Probabilistic Adaptation and Voluntary Attention

by

Sean Griffin

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Psychology

Waterloo, Ontario, Canada, 2019

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

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

Abstract

The following experiments considered the general phenomenon of behavioural adaptation in response to statistical regularities—which we refer to as probability learning (PL). In particular, these experiments focused on spatial PL and its relationship with spatial attention.

Evidence suggests that the set of neural mechanisms responsible for spatial PL might intersect with those which mediate the voluntary expression of spatial attention. Furthermore, inductions of spatial PL are typically successful in the absence of explicit awareness on the behalf of participants. These findings raise the question of whether spatial PL inductions can be used to subtly alter voluntary behaviour by altering attentional biases. If this is the case, spatial PL inductions could have a wide range of applications—for example as tools in skill training and marketing.

We investigated a potential cross-task influence of spatial PL on voluntarily expressed patterns of spatial attention. We used a behavioural task based on the Tse Illusion to measure voluntary shifts in spatial attention (Illusion Task; Tse, Caplovitz, & Hsieh, 2006). We used a feature discrimination task (PL Task) derived from Druker and Anderson (2010) to induce spatial PL.

Experiments 1, 2, and 3 combined the Illusion Task with the PL Task in a pre-test/post-test design. Experiment 3's inclusion of eye tracking permitted us to explore mechanistic hypotheses concerning the cross-task translation of behavioural adaptation.

All three experiments revealed that condition-specific variation of the PL Task's spatial probability distribution produced predictable changes in PL Task performance. Experiment 3 supported the utility of eye-tracking as a tool for understanding the processes underlying spatial PL along with the impact of spatial PL on voluntary attention.

We found that we could reliably induce spatially-specific changes in *involuntary* attention. We produced consistent and robust estimates of the impact of such spatial PL on our feature discrimination task—and, found the effect to be largely driven by changes in eye-movement generation and consequently target acquisition. Finally, we discovered that spatial PL did not influence the expression of voluntary attention in a subsequent task.

Acknowledgements

I couldn't have done this without: the curiosity and the love of learning instilled in me by my mother; the lessons in acceptance and in humility from my father; Kaitlyn's love; and, Britt's expertise as a scientist, teacher, and mentor.

I would also like to extend a sincere thank you to Dr. Derek Besner and to Dr. James Danckert for their feedback and guidance during the final stages of this process.

Dedication

"Plato is dear to me but dearer still is truth." — Aristotle

"Essentially, all models are wrong, but some are useful." — George Box

"Probability theory is nothing but common sense reduced to calculation." — Laplace

"Ignorance is preferable to error and he is less remote from the truth who believes nothing than he who believes what is wrong." — Thomas Jefferson

"But a logically consistent 'normative theory' of rational inference, means necessarily ... a Bayesian theory." — E. T. Jaynes

Table of Contents

List of Figures

Chapter 1

General Introduction

1.1 Purpose

We set out to replicate and build on past work on spatial probability learning (PL). We intended to induce spatial PL in one context in order to reliably manipulate the expression of voluntary spatial attention in a different context. Because spatial PL is implicit, it was of interest to determine the extent to which voluntary behaviour was susceptible to the influence of statistical patterns in the environment. Given these objectives, our spatial PL induction served two functions in each experiment: (1) It was a stand-alone opportunity to replicate past work on spatial PL in the context of feature discrimination; and, (2) it served as our intended manipulation of voluntary attention.

To enhance our understanding of the mechanisms underlying spatial PL we used an eye-tracker in our final experiment. Since spatial PL can be understood as a bias in visual spatial attention, our inclusion of eye-tracking was justified by the tight neuroanatomical and behavioural links between visual spatial attention and gaze. Definitions of key terms and a brief review of research which guided our work follows below.

1.1.1 What is Spatial Probability Learning?

Evidence for spatial PL takes the following canonical form. Targets in a visual search or feature discrimination task are presented across multiple trials. Target locations vary over time according to a probability distribution. This statistical structure renders some regions of the display more likely to contain a target on any given trial than other regions. These regions are denoted high probability and low probability, respectively. Participants come to detect targets which appear in high probability regions more rapidly than targets which appear in low probability regions—this is spatial PL.

1.1.2 Prior Work on Spatial PL

Druker and Anderson (2010) found that participants perform simple feature discrimination more rapidly for targets in high, relative to low, probability locations. In all cases, we induced spatial PL in our experiments with a protocol derived from their work which we refer to as the PL Task from now on.

Jabar and Anderson (2017) demonstrated that spatial PL improves detection of targets, but does not enhance the precision of orientation estimates. Jiang, Sha, and Remington (2015) reported a robust and persistent effect of spatial PL on a visual search task. Geng and Behrmann (2005) found that targets in high, relative to low, probability locations were detected more rapidly.

Carreiro, Haddad, and Baldo (2003) varied the spatial probability of targets and found that RT was faster for targets in high probability locations. Their work showed that the spatial PL effect was present independent of whether or not participants had an explicit awareness of the spatial probability structure of the task.

1.1.3 Visual Spatial Attention: Involuntary and Voluntary

In general, attention is understood as a synonym for the nervous system's optimized allocation of limited resources for information processing (Carrasco, 2011). The conventional perspective on attention is that it is not a unitary construct—rather, it is held that there are multiple types of attention (Anderson, 2011). This complication necessitates an identification of which types of attention were relevant to our investigations.

The spatially-specific task demands of our experiments—and the modality of the sensory input involved—entail that we studied visual spatial attention. For the sake of brevity, throughout the document we mean visual spatial attention when we say attention, unless otherwise noted.

We can turn to the notion of exogeneity of attention to introduce a final degree of precision into our construct identification. According to Carrasco's (2011) review of the preceding 25 years of research on visual attention, exogenous visual spatial attention is an "involuntary system that corresponds to an automatic orienting response to a location where sudden stimulation has occurred", whereas endogenous visual spatial attention is a "voluntary system that corresponds to our ability to willfully monitor information at a given location". These definitions make clear that the mutual exclusivity of exogenous and endogenous attention as constructs hinges on whether attention is deployed voluntarily or involuntarily.

1.1.4 Measuring Voluntary Attention

An investigation of the influence of spatial PL on voluntary attention requires a method of measuring voluntary attention. To this end, we used a task centered on an optical illusion with a subjective perceptual experience which has been reported to change as a function of the locus of one's spatial attention (Tse, 2005; Tse et al., 2006)[1] Specifically, while maintaining fixation on the center, if one selects a circle to pay attention to, its brightness appears to change (see Figure 1.1). Our participants viewed this illusion for a fixed duration. Using circle-specific key presses, they reported the timing and location of their spatial attention as they allocated it in a voluntary fashion to one circle at a time. We refer to this operationalization of voluntary attention as the Illusion Task from now on.

1.1.5 Inducing Spatial PL and Manipulating Attention

Our PL Task was a variant of the implementations reported in Druker and Anderson (2010). Participants performed multiple blocked trials of a binary feature discrimination task. On each trial a target appeared on a display and participants reported its colour with a button press. We recorded the reaction time and accuracy of each report. Crucially, we assigned the locations of targets across trials in accordance with a spatial probability distribution. To infer the presence of spatial PL, we analyzed how performance—in terms of reaction time and accuracy—changed over time in relation to the spatial distribution of the targets.

Because target locations are diffused across the display in a nearly continuous fashion—as opposed to a small discrete set of potential locations—our implementation allows us to disambiguate spatial PL from strict spatial repetition priming, by definition (Maljkovic & Nakayama, 1996). Another strength of the Druker and Anderson (2010) implementation we preserved in our work is the inclusion of a post-questionnaire designed to detect whether

[1]To be clear: we do not analyze, nor do we report on, any results regarding subjective experiences of the illusion.

participants were explicitly aware of the spatial probability distribution of the targets. This allowed us to address questions concerning the implicit vs. explicit nature of spatial PL (see Appendix C).

1.1.6 Spatial Attention, Spatial PL, and Gaze

This section provides a brief overview of research findings that accord with our intuition that adaptation caused by the PL Task might influence voluntary attention in the Illusion Task, due to a common neural bottleneck responsible for changes in attention and gaze.

Recent work has demonstrated that both voluntary and involuntary spatial attention are correlated with the direction of brief eye movements known as microsaccades (Meyberg, Sinn, Engbert, & Sommer, 2017; Pastukhov, Vonau, Stonkute, & Braun, 2013).

Corbetta's (1998) review of a variety of neuroimaging and behavioural studies determined that frontoparietal cortical networks responsible for the direction of attention and gaze alike were largely overlapping, potentially identical, and certainly not independent. Such results accord with the premotor theory of attention of Rizzolatti (1983) which claims that the neural correlates of spatial attention are precisely those involved in the planning of physical movements. Interested readers can also refer to Corbetta and Shulman's (2011) more recent work on spatial neglect and attention networks in the brain.

He and Kowler (1989) found that saccade accuracy is influenced by the presence of distractors and that this impact could be mitigated by increasing the predictability of target locations—i.e., manipulating spatial probability.

Kustov and Robinson (1996) investigated the link between spatial attention and eye movement generation in primate subjects. They used electrical stimulation of the superior colliculus to induce eye movements and found that endogenous and exogenous attentional shifts both caused deviations in the induced eye movements. Furthermore they found that attentional modulation of electrically induced eye movements occurred even when behavioural responses required hand rather than eye movements.

Jiang, Sha, and Remington (2015) suggested that spatial PL is at least partly the consequence of changes to gaze patterns.

Schall and Hanes (1993) summarize central findings on express saccades in humans with the claim that saccadic latency reductions follow from, among other factors, increases in target spatial probability.

Taken together with the result from Jabar and Anderson (2017) that spatial PL involves faster detection of targets, but not necessarily improved perception of targets, it seems plausible that the performance changes we call spatial PL are driven by adaptation in the process responsible for eye movement generation.

1.1.7 Influencing Voluntary Attention with Spatial PL

Our experiments were variations on the following thematic protocol. A pre-test performance of the Illusion Task established the baseline dynamics of each participant's voluntary attention. After pre-test, each participant completed the PL Task—our intended induction of spatially-specific biases in attention. Immediately thereafter, a post-test performance of the Illusion Task served to assess each participant's voluntary attention once more.

Spatial PL depends on adaptation in neural mechanisms involved in the generation of eye-movements. These mechanisms may not be sufficient for the expression of voluntary attention; however, given the tight link between attention and gaze, they are likely necessary. The existence of a shared mechanism makes it possible that spatial biases in gaze induced in the PL Task could exert a residual influence on the expression of voluntary attention in the Illusion Task.

1.1.8 Research Questions

Our first research question was whether we could find evidence for the canonical spatial PL effect through our replication attempts. To this end we analyzed the PL Task datasets and expected to find faster reaction times for targets which appeared in probable, relative to improbable, locations.

Our second research question was whether we could manipulate voluntary attention using spatial PL. To address this question, we compared how participants allocated their voluntary attention during the Illusion Task in pre-test and post-test. We hypothesized that this comparison would reveal the emergence of a spatial bias in voluntary attention that matched the spatial bias induced by the intervening PL Task.

Our third research question was whether the use of eye-tracking would corroborate the assertion that spatial PL is driven by changes in gaze. We altered the PL Task implementation so that targets revealed their colour in a gaze-contingent manner. This allowed us to decompose each RT into two components: localization RT—the time to acquire a target from onset; and classification RT—the time to classify the target's colour after localizing it. We predicted that targets which appeared in probable, relative to improbable, locations, should be localized more quickly.

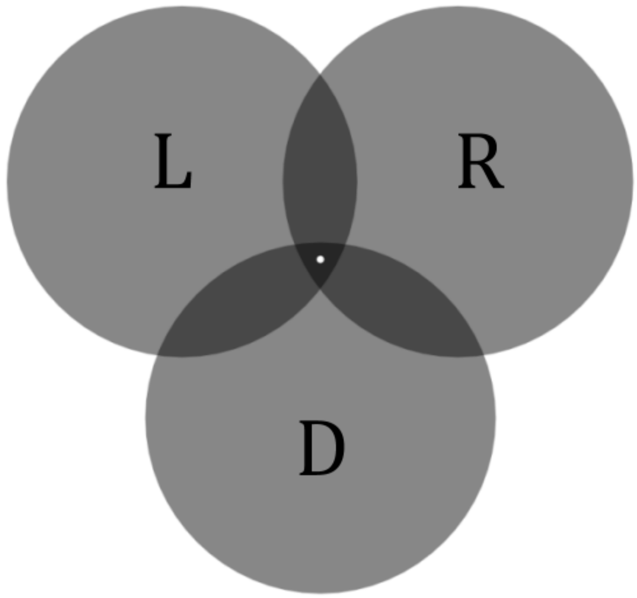

Figure 1.1: The Tse Illusion is depicted to facilitate visualization of the spatial distribution of targets in the PL Task. In this example, we consider participants in the Left condition. Like all participants they began the experiment by performing the Illusion Task—thereby providing us with a pre-test measure of spatial biases in their allocation of voluntary attention. Next, they performed the PL Task. On each trial, they classified the colour of targets which appeared with 70% probability in **L**, 15% probability in **D**, and 15% probability in **R**. After the PL Task, they once again performed the Illusion Task—thereby providing us with a post-test measure of spatial biases in their allocation of voluntary attention. We compared voluntary attention in post-test to pre-test in hopes of finding a translation of the spatially-specific bias for target locations in the PL Task to voluntary attention in the Illusion Task. In this example, we would expect that participants should voluntarily attend to the Left circle more during post-test than during pre-test. We used the PL Task in hopes of biasing their spatial attention to the region of the display that matched the Left circle of the Tse Illusion. We do not analyze nor do we report on any results regarding the subjective experience of the illusion in this work.

Chapter 2

Experiment 1

2.1 Introduction

Experiment 1 was designed to gauge the adequacy of our initial implementations of the Illusion Task and PL Task as experimental adjuncts to the study of voluntary attention and spatial PL. We collected and analyzed data to: (1) Establish the viability of our novel Illusion Task as a behavioural measure of voluntary attention; and, (2) demonstrate that our PL Task was a reliable induction of spatial PL.

Ultimately, Experiment 1 was exploratory in nature; we intended to use its results to guide an eventual refinement of both tasks. Nevertheless, relative to our novel Illusion Task, our PL Task came with broader empirical support (Druker and Anderson, 2010). Thus, it was reasonable to expect that the PL Task results would largely mirror those of its predecessors. Specifically, we predicted that participants would exhibit faster responses for targets which appeared in high-probability (as opposed to low-probability) locations—and, that this effect would not be explained by a speed-accuracy trade-off.

2.2 Method

2.2.1 Participants

Twenty-two undergraduate students from the University of Waterloo provided informed consent prior to participating in this in-lab experiment in exchange for bonus credit in their psychology courses (Median age = 21; 18 right-handed; 5 men). Given the exploratory nature of our study, we based our sample size on the smaller samples used in the prior work we based our tasks on (12 in Experiment 1 of Druker and Anderson (2010) and 16 in Tse (2005)). Self-reports indicated that participant vision was normal (or corrected to normal) in all cases. An institutional ethics review committee approved this experiment. At run-time, each participant was randomly assigned to one of three conditions: Left (n = 7), Right (n = 10), and Lower (n = 5).

2.2.2 Stimuli

From a distance of ∼60 cm, participants viewed all experimental stimuli on a gamma-corrected 19-inch CRT monitor (horizontal refresh rate: 91.12 kHz; vertical refresh rate: 85 Hz).

Illusion Task

The sole stimulus visible during the Illusion Task was Tse's (2006) brightness illusion, centered on a white background. Subtending 0.1 degrees of visual angle, a white dot marked the centroid of the figure and acted as the reference point for participant fixation.

PL Task

Targets were circles with radii equal to 0.5 degrees of visual angle presented on the same white background used for the Illusion Task. Independent of condition or location, on each trial, a target's colour was either blue or purple with equal probability. In contrast, each target's location was determined stochastically in a condition-specific manner (see Figure 1.1). On each trial, targets had a 70% chance of appearing somewhere in the overlap-free regions of the Left, Right, or Lower circles (for participants in the Left, Right, and Lower conditions, respectively). The remaining 30% was divided equally between the other circles.

To be clear, the Tse Illusion stimulus was never visible during the PL Task. Nevertheless, its geometry constrained the locations of PL Task targets, such that the entirety of every target fell exclusively within a region of the monitor corresponding to a unique circle. This degree of geometric continuity between the stimuli of both tasks was implemented in hopes of restricting the spatial specificity of the PL induced by the PL Task, so that a cross-task influence of spatial PL might be observed.

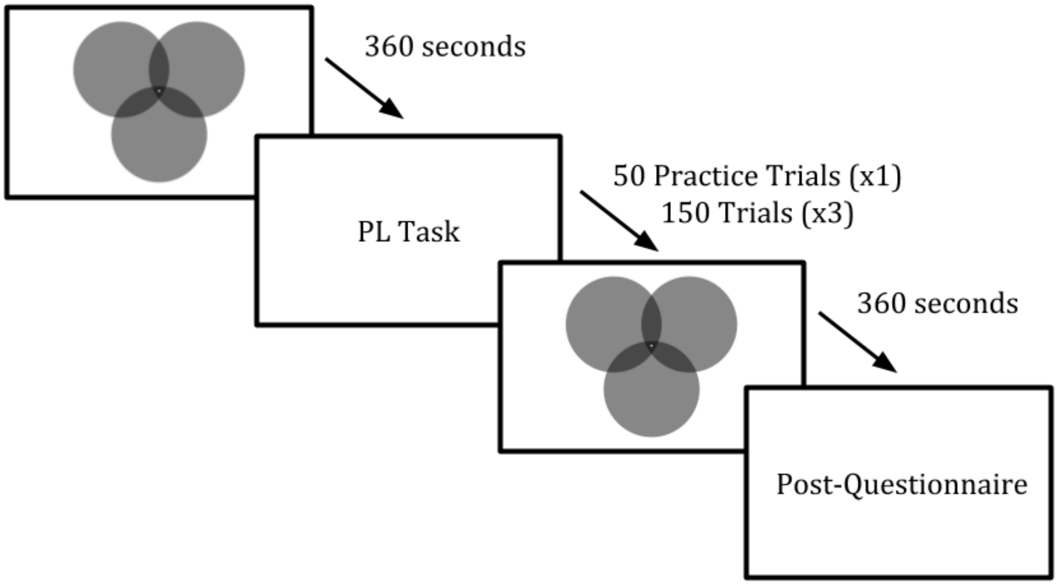

Figure 2.1: Diagram of the protocol for Experiments 1 and 2; see text for details.

2.2.3 Procedure

Throughout the experiment, the arrow keys of a standard computer keyboard captured participant input. A Python script including elements from PsychoPy (Peirce, 2009) randomly assigned participants to conditions, controlled stimulus presentation, registered input, and recorded data. Figure 2.1 depicts the flow of our protocol for Experiments 1 and 2.

Illusion Task: Pre-Test

Participants read the instructions for the pre-test Illusion Task prior to initiating the task with a key press. For the next six minutes, participants maintained fixation on the centroid of the illusion. Participants understood that constant central fixation should be regained as fast as possible if their gaze drifted from fixation for any reason—and, that they were to blink freely in order to reduce eye-strain.

Participants spontaneously allocated their attention to any one of the three circles throughout the six minute viewing period. Whenever they began anew to attend to a

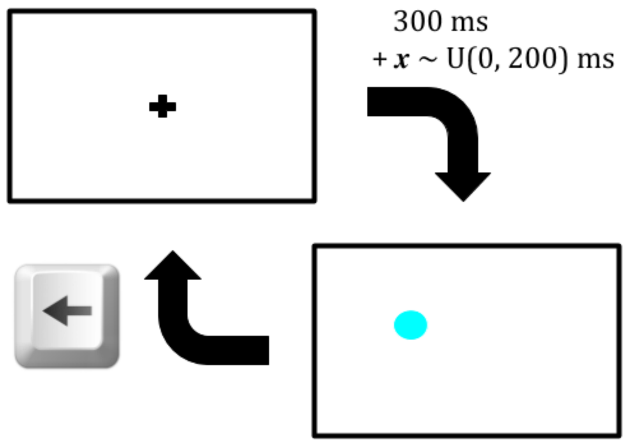

Figure 2.2: Diagram of the PL Task for Experiments 1 and 2; see text for details.

circle, the target (and timing) of their allocation was recorded via the keyboard. The instructions for the PL Task were presented immediately following the end of the pre-test Illusion Task.

PL Task

After reading the instructions, participants initiated the first block of the PL Task with a key press. The appearance of a small black dot in the center of the screen marked the onset of each trial. This fixation dot disappeared after a uniformly random interval of 300 ms to 500 ms. At this point, a target appeared on the screen according to a condition-specific probability distribution (see Figure 1.1 and Figure 2.2).

Participants classified a target's colour with the left and right arrow key. For each participant, this mapping between keys and target colours (e.g., left arrow for blue targets) was determined randomly at the beginning of the experiment, and remained constant for every trial. Receipt of participant input caused the immediate onset of the next trial if any trials remained in the corresponding block. Participants completed a practice block of 50 trials, before completing three blocks of 150 trials. An optional break (maximum duration: 180 seconds) took place immediately after the final trial of all but the final block. Participants could skip these breaks at any time with a key press. Importantly, no break was available between the final block of the PL Task and the onset of the post-test Illusion Task instructions.

Illusion Task: Post-Test

After reading the task instructions, participants initiated the post-test Illusion Task with a key press. Expectations regarding participant performance of the Illusion Task were identical in pre-test and post-test. After six minutes had elapsed, participants were prompted to inform the researcher that they had completed the computer-based portion of the experiment.

Post-Questionnaire

After completing the post-test Illusion Task, each participant responded verbally to a brief questionnaire administered by the researcher. In this way, explicit knowledge of probability effects, with respect to space and targets, was appraised. So too was qualitative feedback regarding subjective experiences of the Tse Illusion collected.

2.3 Results

2.3.1 Analysis Tools

Details on our statistical approach and its justification can be found in Appendix A. All of our modeling was done using R (R Core Team, 2018), specifically, the RStanArm (Stan Development Team, 2018a), and shinystan (Stan Development Team, 2018b) packages.

2.3.2 Data Screening

We retained the data of 19 of the 22 participants for our analyses. Two participants were removed because they misunderstood the task instructions and their experience of the protocol was consequently compromised. In the first case, the participant left the lab room prematurely to alert the experimenter. In the second case, the participant failed to report the onset and target of attentional shifts during the Illusion Task. The final participant was excluded due to unauthorized use of their cell phone during their participation. This left 6, 9, and 4 participants in the Left, Right, and Lower conditions respectively.

2.3.3 Illusion Task

During pre-test, the mean proportion of attentional allocations to the biased region of the illusion was 30% (SEM: 2%)—during post-test, it was 35% (SEM: 3%). This change of 5% is a superficial, pre-inferential indication that the proportion of attention paid to the biased region of the screen increased from pre-test to post-test (see Figure 2.3).

Inference

We fit a series of logistic regression models to address the question of whether this observed change was the result of our PL Task. Model 1 established a baseline goodness-of-fit against which we could compare our more detailed and theoretically motivated models. Its sole predictor was an intercept. To capture within-subject variability and to account for repeated measurements on each participant, we derived Model 2 through the addition of by-participant adjustments to the intercept. Model 2 had marginally better predictive accuracy than Model 1—ΔExpected Log Predictive Density (ELPD): 1.0 (SE: 1.3).

Model 3 encoded our hypothesis. It predicted the probability of attending to the biased region of the illusion using an intercept with by-participant adjustments and a predictor coding for test-time (pre-test or post-test). The improvement in predictive accuracy from Model 2 was minimal—ΔELPD: 0.1 (SE: 1.1).

Of note, the posterior marginal distributions for both the intercept and the effect of test-time had considerable dispersion. This indicates that the data are consistent with a large range of estimates for these parameters.

2.3.4 PL Task

If our intended manipulation of voluntary attention, the PL Task, failed, this could explain our inability to find evidence corroborating our hypothesis that spatial PL would influence voluntary attention. The following analyses spoke to this question while simultaneously addressing our first research question: could we replicate the canonical spatial probability effect?

All non-practice trials were analyzed to compute the accuracy results. Collapsing across conditions, targets with high spatial probability and targets with low spatial probability were classified correctly on a similar proportion of trials: 93.4% and 92.9%, respectively.

Following Druker & Anderson (2010), trials eligible for RT analyses fell in the range [151 ms, 999 ms]; 96.8% of trials were analyzed. Descriptive statistics were computed only on correct trials to facilitate direct comparisons with past work. During modeling, we used all of the eligible trials and explicitly modeled accuracy along with the potential for a spatial probability-by-accuracy interaction. We make the same choices in all of our analysis sections unless otherwise noted.

Collapsing across conditions, targets with high spatial probability, relative to targets with low spatial probability, were classified more quickly— mean median RT_{HIGH}: 443 ms (SD: 48); mean median RT_{LOW}: 458 ms (SD: 48); ΔRT: 15 ms. Figure 2.4 depicts how aggregate RT distributions varied as a function of target location probability.

Inference

As expected our RT distributions were right-skewed. Thus, we model RT on eligible trials with the assumption that it is log-normally distributed (Baayen, 2008).

We began with a model (PL Model 1), which predicted RT using an intercept with by-participant adjustments, the eccentricity of the target relative to screen center, and a factor coding for accuracy. The next model (PL Model 2), was derived from PL Model 1 by including a factor coding for the spatial probability of the target. Comparing the models, we found that including a factor coding for the spatial probability of the target led to an improvement in predictive accuracy—ΔELPD: 10.3 (SE: 4.7).

We fit a final model, (PL Model 3), by adding a target probability by accuracy interaction to PL Model 2. PL Model 2 was preferred according to our model comparison—ΔELPD: -0.8 (SE: 0.7). Inspecting PL Model 3, we found that the 95% central credible interval for the interaction term contained zero. This is evidence that the differences in RT we observed were not the result of a speed vs. accuracy trade-off.

For posterior inferences regarding parameter values, we draw on the best fitting model, PL Model 2, which encoded our a priori hypothesis concerning spatial probability. The median of PL Model 2's posterior marginal distribution over the intercept was 6.03—which translates to 415.72 ms.

The 95% central credible interval for PL Model 2's estimate of the magnitude of the effect of eccentricity did include zero. However, 99.9% of the posterior marginal distribution fell above zero. The median of this posterior marginal distribution was 0.006—which translates to 1.01 x 2 = 2.02 ms.

The 95% central credible interval for PL Model 2's estimate of the magnitude of the effect of accuracy did not include zero. The median of this posterior marginal distribution was 0.06—which translates to 1.06 x 2 = 2.12 ms.

The 95% central credible interval for PL Model 2's estimate of the magnitude of the effect of spatial probability did not include zero. The median of this posterior marginal distribution was -0.01—which translates to 0.99 x 2 = 1.98 ms.

2.4 Discussion

2.4.1 Illusion Task

We found that taking into account within-subject variability in voluntary attention produced a model with marginally improved explanatory merit, relative to our baseline model. However, including information pertaining to test-time barely enhanced our expected out-of-sample predictive accuracy. The median posterior estimate for the intercept was consistent with the expectation that about 1/3rd of attentional allocations would be directed to the biased region of the illusion. However, there was considerable disperion in the posterior marginal distribution, implying that this point estimate is a poor surrogate for the distribution itself. Given the trivial improvement in expected out of sample predictive accuracy we gained from including test-time—and, given that the 95% central credible interval for the posterior estimate of the effect of test-time contained zero, our analyses point to a lack of evidence that the PL Task influenced voluntary attention. Ultimately, the relatively large dispersion of our posterior marginal distributions tells us that there is considerable uncertainty in our conclusions.

2.4.2 PL Task

Our model comparisons showed that taking within-subject differences in baseline responding into account was useful for predicting RT. Moreover, we found that knowing whether or not a participant was correct, and knowing the eccentricity of the target from central fixation, also improved our model. Participants were fastest when targets fell closer to the center of the display in high probability regions. Conversely they were slowest when targets fell further from the center of the display in low probability regions. Correct responses were faster than incorrect responses and this pattern held in regions of both high and low probability regions. We distributed targets such that eccentricity did not systematically covary with spatial probability, ruling out that eccentricity could explain the spatial PL effect. We distributed targets in a diffuse manner across approximately continuous regions of the display rather than a small discrete set of locations, ruling out that strict repetition priming could account for the effect. RT and accuracy was based on feature discrimination; target colour was equally likely to be either colour on each trial regardless of spatial probability. This choice rules out response priming as the source of the spatial PL effect, since responses had nothing to do with target location. Statistically, the crucial test of our hypothesis concerning spatial PL implied that a spatial probability effect was present in the data, even after taking eccentricity and accuracy into account.

The median of the posterior marginal distribution over the magnitude of the effect was 1.98 ms. Given that this marginal posterior distribution did not contain zero, we have evidence for the existence of the effect. Comparatively speaking, the difference is similar in magnitude to the independent impacts of both eccentricity and accuracy.[1] Because we did not find evidence for an accuracy-by-probability interaction, we are confident that the effect, although small, is robust, and is not an artifact of a speed-accuracy trade-off in participant responses.

In terms of our mean median RTs, the reduction in RTs to high probability targets we found—15 ms—is similar to the value reported in work we based our protocol on ∼14 ms (Druker & Anderson, 2010)—although it is smaller than that reported elsewhere—∼30 ms (Jabar & Anderson, 2017). [2]

[1]Interpretation of the magnitude of these effects is aided by considering that a typical latency for a saccade to be initiated to a target following its onset is 200 ms (Carpenter, 1988). Express saccades—with latencies as low as 80 ms to 120 ms—have been shown to be the result of anticipation due to constant temporal offsets of fixation stimuli—and our protocol was designed to mitigate the influence of this 'gap effect' (Schall & Hanes, 1993). More relevant to our protocol then are the findings of Kingstone and Klein (1993)—which place modal minimum saccadic latencies for non-anticipatory saccades around 120 ms to 150 ms. Even so, such latencies represent best-case responses to bright targets that occur in constant locations. In contrast, our protocol was not as conducive to eliciting such low latencies; beyond varying stimulus onset intervals following removal of the fixation marker, we also varied location. With the foregoing in mind, even using the conservative lower bound of 120 ms implies that an independent effect of spatial PL of 2 ms represents a 2 ms / 80 ms = 2.5% reduction in the expected overall saccade latency for high relative to low probability targets. This heuristic interpretation is meant to contextualize a meaningful difference in gaze behaviour which might otherwise appear trivial if the normative time scale familiar from daily life is used as a frame of reference.

[2]We note that this group level difference shouldn't be expected to share the exact magnitude of the

Our PL Task analyses imply the lack of evidence in the Illusion Task dataset that we influenced voluntary attention as intended was not due to a failure of the PL Task to induce spatially-specific biases in attention. Therefore, to reduce our uncertainty in the inferences drawn on the basis of the Illusion Task analyses, we decided to pursue the following next steps. First, we would enhance our confidence in our measure of attentional allocation—i.e., polish the implementation of the Illusion Task. This should eliminate much of the uncertainty in the data pertaining to issues with task compliance, which might obscure the existence of a true effect. Second, we'd simply collect more data. In order to make causal claims pertaining to any potential cross-task translation of the spatial probability induction, we also decided to switch to a mixed factorial design with a control condition.

independent effect of spatial PL on trial-by-trial RT; these are two distinct values. This ΔRT is a difference in second-order summary statistics which reflects many sources of variability and succinctly conveys differences between distributions. That said, we expect the difference largely follows from the cumulative impact of spatial PL across many trials; and this is reasonable given their relative magnitudes. This comparison is included as a sanity check and in the spirit of replication, given that past work did report such a comparison, but did not conduct nor report on the same modeling efforts that we did.

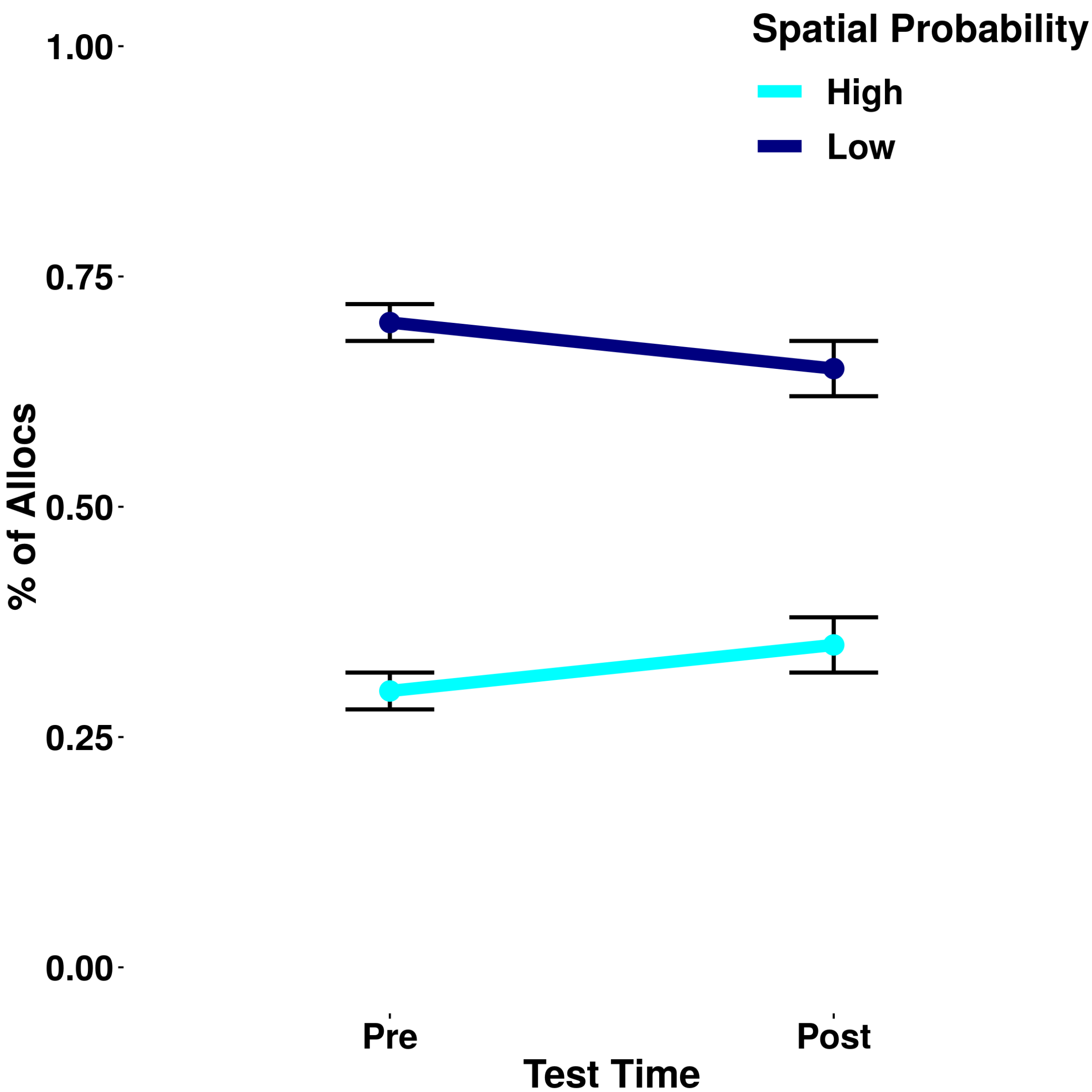

Figure 2.3: Mean (+/- SEM) proportions of attentional allocations to high and low probability regions of the Tse Illusion are depicted as a function of test-time.

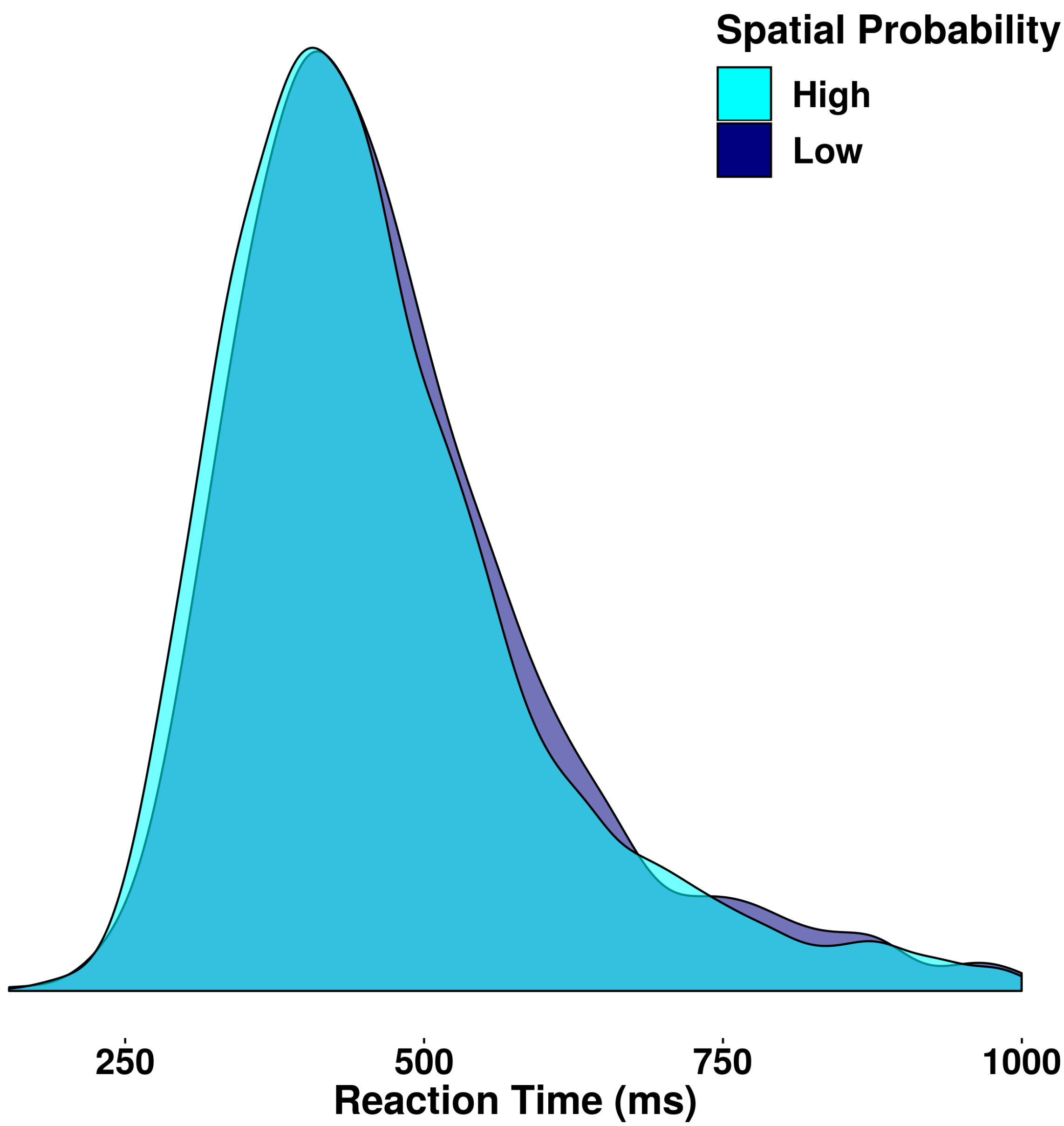

Figure 2.4: RT distributions are depicted as a function of target location probability. In general, targets in probable as opposed to improbable locations were responded to more rapidly.

Chapter 3

Experiment 2

3.1 Introduction

Although we were confident that we could use the PL Task to systematically induce spatial PL, uncertainty remained in our inferences about the Illusion Task. Our misgivings about task compliance were addressed with subtle changes to the instructions for the Illusion Task (see Appendix E). We hoped to prevent the recurrence of any of the misunderstandings that caused us to lose three participants in Experiment 1.

We also chose to enhance our power by studying a larger sample split into two (rather than three) experimental conditions, and justified this decision with recourse to the following considerations. First, this decision was made for efficiency's sake: we wanted to submit the spatial PL effect to a more compelling test while simultaneously investigating the viability of the Illusion Task. Second, with respect to the Illusion Task, we had no a priori reason to believe that this new experimental design would diminish our ability to evaluate the viability of our operationalization. Third, in agreement with past work, Experiment 1's preliminary replication of spatial PL induction had held for participants in all three conditions. This outcome implied that we could choose a single region of the screen instead of using all three without undermining our confidence that the manipulation would work as intended.

Experiment 2's new conditions—which we refer to as the Biased condition and the Uniform condition—were procedurally identical to one another—except for the spatial distributions of their PL Task targets. Targets in Experiment 2's Biased condition appeared according to the distribution used in the Lower condition of Experiment 1 (see 2.2.2). In contrast, targets in the Uniform condition were distributed across all eligible locations with equal probability—in hopes of inducing spatial PL which was not tied to only one region of space. Following our protocol changes our behavioural predictions remained the same. We expected to replicate the spatial PL effect, and to detect an influence of spatial PL on voluntary attention. Specifically, we predicted that participants in the Biased condition would show a bias for allocating attention to the Lower circle in post-test compared to pre-test and that participants in the Uniform condition would not show this attentional bias to the Lower circle.

3.2 Methods

3.2.1 Participants

A new sample of thirty-five right-handed undergraduate students from the University of Waterloo provided informed consent before participating in this in-lab experiment in exchange for bonus credit in their psychology courses (Median age = 19; 8 men). Self-reports indicated that participant vision was normal (or corrected to normal) in all cases. A Python script including elements of PsychoPy (Peirce, 2009) randomly assigned participants to either the Biased condition (n = 19) or the Uniform condition (n = 16).

3.2.2 Stimuli

The monitor, viewing distance, and display settings used in Experiment 1 were also used in Experiment 2 (see 2.2.2). As was the case in Experiment 1, in Experiment 2, conditions were defined by their PL Task spatial probability distributions. Specifically, targets classified by participants in the Biased condition were distributed according to the same structure which defined the Lower condition in Experiment 1. In contrast, every potential location in the sample space was equiprobable for targets classified by participants in the Uniform condition.

3.2.3 Procedure

Participants in Experiment 2 read a revised version of Experiment 1's Illusion Task instructions. The expectation that participants were to generate and report a series of voluntary attentional allocations to different regions of the Illusion was emphasized. The subsidiary requirement that participants maintain fixation on the center of the Tse Illusion at all times was maintained but de-emphasized relative to the instructions about generating the time series data. Experiment 2 was otherwise procedurally identical in all respects to Experiment 1 (2.2.3).

3.3 Results

3.3.1 Data Screening

One participant from each condition failed to complete the protocol in the expected manner, leaving thirty-three out of thirty-five participants for our analyses.

3.3.2 Illusion Task

In the Biased condition: during pre-test, the mean proportion of attentional allocations to the Lower region of the illusion was 38% (SEM: 3%) at pre-test and 38% (SEM: 4%) at post-test. Therefore, we did not find a change in the proportion of attention paid to the Lower region of the screen from pre-test to post-test among Biased condition participants.

In the Uniform condition: during pre-test, the mean proportion of attentional allocations to the Lower region of the illusion was 32% (SEM: 3%) at pre-test and 33% (SEM: 4%) at post-test. This change of 1% is a superficial, pre-inferential indication that, among participants in the Uniform condition, the proportion of attention paid to the Lower region of the screen increased from pre-test to post-test.

Inference

We used logistic regression models to predict the probability of allocating attention to the Lower region of the illusion. To serve as a baseline against which we could compare our full model, we began our analyses by fitting Model 1—which had an intercept as its only predictor. Next, we derived Model 2 from Model 1 through the addition of by-participant adjustments to the intercept. Model 2 had superior predictive accuracy relative to Model 1—ΔELPD: 36.6 (SE: 19.0).

Having captured this within-subject variability with Model 2, we then tested the model encoding our hypothesis, Model 3. Model 3 predicted the probability of attending to the biased region of the illusion using an intercept with by-participant adjustments, a predictor coding for test-time (pre-test vs. post-test), a predictor coding for condition (Uniform vs. Biased), and a test-time by condition interaction. Model 3 represents our hypothesis that only participants in the Biased condition should pay more attention to the Lower region of the illusion in post-test, relative to pre-test, since they receive a spatially-specific bias in the PL Task for the Lower region and Uniform condition participants do not.

Model 3's predictive accuracy was superior to the naive Model 1, but Model 2 was better still—ΔELPD: -3.3 (SE: 1.1).

For posterior inferences about relevant parameters, we draw on Model 3 to explore why these factors failed to boost our expected out of sample predictive accuracy in the way we predicted.

The 95% central credible interval for Model 3's estimate of the intercept did not include zero. The median of this posterior marginal distribution was -0.57—which translates to a baseline propensity for allocating attention to the biased region of 36.1%.

The 95% central credible interval for Model 3's estimate of the magnitude of the effect of test-time, condition, and the test-time by condition interaction all included zero. The medians and the dispersion of the marginal posterior distributions over the effect of test-time and the interaction term imply that we can be confident that these factors offered little in the way of explanatory power to Model 3. The marginal posterior distribution over the value of the effect of condition was shifted further from zero and relatively more dispersed. Thus, Model 3 expressed the baseline between-condition difference in propensity to allocate attention to the Lower region of the illusion but did not represent this difference as significant, which matches our intuition given the standard errors reported above.

3.3.3 PL Task

Accuracy - Descriptive Statistics

All non-practice trials were analyzed to compute the accuracy results. In the Biased condition: 94.6% of targets in the Lower region were classified correctly and 95.2% of targets elsewhere were classified correctly. In the Uniform condition: 96.1% of targets in the Lower region were classified correctly and 94.0% of targets elsewhere were classified correctly.

Reaction Time - Descriptive Statistics

Following Druker & Anderson (2010), trials eligible for RT analyses fell in the range [151 ms, 999 ms]; 2.1% of trials were trimmed. Descriptive statistics for reaction time by-spatial probability and by-condition were computed on correct trials. In the Biased condition: targets in the Lower circle, relative to targets in other locations, were classified more quickly—mean median RT_{LOWER}: 444 ms (SD: 50); mean median $\mathrm{RT}_{NOTLOWER}$: 458 ms (SD: 60); ΔRT: 14 ms. In the Uniform condition: targets in the Lower circle, relative to targets in other locations, were classified less quickly—mean median RT_{LOWER}: 426 ms (SD: 53); mean median $\mathrm{RT}_{NOTLOWER}$: 421 ms (SD: 55); ΔRT: -5 ms. Figure 3.2 depicts how aggregate RT distributions varied as a function of condition and target location.

Inference

As before, we model RT on eligible trials with the assumption that it is log-normally distributed (Baayen, 2008).

PL Model 1 predicted RT using an intercept and PL Model 2 predicted RT using an intercept with by-participant adjustments. The inclusion of by-participant adjustments was justified—ΔELPD: 1646.5 (SE: 53.8).

PL Model 3 predicted RT using an intercept with by-participant adjustments, the eccentricity of the target from central fixation, a factor coding for accuracy, and a factor coding for condition. Compared to PL Model 2, including these factors led to a substantial improvement in expected out-of-sample predictive accuracy—ΔELPD: 114.6 (SE: 16.6).

PL Model 4 encoded our hypothesis. It was an extension of PL Model 3 and featured a factor coding for spatial probability. Comparing against PL Model 3, we found that information pertaining to spatial probability enhanced the explanatory power of PL Model 4—ΔELPD: 6.7 (SE: 3.8).

Our descriptive statistics suggested there might have been a spatial probability by condition interaction, with Biased participants responding more rapidly for targets in the Lower region, and Uniform participants responding more rapidly to targets not in the Lower region. We therefore fit PL Model 5, an extension of PL Model 4 including a condition by probability interaction term. Comparing these models, we found evidence for the interaction—ΔELPD: 16.7 (SE: 5.8).

Our analyses of the Experiment 1 PL Task implied that we had not found evidence for a probability by accuracy interaction. We submitted the Experiment 2 PL Task dataset to the same test, deriving PL Model 6 from PL Model 5 by adding a probability by accuracy interaction term. Including this term was hardly justified—ΔELPD: 0.8 (SE: 2.0). In fact according to PL Model 6 the 95% central credible interval for the probability by accuracy interaction term contained zero. As before this signifies a lack of evidence that the spatial PL effect is the result of a speed vs. accuracy tradeoff.

For posterior inferences regarding parameter values we draw on PL Model 5, which extended our a priori hypothesis concerning spatial probability with a probability by condition interaction term.

The median of the posterior marginal distribution over the intercept was 6.02—which translates to a baseline RT of 411.58 ms.

The 95% central credible interval for the magnitude of the effect of accuracy did not include zero. The median of this posterior marginal distribution was 0.065—which translates to 1.07 x 2 = 2.13 ms.

The 95% central credible interval for the magnitude of the effect of target eccentricity included zero. However, 89% of the interval was greater than zero. The median of this posterior marginal distribution was 0.0016—which translates to 1.00 x 2 = 2.00 ms.

The 95% central credible interval for the magnitude of the effect of condition included zero. The median of this posterior marginal distribution was 0.032—which translates to 1.03 x 2 = 2.06 ms.

The 95% central credible interval for the magnitude of the effect of spatial probability did not include zero. The median of this posterior marginal distribution was 0.007—which translates to 1.01 x 2 = 2.02 ms.

The 95% central credible interval for the magnitude of the condition by spatial probability interaction did not include zero. The median of this posterior marginal distribution was -0.012—which translates to 0.99 x 2 = 1.98 ms.

3.4 Discussion

3.4.1 Illusion Task

Our analyses produced some evidence that contradicted our hypothesis that Biased condition participants would come to allocate more attention to the Lower circle following the PL Task and that Uniform condition participants would not. This outcome agrees with the results of Experiment 1, and came with considerably less uncertainty. At this point, we might conclude that simply taking individual differences in attentional preferences into account is sufficient to explain much of the variability in our participants' behaviour.

Yet, despite our improved instructions, there were still at least two participants who failed to understand the task instructions. Moreover, the relative predictive accuracy of our full baseline model compared to our theoretical model was not so compelling as to convince us that we had conclusively falsified our hypothesis. We also noted that the manner in which we distributed the spatial locations of the targets in the PL Task—which was supposed to be our manipulation of voluntary attention—was variable between-subjects and within-conditions. Finally, we had speculated that the spatial PL induction could persist and influence voluntary attention due to shared underlying neural correlates related to eye-movements. Since we did not enforce fixation and did not supervise our participants it is possible that our intended manipulation wasn't tested in full by Experiment 2. These reflections left us with the impression that it would be premature to conclude that spatial PL did not influence voluntary attention, before testing our hypothesis with an even more rigorous experiment which would include: (1) A built-in task tutorial to make certain that participants understood the Illusion Task; (2) a stronger and standardized manipulation of voluntary attention; and, (3) a larger sample.

3.4.2 PL Task

Our analyses once again supported the claim that our participants exhibited the canonical spatial PL effect. This finding agreed with the results of Experiment 1. In addition we learned that there was evidence for the condition by spatial probability interaction implied by our descriptive statistics. As predicted participants responded more rapidly to targets in the Lower region—but only in the Biased condition. In the Uniform condition, we saw the opposite trend, although it was not as pronounced. We tentatively interpret this result by considering that there are neuroanatomical distinctions in the generation of horizontal vs. vertical eye movements (Blumenfeld, 2002). We add that longer term biases in gaze generation in the horizontal direction (perhaps due to reading) might be relevant. Speculation aside, the predicted change in Biased condition participant performance takes on additional meaning in light of this apparent bias in favour of the non-Lower regions of the screen.

It was reassuring to find such close agreement between our posterior estimates for the effects we tested in Experiment 1 and our posterior estimates for the same effects in Experiment 2. According to the comparison of mean median RT, we also found an effect with nearly the same magnitude as in Experiment 1—again, similar to that reported in the paper that inspired our protocol. At this point, we felt it was worthwhile to alter our protocol to maximize our confidence in our results regarding not only the effect's existence, but also its magnitude.

First, we standardized the sequence of targets participants classified in lieu of randomly generating them on a trial-by-trial basis. This means that for the final experiment, participants in each condition received a condition-specific sequence of PL Task stimuli. In this way, we still used sequences that met our condition-specific probability constraints. However, the standardization meant less varibility between participants within each condition—removing the potential for variability in post-test allocations being attributable to variability in PL Task sequences.

Second, we integrated eye-tracking into our experiment. Using eye-tracking, we would be able to account for heretofore overlooked sources of variability in the dataset. For example, we would be able to enforce central fixation on all trials of the task. As mentioned above, such a change would address our concern about the strength of the PL Task qua manipulation of voluntary attention. But with respect to the spatial PL effect, enforcing central fixation would also ensure that our participants would perform the task in the same way. In particular, it would rule out complications arising from strategic or anticipatory deviations of gaze prior to target onset—which, unaccounted for, might obfuscate our results. We might also filter out trials contaminated by blinks—a source of variability that is relevant given the magnitude of the effect we are studying.

A final benefit of using eye-tracking is that it offers us the opportunity to gain mechanistic insight into spatial PL. How much of the reliable change in feature discrimination could we attribute to spatially-specific changes in localization? What about spatially-specific changes in perceptual processing? By changing the PL Task to involve a gaze-contingent revelation of target colour we set out to pursue these additional questions.

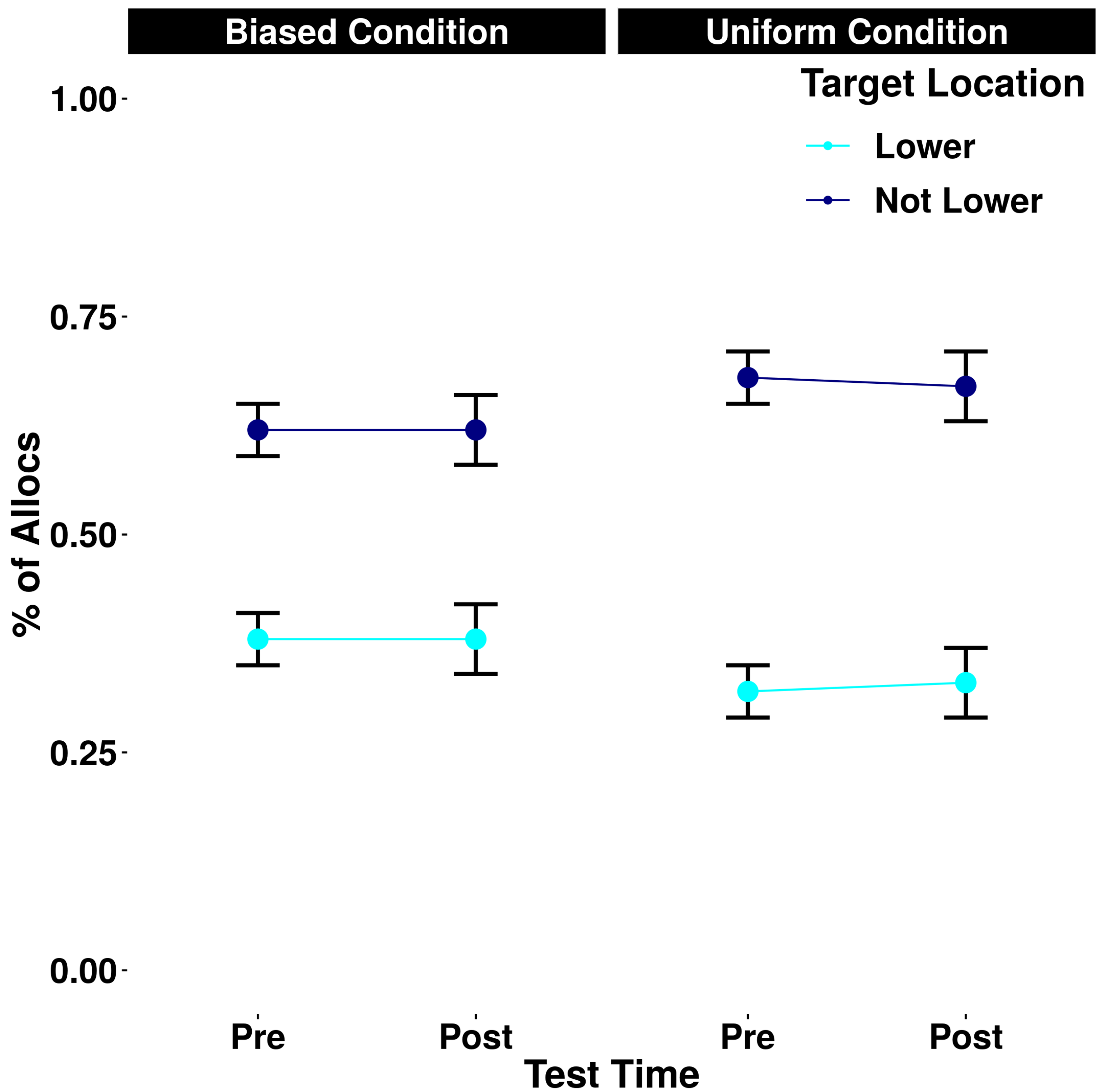

Figure 3.1: Mean (+/- SEM) proportions of attentional allocations to high (e.g., Lower) and low probability regions of the Tse Illusion are depicted as a function of test-time and condition.

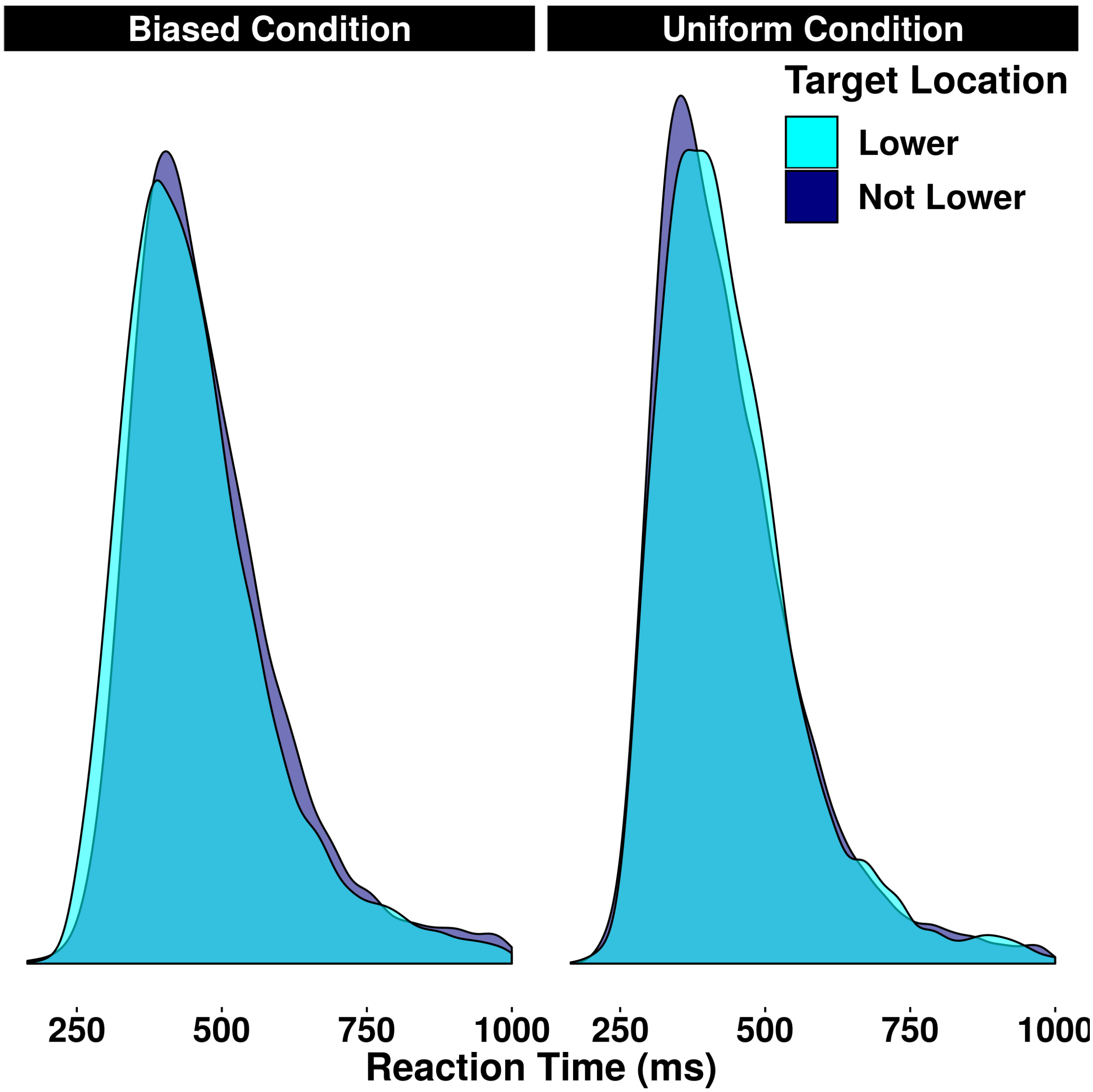

Figure 3.2: RT distributions are depicted as a function of condition and target location. As predicted, only participants in the Biased Condition exhibited comparatively faster RTs for targets in the Lower region.

Chapter 4

Experiment 3

4.1 Introduction

Experiment 3 was designed to replicate the spatial PL effect and to increase the precision of our estimate of its magnitude. The implementation changes we made to the PL Task served this purpose while simultaneously addressing residual uncertainty we had concerning our Illusion Task results. Our prediction that we could manipulate voluntary attention by manipulating involuntary attention remained the same owing to our concerns about participant task compliance in Experiments 1 and 2—and about the strength and consistency of our past manipulations of voluntary attention.

The eye-tracker provided three principal benefits. First, it gave us the ability to exert more control over task compliance by enforcing central fixation at the start of every trial of the PL Task and by enforcing saccade generation on every trial of the PL Task. Second, it gave us the opportunity to filter our data to increase the signal to noise ratio in our favour. Third, it allowed us to investigate a decomposition of RT into two components: localization—the time from target onset until fixation; and classification—the time from fixation until response registration.

Accuracy in our PL Task depends on colour perception. Due to the distribution of cones in the retina, the rate at which participants gather evidence must be a function of the extent to which they foveate targets—and how quickly they do so, following target onset.

Since accuracy requires foveation the retinotopic location and extent of activation in response to every target should be largely the same on each trial regardless of the location of the targets. However the accuracy of saccades is typically 10% of their amplitude—and, based on the accuracy scores we've collected so far, the difficulty of our task is low. Taken together we face the possibility that there could be some spatially-specific perceptual adaptation that takes place during our task. At the same time, such spatially-specific perceptual adaptation could fail to drive a measurable difference in RT given the task difficulty.

So far we've found a difference in overall RT as a function of spatial PL but not a difference in accuracy as a function of spatial PL. This makes it more plausible that changes in RT are driven by changes in how soon a participant is able to foveate a target following its onset. Past work on spatial PL agrees that it involves faster target detection for targets in more probable locations. We therefore predicted that spatial PL should be associated with spatially-specific changes in the localization component of RT.

It is less straightforward to make predictions about the classification component as we defined it. In reality it must be further decomposed into a classification component and a motor response component. Given that we did not have a principled way to gauge this decomposition, it would be difficult to speculate about it without assuming motor responses are constant, which is unreasonable. Furthermore there is less consensus about spatial PL and concomitant changes in perceptual processing—due in part to inter-experimental differences in task demands—i.e., colour discrimination, orientation judgments, and relative contrast judgments (Jabar & Anderson, 2017). We therefore did not make direct predictions about the classification component of RT. Ultimately, the eye-tracker increased the rigour of our experiment, permitted more precise estimation of effect magnitudes, and allowed us to extend our work on spatial PL past a strict replication of the effect's existence.

4.2 Methods

4.2.1 Participants

A new sample of eighty participants provided informed consent prior to taking part in Experiment 3 (Median age = 20; 27 men). Self-reports indicated that participant vision was normal (or corrected to normal) in all cases. Sixteen participants failed to complete the protocol due to incompatibility with the eye-tracker. By incompatibility we mean that calibration was either impossible, or of such low quality that the gaze-contingent nature of the protocol was consistently defective (Biased $n = 7$; Uniform $n = 9$). This left sixty-four participants who were able to complete the protocol in full, with thirty-two in the Biased condition and thirty-two in the Uniform condition.

4.2.2 Stimuli

Participants viewed the display from the same distance of 60 cm as in Experiment 2; however, in Experiment 3, they did so with their heads in a chinrest.

Illusion Task

To reduce boredom and eye-strain we reduced the viewing time of the illusion to 180 seconds from 360 seconds, in both pre-test and post-test.

PL Task

Target colour was now black at onset. When a participant's gaze came within 0.25 degrees of visual angle of a target's center, its colour changed from black to the same blue or purple used in Experiments 1 and 2. Instead of randomly generating the locations and colours of targets on a trial-by-trial basis, we generated a sequence of stimuli for each condition in advance. The same condition-specific sequence was shown to each participant in each condition.

4.2.3 Procedure

Experiments 2 and 3 were procedurally identical except for the following key changes related to our integration of eye-tracking with the protocol. Figure 4.1 depicts the flow of our protocol for Experiment 3.

We determined participant eye dominance at the outset of the experiment and tracked each participant's dominant eye with a SR Research EyeLink 1000 set at a 500 Hz sampling rate.

We added a tutorial to the start of the protocol. The tutorial familiarized participants with the process of calibrating the eye-tracker—which was always done using the eye-tracker's default nine point calibration routine. The importance of keeping their heads still in the chinrest and of alerting the experimenter if at any point they detected a reduction in the responsiveness of the gaze-contingent features of the experiment was underscored.

We used the tutorial to ensure that participants understood the task instructions for both the Illusion Task and the PL Task. For both tasks, starting the Illusion Task, the experimenter read the instructions to the participant in stages. After each stage, participants

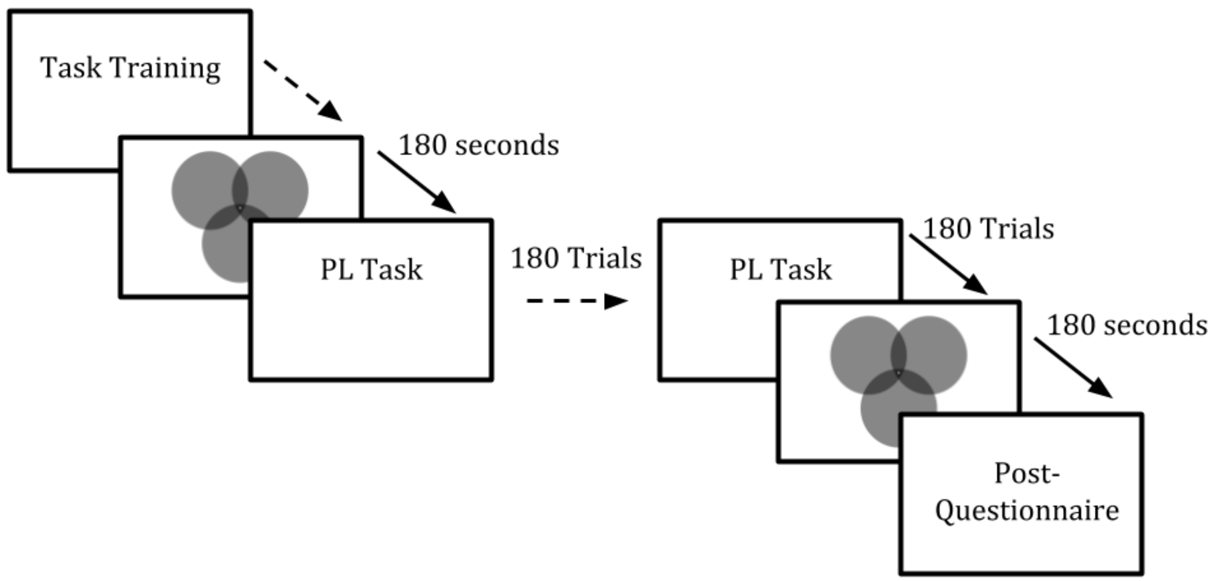

Figure 4.1: Diagram of the protocol for Experiment 3; see text for details. Dotted arrows indicate transitions that involved mandatory breaks and calibration of the eye-tracker.

were made to paraphrase the task expectations in order to demonstrate comprehension. If participants failed to articulate the task demands, the instructions were re-read, and comprehension was once again probed, until participant understanding was ascertained. After articulating their understanding, participants spent a few moments practicing the respective task under the supervision of the experimenter to ensure the understanding was satisfactory in practice. None of the participants failed to demonstrate satisfactory understanding of the task demands of either task.

Participants completed the task tutorial and then took a break while the experimenter explained the structure of the experiment to them. They then recalibrated and completed the Illusion Task pre-test followed by an optional break of up to three minutes. If necessary participants recalibrated and then completed the first block of the PL Task which contained 180 trials. Participants then took a mandatory three minute break before recalibrating and completing a second block of 180 trials of the PL Task. Immediately after the final trial of the PL Task the Illusion Task post-test began. Participants removed their heads from the chinrest before completing a brief post-questionnaire with the experimenter at the end of

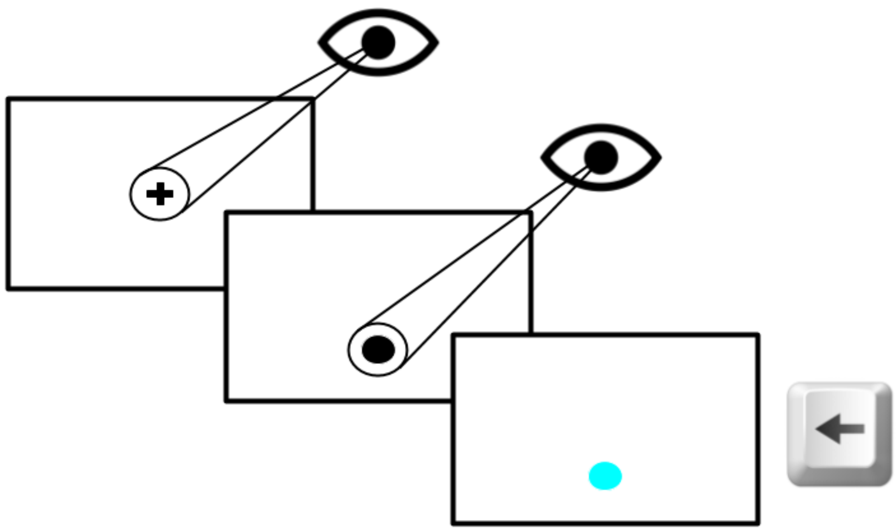

Figure 4.2: Diagram of the PL Task for Experiment 3, depicting the gaze-contingent aspects; see text for details.

post-test.

At any point during the experiment, participants were free to take a break from the chinrest. These impromptu breaks were recorded and taken into account during data analysis.

Participants were instructed to interrupt the PL Task whenever they experienced any latency in the gaze-contingent aspects of the PL Task. Experiencing such latency was an indicator that the quality of the eye-tracking calibration was no longer adequate, largely due to inadvertent head movement in the chinrest. Recalibrations were recorded and taken into account during data analysis.

PL Task trial onset was made gaze-contingent: participant gaze had to be detected within 0.5 degrees of visual angle of central fixation for forty consecutive samples before target onset was triggered. PL Task target colour was revealed in a gaze-contingent manner. Following target onset, participants had to fixate targets to reveal their colour prior to classifying it. Thus latency-related recalibrations were prompted either when participants could not get a trial to start despite maintaining their gaze at central fixation, or when targets stayed black despite participant fixation of the target. Figure 4.2 visualizes the gaze-contingent nature of Experiment 3's PL Task.

4.3 Results

4.3.1 Data Screening

We retained the data of thirty-two participants in each condition for our analyses; exclusions were due strictly to calibration issues.

4.3.2 Illusion Task

In the Biased condition: during pre-test, the mean proportion of attentional allocations to the Lower region of the illusion was 40% (SEM: 3%) and 39% (SEM: 2%) during post-test. This change of 1% is a superficial, pre-inferential indication that, among participants in the Biased condition, the proportion of attention paid to the Lower region of the screen decreased from pre-test to post-test.

In the Uniform condition: during pre-test, the mean proportion of attentional allocations to the Lower region of the illusion was 40% (SEM: 4%) and 38% (SEM: 4%) during post-test. This change of 2% is a superficial, pre-inferential indication that, among participants in the Uniform condition, the proportion of attention paid to the Lower region of the screen decreased from pre-test to post-test. Figure 4.3 visualizes these changes.

Inference

We used logistic regression models to predict the probability of allocating attention to the Lower region of the illusion. To serve as a baseline against which we could compare our full model, we began our analyses by fitting Model 1—which had an intercept as its only predictor. Next, we derived Model 2 from Model 1 through the addition of by-participant adjustments to the intercept. Model 2 was a better fit to the data than Model 1—ΔELPD: 8.9 (SE: 3.5).

Having captured this within-subject variability with Model 2, we then tested the model encoding our hypothesis, Model 3. Model 3 predicted the probability of attending to the biased region of the illusion using an intercept with by-participant adjustments, a predictor coding for test-time (pre-test or post-test), a predictor coding for condition (Uniform or Biased), and a test-time by condition interaction. Model 3 represents our hypothesis that only participants in the Biased condition should pay more attention to the Lower region of the illusion in post-test, relative to pre-test, since they receive a spatially-specific bias in the PL Task for the Lower region and Uniform condition participants do not.

Although its predictive accuracy was higher compared to the naive Model 1, Model 2's was superior—ΔELPD: -2.5 (SE: 2.4).

For posterior inferences about relevant parameters, we draw on Model 3 to explore why these factors failed to boost our expected out of sample predictive accuracy in the way we predicted.

The 95% central credible interval for Model 3's estimate of the intercept did not include zero. The median of this posterior marginal distribution was -0.47—which translates to a baseline propensity for allocating attention to the biased region of 38.5%.

The 95% central credible interval for Model 3's estimate of the magnitude of the effect of test-time, condition, and the test-time by condition interaction all included zero. The medians and the dispersion of the marginal posterior distributions over all of the predictors except for the intercept imply that we can be confident that these factors offered little in the way of explanatory power to Model 3.

4.3.3 PL Task

Accuracy

All non-practice trials were analyzed to compute the accuracy results. In the Biased condition: 97.2% of targets in the Lower region were classified correctly and 95.2% of targets elsewhere were classified correctly. In the Uniform condition: 97.3% of targets in the Lower region were classified correctly and 97.5% of targets elsewhere were classified correctly.

Reaction Time

Trials eligible for our RT analyses fell in the range [151 ms, 1999 ms], did not occur before or after a mid-block recalibration, and did not contain a blink at any point between trial start and target classification; 13.5% of trials were trimmed. We raised the ceiling on our eligibility filter following an inspection of the data which indicated that trials slower than

1000 ms shouldn't be considered as outliers across many participants. Since we filtered out trials contaminated by calibration and blinks while explicitly including accuracy in our models, we did not worry about the legitimacy of these trials. It should be noted that this choice of ceiling was responsible for a slight increase in the baseline response times—however, restricting our analyses to trials in the range [151 ms, 999 ms] only came with a 20 ms reduction in the median baseline RT.

In the Biased condition: targets in the Lower circle, relative to targets in other locations, were classified more quickly—mean median RT_{LOWER}: 689 ms (SD: 97); mean median $RT_{NOTLOWER}$: 720 ms (SD: 86); ΔRT: 31 ms. This pattern held in the localization component—mean median RT $Localize_{LOWER}$: 203 ms (SD: 46); RT $Localize_{NOTLOWER}$: 222 ms (SD: 39); ΔRT Localize: 19 ms. It also held to a lesser extent in the classification component—mean median RT $Classify_{LOWER}$: 476 ms (SD: 71); RT $Classify_{NOTLOWER}$: 482 ms (SD: 71); ΔRT Classify: 6 ms.

In the Uniform condition: targets in the Lower circle, relative to targets in other locations, were classified less quickly—mean median RT_{LOWER}: 708 ms (SD: 82); mean median $RT_{NOTLOWER}$: 664 ms (SD: 67); ΔRT: -44 ms. This pattern held in the localization component—mean median RT $Localize_{LOWER}$: 229 ms (SD: 47); RT $Localize_{NOTLOWER}$: 193 ms (SD: 21); ΔRT Localize: -36 ms. It also held to a lesser extent in the classification component—mean median RT $Classify_{LOWER}$: 463 ms (SD: 53); RT $Classify_{NOTLOWER}$: 457 ms (SD: 55); ΔRT Classify: -6 ms.

Figure 4.4 depicts how aggregate RT distributions varied as a function of condition and target location. Figure 4.5 depicts how the localization and classification components of RT varied as a function of condition and target location.

Inference

As before, we model RT on eligible trials with the assumption that it is log-normally distributed (Baayen, 2008).

PL Model 1 predicted RT using an intercept and PL Model 2 predicted RT using an intercept with by-participant adjustments. The inclusion of by-participant adjustments was justified—ΔELPD: 2475.3 (SE: 70.5).

PL Model 3 predicted RT using an intercept with by-participant adjustments, the eccentricity of the target from central fixation, a factor coding for accuracy, and a factor coding for condition. Compared to PL Model 2, including these factors led to a substantially better fit to the data—ΔELPD: 159.4 (SE: 6.7).

PL Model 4 encoded our pre-Experiment 2 hypothesis. It was an extension of PL Model 3 and featured a factor coding for spatial probability. Comparing against PL Model 3, we found that information pertaining to spatial probability hardly enhanced the explanatory power of PL Model 4—ΔELPD: 1.3 (SE: 2.3).

PL Model 5 extended PL Model 4 by including the probability by condition interaction term we found from the results of Experiment 2. Comparing against PL Model 4, we found that, as was the case in Experiment 2, including the interaction was justified—ΔELPD: 151.9 (SE: 17.7).

Analyzing the PL Task datasets from both Experiment 1 and Experiment 2 found evidence against the existence of a probability by accuracy interaction. We submitted the Experiment 3 PL Task dataset to the same test, deriving PL Model 6 from PL Model 5 by adding a probability by accuracy interaction term. Including this term was not justified—ΔELPD: 0.0 (SE: 1.9). In fact according to PL Model 6 the 95% central credible interval for the probability by accuracy interaction term contained zero. This indicates that the spatial PL effect was not the result of a speed vs. accuracy tradeoff.

For posterior inferences regarding parameter values we draw on PL Model 5, which extended our original hypothesis concerning spatial probability with a probability by condition interaction term.

The 95% central credible interval for the magnitude of the intercept did not include zero. The median of this posterior marginal distribution was 6.45—which translates to a baseline RT of 630.86 ms.

The 95% central credible interval for the magnitude of the effect of accuracy did not include zero. The median of this posterior marginal distribution was 0.041—which translates to 1.04 x 2 = 2.08 ms.

The 95% central credible interval for the magnitude of the effect of target eccentricity did not include zero. The median of this posterior marginal distribution was 0.0155—which translates to 1.01 x 2 = 2.02 ms.

The 95% central credible interval for the magnitude of the effect of condition included zero. The median of this posterior marginal distribution was 0.006—which translates to 1.01 x 2 = 2.02 ms.

The 95% central credible interval for the magnitude of the effect of spatial probability did include zero. The median of this posterior marginal distribution was 0.0019—which translates to 1.00 x 2 = 2.00 ms.

The 95% central credible interval for the magnitude of the condition by spatial probability interaction did not include zero. The median of this posterior marginal distribution was -0.0122—which translates to 0.97 x 2 = 1.94 ms.

We fit the same series of models to predict classification RT and localization RT. In both cases the best model took the same form as PL Model 5. Like PL Model 5 which predicted RT in general, our model predicting the localization component of RT also benefited when we extended it to include the condition by probability interaction term. This result supports our observation that the spatial probability effect is driven by spatially-specific changes in target acquisition.

In contrast, our model predicting the classification component of RT was hardly improved when it was extended to include the condition by probability interaction term. This pattern of results indicates that we found strong evidence in favour of the spatial PL effect being driven by an enhanced, spatially-specific ability to localize targets prior to classifying them.

4.4 Discussion

4.4.1 Illusion Task

Even with our improved manipulation of voluntary attention and larger sample size, we failed to corroborate our hypothesis. We did not find the condition-specific and spatially-specific increase in attentional allocations to the Lower region of the illusion that we were looking for. This is the second experiment which has produced evidence that participants in both conditions showed the same propensity for attending to the Lower region of the illusion in pre-test as they did in post-test. And, compared to Experiment 2, our improved implementation came with a reduction in our posterior uncertainty regarding this inference. Relative to Experiment 2, our modeling offered strong evidence that participants in both conditions had identical propensities to allocate their attention regardless of test-time and condition. We are confident in concluding that spatial PL did not act as a reliable or robust manipulation of voluntary attention—given our modeling assumptions and our operationalizations.

4.4.2 PL Task

Based on an ordered comparison of the expected out of sample predictive accuracies of PL Models 3 through 5, we established strong evidence for the existence of spatial PL. In agreement with Experiment 2, we found that participants responded more rapidly for targets in the Lower region of the display, relative to the other regions—but only if they were members of the Biased condition. In contrast, participants in the Uniform condition showed the opposite pattern, responding more rapidly for targets which did not appear in the Lower region of the screen. In light of the implementation changes we discussed in 3.4 this replicated result strikes us as compelling.

The change in task demands might also account for the significantly slower baseline response time in Experiment 3 relative to Experiments 1 and 2. Perhaps this difference might be due to the exclusion of guessing or parafoveal responding, which we could not detect in a principled manner before now—since Experiment 3 was the first time we strictly forced participants to foveate the targets. Our choice to include legitimate trials in the range [1000 ms to 1999 ms] which were removed from the analyses in Experiments 1 and 2 was only responsible for a 20 ms change in the median baseline response time, and was therefore not the only cause of the comparative slower responding.

The fact that the absolute magnitudes for the effects of the various predictors remained essentially constant between Experiment 2 and Experiment 3 is notable. It implies the changes in behaviour we found were the same in both instances, and supports the conclusion that it was something static about the change in task demands which led to the change in baseline RT. One possibility is that our gaze-contingent revelation of target colour might have introduced a delay into the process of accumulating evidence for either alternative. Further support for this conclusion comes from our analyses of classification RT which showed only meagre evidence that the classification component of RT varied as a function of target location or condition. In fact the median classification RTs in Experiment 3 are on par with the total RTs in Experiments 1 and 2.

As stated in 4.3.3 our modeling efforts also substantiated the claim that the spatial PL effect is predominantly driven by spatially-specific changes in the ability of participants to acquire targets. The evidence for a role played by changes in perceptual processing is much slighter. Admittedly, the feature discrimination task itself was incredibly easy—a glance at the accuracy rates across all of our studies attests to that. Thus it is unclear that we would be able to detect a proportionately large change in classification RTs. We therefore remain agnostic concerning the relationship between spatial PL and perceptual processing while taking seriously the contribution of gaze adaptation.

With the finer curation of our data made possible by use of the eye-tracker we intended to learn more about the magnitude of the spatial PL effect—assuming we found it for a third time—which we did. After enhancing our signal to noise ratio we found a larger difference in mean median RT as a function of spatial probability of 31 ms.

Since we are confident that our estimate is sound, we attribute the non-trivial differences in magnitude to non-trivial differences in protocol demands and implementations. One example of the impact of task demands on model estimates is the fact Experiment 3 produced the strongest evidence so far in favour of target eccentricity being a determinant of RT. We suspect this is was a consequence of the standardization of target sequences within and between conditions, which reduced the within-condition variation in target eccentricity.

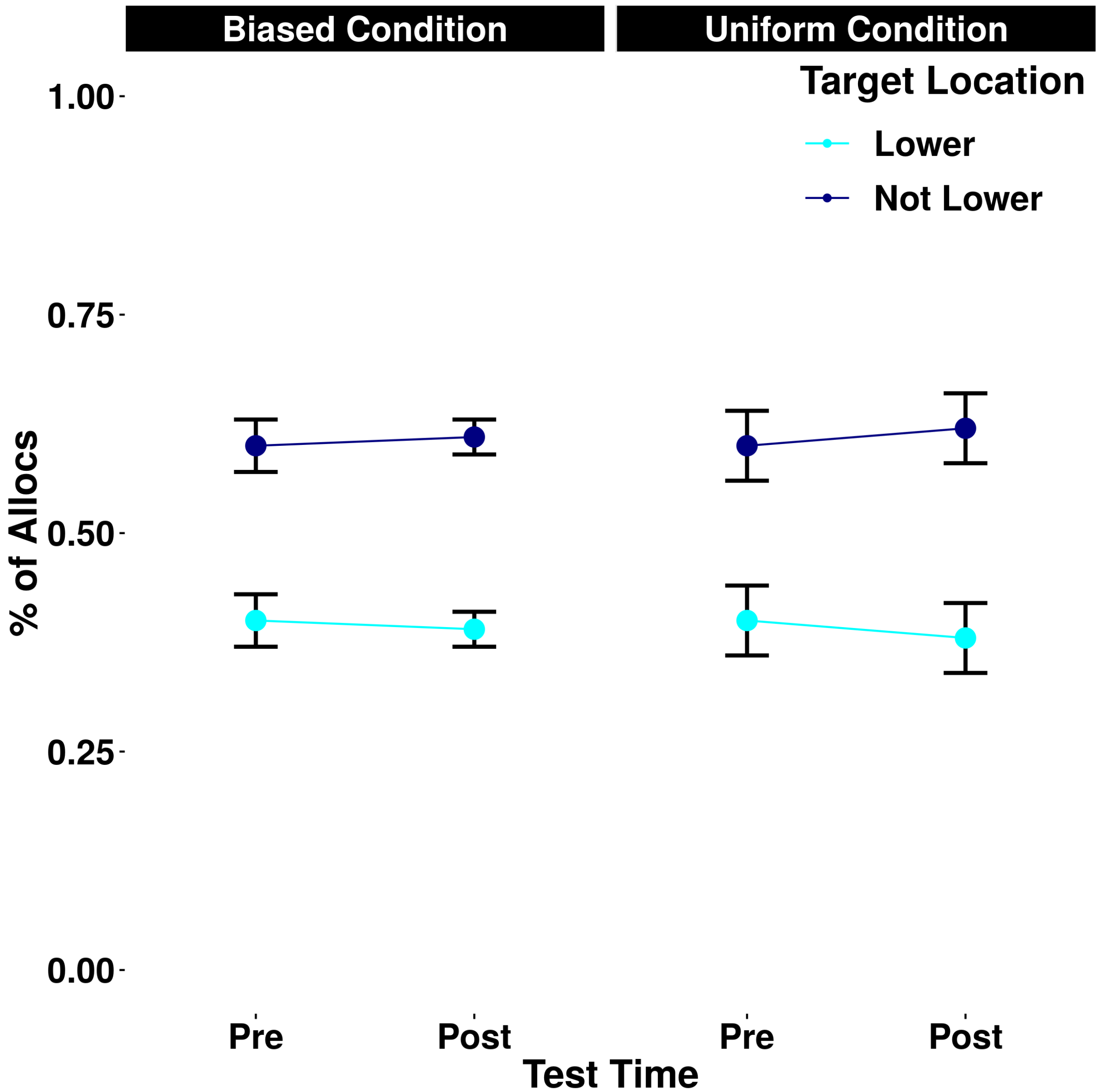

Figure 4.3: Mean (+/- SEM) proportions of attentional allocations to high (e.g., Lower) and low probability regions of the Tse Illusion are depicted as a function of test-time and condition.

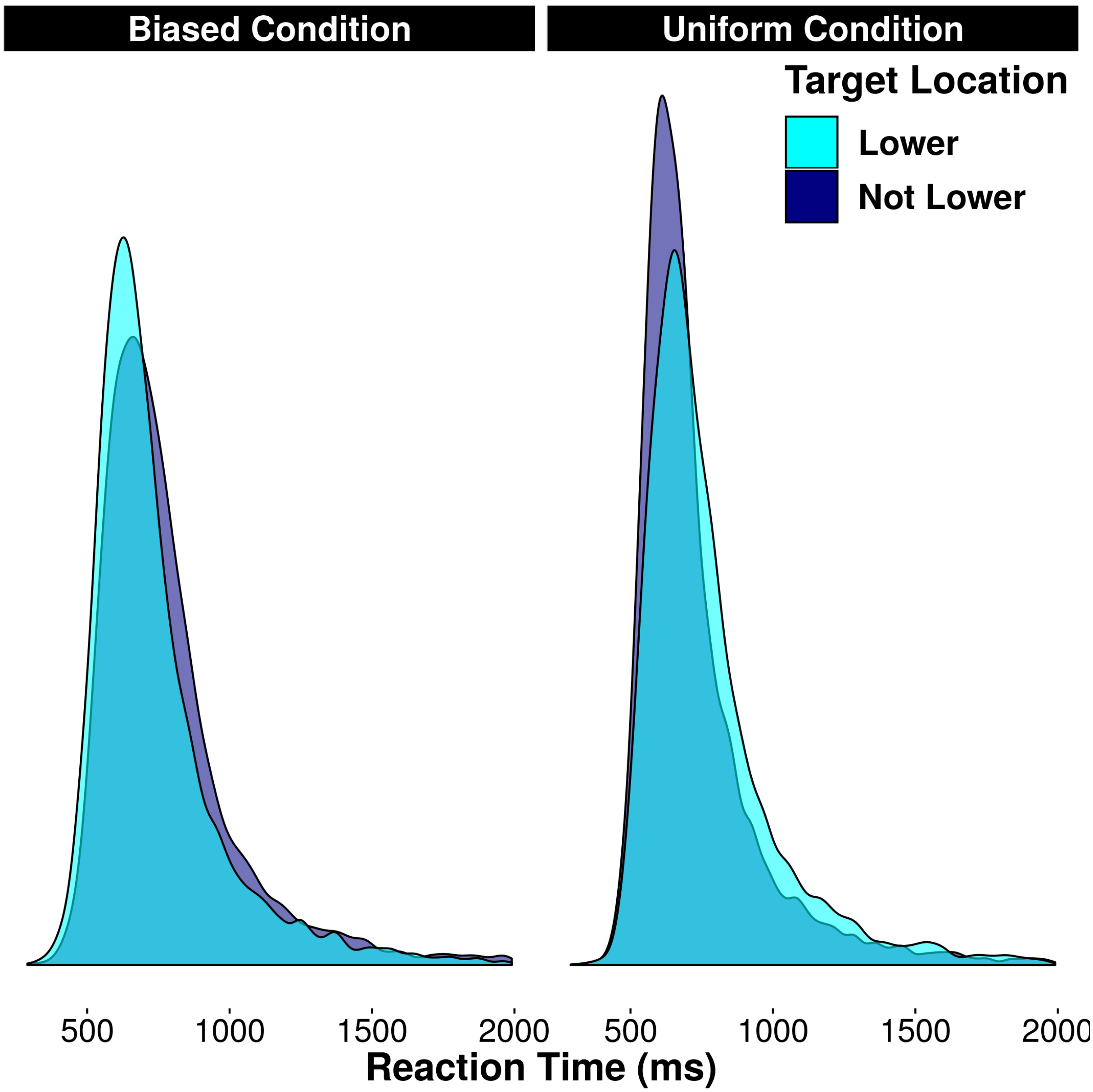

Figure 4.4: RT distributions are depicted as a function of condition and target location. In accordance with the results of Experiment 2, only participants in the Biased Condition exhibited comparatively faster RTs for targets in the Lower region.

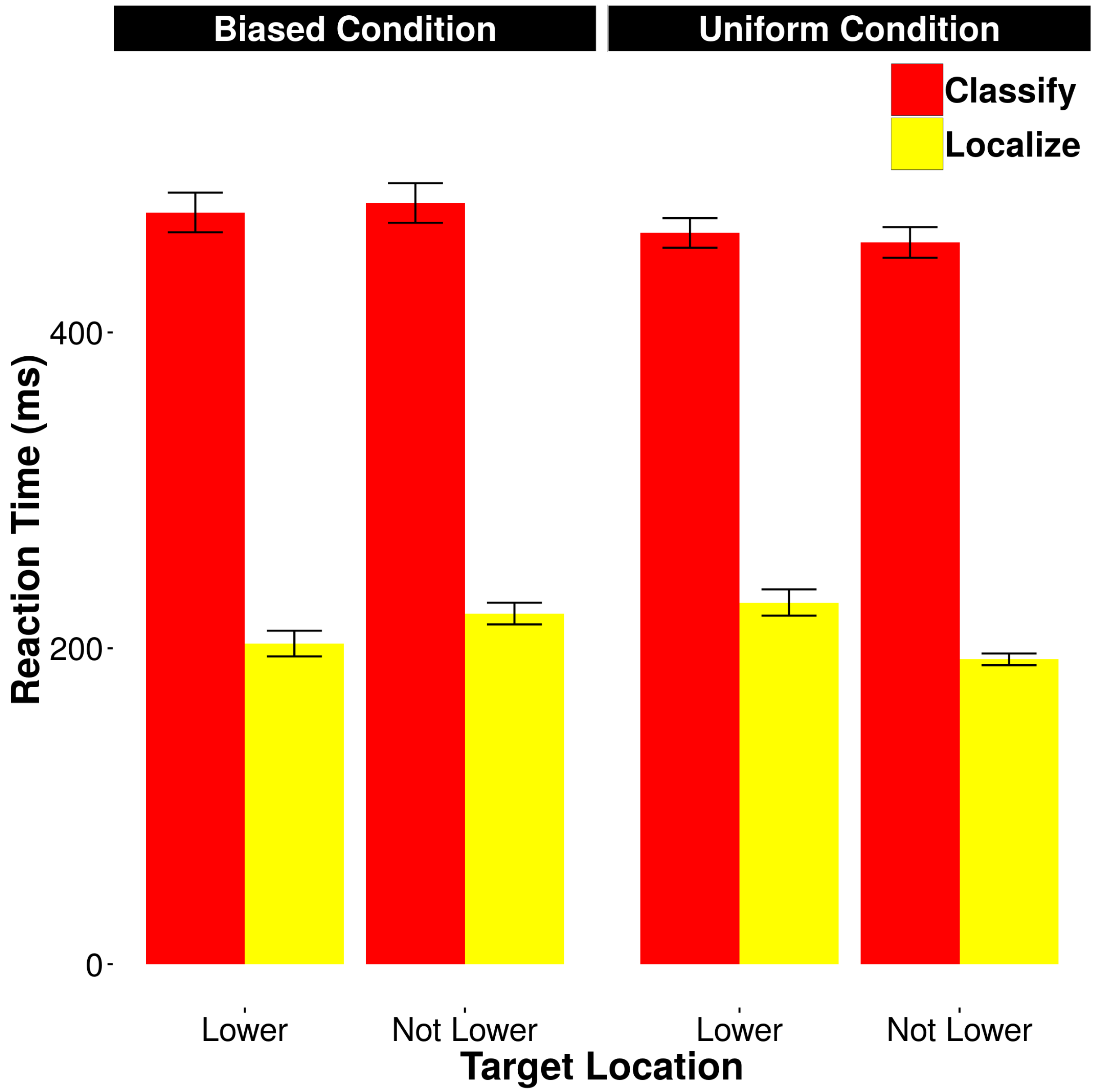

Figure 4.5: Barplots depict the different components of RT in terms of mean median RT (+/- SEM) as a function of target location and condition. Localization takes comparatively less time for targets in the Lower region in the Biased Condition. In the Uniform Condition, the opposite trend holds. Classification RT varies to a lesser extent in relation to spatial probability in both conditions.

Chapter 5

General Discussion

5.1 Purpose

We conducted this research in hopes of replicating and extending past work on spatial PL.

The canonical spatial PL effect involves faster detection of visual targets which is not otherwise explained by strict spatial repetition priming, a speed-accuracy trade-off, or systematic differences in target eccentricity. Contention in the literature regarding the mutual exclusivity of voluntary and involuntary attention led us to investigate whether spatial PL in one context could influence the voluntary deployment of attention in another task. Our suspicion that spatial PL was driven by meaningful changes in eye-movements—along with the shared underlying neural circuitry of voluntary attention and eye-movement generation—justified our attempt to induce spatial PL in order to manipulate voluntary attention. Integrating eye-tracking with our protocol in Experiment 3 allowed us to gain insight into the mechanisms of spatial PL. It allowed us to make our feature discrimination task, which we used to induce spatial PL, gaze-contingent—thereby allowing us to decompose RT into distinct localization and classification components. We learned the extent to which spatial PL was driven by gaze adaptation and enhanced perceptual processing.

5.2 Results

Across three studies, we generated increasingly compelling evidence that we induced spatial PL with our feature discrimination task. In all cases, we arrived at robust estimates of the effect of spatial PL on reaction time and accuracy.

In Experiments 2 and 3 we found evidence for a common malleable pre-existing spatial bias—specifically, that participants would take longer on average to respond to targets in the Lower region of the screen. We found that this relationship could be reversed through spatial PL.

Diffusing target locations across broad regions of space instead of using a handful of repeated locations helped us distinguish the spatial PL effect from strict repetition priming. Dissociating the task demand of colour classification from the probability manipulation of target location allowed us to distinguish the spatial PL effect from response priming. Explicitly modeling accuracy and investigating probability by accuracy interactions helped us rule out the changes as the by-product of a speed-accuracy tradeoff. Removing trials with artifacts (e.g., blinks and unscheduled breaks) enhanced the signal-to-noise ratio in our data. Concerns regarding strategic and anticipatory shifts in gaze were addressed by enforcing the generation of saccades in Experiment 3. Comparing the impact of a uniform versus biased spatial distribution over target locations in Experiments 2 and 3 permitted us to draw causal claims invoking the statistical regularity of visual stimuli as a determinant of behaviour.

Adding a task tutorial and enforcing gaze generation enhanced our confidence in the strength of the spatial PL task as a manipulation of voluntary attention in Experiment 3. In both Experiment 2 and Experiment 3, we found evidence for a lack of influence between spatial PL in one task and voluntary attention in another. It is possible that future analyses of the eye-tracking data might produce evidence in the opposite direction. However, we chose the proportion of attentional allocations as our a priori dependent variable for the work reported here. We therefore conclude that we did not manipulate voluntary attention using spatial PL in the way we predicted we might. Given that the sort of spatial PL we induced was implicit in nearly all cases—in accordance with past work—the implicit learning literature, which has failed to find transfer of implicit learning across tasks, suggests that our original hypothesis may have been misguided despite our legitimate justifications (see Appendix C.2).

Ultimately, we found evidence that spatial PL in our task was largely driven by gaze adaptation. We found weak evidence for a relationship between spatial PL and changes in perceptual processing. Past work has found that spatial PL decreases detection time without improving perceptual processing, but in the context of orientation judgments rather than colour discrimination (Jabar & Anderson, 2017). We remain agnostic about the role of enhanced perceptual processing in the spatial PL effect on the basis of our own findings.

Future work might seek to further clarify this issue. In doing so a mechanistic answer for why spatially diffuse PL appears to involve a global facilitation of RT might be uncovered.

We found this trend among Uniform condition, but not Biased condition, participants in Experiments 2 and 3.

5.3 Conclusion

We found evidence for the canonical spatial probability effect after accounting for repetition priming, response priming, accuracy, task compliance, artifacts, and target eccentricity. As predicted, targets in probable, relative to improbable regions, received more rapid responses. Having replicated past work, and having achieved strong internal replication of the effect's existence and magnitude, we conclude that statistically regular properties of visual stimuli come to meaningfully change our ability to initiate eye movements to specific regions in space.

We also found evidence that despite sharing underlying neural correlates, an induction of a spatially-specific bias in involuntary attention in one task did not transfer to the expression of voluntary attention in a subsequent task. To be clear: our negative result does not conclusively rule out that spatial PL cannot influence the expression of voluntary attention. Rather, we are only certain that we failed to detect such an influence given the operationalization of voluntary attention we used in our protocols. Future analyses of the eye-tracking data gathered during Experiment 3, including changes in the relative amount of time per allocation, may corroborate or contradict our present findings.

Our negative result could have occurred for a number of reasons notwithstanding the adequacy of our choice of DV. First, the residual adaptation that manifests as spatial PL might not mechanistically interact with voluntary attention through a shared neural bottleneck. If it does, its influence might not be strong enough to drive behavioural variability that we would deem significant enough to be a signal amidst the noise. Perhaps the influence is strong enough, but it is transient to the extent that we would only expect systematic differences in voluntary attention during a window of time much shorter than the one we considered here.

It is also useful to contemplate the result through a functional lens. Behavioural adaptation permits a biological system to more economically use its resources as it navigates through the world. The critical issue is that statistical patterns are defined by the time scale at which covariation between stimuli occurs. There are multiple time scales that are relevant to survival, and so ideally, the capacity for adaptation should respect multiple time scales.

The tension is between maintaining sufficient flexibility to reflect relevant environmental changes on short time scales while still guiding one's behaviour such that it respect trends deduced over longer time scales. One general solution to this problem is to include in one's system a mechanism capable of rapid short term adaptation. The history of this mechanism's adaptation can then be integrated over time by other mechanisms, which in turn have their adaptation integrated by other mechanisms. This recurrent hierarchical approach produces generalization at multiple time scales and can produce behavioural adaptation that is guided by information gained from the environment over different epochs. It permits adaptation to longer term trends by providing for rapid adaptation to short term environmental contingencies and condensing these changes into higher order patterns. Such an architecture is a familiar motif in both artificial and biological neural networks.

In this vein, I conclude by speculating that spatial PL does not influence voluntary attention because context-specific state changes in gaze generation originating in cortical circuits flush out residual spatial biases in subcortical structures to facilitate task switching, and the PL Task induction did not constitute a training of biases which were sufficiently generalized to persevere across contexts.

APPENDICES

Appendix A

Science and Statistics

A.1 Hypothetico-Deductive Method

This work is informed by an approach to scientific inquiry known as the hypothetico-deductive method. From this viewpoint, to conduct scientific research is to participate in an iterative refinement of imperfect explanations for natural phenomenon.

A scientific theory is an explanatory account of natural phenomenon which entails a set of testable consequences. An empirical test which could—but does not—contradict the implications of a scientific theory, is said to corroborate it. Crucially, corroboration per se is not the objective of the hypothetico-deductive method. Rather, scientific theories are tested in hopes of falsifying them. We falsify a theory by seeking observations of nature which contradict one or more of the theory's logical consequences.

This approach is counter-intuitive if one conceives of the scientific method as a recipe for proving the truth value of theories. As practitioners of the hypothetico-deductive method we eschew this quest for certainty. We are satisfied instead by repeatedly replacing falsified theories with plausible falsifiable alternatives.

We demand that these refined alternative theories accommodate observations of nature with an ever-increasing degree of success. By construction such alternative theories come to entail sets of testable consequences which are increasingly sophisticated and precise. It follows that our theories are naturally subjected to more stringent testing over time. And, in practice, our interim models of nature become increasingly useful as we continue to apply our method.

A.2 Statistical Tools and Scientific Inquiry

Empirical theory testing—experimental or otherwise—involves a systematic accumulation of observations. In order to arrive at conclusions regarding the status of our theories, we must interpret our data. Statistical tools can facilitate our interpretations and guide our inferences. Thus, to the extent that we rely on them, statistical tools are crucial determinants of the outcome of our program of falsification.

A.3 Selecting Statistical Tools

The importance of data analysis in the process of scientific inquiry obliges us to choose and use our statistical tools with care. Above all we require a methodology which can help us to practice the hypothetico-deductive method. Thus, we seek tools which allow us to:

- Encode theories of arbitrary complexity with unambiguous formal statements;
- Aggregate data from multiple sources;
- Quantify uncertainty about our theories and our data;
- Reason about our uncertainty, theories, and data in a consistent manner; and,
- Specify precise empirically-testable quantitative predictions that follow from our theories.

A.3.1 Null Hypothesis Significance Testing

We might consider using null hypothesis significance testing (NHST) for our purposes. By NHST, we mean the conceptual and methodological hybrid of data analysis techniques created by Fisher, Neyman, and Pearson. Across many disciplines, NHST represents the canonical statistical method to deploy during empirical theory testing.

Recently, the criticism NHST has faced for decades was summarized and extended in an article by Szucs and Ioannides (2017). The authors provided a rigorous and extensive accounting of NHST's conceptual and quantitative shortcomings—which is beyond the scope of this document to recount in full. Their emphatic conclusion—that NHST is unsuitable as a statistical tool for theory testing—accords fully with a recent statement by the American Statistical Association (Wasserstein & Lazar, 2016).

A.4 Bayesian Inference

A.4.1 Models

Consider $\mathbf{y}$, the participant RT data we collected during the PL Task. Some physical process—involving our protocol and our participants—was responsible for generating this data. Our goal is to assess the relative plausibilty of existing theories which make different claims about this generative process.

Such claims nominate candidate exp¡lanatory variables $\mathbf{x}$ upon which the generative process allegedly depends—and some even specify the nature of these dependencies. If we assume that we can use a mathematical model to represent the generative process, then we can use distinct models to encode different theories. By model we mean a function f with parameters $\boldsymbol{\theta}$. Models allow us to quantify the impact of candidate explanatory variables $\mathbf{x}$ on the process of interest. We capture hypothetical dependencies between $\mathbf{y}$ and $\mathbf{x}$ by specifying the functional form of f along with $\boldsymbol{\theta}$—degrees of freedom which permit fine-tuning of f's behaviour. The relative plausibility of existing theories can then be evaluated in light of empirical data by assessing the predictive performance of the corresponding models.

Take for example a theory t_1 that claims that on each trial RT depends exclusively on target eccentricity—the distance between the target and the center of the display. Setting technical details aside, assume further that t_1 deems it appropriate to use a simple linear regression function, f_1, to model the proposed relationship between RT and eccentricity. Then in this case, $\mathbf{y}$ is a list of observed RTs, $\mathbf{x}$ is a list of explanatory variables—an intercept and target eccentricity—and $\boldsymbol{\theta}$ corresponds to f_1's regression coefficients—which parameterize the linear relationship between $\mathbf{y}$ and $\mathbf{x}$ specified by t_1.

During an experiment, we can observe the outcomes $\mathbf{y}$ while systematically varying the state of $\mathbf{x}$—e.g., we can record RT while changing the eccentricity of targets. We can subsequently evaluate f_1 at each value that the explanatory variables in $\mathbf{x}$ took on during the experiment to compute f_1's predictions for RT: $\hat{\mathbf{y}}$. With a loss function of our choice, $\mathcal{L}$, we measure the discrepancy between f_1's predictions for RT, $\hat{\mathbf{y}}$, with the RT we actually observed, $\mathbf{y}$. The process of setting $\boldsymbol{\theta}$ typically involves solving an optimization problem such that the model loss according to $\mathcal{L}$ is minimized. For a simple linear regression like f_1 it is common to choose the mean squared error for $\mathcal{L}$ and to use the method of least squares to solve for $\boldsymbol{\theta}$.

A.4.2 Model Selection

To establish the plausibility of t_1, we assess the performance of f_1 as a surrogate for our generative process of interest. It is intuitive that our assessment of f_1's performance should consider the accuracy of its predictions, $\hat{\mathbf{y}}$. But keep in mind that we set f_1's parameters to optimize its ability to make predictions about $\mathbf{y}$. Thus f_1 will express measurement error or bias inherent to our data that isn't informative about the generative process it is meant to model. And, to the extent that $\mathbf{y}$ constitutes a finite sample from the generative process, f_1 will make predictions impacted by sampling error. Both of these factors might lead us to put too much faith in f_1's fidelity as an approximation to the generative process on the basis of its ability to predict $\mathbf{y}$.

Our work was not conducted in an ideal world. Hence our need to consider how well our models makes predictions about outcomes that were not used to determine the optimal values for its parameters $\boldsymbol{\theta}$—its out-of-sample predictive accuracy. This predictive capacity is our primary criterion for ranking the relative merit of our models. Our choice is pragmatic. We use theories to understand and navigate the natural world. So we prefer theories which make predictions about future states of the world that are reasonable and reliable—and therefore can be expected to guide effective future behaviour.

Since we work within a Bayesian framework, our models issue probabilistic predictions rather than point predictions. We believe that representing uncertainty in our reasoning about states of the world is theoretically and practically necessary. The out-of-sample predictive accuracy of models which make probabilistic predictions can be approximately computed using a process called Bayesian Leave-One-Out cross-validation (LOO-CV). This is the procedure we use to assess and compare our models. The result is an estimate of a model's expected log predictive density (ELPD). The higher a model's ELPD, the better its quality as an approximation to a target process. Other criteria for model comparison—e.g., the Akaike Information Criterion (AIC) or the Deviance Information Criterion (DIC)—use a point-estimate for a given model's parameters to compute its log predictive density. In contrast, LOO-CV accounts for uncertainty about model parameters in a fully Bayesian manner. Unlike the AIC, LOO-CV does not ignore the prior distributions in our models, nor does it assume that the posterior is a multivariate Gaussian distribution. Interested readers are directed to Gelman, Hwang, & Vehtari (2013) and Vehtari, Gelman, & Gabry (2017) for detailed technical discussions of ELPD and LOO-CV that are beyond the scope of this thesis.

A.4.3 On Subjectivity

One complaint about Bayesian inference is that it compromises the integrity of objective analyses through the introduction of subjectivity. This common misconception has been addressed exhaustively elsewhere—we include brief remarks here for clarity's sake. All scientific inquiry is guided by subjective assumptions. The relevant issue is whether or not these assumptions are explicit and amenable to inspection. Sensitivity analysis—the exploration of the inferential consequences of assumptions—is a key feature of Bayesian inference that naturally falls out of the way we represent and operate on our probability models.

The principles which govern encoding and reasoning about theories in the context of Bayesian inference have been mathematically proven to be the unique rules for consistent inference under uncertainty (Jaynes, 2003). This last point is reassuring; however, it does not avoid the problem of "garbage in, garbage out". In other words, the machinery of Bayesian inference is guaranteed to always produce results that are objectively correct—given the input to the process. The conditional nature of this guarantee is critical to acknowledge. Nowhere along the way does a Bayesian data analyst become immunized against subjectivity—but such criticism applies in general and is not unique to Bayesian inference.

Criticism about Bayesian inference might also be aimed specifically at the use of prior distributions to represent existing beliefs. It is interesting to consider that many classical statistical analysis techniques have been proven to be special cases of Bayesian inference with incomplete probability models, or implicit assumptions about prior distributions (Jaynes, 2003; Gelman et al., 2014). Given the universal fallibility of available data analysis methods, we err on the side of transparency and apply the general framework of Bayesian infernce, with an eye on keeping our assumptions explicit and testable.

A.4.4 What to Expect From a Bayesian Data Analysis

We derive models from our theories. Via Bayes' theorem and our use of distributions (rather than classical point-estimates), our models justify conclusions concerning the plausibility of competing theories, given the data and our assumptions. We do not remain fixated on the probability of the data, given our competing hypotheses—a shortcoming of the classical approach.

The result of fitting a given model is a joint posterior probability distribution over that model's parameters—i.e., we learn the probability of every combination of that model's parameters. *Posterior* signifies that a distribution is the result of a process of updating our *prior* beliefs with information contained in a set of observations and our assumptions about how those observations were generated. We can use marginalization to proceed from the joint distribution to a marginal distribution over a single parameter of interest. Marginalization is an antiquated term that originates from the practice of representing frequency data in contingency tables. The following example will help to illustrate it.

Imagine you and a friend are at a bowling alley. You are informed by the manager that you will both be exiled permanently from the establishment unless one of you promptly bowls a strike. You are much more talented than your friend: the chance you will bowl a strike is 50%, whereas your friend will only bowl a strike with 25% probability. Nevertheless your friend insists that you decide who makes the attempt by flipping a fair coin. If it comes up heads, you will bowl. We assume that regardless of the outcome of the coin flip, the chances of bowling a strike remain the same.

The marginal probability of a strike being bowled is the overall probability a strike will be bowled. It is equal to the sum of your chance of bowling a strike (50%) weighted by the probability that you will make the attempt (50%)—and your friend's chance of bowling a strike—25%—weighted by the probability that your friend will make the attempt (50%). Your friend will succeed a quarter of a half of the time (1/8); you will succeed half of half of the time (1/4). This means that the overall probability of success is 37.5% (3/8).

A marginal distribution over a single parameter is a probabilistic prediction of its true magnitude—given its dependence on other parameters in the model and the probabilities with which the other parameters take on their respective values. We use marginalization to compute the probability that a single parameter of interest θ_j takes on a magnitude in a specific interval. For example, when θ_j is a coefficient in a regression function quantifying the magnitude of a effect—perhaps the effect of eccentricity on RT as in f_1—we can compute the probability that the effect is non-zero. In such cases we provide a 95% **credible** interval, which represents the interval within which we expect the true value of the parameter to fall with 95% probability. The dispersion of any probability distribution we summarize with a 95% credible interval represents our uncertainty in our inferred estimates—which can therefore be seen in the relative width of the credible interval.

This honest accounting for uncertainty is largely missing in the classical approach. Moreover, unlike classical confidence intervals, Bayesian credible intervals actually quantify the plausible range of values of the effect of interest. In other words, when an effect's credible interval doesn't contain zero, we have evidence for its existence, and vice versa.

We are interested in assessing the evidence for the existence of spatial PL in our PL Task dataset. We are also interested in assessing the evidence that we influenced voluntary attention by inducing spatial PL. Given these objectives—and our pragmatic interpretation of the hypothetico-deductive method—the Bayesian approach is appropriate for our work. NHST does not allow its users to answer the research questions we are interested in and is therefore inappropriate for our work.

A.4.5 Posterior Distributions

Before we can take any of our posterior inferences seriously, we need to inspect the integrity of the processes which generate our posterior distributions. Outside of special cases, most of the posterior distributions scientists encounter take on forms which cannot be computed exactly. Faced with this challenge and motivated by the benefits of Bayesian inference, a number of clever methods of approximation have been devised over the years. In tandem with advances in computing power, these algorithms are responsible for the recent proliferation of applied Bayesian inference.

We rely on a Markov Chain Monte Carlo (MCMC) algorithm to generate samples from our target posterior distributions. Specifically, an improved version of Hamiltonian Monte Carlo (HMC) known as No U-Turn Sampling (NUTS; Hoffan & Gelman, 2011).

MCMC algorithms address the problem of intractable posterior distributions by taking advantage of the fact that—assuming certain properties about the sampling procedure hold—we can approximate a distribution to a high degree of accuracy from a large enough set of samples drawn from said distribution.

Of course, we don't have an exact representation of the posterior distribution from which to draw samples—otherwise we wouldn't need to approximate it in the first place. What we do have is a sense of the space occupied by the posterior distribution. This information comes from what we do know—our prior beliefs, our model for how our data were generated—and, following Bayes and Laplace, that the posterior is proportional to their product.

It is this constant of proportionality that is the culprit when it comes to rendering our exact computation intractable. In brief: MCMC algorithms take advantage of the behaviour of random walks to explore the space occupied by the posterior distribution in such a way that this constant of proportionality is no longer relevant. NUTS and HMC differ from other MCMC algorithms in terms of how they use technical features of this space to make this random walk behaviour more effective. In the end, we get a chain of samples which we can use to approximate our target posterior distribution.

A.4.6 Sampling Diagnostics

There are a number of diagnostic criteria that allow us to gauge the validity of the posterior distributions we compute. One example is the $\hat{R}$ statistic which compares the performance of multiple sampling chains to gauge whether their behaviour is comparable. Another is the n_{eff} which represents the number of independent samples from the posterior that would provide the same amount of information as the total number of correlated samples drawn by our chains. Since there are no guarantees concerning approximation error and how long a chain must be run, it is best practice to run multiple chains for many thousands of iterations. If multiple independent chains initiated at different random locations converge to the same output—we become increasingly confident in the integrity of our results. As a rule, all of the models we report on in this thesis were based on chains which passed the relevant diagnostic tests and ran for many thousands of iterations. Empirical evidence in support of the sufficiency of the approximation accuracy one should expect under this regime is robust (Gelman et al., 2014).

A.4.7 Interpreting Results

We report the ELPD for each model of interest. To compare a set of models, we compute the differences between their respective ELPDs, which we approximate using Bayesian LOO-CV. We report the standard error of these estimated differences to represent uncertainty. We do not misrepresent the model with the highest ELPD amongst our candidate models for a given dataset as the true or correct model. These models are merely the best of the subset of all possible models that we choose to fit according to convention, past work, and present research questions.

We used t-distibutions with 7 degrees of freedom as weakly informative priors for parameters as suggested in Gelman et al. (2014) in the spirit of robust regression. Sensitivity analyses showed that in no case did our inferences exhibit a meaningful dependency on this choice; these alternative results are not reported here. For parameter estimates, we report 95% credible intervals and the median of the posterior marginal distribution as a quick summary. We take posterior estimate credible intervals for regression coefficients that do not contain zero as support for the claim that there is evidence for the corresponding effect.

We use contrast coding in our regression models for categorical predictor variables. When applicable, factors representing accuracy, condition, test-time, and spatial probability were coded as follows: (-1/+1 for incorrect/correct; Uniform/Biased; pre-test/post-test; low probability/high probability). We report effects associated to categorical predictors after multiplying them by 2 and in doing so represent the effect as the expected difference in the response variable between the two levels of the predictor. A discussion concerning our use of logistic regression and random effects can be found in Appendix B. All of our modeling was done using R (R Core Team, 2018), specifically, the RStanArm (Stan Development Team, 2018a), and shinystan (Stan Development Team, 2018b) packages.

Appendix B

Models and Inference

B.1 Interpretation of Models

B.1.1 What is a Model?

We view a model as a formal statement of ≥ 1 functional relationships between variables of interest. A model's structure encodes assumptions about causal relationships between its constituent variables (Pearl, 2009).

B.1.2 Concrete Illustration: Simple Linear Regression

The familiar model described below by B.1 is the basis of simple linear regression. I introduce it here to provide a concrete illustration of the approach to modeling in this work. A relevant source for the equations is Barr, Levy, Scheepers, and Tily (2013).

$$\begin{aligned} &y_i = \beta_0 + \beta_1 x_1 + \epsilon_i \\ &\text{with} \quad \epsilon_i \sim N(0, \sigma^2) \end{aligned} \tag{B.1}$$

The model predicts each response y_i through a linear combination of two fixed effects, β_0 and β_1, a predictor x_1, and errors ϵ_i. It assumes that each response is conditionally independent of the others, which implies that each one is equally informative with respect to the underlying generative process the model is intended to represent. Equally, the model assumes that the functional relationship between responses and the predictor is correctly specified as linear. It assumes further that no relevant predictors of y have been omitted, and that the predictor was measured without error.

In this model, x_1 is an endogenous variable which systematically varies with y. Error terms, ϵ_i, are understood as observation-level realizations of a collection of strictly exogenous variables which are independent of the predictor. These errors model unspecified external forces which systematically conceal the underlying linear relationship between the parameters of the model and the responses. To solve for the parameters with the method of ordinary least squares, it is adequate to assume that the errors are uncorrelated and distributed with an expected value of zero. Nevertheless, the convention of assuming that the errors are independent realizations from a Gaussian distribution with $\mu = 0$ and variance σ^2 is followed here.

B.1.3 Fixed vs. Random Effects

Fixed effects are thought of as invariant under transformations of the sample—i.e., they are population-level parameters. In particular, the intercept of the regression line is given by β_0, and is interpreted as the baseline expected value of response y in the population, given the model as stated. β_1 quantifies the magnitude of the effect of predictor x_1 on response y in the population.

In contrast, random effects are quantities which are not invariant under transformations of the sample. Variability around the true population-level parameter is introduced by the non-exhaustive nature of the sample with respect to a given random effect. To reflect this variability, estimation of random effects is accomplished by modeling them as samples from a distribution. It is often assumed this distribution is Gaussian with $\mu = 0$. This distribution is parameterized by a covariance matrix which will be described in more detail below.

B.1.4 Motivation for Hierarchical Models

We call a model which includes random effects a multi-level model (MLM) or hierarchical model. The use of hierarchical models is a natural corollary of Bayesian inference—but it is also demanded by the nature of our data collection. Recall that each observation $y_i \in y$ is assumed to be conditionally independent of one another. This assumption is violated, for example, when observations are sampled sequentially from a single participant. Fortunately, hierarchical models with by-participant random effects permit observations belonging to each participant to be clustered together to reflect this dependence. Because our data was gathered via repeated measurement of individual participants it is appropriate to take a hierarchical modeling approach.

Random Intercepts

Recall that β_0 is an estimate of the baseline expected value of the response in the population. When we have reason to believe that individuals in the sample do not possess the same baseline ability to produce the responses, we can introduce by-participant random intercepts to the model. These by-participant random intercepts allow adjustments to the intercept which reflect individual differences with respect to the underlying capacity to produce the responses we observe. A simple linear regression model with by-participant random intercepts can be written as:

$$\begin{aligned} & y_i = \beta_0 + S_0 + \beta_1 x_1 + \epsilon_i \\ & \text{with} \quad \epsilon_i \sim N(0, \sigma^2) \\ & \text{and} \quad S_0 \sim N(0, \tau_0^2) \end{aligned} \tag{B.2}$$

Random Slopes

Recall that β_1 is an estimate of the relative contribution of the predictor x_1 to each response y_i. If individuals in the sample differ in their susceptibility to the effect of x_1, adding by-participant random slopes to the model allows it to express these individual differences. A simple linear regression model with by-participant random slopes can be written as:

$$\begin{aligned} & y_i = \beta_0 + S_1 \beta_1 x_1 + \epsilon_i \\ & \text{with} \quad \epsilon_i \sim N(0, \sigma^2) \\ & \text{and} \quad S_1 \sim N(0, \tau_1^2) \end{aligned} \tag{B.3}$$

Covariance Structure of Random Effects

If we suspect that individual deviations from baseline responding and individual differences in susceptibility to main effects share conceptual justification or otherwise interact, the model should be adapted to reflect this dependency. Such a connection leads us to sample the random intercepts and slopes from a joint distribution. This distribution is assumed to be Gaussian with $\mu = [0, 0]^T$ and is parameterized further by a covariance matrix Σ . We model dependence between the random effects by specifying non-zero covariance between them in Σ. A simple linear regression model with both by-participant random intercepts and by-participant random slopes can be written as:

$$\begin{aligned}
&y_i = \beta_0 + S_0 + S_1\beta_1 x_1 + \epsilon_i \\
&\text{with} \quad \epsilon_i \sim N(0, \sigma^2) \\
&\text{and} \quad S_0, S_1 \sim N([0, 0]^T, \Sigma = \begin{bmatrix} \tau_0^2 & \rho\tau_0\tau_1 \\ \rho\tau_0\tau_1 & \tau_1^2 \end{bmatrix})
\end{aligned} \tag{B.4}$$

The $\tau_0, \tau_1,$ and ρ in B.4 are hyper-parameters of the model. In particular, the dependence we want to capture between random intercepts and random slopes is expressed in the off-diagonal entries. It follows that setting $\rho = 0$ corresponds to the case where the random intercepts and random slopes are conceived of as independent.

B.1.5 Hierarchical Logistic Regression

We use hierarchical linear regression models to analyze the PL Task datasets. For the Tse Illusion Task datasets, we use hierarchical logistic regression. Clarification is provided below.

Motivation

For our Illusion Task analyses, we consider the illusion to be a combination of two distinct regions of interest: a biased region and an unbiased region. During the Illusion Task, each time a participant shifts their attention to a single region, they are making a voluntary choice to allocate their attention to one of these two discrete options. Our concern is the influence of spatial PL on patterns of voluntary attention. Thus, we need to model how changes in this discrete-choice behaviour relate to changes in test-time and condition. Hierarchical logistic regression is the technique we chose for this purpose. It allows us to predict the probability of allocating attention to a given region of the illusion as a function of our explanatory variables.

Hierarchical?

As above, hierarchical implies that a model includes both fixed effects and random effects. In particular, a hierarchical model addresses our need to group multiple observations under individual participants.

Logistic?

Write the probability of an event x_i occuring as $p(x_i)$. By definition, $p(x_i)$ must take on a value that falls in the closed interval [0, 1]—i.e., the range of values $p(x_i)$ can take on is restricted. However, linear regression analyses estimate functions which predict values that fall in an unrestricted range. To bridge this gap, we can transform our data with a pair of mathematical functions: the logit function and the logistic function.

We use the logit function to convert the probability of an event into its log-odds. Given K discrete outcomes, by definition, $\sum_{i=1}^{K} p(x_i) = 1$. It follows that we can write the probability of x_i **not** occuring as $1 - p(x_i)$. The odds (in favour) of x_i, formally $\text{odds}(x_i)$ is defined as:

$$\text{odds}(x_i) = \frac{p(x_i)}{1 - p(x_i)} \tag{B.5}$$

Taking (for instance) the natural logarithm of the odds produces the log-odds of an event:

$$\text{logit}(x_i) = \ln\{\text{odds}(x_i)\} \tag{B.6}$$

The logit function maps values in the closed interval [0, 1] to the real number line. Its inverse, the logistic function, maps values on the real number line back to the closed interval [0, 1]. The logit function permits us to apply a regression analysis to estimate probabilities associated with discrete outcomes. In turn, the logistic function allows us to translate log-odds back into readily-interpretable probability values.

$$\text{logistic}(n) = \frac{\exp(n)}{1 + \exp(n)} \tag{B.7}$$

Appendix C

Probing Explicit Awareness of Spatial Bias

We agree with Druker and Anderson (2010) who remarked that a demonstration of spatial PL is distinct from the question of implicit learning. In this vein, Jiang, Sha, and Remington (2015) found that spatial PL occured whether or not participants were made explicitly aware of the bias in advance of training. We therefore consider the question of awareness as orthogonal to our claims about the existence and magnitude of the spatial PL effect. The awareness results we report here are relevant, however, to our use of the PL Task as a manipulation of voluntary attention, given our pre-test/post-test design.

C.1 Method

In all three experiments we followed the same approach by Druker and Anderson (2010) to probe participant awareness of the spatial biases in the PL Task.

As part of a questionnaire following each of our experiments, we asked participants if they had detected any patterns whatsoever during the PL Task. If they responded in the affirmative, we asked for and recorded their spontaneous description of the pattern. This approach gave participants an opportunity to spontaneously report an explicit awareness of a spatial bias in the distribution of PL Task targets (Unprompted Awareness).

Unless their spontaneously described pattern pertained to the spatial arrangement of the PL Task targets, we proceeded to ask if targets in the PL Task were equally likely to fall anywhere on the display. This question narrowed their focus and provided them an opportunity to report on spatial patterns in particular—especially among participants who detected spurious patterns that were non-spatial. If participants responded that some locations were more probable than others, we asked them to describe the spatial arrangement of the targets. This question helped us gauge whether their response to the spatial probability question was indicative of actual knowledge of the spatial bias in the PL Task, as opposed to a spurious pattern (Prompted Awareness).

We ended with a fully targeted probe if participants reached the end of the questionnaire without making claims regarding either the existence or nature of the spatial distribution of the PL Task targets. These participants were forced to guess where targets were most likely to appear relative to the center of the display (Forced Guess).

C.2 Results

C.2.1 Unprompted Awareness

Across all three experiments only one participant—a member of the Biased condition in Experiment 3—affirmed that they had detected a pattern during the PL Task before spontaneously describing the pattern as the spatial bias of PL Task target locations. Across all three experiments, only one participant—a member of the Uniform condition in Experiment 3—affirmed that they had detected a pattern during the PL Task before spontaneously describing the pattern as a mapping between the locations of PL Task targets and the regions of the display which corresponded to the circles of the Tse Illusion.

C.2.2 Prompted Awareness

In Experiment 1, zero participants who claimed that some target locations were more probable than others accurately described the spatial bias in PL Task targets. In Experiment 2, in the Biased condition, two out of eighteen participants indicated that some target locations were more probable than others before accurately describing the spatial bias in the PL Task target locations. In Experiment 3, in the Biased condition, five out of thirty-two participants indicated that some target locations were more probable than others before accurately describing the spatial bias in the PL Task target locations.

C.2.3 Forced Guess

In Experiment 1, one participant in each condition, when forced to guess, suggested that targets fell in the region of the screen they had been biased towards. In Experiment 2, five participants in each condition guessed that targets fell with higher probability in the Lower region of the screen. In Experiment 3, two participants out of thirty-two in the Uniform condition—and ten participants out of thirty-two in the Biased condition—guessed that targets fell with higher probability in the Lower region of the screen.

C.3 Discussion

We interpret these results, with the limitations of self-report in mind, as evidence that the PL Task induces spatial PL that is implicit in most cases. Furthermore, it seems unlikely that the spatial PL bias acted as a demand characteristic that might have elicited reactivity in participant behaviour during the Illusion Task post-test. These findings accord with the results of Druker and Anderson (2010) and a review of the implicit nature of spatial association learning by Chun and Jiang (1998).

Appendix D

Post-Questionnaire

0) Please briefly describe what your experience of the illusion was like. Which features of the illusion, if any, changed—if anything changed, how, and when?

1) During your first experimental (not practice) viewing of the illusion, approximately how often did the circle to which you chose to pay attention appear darker than its two neighbours? Your response should fall on a scale from 1 to 7, with 1 being "Never" and 7 being "Every single time".

2) During your second experimental (not practice) viewing of the illusion, approximately how often did the circle to which you chose to pay attention appear darker than its two neighbours? Your response should fall on a scale from 1 to 7, with 1 being "Never" and 7 being "Every single time".

3) During your first experimental (not practice) viewing of the illusion, did you purposefully use a strategy to decide the order in which you chose to pay attention to the circles? If you used a strategy, please describe it:

4) During your second experimental (not practice) viewing of the illusion, did you purposefully use a strategy to decide the order in which you chose to pay attention to the circles? If you used a strategy, please describe it:

5) While completing the reaction time task, did you notice any underlying patterns (any kind of pattern whatsoever)? If you detected a pattern, please describe it:

6) Did one colour of circle appear more frequently than the other colour? (e.g., were there more circles of a certain colour overall?) If you detected a difference in frequency, please describe this difference:

7) If you were forced to guess, which colour of circle appeared most often during the reaction time task?

8) Imagine you are at the start of a trial of the reaction time task. You are staring at the small black dot in the center of the screen. It disappears, and a black circle target is about to appear somwhere else on the screen. Was this target equally likely to fall in any location on the screen? (Yes or no).

9) If you were forced to guess, relative to the center of the screen, coloured circles were most likely to fall in which region of the screen?

10) Any final comments or concerns regarding your experience of the experiment?

Appendix E

Illusion Task Instructions

E.1 Experiment 1

The center of the screen will be indicated by a small white square. This square will fall within the center of three overlapping grey circles. Please remember: At all times during this phase of the experiment, it is very important that you stare constantly, and exclusively, at this white square. If your eyes drift at any point, please return to looking at the center square immediately. At any time during the duration of this phase of the experiment, you will be free to choose to pay attention to any one of the three overlapping grey circles.

Paying attention to one of the circles may cause its appearance to change. Specifically, it may become darker, relative to the other two circles. You are free to choose to direct your attention to any of the lighter circles, and are equally free to change circles when or if you feel like it.

If at any point in time, you begin to pay attention to the LEFT circle: Press the LEFT ARROW key. If at any point in time, you begin to pay attention to the RIGHT circle: Press the RIGHT ARROW key. If at any point in time, you begin to pay attention to the BOTTOM circle: Press the DOWN ARROW key.

E.2 Experiment 2

The center of the screen will be indicated by a small white dot. This dot will fall within the center of three overlapping grey circles. Please remember that, although you should blink as often as you feel the need to, during this phase of the experiment, it is very important that you stare constantly at the white dot. If your eyes drift at any point, please return to looking at the white dot immediately. Staring at the white dot is important; however, your main task is as follows:

1) Direct your attention to any one of the circles without moving your eyes off of the white dot.

2) Use the keyboard to indicate the circle to which you have begun to pay attention.

3) Whenever you feel like it, for any reason, select a different circle to pay attention to.

For the duration of this phase of the experiment: You are to proceed in this manner of selecting different circles, and indicating each time you begin to pay attention to a new circle.

Whenever you begin to pay attention to the LEFT circle: Press the LEFT ARROW key.

Whenever you begin to pay attention to the RIGHT circle: Press the RIGHT ARROW key.

Whenever you begin to pay attention to the BOTTOM circle: Press the DOWN ARROW key.

References

[1] Anderson, B. (2011). There is no such thing as attention. *Frontiers in Psychology,* 2(246), 1-8.

[2] Baayen, R. (2008). Analyzing linguistic data: A practical introduction to statistics using R. Cambridge, UK: Cambridge University Press.

[3] Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structures for confirmatory hypothesis testing: Keeping it maximal. *Journal of Memory and Language,* 68(3), 255-278.

[4] Carpenter, R. H. S. (1988). Movements of the eyes (2nd ed.). London: Pion.

[5] Carrasco, M. (2011). Visual attention: The past 25 years. *Vision Research,* 51(13), 1484-1525.

[6] Carreiro, L. R. R., Haddad Jr., H., & Baldo, M. V. C. (2003). The modulation of simple reaction time by the spatial probability of a visual stimulus. *Brazilian Journal of Medical and Biological Research,* 36, 907-911.

[7] Corbetta, M. (1998). Frontoparietal coritical networks for directing attention and the eye to visual locations: Identical, independent, or overlapping neural systems? *Proceedings of the National Academy of Science,* 95, 831-838.

[8] Corbetta, M., & Shulman, G. L. (2011). Spatial neglect and attention networks. *Annual Review of Neuroscience,* 34, 569-599.

[9] Chun, M. M., & Jiang, Y. (1998). Contextual cueing: Implicit learning and memory of visual context guides spatial attention. *Cognitive Psychology,* 36(1), 28-71.

[10] Druker, M., & Anderson, B. (2010). Spatial probability aids visual stimulus discrimination. *Frontiers in Human Neuroscience,* 4(63).

[11] He, P., & Kowler, E. (1989). The role of location probability in the programming of saccades: Implications for Center-of-Gravity tendencies. *Vision Research,* 29(9), 1165-1181.

[12] Hoffman, M. D., & Gelman, A. (2011). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. arXiv:1111.4246

[13] Gelman, A., Hwang, J., & Vehtari A. (2013). Understanding predictive information criteria for Bayesian models. arXiv:1307:5928.

[14] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). Bayesian data analysis (3rd ed.). Boca Raton, FL: CRC Press.

[15] Geng, J. J., & Behrmann, M. (2005). Spatial probability as an attentional cue in visual search. *Perception & Psychophysics,* 67, 1252-1268.

[16] Jabar, S. & Anderson, B. (2017). Not all probabilities are equivalent: Evidence from orientation versus spatial probability learning. *Journal of Experimental Psychology Human Perception & Performance,* 43(5).

[17] Jaynes, E. T. (2003). Probability theory: The logic of science. G. L. Bretthorst (Ed.). Cambridge University Press.

[18] Jiang, Y. V., Sha, L., & Remington, R. (2015). Modulation of spatial attention by goals, statistical learning, and monetary reward. *Attention, Perception & Psychophysics,* 77(7).

[19] Maljkovic, V., & Nakayama, K. (1996). Priming of pop-out: II. Role of position. *Perception & Psychophysics,* 58, 977-991.

[20] Meyberg, S., Sinn, P., Engbert, R., & Sommer, W. (2017). Revising the link between microsaccades and the spatial cueing of voluntary attention. *Vision Research,* 133, 47-60.

[21] Pearl, J. (2009). Causality: Models, reasoning, and inference. (2nd ed.). New York, NY: Cambridge University Press.

[22] Peirce, J. W. (2009). Generating stimuli for neuroscience using psychopy. *Frontiers in Neuroinformatics,* 2, 1-8.

[23] Popper, K. R. (1979). Objective knowledge: An evolutionary approach. Oxford University Press.

[24] Kingstone, A., & Klein, R. M. (1993). What are human express saccades? *Perception & Psychophysics,* 54(2), 260-273.

[25] Kustov, A. A., & Robinson, D. L. (1996). Shared neural control of attentional shifts and eye movements. *Nature,* 384(6604), 74-77.

[26] Pastukhov, A., Vonau, V., Stonkute, S., & Braun, J. (2013). Spatial and temporal attention revealed by microsaccades. *Vision Research,* 85, 45-57.

[27] *R Core Team. (2018).* R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.

[28] Rizzolati, G., Riggio, L., Dascolo, I. & Umilta, C. (1987). Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia,* 25, 31-40.

[29] Schall, J. D., & Hanes, D. P. (1993). Saccade latency in context: Regulation of gaze behaviour by supplementary eye field. *Behavioral and Brain Sciences,* 16, 553-610.

[30] *Stan Development Team. (2018a).* RStanArm: Bayesian applied regression modeling via Stan. R package version 2.17.4.

[31] *Stan Development Team. (2018b).* ShinyStan: Interactive visual and numerical diagnostics and posterior analysis for Bayesian models. R package version 2.5.0.

[32] Szucs, D., & Ioannidis, J. P. A. (2017). When Null Hypothesis Significance Testing is Unsuitable for Research: A reassessment. (2017). *Frontiers in Human Neuroscience,* 11(390).

[33] Tse, P. U. (2005). Voluntary attention modulates the brightness of overlapping transparent surfaces. *Vision Research,* 45, 1095-1098.

[34] Tse, P. U, Caplovitz, G. P., & Hsieh, P. -J. (2006). Microsaccade directions do not predict directionality of illusory brightness changes of overlapping transparent surfaces. *Vision Research,* 46, 3823-3830.

[35] Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's Statement on p-Values: Context, process, and purpose. *The American Statistician,* 70(2), 129-133.

[36] Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing,* 27(5), 1413-1432.