the first baseline, may make eye movements less likely in this design. Another consideration for this design is how the relationship between cue potency and location effects may play a role in their findings. The results from the non-evaluated control condition from Normand et al. (2014) show that manipulating the cue’s color led to different sized cueing effects, suggesting that the two cue displays were not equipotent, replicating earlier work (e.g., Folk & Remington, 1998). Through five experiments we have shown that even small changes in
cue potency can lead to large changes at the level of reaction time and it is possible that the same processes underlie the results of Normand et al.’s second experiment.

The current work has highlighted the importance of understanding task characteristics that may seem like minutiae, when, in actuality, small decisions regarding stimulus conditions can lead to large changes in outcome measures when participants are faced with evaluation. Too often tasks are chosen because they are convenient or because they have been used in the particular line of work before, and, frequently on these tasks, only a single measure is considered (e.g., number of RAT items solved, terminal reaction times), perhaps supplemented by some self-reports. In our work, we have found that one can find what looks to be strong support for a given account when only terminal reaction times are used, but when we look in more detail at what is going on, by using eye-tracking for example, we find that the process that produces these terminal reaction times is completely different from what the account would suggest.

Finally, a substantial body of previous work has found support for the mere effort account, using two social threats, evaluation and stereotype threat, and a range of tasks, both cognitive (anagrams, GRE quantitative problems, horizontal subtraction problems, modular arithmetic, antisaccade task, Stroop, SART, visual attention) and physical (virtual ball bouncing). Mere effort also has the potential to integrate a range of research areas that focus on motivated task performance in addition to social facilitation/social loafing and stereotype threat, (e.g., goal setting, intrinsic motivation/creativity, and achievement goal theory). However, the current work suggests that, even if the mere effort account ultimately does not prove to be up to this task, whatever turns out to be the explanation for the effect of the potential for evaluation on task performance, the effects of the motivation to do well will turn out to be a crucial part of it. Most researchers agree that the potential for evaluation and stereotype threat motivate participants to
do well, but it is also the case that motivation is virtually always accorded a secondary or peripheral role in the explanation that is then offered for the performance effects (e.g., Schmader, Johns, & Forbes, 2008). Our research strongly argues that motivation will play a central role in whatever account ultimately explains the effect of social threats, like the potential for evaluation and stereotype threat, on task performance.
References


Figure 1. Sequence of events on an invalid trial
Figure 2. Reaction times by cue type and evaluation condition for Evaluation Experiment 2.
Figure 3. Percentage of eye movements to uncued locations in Evaluation Experiment 2 by display location
Figure 4. Reaction times by cue type and display location condition for Evaluation Experiment 2.
Figure 5. Reaction times by cue type, display location, and evaluation condition for Evaluation Experiment 2.
Figure 6. Percentage of eye movements split by location and cue type for Baseline 3.
Figure 7. Percentage of uncued eye movements split by evaluation condition and location for Experimenter Evaluation Experiment 3.
Figure 8. Percentage of cued eye movements split by evaluation condition and location for Experimenter Evaluation Experiment 3.
Figure 9. Invalid reaction times split by evaluation condition and eye movement for Experimenter Evaluation Experiment 5.
Figure 10. Valid reaction times split by evaluation condition and eye movement for Experimenter Evaluation Experiment 5.