Adapting to Change: The Role of Priors, Surprise and Brain Damage on Mental Model Updating

by

Alexandre Leo Stephen Filipowicz

A thesis

presented to the University of Waterloo

in the fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Psychology

Waterloo, Ontario, Canada, 2017

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Dr. Suzanna Becker Professor
Supervisor(s)	Dr. James Danckert Professor
	Dr. Britt Anderson Associate Professor
Internal Member	Dr. Stephanie Denison Assistant Professor
Internal-external Member	Dr. Paul Marriott Professor

Author's Declaration

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

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

Abstract

To make sense of the world, humans build mental models that guide actions and expectations. These mental models need to be receptive to change and updated when they no longer accurately predict observations from an environment. Although ubiquitous in our everyday lives, research is still uncovering the factors that guide mental model building and updating. A significant challenge arises from the need to characterize how mental models can be both robust to noisy, stochastic fluctuations, while also being flexible to environmental changes. The current thesis explores this trade-off by examining some of the main components involved in updating. Chapter 2 proposes a novel task to measure the influence of prior mental models on the way new information is integrated. Chapter 3 tests the role of unexpected, 'surprising' events on our ability to detect changes in the environment. Chapter 4 measures the strategies used to explore new mental models, after a change has been detected, and how specific forms of brain damage influence these strategies. The results from this thesis provide novel insights into the behavioural and neural mechanisms that underlie mental model updating. The last chapter situates these results in existing literature, and suggests directions for future research.

Acknowledgements

I would like to begin by thanking my supervisors, Dr. James Danckert & Dr. Britt Anderson, for their tremendous support and guidance throughout my graduate studies. You have both been instrumental in developing my skills as a researcher and I could not have asked for better academic mentors. I would also like to thank the members of my defense committee, Dr. Suzanna Becker, Dr. Paul Marriott, and Dr. Stephanie Denison, for offering to read and comment on my thesis work. Your time and effort are deeply appreciated.

To the present/past members of the DAAG Lab and Brittlab, thank you for your years of comments, advice, and camaraderie. I would particularly like to thank Nadine Quehl for all of her help with patient recruitment, and Dr. Elisabeth Stöttinger (my unofficial third supervisor), Dr. Derick Valadao, and Syaheed Jabar for the stimulating (and often challenging) discussions that were crucial in developing to the work presented in this thesis.

To my parents, Christine and Stephen, thank you for always encouraging my interests and providing me with an enriching environment in which to grow. To my brother Josef, thank you for keeping me in check. To my children William and Madeleine, thank you for lighting my life and being there when I needed you most.

And finally and most importantly, to my wife Candice, thank you for your incredible love and support throughout these often trying times. None of this would have been possible without you.

For June.

Table of Contents

List of Tables

List of Figures

“La science ne connaît pas de pays, parce que la connaissance appartient à l'humanité, et elle est la torche qui illumine le monde./ Science knows no country, because knowledge belongs to humanity, and is the torch that illuminates the world.”

- Louis Pasteur

Chapter 1: Introduction

Humans have a remarkable ability to compress complex, noisy, and dynamic information into coherent *mental models* of the world (Johnson-Laird, 2013; Johnson-Laird, 2004; Tenenbaum, Kemp, Griffiths, & Goodman, 2011). We use these models to inform everyday decisions: if I leave at 7:00 a.m. will I make it to work on time? Should I invest in a particular stock?

However, the world is noisy and the information we receive from our environment often exceeds our perceptual capacities (Barlow, 1961; Summerfield & Tsetsos, 2015; Bach & Dolan, 2012). Given these perceptual limitations, our mental models are, by necessity, incomplete representations of larger concepts (Johnson-Laird, 1983; Tenebaum et al., 2011, Bach & Dolan, 2012). Therefore, an equally important ability involves *updating* mental models when faced with new information. That is, mental models are not static – they need to be receptive to environmental changes and adapt accordingly.

For example, consider a batter in baseball. To predict the next pitch, the batter relies on a mental model built from prior observations of the pitcher's throwing patterns. Using this mental model, the batter may predict the next pitch will be a fastball. If the pitch is a fastball, the batter's model is confirmed and no updating is required. If, instead, a curveball is thrown, something unexpected has happened in the context of the batter's mental model. Here the batter must determine whether the unexpected pitch represents noise (no updating required), or conversely, an important data point that needs to either be integrated into the old model (fine tuning) or used to create a new mental model.

Given the importance of updating in human learning and decision making, considerable research has attempted to understand its component processes (Wilson, Nassar, & Gold, 2010; Glaze, Kable, & Gold, 2015; Collins & Koechlin, 2012; Behrens, Woolrich, Walton, & Rushworth, 2007; Stöttinger, Filipowicz, Danckert, & Anderson, 2014; Danckert, Stöttinger, Quehl, Anderson, 2012; McGuire, Nassar, Gold, Kable, 2014; O'Reilly et al., 2013). From this research, updating can be characterized as having three main stages: first a mental model must be *built* in order to compare it with events from the environment (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Griffiths & Tenenbaum, 2006; Tenenbaum et al., 2011; Green, Benson, Kersten, & Schrater, 2010). Next, to update a model, it is necessary to *detect mismatches* between a current mental model and the observations it is meant to predict (O'Reilly et al., 2013; Nassar et al., 2010; Glaze et al., 2015; Wilson et al., 2010). Finally, once a mismatch has been detected, it becomes necessary to *explore* alternative mental models, which can explain the discrepancies between what was expected and what is observed (Collins & Koechlin, 2012; Donoso, Collins & Koechlin, 2014; Stöttinger, Filipowicz, Danckert, et al., 2014). Once a new, more accurate mental model has been identified, this model is then compared with observations and the cycle continues.

Although previous research agrees on these main updating components, questions still remain about how these different processes operate. A large body of research demonstrates that humans are extremely proficient at learning from the statistics in their environment (Saffran, Aslin, & Newport, 1996; Nissen & Bullemer, 1987; Jueptner, Stephan, et al., 1997; Jueptner, Frith, Brooks, Frackowiak, & Passingham, 1997; Orbán, Fiser, Aslin, & Lengyel, 2008; Toni, Krams, Turner, & Passingham, 1998; Turk-Browne,

2005). This ability has led some researchers to propose that the mechanisms that drive mental model building and updating follow optimal rules of probability (Chater, Tenenbaum, & Yuille, 2006; Costello & Watts, 2014; Fiser, Berkes, Orbán, & Lengyel, 2010; Griffiths & Tenenbaum, 2006; Lewandowsky, Griffiths, & Kalish, 2009; Nassar, Wilson, Heasly, & Gold, 2010; Tenenbaum et al., 2011). Others note that this *optimal* interpretation of mental model building and updating is difficult to reconcile with research demonstrating consistently *suboptimal* use of probabilistic information (Bowers & Davis, 2012a; Jones & Love, 2011; Nickerson, 1998; Dawson & Abkes, 1987; Gigerenzer & Gaissmaier, 2011; Hilbert, 2012; Tversky & Kahneman, 1981; Tversky & Koehler, 1994; Kahneman, 2011).

The current thesis seeks to clarify some of these disparate findings related to mental model building and updating. Each chapter examines one of the of the three main updating component processes. Chapter 2 introduces a novel task to measure how *prior beliefs* influence mental model *building*, while addressing some of the limitations evident in previous studies. Chapter 3 goes on to measure how unexpected, or *surprising* events influence our ability to detect changes in our environment. Lastly, Chapter 4 examines different strategies used to *explore* alternative mental models, after a change has been detected, and how specific forms of brain damage influence these processes.

Together, the results from this thesis provide insights into the processes involved in the different stages of updating. The final chapter summarizes the primary results from this thesis while discussing limitations and suggestions for future research.

Chapter 2: The Influence of Priors on Mental Model Building[1]

2.1. Introduction

To measure updating, it is important to accurately represent the mental model a person is using to interpret information in the environment. The importance of this is not trivial: research shows that the models we use to interpret events can have a significant impact on decision making (Green, et. al, 2010; Hogarth & Einhorn, 1992; Lee & Johnson-Laird, 2012; Park & Sloman, 2014; Patrick & Ahmed, 2014; Stöttinger, Filipowicz, Danckert, et al., 2014; Collins & Koechlin, 2012). But how do we measure a participant's mental model? Chapter 2 introduces some of the challenges faced when representing participant expectations, and presents a novel method to address limitations found in previous research.

A common way to represent a person's expectations is to examine response trends over multiple trials (Jueptner, Stephan, et al., 1997; Jueptner, Frith, et al., 1997; Nissen & Bullemer, 1987; Robertson, Tormos, Maeda, & Pascual-Leone, 2001; Toni et al., 1998; Vulkan, 2000). Participant responses are aggregated over multiple trials to measure how closely they represent observed events. However, while such measures enable exploration of how closely participants manage to *match* task contingencies, they give only limited information as to the *mental models* that drive responses. As highlighted by Stöttinger and colleagues (2014), data trends from individual responses alone can result from a

[1] A version of this chapter has been published as Filipowicz, A., Valadao, D., Anderson, B., & Danckert, J. (2014) Measuring the influence of prior beliefs on probabilistic estimations. In P. Bello, M. Guarini, M. McShane, & B. Scassellati (Eds.), *Proceedings of the 36th Annual Conference of the Cognitive Science Society* (pp. 2198-2203). Austin, TX; Cognitive Science Society. It is reproduced here with permission.

number of different mental models that may be unknown to the experimenter (Stöttinger, Filipowicz, Marandi, Quehl, Danckert, & Anderson, 2014).

Recent work has made use of probabilistic methods to gain a detailed representation of participant mental models. A large body or research demonstrates that humans have a keen ability to learn from statistics in the environment (Nissen & Bullemer, 1987; Mayr, 1996; Jueptner, Stephan, et al., 1997; Jueptner, Frith, Brooks, Frackowiak, & Passingham, 1997; Orbán et al., 2008; Toni, Krams, Turner, & Passingham, 1998; Filipowicz, Danckert, & Anderson, 2014; Saffran et al., 1996; Turk-Browne, 2005). This ability has led some researchers to imply that cognitive processes, such as mental model building and updating, follow optimal rules of probabilities (Chater et al., 2006; Costello & Watts, 2014; Fiser et al., 2010; Griffiths & Tenenbaum, 2006; Lewandowsky et al., 2009; Nassar et al., 2010; Tenenbaum et al., 2011). Under these frameworks, mental models are represented as *probability distributions*, where events in an environment are expected based on how frequently they were experienced in the past (Griffiths & Tenenbaum, 2006; Knill & Pouget, 2004; Tenenbaum et al., 2011; O'Reilly et al., 2013; O'Reilly, 2013; Summerfield & Tsetsos, 2015; Collins & Koechlin, 2012; Costello & Watts, 2014; Behrens et al., 2007).

One prominent framework proposes that mental models are built and updated following approximations of Bayes' theorem:

$$P(M \mid E) = \frac{P(M) \times P(E \mid M)}{P(E)} \qquad (2.1)$$

where $P(M|E)$ – the *posterior probability* – represents the probability that a mental model M is true given the evidence E; $P(M)$ – the *prior probability* – represents the probability that a mental model is true before observing E; $P(E|M)$ – the *likelihood* – represents the probability that E would have been obtained from the mental model M; and $P(E)$ represents the probability of the evidence E.

Tracking participant mental models is at the core of Bayesian models of cognition, and research demonstrates that in certain circumstances, the behavioural and neural processes underlying updating behave in Bayesian-like ways (Doya, Ishii, Pouget, & Rao, 2007; Nassar et al., 2010; Wilson et al., 2010; O'Reilly et al., 2013; Griffiths & Tenenbaum, 2006; Knill & Pouget, 2004; Tenenbaum et al., 2011). This approach, however, has been met with criticism. It is at odds with a large body of research in decision making demonstrating *suboptimal* use of probabilistic information (Nickerson, 1998; Dawson & Abkes, 1987; Gigerenzer & Gaissmaier, 2011; Hilbert, 2012; Tversky & Kahneman, 1981; Tversky & Koehler, 1994; Kahneman, 2011). Furthermore, critics have noted that some Bayesian models rely too heavily on specific assumptions to obtain an optimal fit (Bowers & Davis, 2012a, 2012b; Jones & Love, 2011; see Griffiths, Chater, Norris, & Pouget, 2012 for a response).

One main point of contention concerns the ways in which Bayesian models characterize *priors* – the mental models participants bring to a task *before* observing any new events. As demonstrated by Bowers & Davis (2012a), the success of a Bayesian model depends heavily on which prior is selected, and unprincipled prior selection can be used to fit nearly any pattern of data. A common approach to characterize priors rests on the assumption that participants begin an unfamiliar task in a state of maximal

uncertainty, where all options are equally probable (Bestmann et al., 2008; Harrison, Duggins, & Friston, 2006; Mars et al., 2008; Strange, Duggins, Penny, Dolan, & Friston, 2005). A participant's expectations regarding each event's occurrence are then updated over time relative to the frequency with which distinct events are observed. Although plausible, the notion that *all* participants begin a task without *any* biases or expectations is often unaccompanied by any empirical evidence.

Another approach is to characterize a participant's beliefs by fitting their responses to distributions approximating the events they are estimating (Griffiths & Tenenbaum, 2006; Lewandowsky et al., 2009; Nassar, Rumsey, Wilson, Prikh, Heasly, & Gold, 2012; Nassar et al., 2010; McGuire et al., 2014; O'Reilly et al., 2013). A participant's prior is assumed to match the overall shape of the distribution of events they are estimating, and the prior's parameters are updated as new events occur. As with assumptions of maximal uncertainty, participant priors are rarely measured *before* observing any events. As already noted, this method has also been criticized as it relies heavily on the selected prior chosen for a given model – a prior selected by the researchers (Bowers & Davis, 2012a; Jones & Love, 2011).

Regardless of how they are characterized, methods using uniform or parametric priors assume that participant priors are *homogeneous* – that is, all participants start a novel task with the *same* prior mental models. This is an important assumption, but one that is not often explicitly tested. Indeed, certain statistically *suboptimal* behaviours have been accounted for through a better understanding of idiosyncratic participant priors. For example, studies have shown that the statistically suboptimal behaviour associated with *probability matching* (Vulkan, 2000; Koehler & James, 2009) depends on participant

beliefs regarding the mechanisms generating the events they are predicting (Green et al., 2010), and the strategies used to interpret stochastic events (Otto, Taylor, & Markman, 2011). Additionally, simple manipulations of participant priors can reduce or eliminate these suboptimal behaviours (Shanks, Tunney, & Mccarthy, 2002). This shows that the mental models participants hold *before* observing events, and the way these beliefs evolve *after* observing new events strongly impact decisions (Stöttinger Filipowicz, Danckert, et al., 2014; Collins & Koechlin, 2012; Lee & Johnson-Laird, 2013).

In summary, questions remain about some of the assumptions used to measure mental models. The current chapter seeks to address these limitations by directly measuring both the priors participants bring to a task, and how new information is integrated over time. Using a task based on the game 'Plinko', participants estimated the likelihood that balls would fall in a range of slots. Estimates were made in the form of probability distribution that could be changed with new information. In this task, estimates were provided before any events were observed, making it possible to measure participant priors, whether these priors differ between individuals, and how these priors influence learning.

2.2. Experiment 2.1: Methods

Participants

One hundred and nineteen University of Waterloo undergraduates (76 female, mean age = 19.57 years, SD = 2.22 years) participated in Experiment 2.1 in exchange for course credit. The University of Waterloo's Office of Research Ethics approved the study protocol and participants gave informed consent before participating.

Plinko task environment

Mental models were measured using a computerized version of the game 'Plinko'[2]. The task environment was programmed in Python using the PsychoPy library (Pierce, 2009). Unless otherwise specified, the general Plinko task environment described below was identical in Experiment 2.1 from the current chapter, and Experiments 3.1 and 3.2 in chapter 3.

During the task, a red ball fell through a triangle of pegs into one of 40 possible slots. The triangle consisted of 29 rows of black pegs that increased in number from the top to the bottom of the triangle (*i.e.*, the top row contained 1 peg and the bottom row contained 29 pegs). A rectangle was located below the 40 slots spanning their width. Participants were instructed to make their responses in this space (Figure 2.1).

[2] Based on the game found on the American game show *The Price is Right* (http://www.thepriceisright.com).

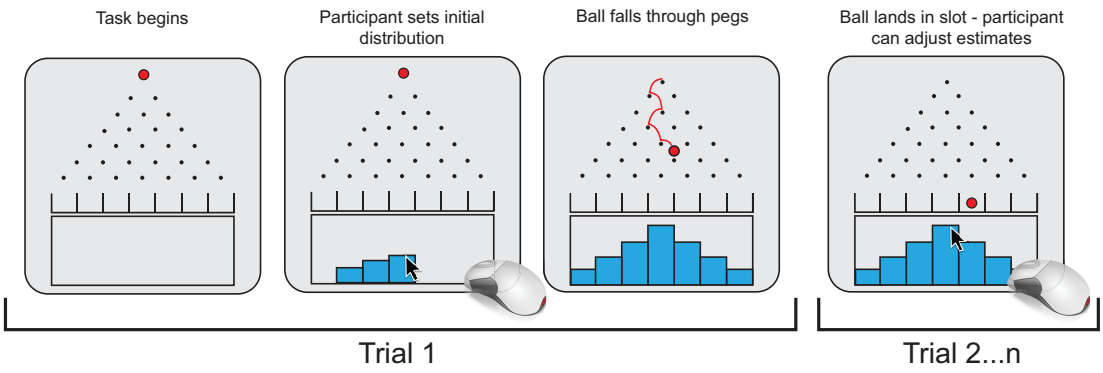

Figure 2.1. Schematic of Plinko Task. Participants were informed that a red ball would fall through a triangle of pegs and land in one of a number of slots below (note: 7 slots are pictured here for illustrative purposes). They were instructed to draw bars using the computer mouse to indicate the most likely locations the ball would land in – higher bars indicate an expectation of higher likelihood. They drew their first set of bars before seeing any ball drops, and could adjust their bars at the start of each subsequent trial.

Bars were drawn using the computer mouse: the height of the bars could be adjusted by holding down the left mouse button and dragging the cursor. The height of the bar would match the position of the cursor within the limits of the rectangle below the slots. Participants could also erase any single bar by right-clicking with the cursor on the bar they wished to delete, or by clicking the backspace key to delete all bars on screen. The bars were not assigned any value; participants were simply told that taller bars represented a higher probability that a ball would fall in a slot, shorter bars a lower probability, and no bars represented zero probability. Participants were informed that they had the option of adjusting their bars at the start of every trial and that they had to have at least one bar on screen before proceeding with the trial. Once participants had indicated their likelihood estimates, they pressed the spacebar to proceed with the trial (Figure 2.1).

After completing the task, participants completed a brief questionnaire, asking if they had noticed any structure to the locations of the ball drops, if they had noticed that the structure had changed at any point during the task, and to elaborate on any strategies they used to adjust their bars.

Experimental conditions

Participants were exposed to a sequence of ball drops generated from one of three distributions (note that the mean and variance refer to slot numbers): 1) a Gaussian[3] with a mean of 18 and standard deviation of 6 (Wide Gaussian 1), 2) a Gaussian with a mean

[3] It is important to note that the distributions of ball drops participants were exposed to, along with the distributions participants drew as a part of the Plinko task, were all discrete in nature, and thus not true Gaussian distributions. The term "Gaussian", used here and later in this thesis to describe sequences of ball drops or participant responses, refers to discrete *approximations* of Gaussian distributions.

of 17 and standard deviation of 6 (Wide Gaussian 2), 3) a Gaussian with a mean of 17, and a standard deviation of 1.9 (Narrow Gaussian). Participants in each separate distribution condition were exposed to identical sequences of events, and provided their estimates *before* seeing any ball drops. They could then modify these estimates as they saw new events. Participants were not informed that there was any structure to the distribution of ball drops.

Measuring accuracy

Participant bars were normalized by dividing the height of each bar by the sum of the heights of all bars drawn on screen. This normalization process provided a discrete probability distribution on each trial for every participant.

Accuracy A was measured as the proportion of overlap between a participant's drawn distribution on any trial and the discrete distribution of ball drops they were being presented with. The proportion of overlap was calculated by summing the minimum probability value x of every slot i between a participant's distribution P on any trial t and a computer's distribution C for any block of trials j:

$$A_t = \sum_{i=1}^{40} \min\left(P(x_{it}), C(x_{ij})\right) \qquad (2.2)$$

Participant accuracy could range between 0 and 1, with 0 indicating no overlap between the participant's distribution and the computer's distribution for a particular block, and 1 indicating perfect overlap.

Measuring learning rates

Once participant accuracy was calculated on each trial, standard exponential learning curves were fit to participant accuracy scores over time (*e.g.*, Estes, 1950; Healthcote, Brown, & Mewhort, 2000; Ritter & Schooler, 2001):

$$\hat{A}_t = a_\infty - (a_\infty - a_0)e^{-\alpha t} \qquad (2.3)$$

where t denotes the trial number, $\hat{A}_t$ a participant's estimated accuracy on trial t, a_0 a participant's starting accuracy, a_∞ asymptotic accuracy, and α a constant rate coefficient to capture how quickly participants reached their asymptote from their starting accuracy. This function was fit to each participant's accuracy scores using a nonlinear least squares function in the R statistical package ('nls' function; R Core Team, 2014). Given that participant accuracy could only range between 0 and 1, the function's lower and upper limits for a participant's minimum starting accuracy and maximum asymptote value were set to 0 and 1 respectively. To characterize participant performance over time, the estimated starting accuracy was used to represent a participant's starting accuracy value, learning rate to capture how quickly they reached their asymptote from their starting value, and a participant's estimated accuracy on the last trial of the distribution they were estimating to indicate a final level of accuracy achieved.

The estimated last trial accuracy was obtained by computing a participant's accuracy using equation (2.3), with the participant's three fit parameters (asymptote, starting accuracy, and learning rate), and the value for t set to the total number of trials in the distribution they were estimating. Since participants were exposed to each distribution

for 100 trials, the value of *t* was always set to 100. This last parameter was used instead of the asymptote value because the asymptote reflects a participant's maximal estimated accuracy after an unspecified number of trials, not necessarily the estimated accuracy at the end of a finite block of trials. For example, a participant with a fitted starting value of .5, an asymptote of .95, and a learning rate of .01 would only reach asymptotic performance after more than 1000 trials, whereas their performance after 100 trials, the length of each of the distributions participants were asked to estimate in the current experiment, would be a closer to .78.

Characterizing participant prior mental models

To characterize prior mental models, the distributions participants drew on trial one, before having seen any ball drops, were classified into different 'prior groups' based on distribution shape. Prior groups were first categorized using a density clustering algorithm (Rodrigez & Laio, 2014). This clustering method provided the general ability to separate between participants with uniform priors (*i.e.*, where equal or near equal probability was assigned to each slot) and Gaussian priors (*i.e.*, when participant estimates approximated Gaussian distributions). However, as described below, a large proportion of participants started the task with a type of *bimodal* prior, where low probability was assigned to middle slots, and high probability to the left and right of central slots. In some cases, the number of slots separating the two modes was very small, and the clustering algorithm failed to reliably distinguish bimodal priors from Gaussian priors. As such, to better capture the variability in participant priors, a qualitative approach was used to classify participants into different prior groups.

2.3 Experiment 2.1: Results

Participants exhibit heterogeneous prior mental models

Initially two types of prior mental models were expected. Those familiar with statistics may know that the distribution of ball drops in a fair Plinko game should be approximately Gaussian, with a mean centered on the ball's starting position (Galton, 1889). This assumption fits with research assuming participant priors match actual event distributions (Griffiths & Tenenbaum, 2006; Nassar et al., 2010). Alternatively, those unfamiliar with the task may choose a state of maximal uncertainty, selecting a uniform distribution as their prior. This assumption fits with research characterizing participant priors as uniform across all possible options (Harrison et al., 2006; Mars et al., 2008; Strange et al., 2005).

Of the 119 participants in both studies, 36 (30%) participants began with a Gaussian-like distribution: 33 participants drew Gaussian distributions centered around the middle of the row of slots whereas three drew Gaussian distributions centered to the right of the row of slots with either a negative or positive skew (labeled as 'Skewed'). Of the remaining 83 (70%) participants, 11 (9%) began with a uniform distribution, 33 (28%) indicated a bimodal distribution, and 39 (33%) began with 'jagged' distributions, where initial estimates were minimally composed of a few interspersed bars (in some cases only one bar; Figure 2.2). This conclusively demonstrates that, far from being homogeneous, participants come to the task with varying prior beliefs about where the balls would fall.

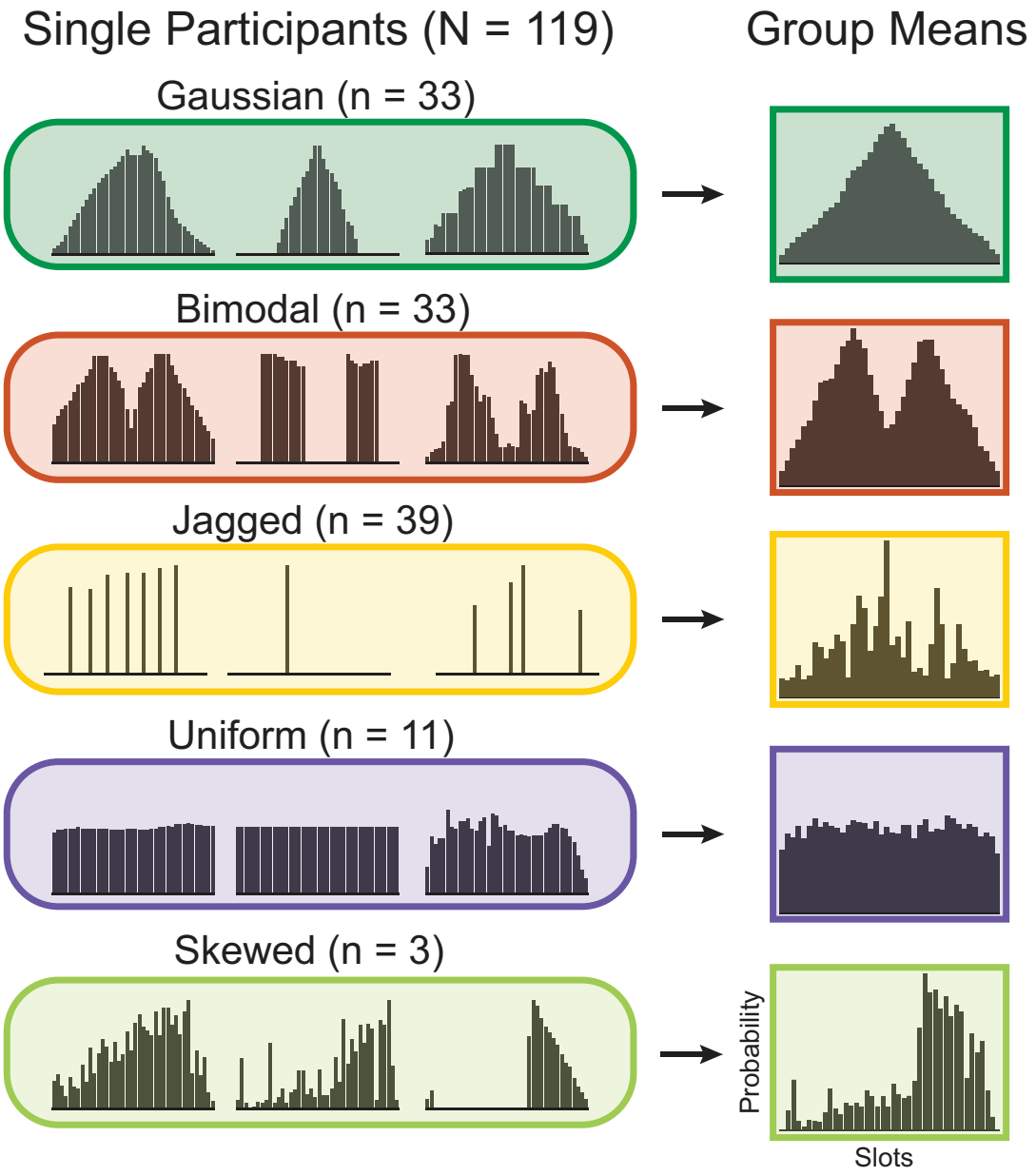

Figure 2.2. Participant Prior Distribution Groups. Estimates participants provided *before* seeing any ball drops were categorized based on their shape. The left panels provide individual exemplars of responses on the first trial whereas the right panels displays group mean estimates for each category. The number of participants (n) in each group are included next to the group titles.

Prior group performance when exposed to Wide Gaussian distributions

Performance in each prior distribution group was first examined for the two wide Gaussian distributions. Of the 41 participants exposed to the first wide Gaussian, 17 indicated they expected to see a Gaussian, seven indicated they expected a Bimodal distribution, and 13 indicated they expected a Jagged distribution. Of the 39 participants exposed to the second wide Gaussian, nine participants indicated they expected to see a Gaussian, 14 indicated they expected a Bimodal distribution, and 12 indicated they expected a Jagged distribution.

Repeated measures ANOVAs with Trial Accuracy as a dependent variable, Trial Number (1 to 100) as a within subject factor, and Prior Group (Gaussian, Bimodal, and Jagged) as a between subject factor revealed main effects in both wide distribution conditions of Trial Number (Wide Gaussian 1: $F(99,3366)=18.157$, $MSE=.063$, $p<.001$; Wide Gaussian 2: $F(99,3168)=23.908$, $MSE=.068$, $p<.001$) and Prior Group (Wide Gaussian 1: $F(2,34)=7.575$, $MSE=3.457$, $p<.002$; Wide Gaussian 2: $F(2,32)=7.33$, $MSE=3.505$, $p<.003$), along with Trial Number x Prior Group interactions (Wide Gaussian 1: $F(198,3366)=4.135$, $MSE=.014$, $p<.001$; Wide Gaussian 2: $F(198,3168)=4.225$, $MSE=.012$, $p<.001$). These results indicate that learning rates differed across the three main groups when exposed to wide Gaussian distributions.

To examine these interactions more closely participant exponential curve fit parameters were compared between the three prior groups using one-way ANOVAs on each parameter. For both wide Gaussian distributions, estimated starting values differed between prior distribution groups (Wide Gaussian 1: $F(2,34)=41.33$, $MSE=.600$, $p<.001$; Wide Gaussian 2: $F(2,32)=26.9$, $MSE=.393$, $p<.001$). Tukey's Honest Significant

Difference (HSD) post-hoc tests confirmed that participants in the Jagged group started with lower accuracy than both the Gaussian and Bimodal groups (Wide Gaussian 1 and 2: all ps<.001), and that there were no differences between the starting values of the Gaussian and Bimodal groups (Wide Gaussian 1: p=.195; Wide Gaussian 2: p=.415). Learning rates did not differ between prior distribution groups (Wide Gaussian 1: $F(2,34)=0.284$, MSE=.023, p=.754; Wide Gaussian 2: $F(2,32)=1.368$, MSE=.029, p=.269), and although estimated last trial accuracy tended to be lower in the Jagged group compared to the Gaussian and Bimodal groups, their estimated last trial accuracy was only found to be significantly lower when exposed to the first wide Gaussian ($F(2,34)=3.911$, MSE=.019, p<.03; Tukey's HSD - Gaussian-Jagged: p<.05, Bimodal-Jagged: p=.07; Bimodal-Gaussian: p=.927) but not when exposed to the second wide Gaussian ($F(2,32)=1.734$, MSE=.011, p=.193; Figure 2.3; Table 2.1).

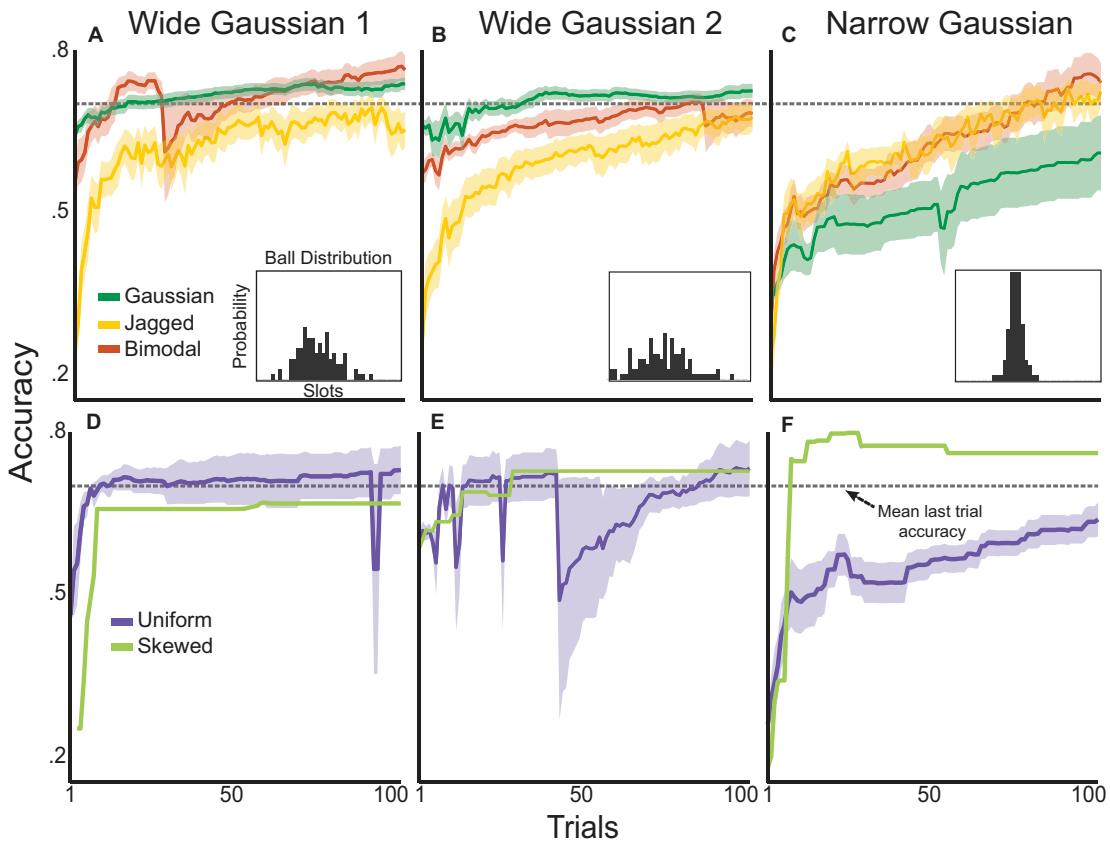

Figure 2.3. Prior distribution group performance for each distribution (Wide Gaussian 1, Wide Gaussian 2, Narrow Gaussian). Graphs A-C display performance of the three largest groups (*i.e.*, Gaussian, Jagged, and Bimodal); Graphs D-F display performance of the two smaller prior distribution groups (*i.e.*, Uniform and Skewed). The histograms in graphs A-C represent the distribution of ball drops participants were exposed to in each condition. The solid lines represent mean accuracy scores, and shading represents ± 1 standard error. The dotted lines represent the mean last trial accuracy averaged across all participants in all conditions.

Table 2.1. Mean performance parameters for prior distribution groups.

	Wide Gaussian 1			Wide Gaussian 2			Narrow Gaussian		
Prior Groups	SV	LR	LTA	SV	LR	LTA	SV	LR	LTA
Gaussian	.63	.14	.73	.62	.06	.72	.37	.02	.59
Bimodal	.56	.22	.75	.53	.05	.69	.33	.07	.72
Jagged	.27	.21	.67	.22	.14	.66	.28	.08	.70
Uniform	.47	.17	.70	.59	.03	.68	.32	.05	.62
Skewed	.00	.19	.67	.56	.07	.73	.00	.20	.77

Note. SV: Start Value; LR: Learning Rate; LTA: Last Trial Accuracy

Performance from participants in the Uniform and Skewed groups was qualitatively examined given their small size. Of the participants exposed to the first wide Gaussian, three participants indicated they expected a uniform distribution and one expected a positively skewed Gaussian with a mean to the right of center. Participants in the Uniform condition had priors that covered a large number of slots leading to a starting accuracy that was slightly lower than the starting accuracy of the Bimodal and Gaussian groups (mean starting value = .47). However, participants in this group adjusted their distributions rapidly (mean learning rate Uniform group = .17) to match the higher probability of ball drops occurring near the center-left while also lowering or deleting bars under peripheral slots. By the end of the block, participants with Uniform priors ended with similar estimated last trial accuracy (mean = .70) as participants in the other groups (Figure 2.3D).

Of those exposed to the second wide Gaussian, three participants indicated a Uniform prior and one indicated they expected a negatively skewed Gaussian with a mean to the right of center. The performance of the three participants with a Uniform prior was difficult to fit using exponential curves because at certain points throughout the task two of the participants made large changes to their distributions, events not well fit by exponential curves (Figure 2.3E). This strategy was best characterized as 'wiping the slate clean' (see spikes on trials 14, 26 and 43 in Figure 2.3E), followed by rapidly improving estimates of the distribution of ball drops being shown – a strategy shown to be quite successful for one participant who finished with a final accuracy of .81.

In both wide distribution conditions, participants with Skewed priors tended to start with low estimated accuracy (mean start value = .29) given that the bulk of their

expected distributions were to the right of center, whereas the computer's distribution fell slightly left of center. Nevertheless, this group rapidly reached an accuracy level similar to the Gaussian and Bimodal groups (Figure 2.3D, E).

Prior group performance when exposed to Narrow Gaussian distribution

The influence of priors was next examined when exposed to the narrow Gaussian distribution. Once again, performance differences were first examined between the three main prior distribution groups. Of the 39 participants exposed to the narrow Gaussian, seven indicated they expected to see a Gaussian, 12 indicated they expected a Bimodal distribution, and 14 indicated they expected a Jagged distribution. A factorial ANOVA with Trial Accuracy as a dependent variable, Trial Number (1 to 100) as a within subject factor, and Prior Group (Gaussian, Bimodal, Jagged) as a between subject factor revealed no significant Trial Number x Prior Group interaction ($F(198,2970)=1.117$, $MSE=.006$, $p=.133$), and no main effects of Prior Group ($F(2,30)=2.287$, $MSE=2.717$, $p=.119$). Separate one-way ANOVAs revealed no significant differences between prior groups in estimated start values ($F(2,30)=.829$, $MSE=.020$, $p=.446$), learning rate ($F(2,30)=.667$, $MSE=.009$, $p=.520$), or estimated last trial accuracy, despite the Gaussian group finishing with a nominally lower mean last trial accuracy than the other groups ($F(2,30)=2.04$, $MSE=.037$, $p=.148$; Figure 2.3C; Table 2.1).

Five participants indicated that they expected a Uniform distribution, and one reported expecting a negatively Skewed distribution with a mean to the right of center. The five participants with uniform priors performed much like those in the larger groups, with similar starting values, learning rates, but slightly lower last trial accuracy (Table

2.1; Figure 2.3F). The one participant reporting a Skewed prior performed particularly well: despite an estimated starting value of 0 (*i.e.*, no overlap between their estimate and the computer's distribution), within 8 trials this participant reached an accuracy of .75 and remained within this range for the rest of the task (Figure 2.3F).

Evidence was also found for an opposite effect, such that the prior expectations of some participants seemed to *impair* their ability to adjust to the narrow distribution. Of the seven participants indicating Gaussian priors, three ended with raw accuracy scores between .41 and .45, the lowest among all participants exposed to the narrow distribution. These participants began with Gaussian priors centered on the middle of the array (mean = 20.93) with large variance (average SD = 8.79; *i.e.*, expected ball drops spanning 16 slots around a mean of slot 21). Although they did shift the means of their estimates to better match the computer's distribution (average last trial mean=17.24), they did not adjust *the variance* of their estimates to reflect the narrow variance in the computer's distribution (average last trial SD=7.22). In other words, although these participants correctly estimated the area of highest probability of the computer's distribution, they did not adapt to the *absence* of events in the outer limits of their estimates.

These results were not true of all participants with a Gaussian prior: the other four participants in the Gaussian group started with similarly wide estimates (average first trial mean = 20.63; average first trial SD = 8.17), but ended with higher raw accuracy scores, ranging from .63 to .90. That is, the accuracy achieved by the latter group was derived from the fact that they *narrowed* the variances of their distribution (average last trial SD = 2.78).

2.4. Experiment 2.1: Discussion

The goal of Chapter 2 was to develop a task that allows for a detailed representation of participant mental models, and test some of the assumptions of previous updating models. Experiment 2.1 was meant to capture the prior mental models participants bring to a task, and examine how these priors influence their ability to learn distributions of events.

Contrary to methods that assume homogeneous participant priors, the mental models participants brought to a task differed between participants. In this particular task, priors tended to cluster into three main groups: Gaussian, Bimodal, and Jagged. Priors from the first two groups were likely based on information gleaned at the start of the task. While drawing their initial distributions, participants could see the triangle of pegs the ball would fall through, and could see that the ball's start position was centered on the top peg. Given that the expected distribution of a Plinko table is known to approximate a Gaussian distribution (Galton, 1889), participants in the Gaussian group may have used this knowledge to draw their distributions. This is consistent with some participant responses in the post-experimental questionnaire. Thirteen of the participants in the Gaussian prior group reported expecting the ball pattern to be "shaped like a bell-curve" or "normally distributed". Participants in the Bimodal group seemed to behave in a similar way, but using a more direct assessment of their environment to set their bars. Three of the participants with Bimodal priors mentioned that the ball's starting position would make it more likely for it to fall left or right of center, and less likely to land *in* the center.

Conversely, the estimates made by the Uniform and Jagged groups seemed to have resulted from uncertainty about task contingencies. This is most easily explained in the Uniform group: given that participants had never been exposed to this task and had not observed any events, they assigned equivalent probability to all possible outcomes – a strategy that matches the general assumption made by researchers who represent maximal uncertainty using uniform priors. Participants indicating Jagged priors also seemed to be expressing a form of uncertainty, but in a different way than participants with Uniform priors. Six participants in the Jagged group reported drawing a few bars on screen for the sole purpose of getting more information about the task. After having observed the first trial, the majority of participants in the Jagged group that provided information about the strategies they used to adjust their distributions (30 of 38) indicated that they used a type of frequency matching strategy, where they adjusted bars every few trials to match where the ball had landed. Although participants in this group generally started each distribution with a low starting accuracy, they eventually reached the same level of accuracy as the groups starting with higher accuracy, although in some circumstances (*i.e.*, estimating the wide Gaussian), taking longer to do so.

At an individual level, some priors were found to help or hinder learning. Some participants that started with priors that contrasted strongly with the computer's distribution seemed to update rapidly (*e.g.*, participants with 'Skewed' priors). This suggests that strong mismatches between participant priors and the events they must predict may influence the efficiency of updating mental models. There were also situations in which participants had more trouble adapting their distributions to reflect the *absence* of predicted events. When exposed to wide-Gaussian distributions, participants

with wide-Gaussian priors were at an advantage, starting the task with a high degree of accuracy. However, when this same group was exposed to a narrow-Gaussian distribution, approximately half of the participants had difficulty representing the narrowed variance of the distribution they observed. These types of effects on updating are explored more thoroughly in Chapter 3.

Taken together, these findings demonstrate that participants start a task with a number of different priors, and that these priors influence how effectively they learn the probabilistic contingencies underlying events. Some participants may start with priors based on prior knowledge or educated guesses, while others may start with few to no assumptions, choosing instead to seek evidence of the events they are predicting. In either case, using a method that requires fewer assumptions, the current results do not support the notion that participants start a task with homogenous priors. They also highlight the importance of capturing idiosyncratic priors that participants bring to a task if we are to appropriately characterize the factors driving their learning and updating behaviour.

Chapter 3: The Role of Surprise in Change Detection[4]

3.1. Introduction

In addition to building accurate mental models of the world, it is equally important for these models to be updated when faced with environmental changes. One of the difficulties for an observer is to know when new observations reflect a relevant change in the environment, and when they reflect random variation: if a bus comes 10 minutes late, does this reflect a change to the schedule, or simply a 'one-off' event? Predicting *when* and to *what extent* a model has been updated continues to be a difficult challenge (Glaze et al., 2015; O'Reilly, 2013). Whereas Chapter 2 explored the influence of priors on the way mental models are built, Chapter 3 examines some of the factors that drive updating when the environment itself is changing.

Research demonstrates that we pay close attention to *surprising* information when judging probabilities (Fisk, 2002), and the concept of *surprise* plays an important role in current studies of updating (Nassar et al., 2010; McGuire et al., 2014; O'Reilly et al., 2013; Mars et al., 2008). In the context of learning and decision making, surprise describes our reaction to an unexpected and/or novel event, particularly one that is *contrasted* with another, more expected event (Teigen & Keren, 2003). This definition implies that surprise is a subjective experience, one that depends on a person's current expectations. Thus, surprise can only properly be measured insofar as a person's *expectations* can be measured. Indeed, updating does not occur in a vacuum – it depends largely on the mental model an observer is using to interpret their environment

[4] A version of this chapter has been published as Filipowicz, A., Valadao, D., Anderson, B., & Danckert, J. (2016) Rejecting Outliers: surprising changes do not always improve belief updating. *Decision*. Advanced online publication. DOI: 10.1037/dec000007. It is reproduced here with permission.

(Stöttinger Filipowicz, Danckert, et al., 2014; Collins & Koechlin, 2012; Lee & Johnson-Laird, 2012).

Previous research measuring the effect of surprise on updating has generally approximated mental models based on participant responses. As highlighted in Chapter 2, these approximations are often obtained by building ideal observers (O'Reilly et al., 2013; Mars et al., 2008), or by fitting participant responses to computational models (*e.g.*, Bayesian change-point models: Nassar et al., 2010, 2012; McGuire et al., 2014). A measure of surprise is then obtained by measuring the discrepancy between these mental model approximations, and the observations the mental model is attempting to predict – the larger the discrepancy, the higher the calculated surprise of the event. This research has consistently found that participants update more quickly with increasing discrepancies between their predictions and current observations, suggesting that surprise is positively and linearly related to updating (Nassar et al., 2010, 2012; McGuire et al., 2014).

There are, however, some questions related to the ubiquity of this relationship. Do we *always* update when faced with surprising events? As outlined in Chapter 2, Bayesian models of updating inherently assume that *all* new information is integrated equally in a mental model. As such, these models should, by design, *always* find a positive monotonic relationship between surprise and updating. In some cases however, we treat discrepant information with a sort of skepticism, and discount it when building a representation of our environment (De Gardelle & Summerfield, 2010). For example, when attempting to classify an array of objects based on color, participants were found to base their responses more on coherent objects in the array while rejecting the contribution of items

that deviated strongly from the rest (De Gardelle & Summerfield, 2010). Although these rules have primarily been found in studies of human perception, some researchers have argued that these tendencies are also present in decision making, leading us to sometimes treat highly surprising events as a type of 'outlier' (Summerfield & Tsetsos, 2015). Similarly, studies relating to heuristic decision making strategies find consistent examples of *confirmation biases,* situations in which surprising information is ignored in favor of information that supports a current mental model (Nickerson, 1998; Hollard & Massoni, 2015; Miller, Spengler & Spengler, 2015). This suggests that rather than blindly integrating *any* surprising information, there may be situations in which we can be *resistant* to highly surprising changes.

The current chapter explores the relationship between surprise and updating in more detail. Using the Plinko task introduced in Chapter 2 to capture detailed measures of mental models, participants were exposed to two experiments in which distributions of ball drops changed at an unannounced point. Experiment 3.1 was meant to gain a general sense of how effectively participants could use the Plinko task to detect and update to changes. Experiment 3.2 specifically explores the relationship between surprise and updating. In contrast to prior work, updating was not *always* positively related to the degree of surprise. Instead, there were some situations in which surprise and updating were *negatively* correlated, such that, rather than integrate highly surprising events, participants devalued them.

3.2. Experiment 3.1: Methods

Participants

Thirty-seven undergraduates from the University of Waterloo (21 female; mean age = 20.07 years, SD = 2.06 years) participated in Experiment 3.1, with 18 participants assigned to a *break* condition (10 female, mean age = 19.55, SD = 1.50 years), and 19 participants to a *continuous* condition (11 female, mean age = 20.58 years, SD = 2.06 years; conditions explained below).

Experimental conditions

All participants completed a Plinko task in which they saw sequences of 400 ball drops that, unbeknownst to them, changed every 100 trials. The sequences were generated from four distinct probability distributions: 1) a Gaussian distribution with a mean of 18 and standard deviation of 6 (Wide Gaussian; note: the mean and variance measures refer to slot numbers), 2) a Gaussian distribution with a mean of 31 and standard deviation of 2 (Narrow Gaussian), 3) a bimodal distribution generated by mixing two Gaussians with different means (10 and 28 respectively) and the same standard deviation of 3 (Bimodal), and 4) a Weibull distribution with a shape parameter of 1 and scale parameter of 6 (Skewed). All participants were exposed to the same order of distributions: Wide Gaussian, Narrow Gaussian, Bimodal, and Skewed.

Participants were assigned to either a *break* condition or a *continuous* condition. In the *break* condition, participants were exposed to each distribution in separate blocks. The purpose of this condition was to determine how effectively participants could learn each of the different distributions they were exposed to, and to limit the influence of

previous estimations. At the end of each block, participants were given an accuracy score that reflected their performance on the just completed distribution (*i.e.*, their accuracy score after the most recent 100 trials). They then pressed a computer key when ready to begin the next block. At the start of each new block, a participant's probability estimates (*i.e.*, the bars they had drawn) were erased. New estimates were made (*i.e.*, they were required to draw a new set of bars) before being exposed to the next distribution of ball drops.

In the *continuous* condition, participants were exposed to the same order and sequences of ball drops as in the break condition, but were not given any breaks, accuracy feedback or other indicators between each of the distributions. The purpose of this condition was to have participants update to the same distribution changes as in the break condition, but relying solely on the information provided by new ball drops to update their estimates. In both conditions, participants were not given any information about the distributions they would be estimating and were not informed that any changes to the distributions would occur.

Computing surprise

As highlighted by Tiegen & Keren (2003), an event is only surprising insofar as it is contrasted with other, more expected events. Therefore, attempts to measure surprise need to take into account not only the probability assigned to an event, but how this probability contrasts with the probability assigned to other possible events. Information theory provides a definition of surprise that accounts for this contrast: the surprise, or 'surprisal', is defined as *the negative log probability* of an event occurring under a

distribution of expectations (cf. Shannon, 1948; Attneave, 1959). This quantity represents the information contained by a particular event: the higher the surprisal, the more information is contained by the event that cannot be accounted for by a current mental model. This method has commonly been used to quantify surprise in learning tasks, primarily when participant expectations are characterized as continuous probability distributions (O'Reilly et al., 2013; Mars et al., 2008; Strange et al., 2005; Doya et al, 2007).

Given that participant distributions in the Plinko task were both discrete, and that participants were not required to draw bars under every slot (*i.e.*, potentially leaving some slots with a value of 0), a pure measure of negative log odds could not be computed to characterize surprise. Instead, two complementary measures of surprise were used and compared to account for the discrete nature of participant distributions.

The first measure used a modification of participant slot probability values to make them compatible with a measure of negative log probability. This was done by replacing slot values equal to 0 with an arbitrarily small number $1x10^{-10}$, which approaches 0, but still has a natural logarithm.

As a second measure, *weighted empirical log odds* were used to compute surprise (Cox & Snell, 1989). The advantage of this measure is that it can be applied directly to discrete distributions, and does not require any modifications to participant responses. Although not entirely identical, odds and probabilities both provide information about likelihoods, and make predictions about how expected certain events are to occur.

To compute the *weighted* empirical log odds of a given probability value the empirical log odds (Elog) need to computed. The Elog E of any probability p is defined as:

$$E_p = \ln\left(\frac{p+0.5}{1-p+0.5}\right) \tag{3.1}$$

If the frequency r is known of a specific event compared to the total number of events n that were used to compute p, E can also be defined as:

$$E_r = \ln\left(\frac{r+0.5}{n-r+0.5}\right) \tag{3.2}$$

Elog can be weighted depending on the number of observations n that are used to compute p. The variance V of Elog is defined as:

$$V = \frac{(n+1)(n+2)}{n(r+1)(n-r+1)} \tag{3.3}$$

of which the inverse can be used to compute a weight W:

$$W = \frac{1}{V} \tag{3.4}$$

The weighted empirical log odds $\hat{E}$ for any probability value in slot i can be computed as

$$\hat{E}_i = E_i \times W_i \tag{3.5}$$

Since each distribution consisted of 100 ball drops, the weights for the wElog calculation were computed using variance as expressed in equation (3.5) with $n = 100$ for each distribution.

Using these two measures to quantify surprise, a 'surprise factor' S was computed for a shift from any first distribution j to any second distribution k as the sum of the ratio of the surprise s of each slot i (computed using either NLP or wElog) of each of the two distributions:

$$S_{jk} = \sum_{i=1}^{40} \frac{s_{ij}}{s_{ik}} \tag{3.6}$$

This formula computes shifts that *include* unexpected events to be more surprising than shifts that *omit* previously observed events. This quantification of surprise was used to calculate how surprising each distribution shift was to each participant. With this calculation of surprise it is important to note that *higher* levels of surprise are represented by *larger* values of S when calculated using NLP, and *smaller* values of S when calculated using wElog.

3.3. Experiment 3.1: Results

Updating accuracy is worse with no breaks

Performance differences were examined between the two conditions. Note that the objective evidence, the timing, and location of ball drops were identical in both conditions. If participants were simply using the ball drops as evidence and serially updating, there should be no difference between conditions. A mixed factorial ANOVA with Trial Accuracy as a dependent variable, Distribution Trial Number (1 to 100) and Distribution Type (Wide Gaussian, Narrow Gaussian, Bimodal, Skewed) as within subject factors, and Condition (Break, Continuous) as a between subject factor revealed main effects of Trial Number ($F(99,3465)=116.38$, $MSE=.007$, $p<.001$), Distribution Type ($F(3,105)=17.82$, $MSE=.64$, $p<.001$), and Condition ($F(1,35)=49.88$, $MSE=2.30$, $p<.001$), and a three-way Trial Number x Distribution Type x Condition interaction ($F(297,10395)=5.063$, $MSE=.004$, $p<.001$), indicating that both the rate and accuracy with which each distribution was learned varied between distribution types across the two conditions.

Separate factorial ANOVAs were computed to compare changes in trial accuracy between conditions for each distribution type. No main effects of Condition and no Trial Number x Condition interaction were expected for performance in the Wide Gaussian distribution given that this was the first distribution all participants were exposed to. However, main effects of Condition and Trial Number x Condition interactions were expected in each of the subsequent distributions.

As expected, when comparing performance on the Wide Gaussian, there was a main effect of Trial Number ($F(99,3465)=18.137$, $MSE=18.137$, $p<.001$), but no main

effect of Condition ($F(1,35)=1.85$, $MSE=.423$, $p=.182$) and no Condition x Trial Number interaction ($F(99,3465)=.504$, $MSE=.004$, $p>.99$). However, in each of the subsequent distributions, in addition to main effects of Trial Number (for each distribution type all $Fs>55$, all $ps<0.001$), and Condition (across all distributions all $Fs>28$, all $ps<.001$), there were Trial Number x Condition interactions (across all distributions all $Fs>5$, all $ps<0.001$; Figure 3.1A).

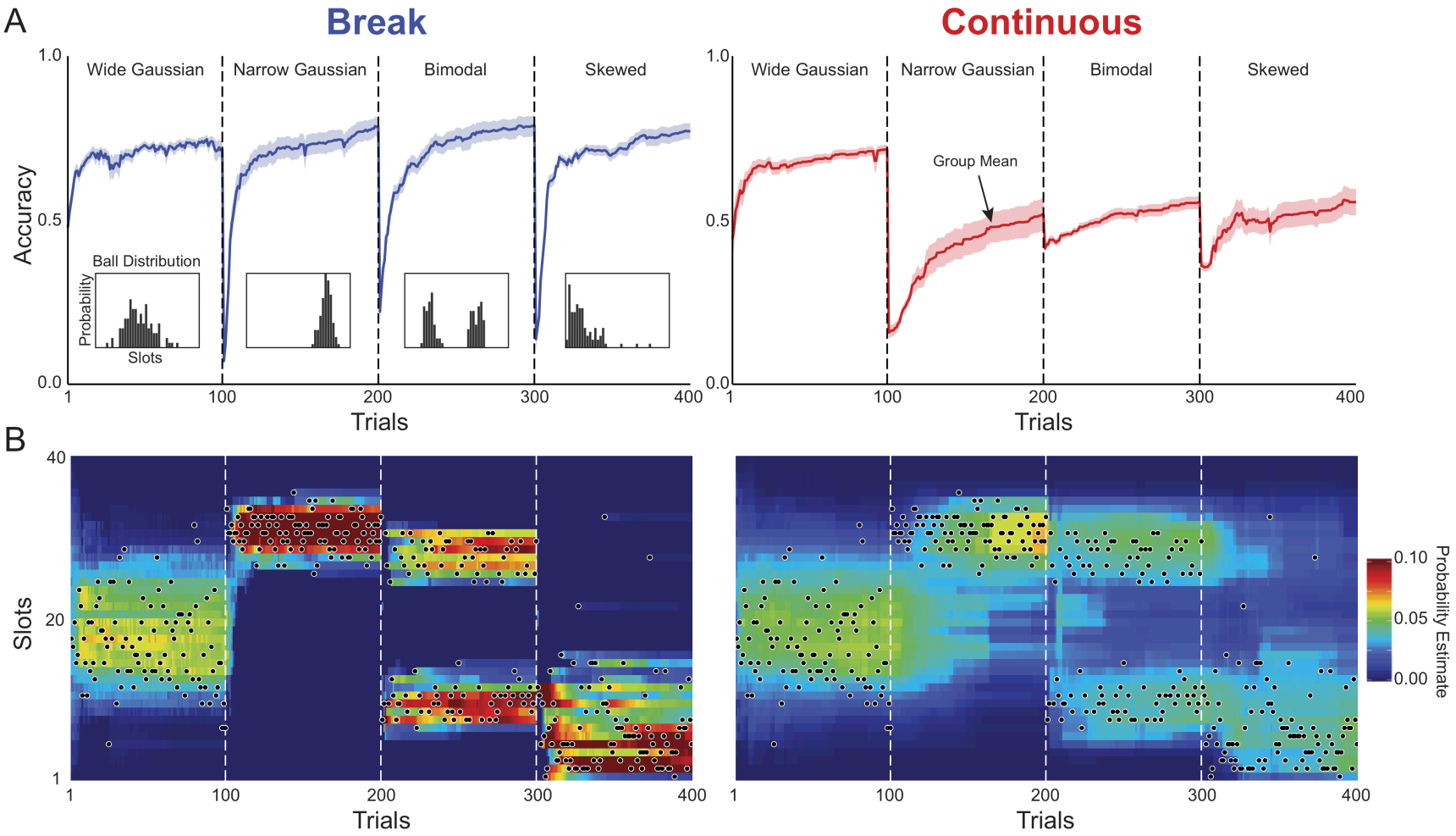

Figure 3.1. Participant learning performance between break and continuous conditions. Panel A displays learning performance between participants assigned to the *break* condition (blue lines) and those in the *continuous* condition (red lines). The four probability distributions participants were exposed to (*i.e.*, Wide Gaussian, Narrow Gaussian, Bimodal, and Skewed) are inset in the left graph. The solid lines represent mean accuracy for each condition while shading represents ± 1 standard error. Panel B displays the mean probability values participants assigned to each slot on each trial. The black dots represent the slot in which a ball fell on each trial.

To explore these differences further, fit parameters were compared between conditions. Since no performance differences were observed between conditions when participants were exposed to the Wide Gaussian distribution (*i.e.*, the first distribution), no differences were expected between fit parameters for this distribution. Corrected t-tests showed no differences between conditions in the Wide Gaussian distribution on estimated last trial accuracy, learning rate, or starting values (all ps > .21). However, last trial accuracy and learning rates differed between conditions in all subsequent distributions (all ps < .012), with steeper learning rates and higher last trial accuracy in the break vs. the continuous condition. Starting values were different between conditions in the Bimodal and Skewed distributions, with starting values higher in the continuous than in the break condition (all ps < .001), but did not differ in the Narrow Gaussian distribution (p = .15; Table 3.1).

Table 3.1. Experiment 3.2 estimated performance parameters.

	Wide Gaussian			Narrow Gaussian			Bimodal			Skewed		
Conditions	SV	LR	LTA	SV	LR	LTA	SV	LR	LTA	SV	LR	LTA
Break	.47	.12	.73	.06	.14	.75	.17	.14	.78	.13	.13	.75
Continuous	.41	.23	.71	.12	.04	.52	.41	.03	.55	.31	.06	.55

Note. SV: Start Value; LR: Learning Rate; LTA: Last Trial Accuracy

Participants can represent a number of distinct distributions

Next, performance was examined to determine how accurately participants could represent different distributions. In the *break* condition, repeated measures ANOVAs comparing learning curve parameter fits between the four Distribution Types (Wide Gaussian, Narrow Gaussian, Bimodal, and Skewed) revealed no differences in estimated last trial accuracy ($F(3,54)=1.15$, $MSE=.006$, $p=.338$) or mean learning rates ($F(3,54)=.31$, $MSE=.009$, $p=.818$; Table 3.1). There were differences between mean start value estimates, primarily due to initial estimates in the Wide Gaussian distribution matching the ball drops they would be exposed to with higher accuracy ($F(3,54)=22.2$, $MSE=.028$, $p<.001$). These results demonstrate that participants can effectively represent a number of different distributions with equivalent accuracy (Figure 3.1A).

Participants in the continuous condition demonstrate more hysteresis

The same analysis was conducted on performance in the *continuous* condition. If participants are able to abandon their beliefs when faced with environmental change, last trial accuracy should be equivalent across all distributions, similar to the results from the *break* condition. If, however, previous events exerted an influence on updating, this should be reflected in differences in performance metrics between distributions. A repeated measures ANOVA on performance parameters in the *continuous* condition demonstrated significant differences in estimated last trial accuracy ($F(3,51)=9.714$, $MSE=.130$, $p<.001$), leaning rate estimates ($F(3,51)=4.874$, $MSE=.154$, $p<.005$), and starting values ($F(3,51)=9.714$, $MSE=.130$, $p<.001$; Figure 3.1A; Table 3.1). Taken together, these results demonstrate that, when given a break between distributions,

participants learned each of the four distinct distributions with equivalent accuracy; however, when switched continuously between distributions, participants were unable to adapt their distributions to the same level of accuracy.

These performance differences between the two conditions seemed to stem from participants in the *continuous* condition having their responses influenced by the previous events they had observed – in other words, the continuous group demonstrated a kind of *hysteresis* (cf., Hock, Bukowski, Nichols, Huisman, & Rivera, 2005) in which the just learned distribution exerted an influence on their capacity to learn new distributions. As is evident in Figure 3.1B, when switched to new distributions, participant estimates from the previously observed distribution lingered for a substantial number of trials after the switch. In contrast, participants in the break condition show little influence from previous events, primarily representing the probability of slots that had more recently received ball drops.

Surprise positively predicts updating in the continuous condition

Next, analyses were performed to examine the influence of surprise on a participant's ability to update in the continuous condition. Surprise factors were calculated with both the modified NLP and wElog measures, using the distributions participants had drawn *before* being shifted to a new distribution as the numerator in equation 3.6 (*i.e.*, distribution *j*), and the new distribution participants would be subsequently exposed to as the denominator (*i.e.*, distribution *k*). To estimate the influence of surprise on updating, these surprise factors for each shift were compared with how accurately participants managed to subsequently represent the new distribution.

This comparison was measured using linear mixed effects models (one using the NLP measure and one using wElog to compute the value of s in equation 3.6; linear mixed effects models were analyzed using the lme4 R package; Bates, Mächler, Bolker, & Walker, 2015) that measured the influence of a surprise factor, *prior* to observing a new distribution, with a participant's mean accuracy on this new distribution. These models included random intercepts and random slopes for each subject. Significance was measured by comparing the models to null models that assumed no influence of surprise on mean accuracy.

As predicted by previous research, surprise was positively related to updating both using the NLP surprise metric (b = .001 ± SEM .001; $\chi^2(1) = 4.12$, $p < .05$, $R^2 = .06$) and the wElog surprise metric (b = -.01 ± SEM .003; $\chi^2(1) = 15.56$, $p < .001$, $R^2 = .28$; Figure 3.2).

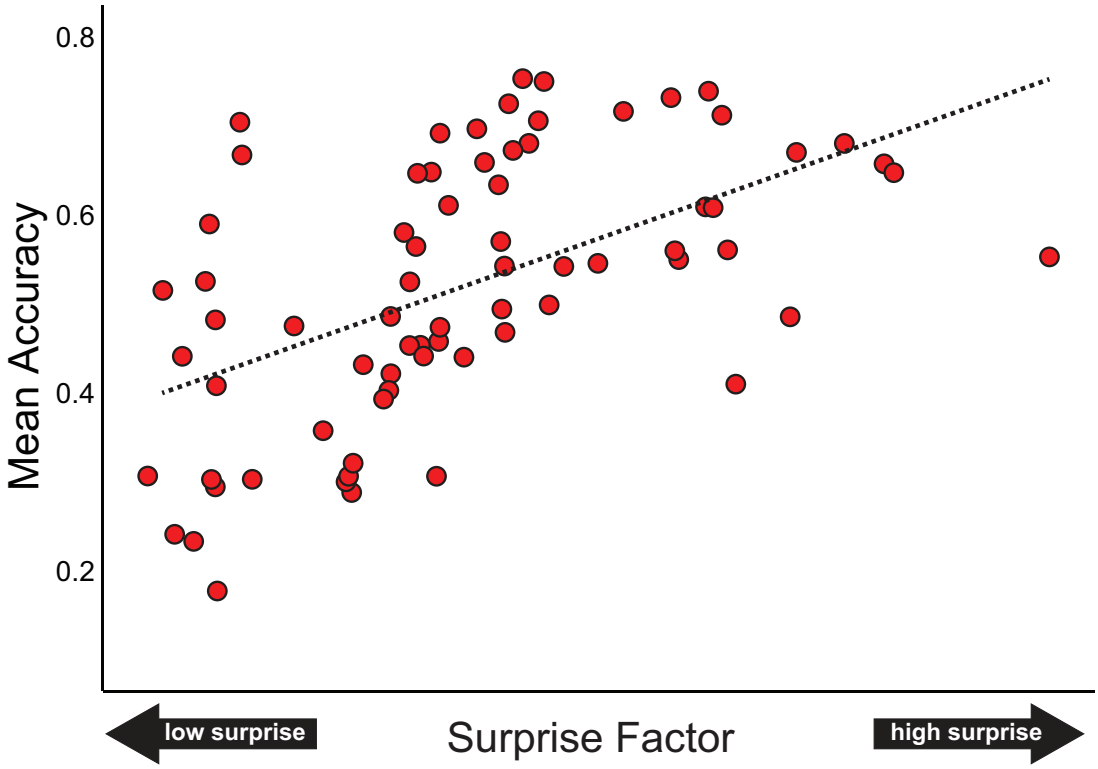

Figure 3.2. Relationship between surprise factor (wElog) and updating in the continuous condition. A surprise factor was calculated for each shift in the continuous condition and compared to how accurately participants managed to represent each new distribution. Surprise was positively related to updating accuracy in the continuous condition. The dotted line represents the best fitting regression line.

3.4. Experiment 3.1: Discussion

Experiment 3.1 was meant to examine how participants updated in a Plinko environment where distributions of ball drops changed. When distribution changes followed breaks, participants were quickly and efficiently able to represent a number of different distributions with equivalent accuracy. This was in part due to there being very little residual influence from the distributions they had just seen. Conversely, participants in the continuous condition represented distributions that followed a shift less accurately. This poorer performance cannot simply be attributed to difficulties representing different types of distributions – participants in the break condition managed to represent all four distinct distributions with equivalent accuracy. Instead, the lack of breaks in the continuous conditions provoked a type of *hysteresis*, where participants were slower to abandon representations of observations they had previously seen.

Previous research suggests that more *surprising* shifts could improve a participant's ability to represent environmental changes. When the surprise of each shift was measured using a calculated surprise factor, as predicted by previous research, this value was positively correlated with updating accuracy. This initial result supports the notion that updating and surprise are positively correlated – the more unexpected a shift, the more participants abandon their representations of previous events in favor of new observations.

However, the distributions in Experiment 3.1 differed in a number of different respects. In all cases, the means, variances, shapes, and degree of overlap between distributions differed, making it difficult to assess the specific contribution of surprise per se on a participant's ability to detect changes and update accordingly. Experiment 3.2

addresses this limitation by providing a more controlled environment to measure the influence of surprise on updating. Distribution switches were controlled to vary on a single dimension (*i.e.*, change in mean or in variance), all providing different levels of surprise, while holding other dimensions constant (*i.e.*, shape and overlap).

3.5. Experiment 3.2: Methods

Participants

Seventy-eight undergraduates (54 female, mean age = 19.64 years, SD = 1.59 years) from the University of Waterloo participated in Experiment 3.2 in exchange for course credit. The University of Waterloo's Office of Research Ethics approved the study protocol and participants gave informed consent before participating.

Experimental conditions

All participants performed a Plinko task in which they were exposed to a first Gaussian distribution of 100 ball drops, then switched to a second Gaussian distribution of 100 balls drops without any cues to indicate that a switch had occurred. Participants began the task by being exposed to either a wide Gaussian (mean=17, SD=6), or a narrow Gaussian distribution (mean=17, SD=1.9). After 100 trials of this first distribution, participants were then switched either to a wide distribution (mean=25, SD=6), or a narrow Gaussian (mean=19.5, SD=1.9). This resulted in four between subject conditions: two mean shift conditions (narrow-narrow, and wide-wide), and two variance shift conditions (narrow-wide, and wide-narrow; Figure 3.3).

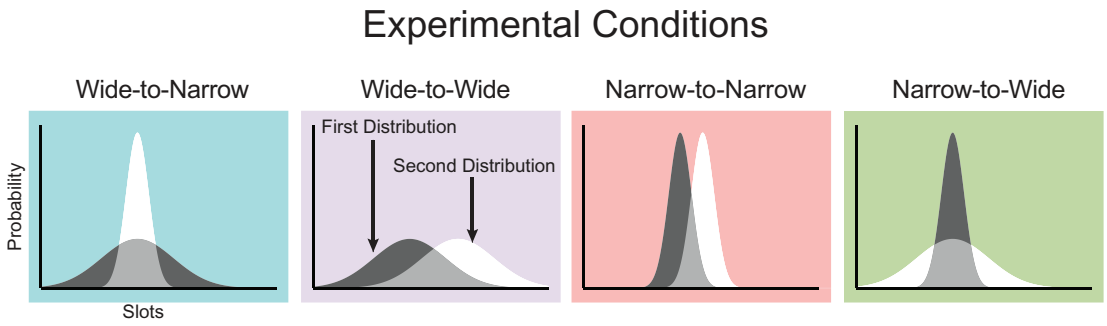

Figure 3.3. Experimental conditions – Experiment 3.2. Participants were assigned to one of four switch conditions. Participants saw a first distribution of 100 ball drops that was generated from either a wide or narrow Gaussian distribution. They were then switched to a second Gaussian distribution of 100 ball drops that either changed in mean while holding the variance constant (*i.e.*, wide-to-wide and narrow-to-narrow conditions) or changed in variance while holding the mean constant (*i.e.*, wide-to-narrow and narrow-to-wide).

In all conditions, the magnitude of each shift condition was kept equivalent, while manipulating the amount of surprise expected from each shift. Shift magnitude is defined as the percent overlap between the first and second distribution of ball drops. Overlap was computed using the same formula used to compute participant accuracy (eq. 2.2). The overlap between the continuous distributions in each of these switches was nearly identical, with roughly 50% overlap between the first and second distributions. The discrete distributions generated from these continuous distributions were also nearly identical in their overlap, with overlap between the first and second distributions ranging between 44% and 46%.

Although equivalent in their overlap, these distribution shifts varied in their calculated surprise factor. Using Equation 3.6, the surprise factor s for each shift condition was computed using both the modified negative log probability measure (NLP), and the weighted empirical log odds measure (wElog). As is evident in Table 3.2, both measures predict a similar trend for the surprise factors of each switch condition, with the highest surprise factor being predicted for the narrow-wide condition, mid-range surprise for both mean shifts (narrow-narrow and wide-wide), and the lowest surprise for the wide-narrow condition.

Table 3.2. Comparison of expected surprise measures for each switch type.

Surprise Measure	Narrow-Wide	Narrow-Narrow	Wide-Wide	Wide-Narrow
NLP	124.80	42.71	39.98	26.99
wElog*	31.51	44.86	52.00	60.02

Note. NLP: Negative Log Probability; NLO: Negative Log Odds; wElog: Weighted Empirical Log Odds; *Lower values indicate higher predicted surprise.

3.6. Experiment 3.2: Results

Updating is worst for low surprise shifts

Updating accuracy was first examined between the different surprise conditions. Separate mixed factorial ANOVAs were run for each distribution participants were exposed to (First or Second distribution), with Trial Accuracy as a dependent measure, Trial Number as a within subjects factor, and Condition as between subjects factor.

When examining performance between conditions in the first distribution, there were significant main effects of Condition ($F(3,74) = 5.852$, $MSE = 4.712$, $p < .002$) and Trial Number ($F(99,7326) = 64.749$, $MSE = .015$, $p < .001$), and a Trial Number x Condition interaction ($F(99,7326) = 3.227$, $MSE = .015$, $p < .001$), indicating that there were overall differences between groups in mean accuracy over the course of the first distribution, and that the rate at which participants managed to learn the first distribution varied between switch conditions. Performance in the second distribution also yielded main effects of Condition ($F(3,74) = 10.41$, $MSE = 11.87$, $p < .001$) and Trial Number ($F(99,7326) = 18.276$, $MSE = .038$, $p < .001$), but no Trial Number x Condition interaction ($F(99,7326) = .918$, $MSE = .004$, $p = .837$), suggesting that although mean accuracy differed between switch conditions, their accuracy changed at a similar rate over the course of the second distribution (Figure 3.4).

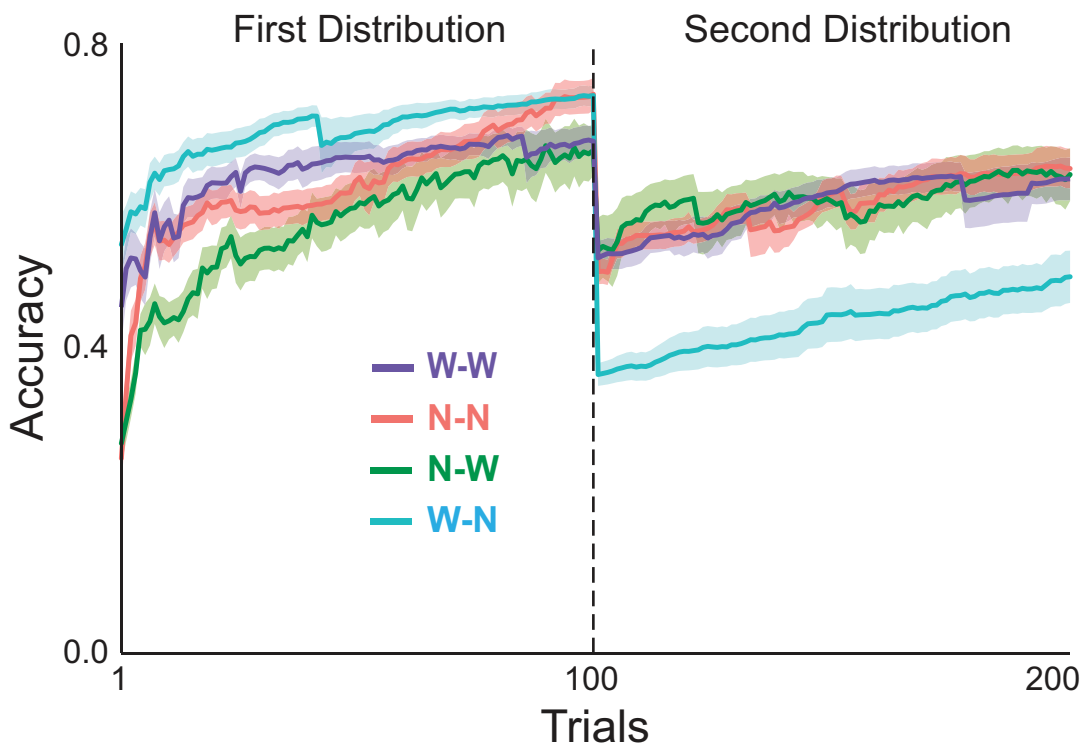

Figure 3.4. Accuracy performance for each surprise condition. Although all groups managed to learn the first distribution with equivalent accuracy, participants exposed to the low surprise, wide–narrow shift (*i.e.*, W-N; blue line) finished the second distribution with the lowest accuracy of all four groups. There were no accuracy differences between participants in the medium-surprise, narrow–narrow and wide–wide conditions (*i.e.*, N-N and W-W; purple and pink lines, respectively) and the high surprise, narrow–wide condition (*i.e.*, N-W; green line). The lines represent group means for each respective condition on each trial and shading represents ±1 standard error of the mean.

When exposed to both distributions, post hoc paired samples t-tests indicated that raw accuracy values increased from the first trial of each distribution (mean accuracy first distribution: .38, second distribution: .48), to the last trial (mean accuracy first distribution: .70, second distribution: .60; all *p*s < .001), indicating that participants were effectively learning the distributions they were observing. This demonstrates that the changes participants made to their distributions, rather than being random, were directly related to the events they observed in the task environment.

One-way ANOVAs and Tukey HSD post-hoc tests were used to examine how participant performance fit parameters differed between different switch conditions. In the first distribution, start values differed between switch conditions ($F(3,74) = 5.591$, $MSE = .181$, $p < .002$), with higher start accuracy in the wide-narrow condition of the first distribution than both the narrow-narrow and narrow-wide conditions (all *p*s <.04), and higher start accuracy in the wide-wide condition than in the narrow-wide condition ($p < .04$). However, there were no differences between conditions when examining their learning rates ($F(3,74) = .302$, $MSE = .005$, $p = .824$) or estimated last trial accuracy ($F(3,74) = 1.953$, $MSE = .021$, $p = .128$). This indicates that although the groups differed in their starting accuracy, participants in all switch conditions learned the first distribution they were exposed to with a similar level of accuracy by the end of the 100 trials (Figure 3.4; see Table 3.3 for mean performance parameters).

Table 3.3. Experiment 3.2 estimated performance parameters.

	First Distribution			Second Distribution		
Switch Type	SV	LR	LTA	SV	LR	LTA
Narrow-Narrow	.35	.07	.71	.50	.05	.63
Wide-Wide	.46	.08	.67	.51	.02	.63
Narrow-Wide	.30	.05	.65	.51	.04	.62
Wide-Narrow	.51	.07	.72	.35	.01	.48

Note. SV: Start Value; LR: Learning Rate; LTA: Last Trial Accuracy

When comparing participant performance parameters for the second distribution, there were differences in switch condition starting values ($F(3,74) = 6.308$, $MSE = .117$, $p < .001$) and estimated last trial accuracy ($F(3,74) = 5.811$, $MSE = .098$, $p < .002$). However, as indicated by the lack of Trial Number x Condition interaction in the overall ANOVA, learning rates did not differ between switch conditions ($F(3,74) = 1.451$, $MSE = .006$, $p = .235$).

As predicted, the wide-narrow switch group, which had the lowest computed surprise, had the lowest overall estimated last trial accuracy when compared to all other switch groups (all ps $< .01$) and was also the group with the lowest overall start value (all ps $< .007$). Additionally, as expected, participants in the mean shift conditions (wide-wide, narrow-narrow) did not differ on any fit parameters (all ps $> .50$). However, contrary to expectations, participants in the high surprise condition (narrow-wide), although performing better than participants in the low surprise condition, did not perform any better on any parameters than participants exposed to mean shifts (all ps $> .74$; Table 3.3).

High surprise shifts do not always lead to better updating

To understand why participants in the high surprise condition did not show any clear updating advantages over participants in the medium-surprise conditions, the same relative surprise factors computed in Experiment 3.1 were compared with updating accuracy. These individual surprise factors were computed using the distributions participants had drawn after observing all 100 trials of the first distribution as the numerator in equation 3.6, and the second computer distribution they would be exposed

to on the next 100 trials as the denominator. Each participant's surprise factor was then compared with their estimated last trial accuracy on the second distribution as an estimate of how accurately they managed to update.

As is evident in Figure 3.5, participants with both the lowest *and* highest levels of estimated surprise seemed to perform more poorly than participants with mid-range surprise values. Non-linear regressions comparing the two measures of surprise and estimated last trial accuracy on the second distribution found significant quadratic relationships when using both NLP (b^2 = -.0004, b = .0029; $t(75)$ = -5.458, p <.001, R^2 = .27) and wElog (b^2 = -.0006, b = .0058; $t(75)$ = -5.575, p <.001, R^2 = .48; Figure 3.5).

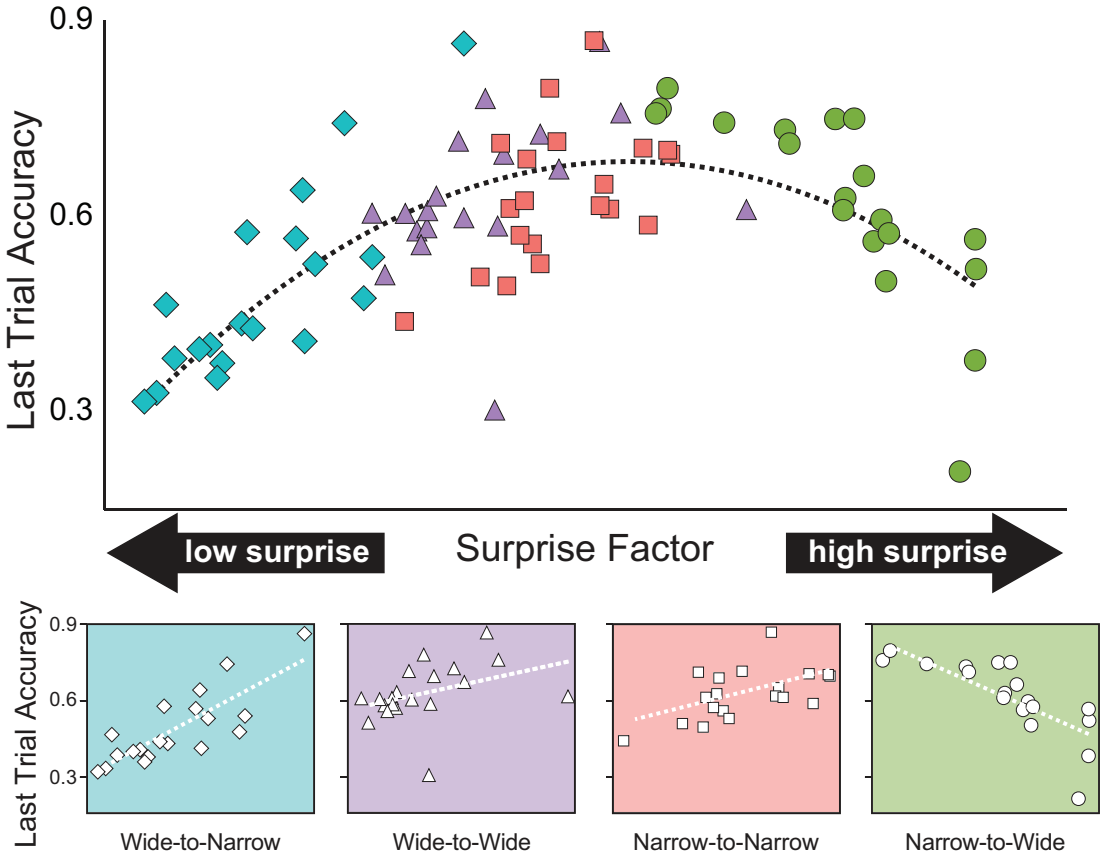

Figure 3.5. Relationship between surprise factor (wElog) and updating in Experiment 3.2. Surprise factor calculated using weighted empirical log odds and updating accuracy plotted for each participant. Updating accuracy was quadratically related to surprise factor, with both high and low surprise factor values predicting poor updating accuracy. This trend was due to a positive relationship between surprise factor and updating accuracy in the low- and medium-surprise conditions (wide-to-narrow, wide-to-wide, and narrow-to-narrow) and a negative relationship between surprise factor and updating accuracy in the highest surprise condition (narrow-to-wide). The dotted lines in the panels below the main figure represent the regression line for each individual surprise condition.

This quadratic trend seemed driven by a negative relationship between surprise and accuracy in the narrow-wide switch condition compared to all other conditions. To test this, two separate linear regressions were computed: one for participants in the low-mid surprise groups (wide-wide, narrow-narrow, and wide-narrow), and one for participants in high surprise group (narrow-wide). For the low-mid surprise groups, similar to participants in the *continuous* condition in Experiment 3.1, higher surprise values were related to better estimated last trial accuracy using both the NLP measure (b = .003; $t(57) = 3.941$, $p < .001$, $R^2 = .20$) and wElog (b = -.012; $t(57) = -6.940$, $p < .001$, R^2 = .45). In contrast, a linear regression comparing surprise and estimated last trial accuracy in the narrow-wide condition, demonstrated the *opposite* relationship, with higher surprise predicting lower estimated last trial accuracy (NLP: $b = -.002$; $t(17) = -5.456$, $p < .001$, $R^2 = .62$; wElog: $b = .022$; $t(17) = 5.206$, $p < .001$, $R^2 = .59$; Figure 3.5).

To better understand what was driving these condition differences, a closer look was taken to see how participants in each condition integrated varying levels of surprising information. The nature of our high surprise shift was to increase the variance of the second distribution relative to the first, exposing participants to balls falling in previously unused slots. Participants in the high surprise condition should therefore see a higher number of surprising events when switched to the second distribution, largely driven by balls falling in slots under which they had not drawn any bars (zero-probability slots), and should thus make larger and more frequent changes to their distributions. To examine this, the 'change magnitude' on each trial was computed as one minus the proportion of overlap between a participant's distribution on the current trial and their distribution on the previous trial (with a value of 0 indicating that the participant had

made no change to their distribution). To measure the frequency of changes, the number of instances in which participants made changes to their distributions was also counted (*i.e.*, the number of instances where change magnitude was > 0).

When learning the first distribution, participant mean change magnitude did not differ between conditions (wide-narrow = .015, wide-wide = .031, narrow-narrow = .021, narrow-wide = .054; ANOVA $F(3,74) = .901$, $MSE = .006$, $p = .445$), and while participants first exposed to wide Gaussians made nominally more frequent changes to their distributions, the mean number of changes did not differ significantly between conditions (wide-narrow = 41.6, wide-wide = 43.3, narrow-narrow = 28.5, narrow-wide = 27.8; ANOVA $F(3,74) = 2.15$, $MSE = 1336.6$, $p = .101$).

When switched to a second distribution, as expected, the mean number of balls falling in zero-probability slots differed across conditions (Mean ball drops in zero-probability slots: wide-narrow = .42, wide-wide = 4.60, narrow-narrow = 2.35, narrow-wide = 20.32; ANOVA $F(3,74) = 10.85$, $MSE = 1578.9$, $p < .001$) with participants in the narrow-wide condition experiencing more ball drops in zero-probability slots after a switch had occurred than any of the other conditions (all ps < .001). However, despite the fact that participants in the high surprise condition experienced more surprising events after a switch in distributions, mean change *magnitude* did not significantly differ between conditions (Mean change magnitude: wide-narrow = .006, wide-wide = .006, narrow-narrow = .009, narrow-wide = .053; ANOVA $F(3,74) = 1.513$, $MSE = .011$, $p = .218$), and participants in all conditions made the same average number of changes to their distribution (Mean number of distribution changes: wide-narrow = 27.1, wide-wide

= 28.6, narrow-narrow = 16.9, narrow-wide = 28.4; ANOVA $F(3,74) = .824$, $MSE = 626.1$, $p = .485$).

These results suggest that the negative correlation between surprise and updating observed in the narrow-wide condition could be due to participants choosing not to integrate highly surprising events. If this were the case, the *variance* of the last distributions drawn by participants should be narrower than the wide distribution presented to them after the switch. This was indeed the case. On average, participants' estimates had smaller standard deviations (SD: 6.03 slots) than the actual standard deviation of the discrete wide Gaussian distribution presented to them (Computer SD: 6.99 slots; $t(18) = -2.10$, $SE = .46$, $p < .05$).

Following from this observation, the next analysis measured whether or not the tendency to devalue surprising events generalized across all subjects in our experiment – in other words, do participants tend to discount surprising events in all four conditions, not just the narrow-wide condition? To do this, the mode of a participant's distribution was calculated on each trial by identifying the slot with the highest assigned probability value. On trials where participants had multiple bars with the same highest probability value, the mean position of these bars was used as a proxy for the mode. For each trial, the absolute difference between a participant's mode and the location of a ball drop on the same trial was calculated. Next, the proportion of changes participants made to their estimates of the underlying distribution was computed when a ball fell at different distances from their mode. As is evident from Figure 3.6, participants made the fewest changes to their distributions when balls fell either near *or* far away from the mode of their distribution. Fewer changes are to be expected at the mode given that these events

represent confirmatory evidence. However, the relatively few changes made when balls fell far from the mode is suggestive of discounting of outliers (Figure 3.6).

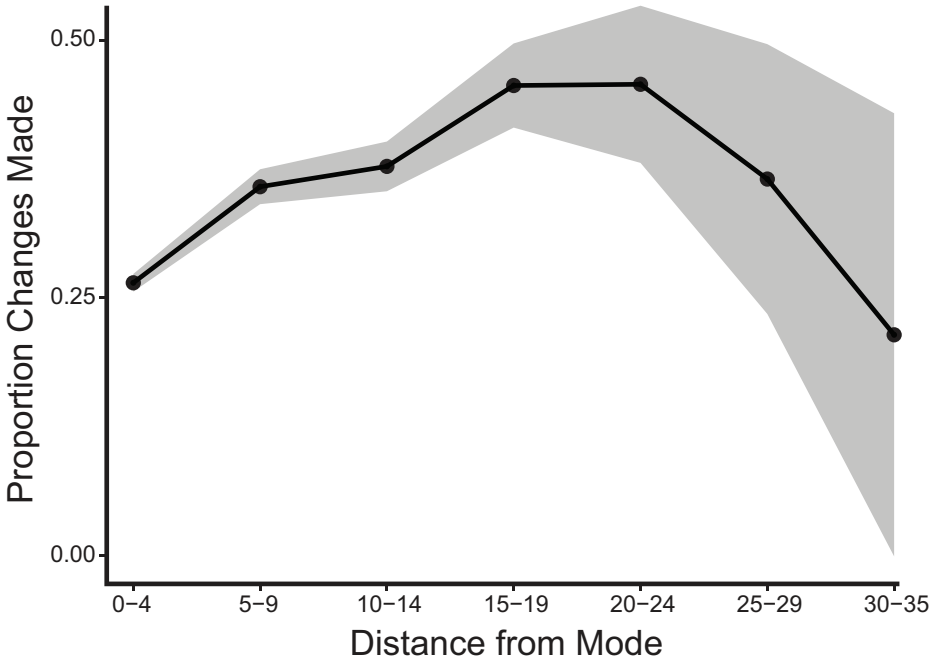

Figure 3.6. Proportion of changes made compared to distance from mode. On each trial, participants' mode was identified as the slot with the highest estimated probability. The absolute distance was calculated as the distance a ball fell from the participants' mode on trial t, whereas the proportion of change was calculated using the changes participants made to their distributions on trial t +1. The black dots represent mean proportion values, whereas shading represents ±95% confidence intervals.

When participant post-questionnaire responses were examined, a chi-square test of independence did not find any performance differences between participants that reported detecting a change to the ball distributions compared to those who had not detected a change (proportion detected: wide-narrow = .53, wide-wide = .53, narrow-narrow = .85, narrow-wide = .76; $\chi^2(3, N = 71) = 6.714$, $p = .082$). However, approximately one quarter of the participants reported that they made adjustments based on observations from the last few trials (in some cases, using the word "outlier" to refer to unexpected ball drops that they chose not to integrate in their estimates). A mixed effects logistic regression was used to measure the influence of factors that contributed to the likelihood that participants made changes to their distributions, including the contribution of the surprise experienced on previous trials. The factors included were the participant's current trial, and the surprise participants had experienced on prior trials. As is evident from Table 3.4, the surprise value from the immediately preceding trial (n – 1) had the greatest influence on the likelihood that a participant would make an adjustment to their distribution. However, similar to participant reports, trials as far as three trials back (n – 3) made additional, independent, statistically significant contributions (Table 3.4). When the same logistic regression was performed on each separate surprise condition, this effect was primarily present in participants in the high surprise condition, a condition where surprising events were more commonly observed, whereas participants in the low and medium surprise conditions relied mostly on the surprise from trial n – 1 (Table 3.4).

Table 3.4. Influence of trial and surprise on likelihood of updating.

Predictor	All Conditions	Wide-Narrow	Wide-Wide	Narrow-Narrow	Narrow-Wide
Trial	-.009 ***	-.010 ***	-.009 ***	-.010 ***	-.007 ***
n – 1	.067 ***	.055 ***	.045 **	.082 ***	.074 ***
n – 2	.007 *	.017 *	.006	-.016	.010 *
n – 3	.009 **	.005	.003	.006	.016 ***

Note. Predictors include the general influence of trials, along with the influence of surprise on previous trials (up to trial n – 3). The values in the cells represent beta weights for each predictor. *** $p < .001$, ** $p < .01$, * $p < .05$

3.7. Experiment 3.2: Discussion

The goal of Experiment 3.2 was to explore how the surprise of an environmental shift influences mental model updating. The first result suggests that switches signaled by the *absence* of previously observed events (low surprise), although similar in magnitude, were learned less accurately than shifts signaled primarily by the *presence* of new events. This result mirrors performance from some of the participants in Experiment 1 that had started with wide Gaussian priors, but had difficulty adjusting the width of their estimates when presented with a narrow Gaussian distribution. Additionally, consistent with previous research and similar to findings from Experiment 3.1, there were some situations – the low and medium surprise conditions – in which updating was positively correlated with surprise. Taken together, these results support previous research suggesting that, under certain circumstances, surprising observations lead to more efficient updating (Nassar, 2010; McGuire et al., 2014). However, in addition to these positive correlations, the opposite trend was found in the highest surprise condition, where higher levels of surprise predicted *poorer* updating performance (Figure 3.5). These results demonstrate that although surprise can play an important role in the updating process, highly surprising events do not always predict better updating.

It is possible that the negative correlation between surprise and updating in the high surprise condition stems from a form of outlier devaluation, in which participants chose not to integrate highly surprising events into their estimates. This notion is supported by studies demonstrating that participants tend to discount highly discrepant exemplars when categorizing events (De Gardelle & Summerfield, 2011; Summerfield & Tsetsos, 2015; Wei & Stocker, 2015). Participants in the highest surprise condition

seemed to take more previous events into consideration when making changes to their distributions, rather than relying solely on the last trial seen, as was primarily the case in the other conditions. When asked on the post-experimental questionnaire about the strategies participants used to update their estimates, approximately one quarter of the participants reported waiting for the same event to occur a number of times in quick succession before committing to a change (some using the word "outlier" to describe unexpected events that they chose not to integrate).

This last finding is particularly important, as outlier devaluation is not a part of any current model of dynamic mental model updating. Current models suggest that *any* surprising event should increase a person's propensity to update (Nassar et al., 2010; McGuire et al., 2014; O'Reilly et al., 2013). These models were built to fit tightly controlled experimental environments, where changes are expected, are very similar in their nature (*e.g.*, consist primarily of mean shifts), and where participants are encouraged to make changes to their beliefs on a trial-by-trial basis. When some of these constraints are removed, surprising information can be treated differently than predicted by these highly controlled environments.

Even though participants in this experiment were not made explicitly aware that changes would occur, some parallels can be drawn with previous research. In the mean shift conditions, which most closely match previous work, surprise was positively related to updating. However, this was not the case in one of the variance conditions, suggesting that the *type* of change participants observe can have implications in the way they treat surprising information. This is not to suggest that *all* surprising information is devalued,

merely that the *type* of change needs to be considered when attempting to measure the influence of surprise on updating.

3.8. Discussion

The goal of Chapter 3 was to examine some of the factors that lead us to detect changes in the environment. The results from Experiment 3.1 and 2.2 demonstrate that in some situations, surprise can be positively related with updating; however, some highly surprising changes can be also be rejected or devalued.

One explanation for these results is that we focus only on the features we believe we need, rather than encoding all events equally. Proponents of the 'efficient coding hypothesis' suggest that we weight perceptual events in proportion to the probability of their occurrence (Barlow, 1961; Wei & Stocker, 2015). In essence, we focus primarily on *modal* elements of a distribution. This hypothesis helps integrate the results from Experiments 3.1 and 2.2 with previous research. Participants in the continuous condition of Experiment 3.1, and those from Experiment 3.2 exposed to changes signaled by mean shifts tend to update depending on the level of surprise prompted by the shift. In all of these conditions, the *modal* elements changed between distributions. These change properties are also found in many of the previous studies that have reported positive relationships between surprise and updating (Nassar et al., 2010, 2012; O'Reilly et al., 2013; McGuire et al., 2014). However, in the variance shift conditions of Experiment 3.2, although the *magnitude* of the distribution's change was similar to that of the mean shifts, the *modal* elements of each distribution remained the same. As a result, participants in these conditions were either worse overall, or tended not to give as much weight to highly

unexpected events. The efficient coding hypothesis provides a plausible explanation for these results, and could potentially apply to the way in which we use information to inform, and update, our mental models (Summerfield & Tsetsos, 2015).

The results from this chapter are also reminiscent of the 'Goldilocks effect' from infant learning studies (Kidd, Piantadosi, & Aslin, 2012; Kidd, Piantadosi, & Aslin, 2014). This research finds that infant attention is drawn to events that are neither too simple (*i.e.*, already known, providing low surprise), nor too complex (*i.e.*, highly unknown, or providing high surprise), choosing instead to focus on events providing a moderate level of complexity. Mental model updating could potentially operate within a similar trade-off: maximizing information gain (*i.e.*, integrating surprising information) while also minimizing mental model complexity (*i.e.*, devaluing or ignoring information that is *too* surprising and difficult to integrate into an existing model). Computational accounts that take this kind of approach, trading off being receptive to change while also being robust to overly complex information, could potentially provide a better characterization of the processes underlying mental model updating.

Taken together, the results from this chapter provide important insights into the factors that can lead us to detect environmental change. A better understanding of the situations that lead people to integrate or ignore surprising information will help develop more accurate models of dynamic updating.

Chapter 4: The Influence of Brain Damage on Exploration

4.1. Introduction

Chapters 2 and 3 examined aspects of the first two components required for updating: building a mental model, and detecting relevant changes in the environment. This next chapter examines the mechanisms involved in *exploring* alternative mental models once a discrepancy between a current model and observations has been detected. The motivation for this chapter comes from updating deficits observed in patients with right hemisphere brain damage. Previous research has suggested that these deficits could be classified as problems of *exploration,* rather than difficulties with either mental model building, or change detection (Danckert et al., 2012; Sepavhand , Stöttinger, Danckert, & Anderson, 2014; Filipowicz, Anderson, & Danckert, 2016). The aim of the following study was to characterize the exploratory processes used by different groups of brain damaged patients to examine this relationship in more detail. This introduction begins by highlighting the updating deficits observed in right brain damaged (RBD) patients, and then introduces a computational approach to characterize exploratory processes.

A growing body of research demonstrates that right hemisphere lesions impair mental model updating (Danckert et al., 2012; Stöttinger, Filipowicz, Marandi, 2014; Geng & Vossel, 2013; Decety & Lamm, 2007; Vocat, Saj, & Vuilleumier, 2013; Filipowicz et al., 2016). One of the first studies to test this notion directly had RBD patients play the children's game 'rock-paper-scissors' (RPS) against a computer opponent that adopted changing play strategies (Danckert et al., 2012). The computer began by playing a uniform strategy, in which each option ('rock', 'paper', and 'scissors') was played with equal frequency, before switching to a moderately biased

strategy (50% rock), and finally a strongly biased strategy (80% paper). Switches occurred at unannounced points throughout the task and patients were instructed to win as often by possible by attempting to exploit any strategies they may notice in the computer's play. Both healthy controls and left brain damaged (LBD) patients managed to update to the task's changing contingencies, with some LBD patients actually *outperforming* healthy controls. In contrast, even when shifted to a strong bias, in which 'paper' was played 80% of the time by the computer, RBD patients had much more difficulty updating, with many patients continuing to play randomly during this highly biased portion of the task (Danckert et al., 2012).

Although providing evidence for an updating deficit in RBD patients, it was unclear from this study whether the observed problems were due to a general inability to *learn* environmental statistics, or were more specifically related to a problem of updating. To address this question, Stöttinger and colleagues (2014) had RBD patients perform a very similar 'rock-paper-scissors' task, this time having the computer start play with a strongly biased strategy (80% paper), before switching to a different highly biased strategy (80% rock). RBD patients had less difficulty exploiting the first bias, but more difficulty updating to the second bias (Stöttinger, Filipowicz, Marandi, et al., 2014). If the deficits observed in the original study were due to a general problem with learning probabilities, RBD patients should not have been able to learn either strategy.

The results discussed thus far dealt with probabilistic information. In a second experiment, patients performed a perceptual updating task to test whether or not the updating deficits observed in RBD patients were generic. In this task, patients were presented with a series of images that gradually morphed from one object (*e.g.*, a swan)

to another (*e.g.*, a cat) passing through a midpoint of perceptual ambiguity. Patients were informed that the images would be changing and were asked to tell the experimenter what they saw on each image. Updating in this task was measured as the number of images a patient saw before identifying the second object. They found that RBD patients required more images to identify a second object than either healthy controls, or LBD patients. This was not due to a failure to *detect* changes in the morphing images. The RBD patients reported differences in images as they were changing, but interpreted those changes in the context of their initial perceptual representation of the first object they had seen. Additionally, RBD patient updating performance on this picture morphing task was correlated with performance on the RPS task, such that poor performance on one task predicted poor performance on the other. No such correlations were found for either HCs or LBD patients (Stöttinger, Filipowicz, Marandi, et al., 2014).

These results support the general claim that the right hemisphere plays an important role in mental model updating. However, questions remain regarding the parts of the updating process impaired by RBD. Hints can be gleaned from previous research. Right hemisphere patients were able to learn a strong bias in their opponent in the 'rock-paper-scissors' game when it was the first strategy they were exposed to, suggesting the observed deficits cannot simply be attributed to problems *building* mental models. Additionally, problems in updating mental models in RBD do not seem related to an inability to detect environmental changes – RBD patients noticed and reported changes to the gradually morphing images. Instead, RBD updating deficits seem better described as difficulties *exploring* alternative mental models once a current model is deemed to be no longer valid. The RBD patients in Danckert and colleague's (2012) original study who

presented with updating difficulties were found to explore a very limited space of possible 'rock-paper-scissors' strategies in response to a change in the computer's play. The few RBD patients who *did* manage to update were those who explored the largest area of strategy space (Danckert et al., 2012). A subsequent computational analysis of this data found that RBD patients had difficulty settling on a new effective play strategy to match the computer's plays (Sepavhand et al., 2014). This was due in part to a tendency to quickly abandon a strategy when it was not supported by task outcomes. Given the stochastic nature of the 'rock-paper-scissors' task, this approach to strategy selection would make it difficult to find new, reliable strategies.

The work discussed above suggests that RBD primarily hinders the ability to explore novel hypotheses when a current mental model fails to accurately represent regularities in the environment. To examine this possibility, it becomes necessary to characterize the exploratory policies used by RBD patients to determine how they differ from LBD patients and healthy controls. Accurately describing these exploratory policies has been challenging to researchers interested in mental model updating (Collins & Koechlin, 2012; Wilson et al., 2010; Glaze et al., 2015; McGuire et al., 2014; O'Reilly, 2013; Behrens et al., 2007; O'Doherty, Dayan, Schultz, Deichmann, Friston, & Dolan, 2004). One particular challenge is appropriately characterizing the trade-off between either *incrementally changing* or *abandoning* a mental model when faced with model discrepant information.

Computational models of reinforcement learning (RL) propose that mental models change on an observation-by-observation basis, incrementally changing with the feedback received from the environment (O'Doherty et al., 2004; Daw et al., 2006;

Sutton & Barto, 1998). Although computationally efficient, RL updating is relatively slow – a current mental model needs to be 'unlearned' before a new one can be learned. Reinforcement learning models are also prone to integrating *any* feedback, making them vulnerable to noisy and stochastic environments. While these classes of models seem to explain some of the driving principles behind the way mental models are *built* (Daw et al., 2006), they do not provide a full account of the way they are *updated.*

Recent research has demonstrated that humans generally update more rapidly than predicted by strict RL models (Collins & Koechlin, 2012; Sepavhand et al., 2014; Donoso et al., 2014). Collins and Koechlin (2012) developed a computational model of adaptive behaviour and executive control (dubbed the PROBE model) that updates a mental model by evaluating whether or not it should be *abandoned*, rather than 'tweaked' or modified. The PROBE model proposes that at any given point, a mental model is used to interact with an environment. If this model is supported by observations, it is considered *reliable*, and the agent is considered to be *exploiting* this reliable mental model. In addition to considering this mental model, the reliability of alternative, *counterfactual* mental models are also considered and compared with current observations. These alternatives are based on mental models that were previously deemed reliable. When a current mental model no longer matches what is observed in the environment, signaling a need to adopt a different model, the agent enters an *exploration* phase, where information from previous mental models are used to find the mental model that best exploits the observed changes.

Whereas RL models assume incremental changes to an *existing* mental model, the PROBE model explores the possibility of *alternative* candidate models. Although

complex, the PROBE model has been validated as a more complete characterization of the processes involved in updating than many other classes of computational models of updating (Collins & Koechlin, 2012; Donoso et al., 2014).

The goal of the current study was to test whether updating deficits in RBD could be attributed to difficulties with exploring new mental models. Both RBD and LBD patients performed an updating task in which they learned stimulus-action mappings that changed at unannounced points throughout the task. Their performance was fit using both an RL model and the PROBE model to characterize the exploratory strategies used by the different patient groups. In addition to these behavioural differences, patient brain scans were also compared to examine the contribution of specific brain lesions to any updating deficits observed.

Neuroimaging studies of updating, including a study of the PROBE model, have identified a network of frontal brain regions involved in the three different aspects of updating. In this network, building and maintaining a mental model is driven by a connection between the ventral striatum and ventromedial prefrontal cortex (Daw et al., 2006; Kolling, Behrens, & Mars, 2012; Donoso et al., 2014). The decision to change is represented in the medial prefrontal cortex, including the dorsal anterior cingulate (O'Reilly et al., 2013; Behrens et al., 2007; McGuire & Kable, 2015; McGuire et al., 2014; Donoso et al., 2014), and the exploratory component of updating, particularly the evaluation of alternative mental models, is driven by anterior and lateral prefrontal regions (Donoso et al., 2014; Koechlin & Summerfield, 2007; Daw et al., 2006; Domenech & Koechlin, 2015). In addition to these frontal regions, the inferior parietal lobule has also been consistently implicated in the networks involved in updating

(Danckert et al., 2012; Stöttinger et al., 2015; O'Reilly et al., 2013; McGuire et al., 2014). Its role seems primarily involved in representing mismatching or surprising events with activity during updating closely linked to activity in the medial prefrontal cortex (O'Reilly et al., 2013).

Since updating deficits in RBD seemed related to problems of exploration, and that these functions have been linked to regions of frontal cortex, an attempt was made to include patients whose primary site of damage was found in regions of frontal cortex. These patients performed an updating task to try and identify the parts of the updating process that are impaired, and the brain regions associated with these deficits.

4.2. Experiment 4.1: Methods

Participants

Four groups of participants were recruited to participate in our study: RBD patients, LBD patients, healthy older controls (HC-O) and healthy younger controls (HC-Y).

Patients with a documented history of stroke extending to the right or left frontal cortex were recruited from the Neurological Patient Database at the University of Waterloo. A total of 20 patients participated in the study (11 RBD and 9 LBD). Data from three patients were excluded. Two patients (one LBD and one RBD patient) could not correctly repeat the task instructions, and were unable to complete the initial trials of the task. One RBD patient was subsequently excluded when no discernable lesion could be identified on the available brain scans. The final patient sample comprised 17 patients (9 RBD, 8 LBD). All patients were tested at least three months post-stroke. Complete

patient information can be found in Tables 4.1 and 4.2, and lesions traces are displayed in Figures 4.1 and 4.2.

Table 4.1 Patient demographics for RBD group

ID	Sex	MoCA	Age	Education (Years)	Months Since Stroke	Lesion	Lesion Volume (voxels)
228*	F	27	85	14	68	F, T, Ins†, BG, RO	17,450
384*	M	27	41	12	35	F, T, Ins†, BG, RO, Cereb	13,441
423	M	26	62	10	53	F, T, P, O, Ins†, BG, RO	31,075
449	M	26	69	18	48	F, T, O, Ins, BG, RO	2,318
478	F	19	56	17	48	F, Ins	849
632	F	27	55	17	32	F	1,578
649*	M	27	70	14	13	F	644
771	M	23	60	12	14	F	145
874	M	27	55	15	3	F, T, Ins†, BG, RO	12,131

Table 4.2 Patient demographics for LBD group

ID	Sex	MoCA	Age	Education (Years)	Months Since Stroke	Lesion	Lesion Volume (voxels)
73*	M	13	75	9	3	F, T, Ins†, BG, RO	33,464
110	M	26	63	12	84	F, Ins, BG	4,131
269*	F	22	75	12	57	F, T, Ins†, BG	1,833
414*	M	17	74	12	54	F, T, Ins†, BG, RO, Th	16,566
799*	M	25	80	11	7	F, Ins†, Cereb	2,400
872	M	28	71	10	6	F, P, BG	6,963
894	M	21	62	12	3	F, T, Ins	1,703
898*	M	22	66	8	3	F, T, Ins†, BG, RO	9,352

Note. Sex: F = female, M = male; Fr = frontal; T = temporal; P = parietal; O = occipital; Ins = insula († indicates the inclusion of *anterior* insular damage); BG = basal ganglia; RO = rolandic operculum; Cereb = cerebellum; Th = thalamus; ID numbers with a * indicate patients that were identified as 'Poor Updaters'.

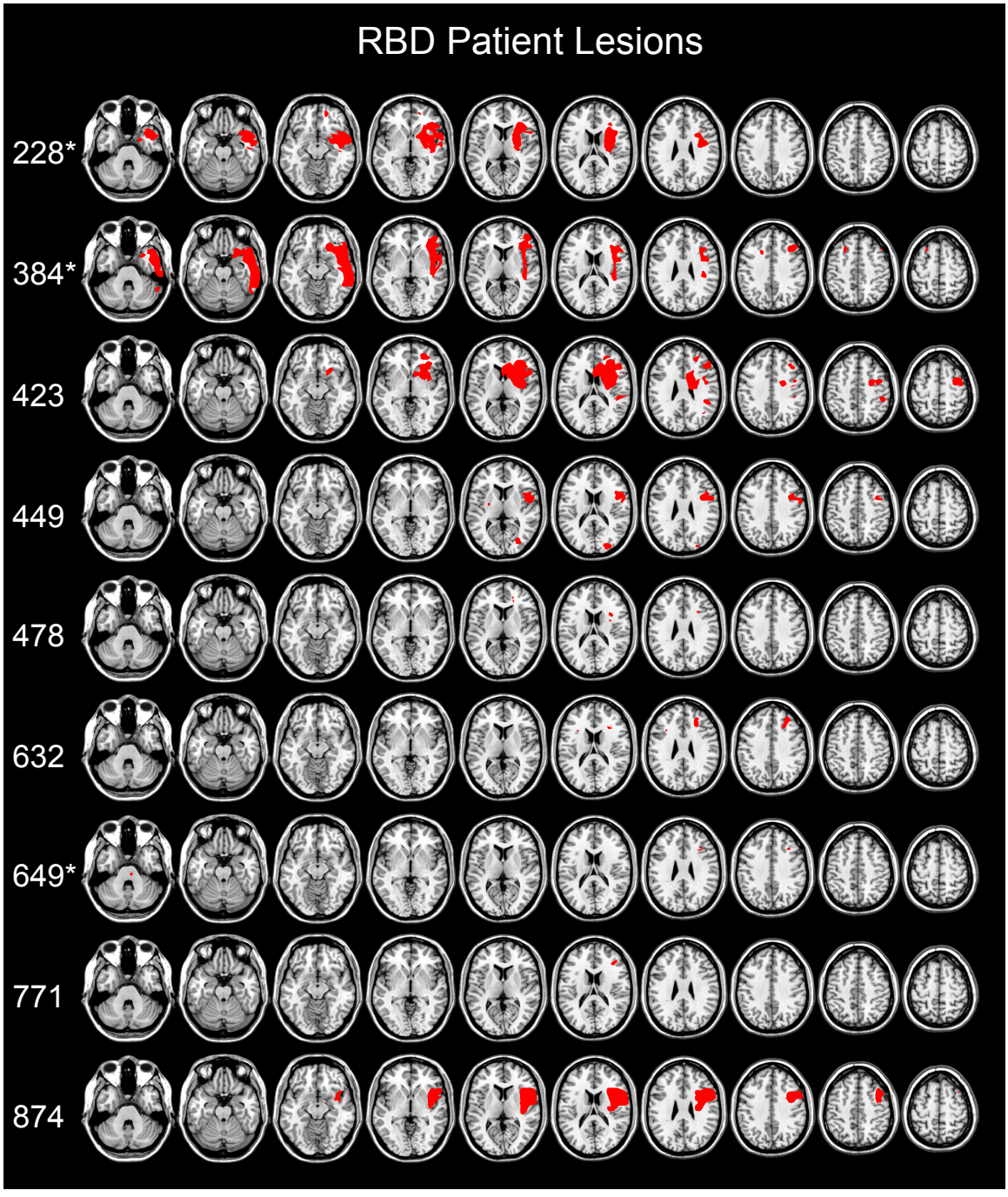

Figure 4.1. Traces of right brain damaged (RBD) patient lesions. Lesion traces for all nine RBD patients superimposed on an MNI template. Patient numbers marked with a * indicate patients classified as 'Poor Updaters'.

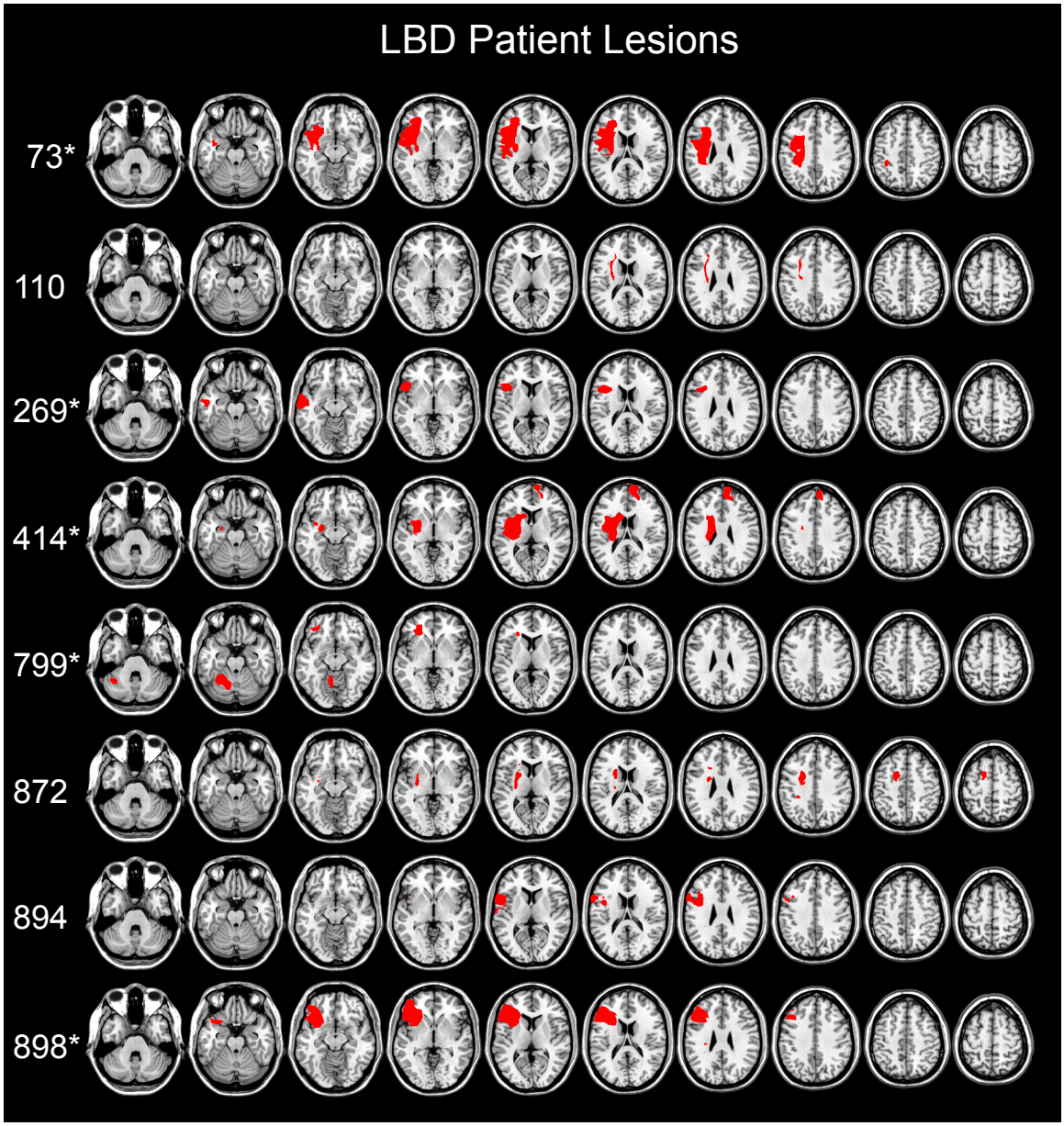

Figure 4.2. Traces of left brain damaged (LBD) patient lesions. Lesion traces for all eight LBD patients superimposed on an MNI template. Patient numbers marked with a * indicate patients classified as 'Poor Updaters'.

Eighteen older healthy controls (HC-O) were recruited from the Waterloo Research in Aging Participant Pool and had no history of neurological or psychiatric illness (11 females, mean age = 73.72, age range = 63-87). Patients and HC-O participants were paid $10 per hour of study.

In addition to older controls, a group of young healthy controls (HC-Y) comprised of 34 university undergraduates (24 female, mean age = 20.44 years, age range = 17-25) participated in exchange for course credits, and an additional $5 for completing the second session. This group was meant to serve as a point of comparison with the demographic groups tested in previous studies using the PROBE task, especially given that this task has never been administered to elderly populations (Collins & Koechlin, 2012; Donoso et al., 2014).

Each participant's testing required two days. All patients and HC-O participants performed the Montreal Cognitive Assessment Task (MoCA; Nasreddine et al., 2005) as a measure of overall cognitive function on the first day of testing, before the start of the first behavioural session.

The experimental protocol was approved by the University of Waterloo's Office of Research Ethics, and all participants provided written informed consent prior to participation.

Lesion tracing and analysis

Patient brain scans were obtained through the University of Waterloo's Neurological Patient Database. The most recent available clinical CT (15) or MRI (2) scans were used to characterize patient brain lesions. All brain images were aligned and

centered on the anterior commissure using SPM 8. Lesions were then traced by hand using MRIcron (Rorden, Karnath, & Bonilha, 2007) by a single investigator (AF). Traced lesion files and scans were then spatially normalized using the Clinical Toolbox for SPM (Rorden, Bonilha, Fridriksson, Bender, & Karnath, 2012). This toolbox makes it possible to spatially normalize brain images across different scanning modalities (*e.g.*, normalize CT and MRI scans into the same space). These normalized images were then compared across different patient groups. Brain damaged regions were identified by overlaying individual normalized brain lesions on the Automated Anatomical Labeling map (Tzourio-Mazoyer et al., 2002), and using MRIcron's Descriptive tool to obtain a summary of the location and size of participant lesions. All analyses were conducted using MRIcron and Non-Parametric Mapping software (NPM; Rorden et al., 2007).

PROBE Task

The behavioural task participants performed was identical to the task administered by Donoso and colleagues (2014), except that the breaks participants received differed slightly from the original protocol. In this task, participants are required to learn changing stimulus-action mappings by using feedback. A specific rule required one of four buttons (the 'j', 'k', 'l', and ';' keys on a QWERTY keyboard) to be pressed for each of three number stimuli. The set of three numbers could either be {1,3,5} or {2,4,6} with the order of the sets counterbalanced between participants.

Stimuli were presented on a computer screen. White boxes were shown on a black background (Figure 4.3). On each trial, one numeral appeared in all four boxes and the participant chose one of four buttons to press. Feedback indicated whether or not the

correct button had been pressed. If the correct button was pressed, the number would change to green in the box corresponding to the button pressed; if the incorrect button was pressed, a red 'X' would appear in the box corresponding to the button pressed (Figure 4.3). Feedback in this task was noisy with 10% of trials giving participants incongruent performance feedback (*i.e.*, on 10% of trials, participants would be given either a red 'X' for a correct response, or a green number for incorrect responses; Figure 4.3).

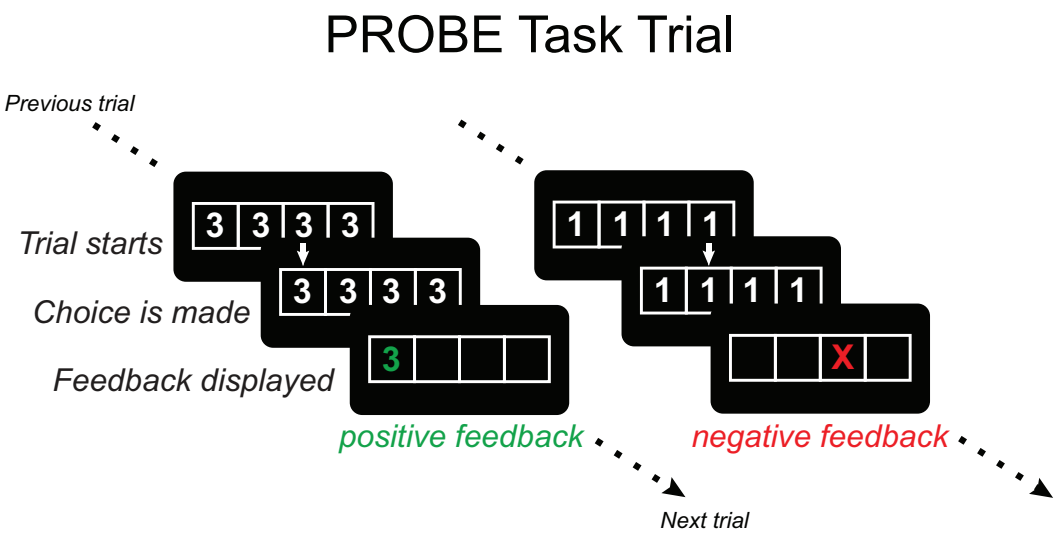

Figure 4.3. Schematic of PROBE task. Participants learn mappings between four keyboard buttons and three number stimuli. On any given trial, a specific stimulus-response rule applied, such that each number stimulus was associated with one button press (with one button not associated with any stimuli). Participants learned these rules through task feedback. On each trial, participants saw four white boxes on a black background. These boxes corresponded to the four buttons they could press on each trial. At the start of a trial, one of the three number stimuli appeared in all four boxes. This number remained on screen until participants pressed the button they thought was associated with the number on screen. If participants pressed the correct button, the number would turn green in the box corresponding to the participant's button press (positive feedback). If an incorrect button was pressed (*i.e.*, a button other than the one associated with the number on screen), a red 'X' appeared in the box corresponding to the participant's button press (negative feedback). Rules changed at unannounced points every 33-48 trials. Additionally, incongruent feedback was pseudo-randomly provided on 10% of trials, such that correct participant responses were given negative feedback (*i.e.*, a red 'X'), and incorrect responses were given positive feedback (*i.e.*, the number would turn green). Participants were aware that rules would change throughout the task and that feedback would be noisy. They were instructed to make as many correct responses as possible by learning task rules, and detecting when rules changed.

Each stimuli-keypress rule remained in place for 33-48 trials. None of the correct responses from a current rule overlapped with the correct responses in the rule that followed. Participants were informed that there was a rule that would determine the key press-stimulus mappings, and that they had to learn this rule using the feedback they received from their responses. Participants were made aware that rules would change at unannounced points throughout the task, and that they would have to adapt their responses to reflect these changes. Participants were also told that the feedback would be noisy, and that the computer would sometimes "lie" to them. However, the frequency of the rule changes, and the frequency of the noisy feedback were not explicitly communicated to participants.

All participants performed two experimental sessions over two separate days: a *recurrent* session and an *open* session. In each session, participants were exposed to 24 rules split into 6 blocks (*i.e.*, each block containing 4 rules to learn). Between each block participants were given a short, self-terminated break. Participants were informed that the task was meant to be continuous, and that the breaks were only meant for their comfort. They were informed that whichever rule applied at the end of one block would also apply at the beginning of the next block. When participants started the next block, they were exposed to 6-9 trials of the rule that had applied at the end of the last block before being switched to a new rule. In Donoso and colleagues' original study, participants were asked to memorize the rule that applied at the end of each block, and used this memory to start the next block of trials. An initial pilot study demonstrated that patients had difficulty keeping a memorized rule for the duration of their self-timed break. To facilitate memory for a rule, at the end of each block participants were asked to indicate what they believed

to be the current correct rule, without any feedback indicating whether or not their reports were correct. At the start of the next block, they were reminded of the rule they believed applied at the end of the last block, and were instructed to keep using this rule until they believed that it no longer applied.

In the *recurrent* session, participants were exposed to three rules that repeated pseudo-randomly throughout the session. In this session, each of the three rules was non-overlapping with the other rules (*i.e.*, no two rules had the same response for any given number). In the *open* session, participants were exposed to 24 distinct rules arranged pseudo-randomly between participants. Although some rules in the open session had overlapping responses, no current rule had any stimulus-responses mappings that overlapped with either the rule that preceded it, or the rule that followed. The order in which the recurrent and open sessions were completed was counterbalanced across participants, and participants were not informed of any differences between the two sessions.

Similar to previous studies, at the end of each session, participants completed a post-session questionnaire. They were presented with 6 rules that they were told may or may not have applied throughout the session. When they saw each rule, they were asked to rate how certain they were that they either had or had not seen each rule throughout the session. The ratings ranged from 1 ("I am certain that I DID NOT see this rule") to 5 (I am certain that I DID see this rule"), with 3 indicating uncertainty ("Not sure"). This post-test was meant to identify participants that could correctly remember any of the three rules that applied in the *recurrent* session. A memory score was calculated by subtracting the sum of the certainty scores for rules participants *had* seen from the sum of

the certainty scores for the rules they *had not* seen, and dividing the sum of these differences by the number of rules they were asked to recall (6). Participant scores could range between -2 and 2, with scores greater than 0 indicating more reported certainty for rules that had repeated.

Computational Models

Participant performance was characterized using two computational models of adaptive behaviour. Both models (PROBE and RL) are identical to the ones presented in Collins and Koechlin (2012) and by Donoso and colleagues (2014). It is worth noting that contextual elements (*e.g.*, background colour changes that could indicate rule changes) are also considered in the original PROBE model (Collins & Koechlin, 2012). Given that no contextual elements were present in the current study, these were not included in the version of the PROBE model used below (similar to Donoso et al., 2014). The following two sections provide overviews of each computational model. Full mathematical descriptions can be found in Appendix 1.

PROBE model

The PROBE model characterizes updating as a forward inference process that arbitrates between knowing when to exploit a reliable mental model, and when to explore new, alternative mental models. Mental models are characterized as stimulus-action mappings that are learned from action outcomes. At any given moment, a single *actor* mental model drives behaviour, and its stimulus-action mappings are updated with new trial outcomes. In addition to these mappings, the *reliability* of a mental model (*i.e.*, the

likelihood that it this model is correct given observations in the environment) is also monitored and updated with each new outcome. In addition to the actor, participants are assumed to have a working memory "buffer" that concurrently monitors the reliability of a limited number of alternative, counterfactual mental models. These counterfactuals are composed of mental models that were previously used at some point in the past. In order for a mental model to become an actor, it needs to become more reliable than any other monitored mental model. When an actor is reliable, the PROBE model is said to be in a state of *exploitation*. The PROBE model also assumes that all past reliable mental models are stored in a long-term memory component, but their reliability is not actively monitored.

If an actor becomes unreliable (*i.e.*, when environmental contingencies change), the reliabilities of the monitored counterfactuals mental models are examined. If a counterfactual is found to be reliable, the previous actor is placed in the buffer, and the reliable counterfactual becomes the new actor. If no monitored counterfactual is reliable, the PROBE model enters a state of *exploration*. In this state, the unreliable actor is added to the buffer and a new *provisional actor* is created based on a mixture of the mental models stored in long-term memory. This provisional actor drives behaviour during the exploration stage. It is initially unreliable and is *probed* by adjusting its stimulus-action mappings in response to new outcomes. There are two ways for the PROBE model to exit exploration. If the provisional actor eventually becomes reliable, it is *confirmed*, and becomes the new actor, effectively ending the exploration phase. Once confirmed, the new actor is stored in long-term memory and the old actor is kept in the memory buffer, replacing the buffer's least recently used mental model. If, however, during the

exploration phase one of the monitored counterfactual mental models becomes reliable, the provisional actor is *rejected*, and the PROBE model exits exploration with the reliable counterfactual as its new actor.

Overall, the PROBE model contains 7 free parameters that can vary between subjects:

Buffer capacity (N). An integer value that corresponds to the number of mental models, including the actor, that can concurrently be monitored. It is worth noting that the provisional actors are added to this buffer during periods of exploration.

Perceived volatility (τ). A parameter ranging between 0 and 1 that represents the likelihood that an environmental change will occur at any given point throughout the task.

Recollection entropy (η). A continuous parameter ranging between 0 and 1 that determines how much a provisional actor's mappings rely on mappings stored in long-term memory (with values closer to 1 indicated less influence from long-term memory).

Prior reliability bias (θ). Participants may be more or less inclined to confirm rather than reject newly created provisional actors. The θ parameter scales the deviation of a probe actor's prior reliability from an uninformative prior to the reliability threshold of 0.5 required to confirm an actor (see Appendix 1 for full details).

Learning rate (α). Standard reinforcement learning rate parameter that can range between 0 and 1, and determines the weight given to an outcome when an actor updates its stimulus-action mappings.

Inverse temperature (β). Standard continuous reinforcement learning parameter that scales the 'greediness' of an actor's action selection. Higher values of β indicate

higher biases towards actions that maximize reward output.

Noise (ε). Standard continuous reinforcement learning parameter that can vary between 0 and 1, and accounts for lapses in subject responses. The larger this parameter, the more lapses unaccounted for by the model.

Reinforcement Learning (RL) Model

Participant performance was also fit using a standard reinforcement learning model (RL) that uses past outcomes to build and update stimulus-action mapping. This model does not compute mental model reliabilities or monitor counterfactuals. Instead, updating is characterized as continuous adjustments made to an existing model that follow reinforcement learning rules (see Appendix 1). This RL model contains three free parameters: α (learning rate), β (temperature), and ε (noise).

Model fitting methods

The set of best fitting parameters for both models was obtained by finding the set of parameters that maximized the log likelihood of the model's responses compared to each participant's responses. To find the best fitting RL parameters for each subject, maximum likelihood estimates were obtained by sampling from the parameter space likelihood function using a slice sampling algorithm (Neal, 2003). Four independent slice sampling Markov chains were launched from independent starting points in parameter space to sample likelihood values from a broad range of parameter values. A gradient ascent was then run on the highest likelihood value obtained over the four chains, and the

parameter set corresponding to the maximum likelihood estimate was used as the best fitting RL parameters for each subject.

A similar method was used for the PROBE model with an added step (identical to the one used by Donoso et al., 2014). The buffer capacity (N) in this model is a discrete parameter, while the remaining six parameters are continuous. To identify the best buffer capacity, the best sets of parameters were identified for buffer sizes ranging from 1-5, with the best fitting set of parameters determining specific buffer sizes for each individual participant. These parameters were obtained for each buffer size using the slice sampling and gradient ascent methods outlined above. The buffer size with the set of parameters providing the highest likelihood values determined both the discrete buffer size parameter, and the set of continuous parameters used for each subject.

4.3. Experiment 4.1: Results

Demographic differences

Patient and HC-O demographic information was compared using independent t-tests to determine whether any of these factors differed between participant groups. The first set of analyses compared participants from the HC-O group to all patients involved in the study. The HC-O group were significantly older than the patient group (HC-O mean age (95% CI): 73.7 years, (70.4-77.1); patient mean age: 66.0 years, (60.8-71.1); $t(27.72)=2.47$, $p < .03$), had significantly more years of education (Mean years of education (95% CI) – HC-O: 15.4 years (14.4-16.4), Patients: 12.3 years (11.0-13.5); $t(31.39)=3.74$, $p <.001$), and, as would be expected, had significantly higher MoCA

scores (Mean MoCA scores (95% CI) – HC-O: 27.3 (26.5-28.2), Patients: 23.7 (21.7-25.7); $t(21.82)=3.25$, $p < .04$).

Next, demographic and lesion information was compared between LBD and RBD patients. RBD patients had significantly more years of education than LBD patients (Mean years of education (95% CI) – RBD: 13.7 years (11.9-15.4), LBD: 10.8 years (9.7-11.8); $t(13.15)=2.76$, $p < .02$). There were also trending but non-significant differences in patient group age (Mean age (95% CI) – RBD: 61.7 years (53.5-69.8), LBD: 70.8 years (66.3-75.3); $t(12.33)=1.93$, $p = .08$) and MoCA scores (Mean MoCA scores (95% CI) – RBD: 25.4 (23.7-27.2), LBD: 21.8 (18.4-25.1); $t(10.728)=1.89$, $p = .09$). There were no significant differences in lesion volume between patient groups (Mean lesion volume (95% CI) – RBD: 8,848 voxels (1909-15787), LBD: 9,552 voxels (2008-17095); $t(15.67)=.13$, $p = .89$) or in time post-stroke (Mean time since stroke (95% CI) – RBD: 34.9 months (20.8-49.0), LBD: 27.3 months (4.8-49.8); $t(11.95)=.56$, $p = .59$).

Behvavioural Performance

The first analysis examined performance differences between all four participant groups. To measure rule learning, the proportion of correct responses for each group was computed for each trial. The trial number in this analysis was relative to the last rule switch. Group performances were compared using linear mixed-effects modeling, with Proportion Correct as a dependent variable, Participant Group (HC-Y, HC-O, LBD, and RBD) as a fixed factor, and Participant and Trial as random intercept factors. This model was compared to a null model, which assumed no effect of Participant Group, and the likelihood ratios of these models were used to determine whether or not Participant

Group was a significant predictor of correct response rate. As is evident from Figure 4.4, there was a significant main effect of Participant Group ($\chi^2(3) = 18.587$, $p < .001$).

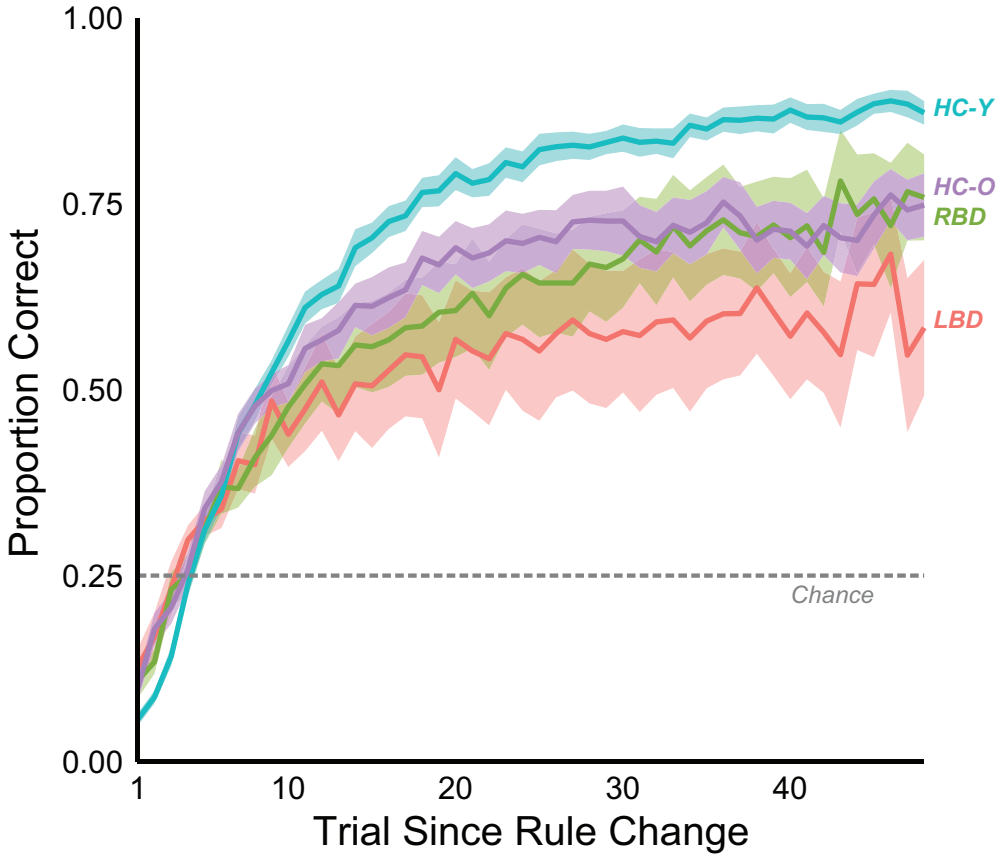

Figure 4.4. Participant correct response rates on the PROBE task. Participant correct response rates were averaged over all task trials since the last rule change and compared between participant groups. Healthy younger adults (HC-Y - cyan) had higher correct response rates than any other group. Performance did not differ between older adults (HC-O – purple), right brain damaged (RBD - green) and left brain damaged (LBD - pink). Solid lines represent group means and shading represents ± one standard error of the mean.

Pairwise comparisons showed that HC-Y participants performed best overall (Mean correct response rate (95% CI) – HC-Y = .70 (.69-71)), outperforming older controls (HC-O = .61 (.60-.63); $\chi^2(1) = 8.641$, $p < .004$), LBD patients (LBD = .51 (.49-.54); $\chi^2(1) = 16.222$, $p < .001$), and RBD patients (RBD = .58 (.56-.61); $\chi^2(1) = 8.710$, $p < .004$). However, correct response rate did not differ significantly between healthy elderly controls and LBD patients ($\chi^2(1) = 2.780$, $p = .095$) or RBD patients ($\chi^2(1) = .360$, $p = .549$). Additionally, there were no significant differences in correct response rate between RBD patients and LBD patients ($\chi^2(1) = .72$, $p = .396$). Overall, this first analysis suggests that there were no differences between RBD and LBD patients and healthy aged matched controls in the proportion of correct responses.

Updating Performance in Recurrent vs Open Sessions

Updating performance was next compared between the *recurrent* and *open* sessions within each participant group. The long term memory and counterfactual rule monitoring components of the PROBE model predict that participants should be quicker to update to rules they have learned in the past. This difference is not predicted by RL models. Supporting this prediction, previous studies found small but significant updating advantages in the recurrent session compared to the open session (Collins & Koechlin, 2012; Donoso et al., 2014).

Updating performance between the two sessions was compared for each individual participant group using linear mixed effects models with Proportion Correct as a dependent variable, Session (recurrent or open) as a fixed effect, random intercepts for

Subject and Trial Number, and a random slope for Session. These models were compared to null models that assumed no Session effects.

Given prior results (Collins & Koechlin, 2012; Donoso et al., 2014), it was predicted that participants in the HC-Y group would differ between the recurrent and open sessions, with better performance in the former. When all trials were considered together, contrary to results from previous studies, there were no performance differences between the two sessions in the HC-Y group (Mean correct response rate (95% CI): Recurrent = .71 (.70-.72), Open = .70 (.69-.72); $\chi^2(1) = .341$, $p = .559$). When performance was examined more closely, there was an effect of session, but only when comparing the later portion of each session. When comparing the first half of each session (*i.e.*, learning the first 12 rules), there were no overall updating advantages (Recurrent = .70 (.69-.71), Open = .72 (.71-.73); $\chi^2(1) = 1.320$, $p = .251$). However, there was a small but significant advantage for the recurrent condition when examining the second half of each session (Recurrent = .72 (.71-.73), Open = .69 (.67-.70); $\chi^2(1) = 6.883$, $p < .009$; Figure 4.5). These results demonstrate that performance of the recurrent session improved to a slightly higher level by the end of the task when compared with performance in the open session.

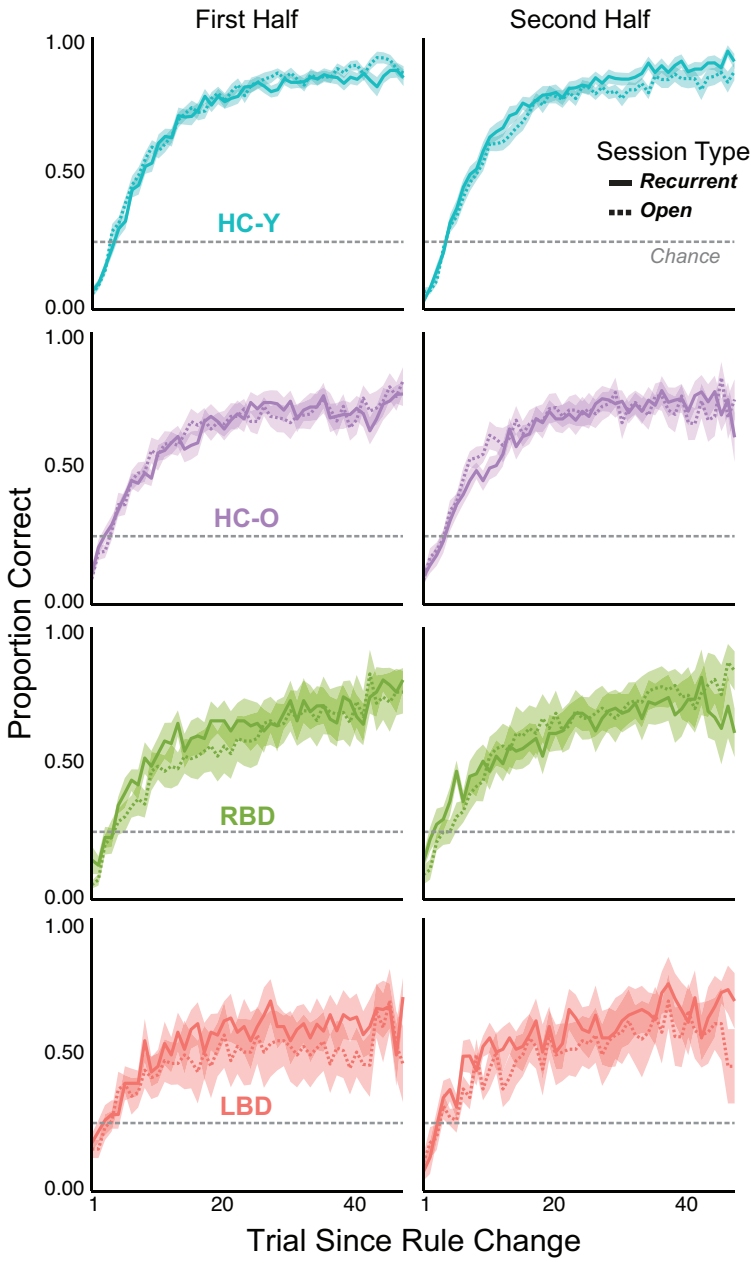

Figure 4.5. Participant performance between the recurrent and open sessions. When split between rules learned in the first half of each session (*i.e.*, first 12 rules) and second half (*i.e.*, last 12 rules), healthy younger adults (HC-Y - cyan) had the same correct response rates between the recurrent (dashed lines) and open sessions (solid lines), but performed better in the recurrent session in the second half. Left brain damaged (LBD - pink) patients performed better in the recurrent session in both the first and second half of the experiment. Right brain damaged (RBD - green) patients and healthy older adults (HC-O - purple) did not show any session differences throughout the task. Solid and dashed lines represent group means and shading represents ± one standard error of the mean.

Performance between sessions was next examined in the HC-O and patient groups. Similar to HC-Y, HC-O participants did not show any overall differences between sessions (Mean correct response rate (95% CI): Recurrent = .62 (.60-.63), Open = .62 (.60-.64); $\chi^2(1) = .101$, $p = .751$). Unlike HC-Y participants, when performance was split into the first and second half of each session, HC-O participants did not show any benefits of the recurrent session in either the first (Recurrent = .62 (.60-.63), Open = .62 (.60-.63); $\chi^2(1) = .003$, $p = .959$), or second half of each session (Recurrent = .62 (.60-.63), Open = .63 (.62-.65); $\chi^2(1) = .286$, $p = .593$). These results show that participants in the HC-O group did not demonstrate any performance benefits from repeating strategies in the recurrent session.

Performance between sessions was next compared between LBD and RBD patients. Over all trials, all eight LBD patients performed better in the recurrent session than in the open session (Mean correct response rate (95% CI): Recurrent = .55 (.53-.57), Open = .48 (.46-.51); $\chi^2(1) = 4.985$, $p > .03$). This difference was present when comparing both the first (Recurrent = .54 (.52-.57), Open = .47 (.44-.50); $\chi^2(1) = 3.950$, $p < .05$), and second half of each session (Recurrent = .56 (.53-.59), Open = .49 (.46-.52); $\chi^2(1) = 6.316$, $p < .02$).

RBD patients did not show any overall benefits in the recurrent session (Recurrent = .59 (.57-.62), Open = .58 (.55-.61); $\chi^2(1) = 0.216$, $p = .643$). There was a nominal but non-significant advantage for the recurrent session when the comparing the first half of each session (Recurrent = .60 (.58-.62), Open = .56 (.53-.58); $\chi^2(1) = 1.902$, $p = .168$), and no advantage when comparing the second half of each session (Recurrent = .59 (.57-.61), Open = .61 (.58-.64); $\chi^2(1) = .284$, $p = .594$). These results demonstrate that while

there may not have been any gross updating performance differences between patient groups when contrasting performance over all sessions, LBD patients benefitted more from the repeating rules in the recurrent session than did RBD patients.

To see how well participants from each group managed to explicitly remember the repeating strategies, scores from the post-experimental questionnaire in the recurrent condition were compared. A one way ANOVA with memory scores as a dependent variable and participant group (RBD, LBD, HC-O, and HC-Y) as a predictor revealed that memory scores differed between participant groups ($F(3,65)=4.492$, $MSE = .251$, $p < .007$). A subsequent Tukey HSD post-hoc comparisons of group means revealed that the HC-Y memory scores (Mean memory scores (95% CI): HC-Y = .60 (.40-.80)) were higher than HC-O scores (HC-O = .17 (-.05-.38); $p = .02$) and LBD scores (LBD = .08 (-.00-.17); $p < .05$), but not significantly different from RBD patients (RBD = .24 (.02-.46); $p = .23$). Memory scores did not differ among any other participant groups (all remaining ps > .91).

Good and Poor Updaters

Although aggregate performance was not found to differ between RBD and LBD patients, *individual* patient performance was heterogeneous. Figure 4.6 shows patient performance (collapsed across hemispheres). Two patient groups were apparent: one group of patients (6 RBD and 3 LBD) had accuracy scores similar to (or better than) healthy controls; the other group (5 LBD and 3 RBD) performed well below this range. Patients with task performance better than the lower bound of the 95% confidence

interval from the HC-O group were classified as ‘Good Updaters’. Patients with accuracy performance below this mark were classified as ‘Poor Updaters’.

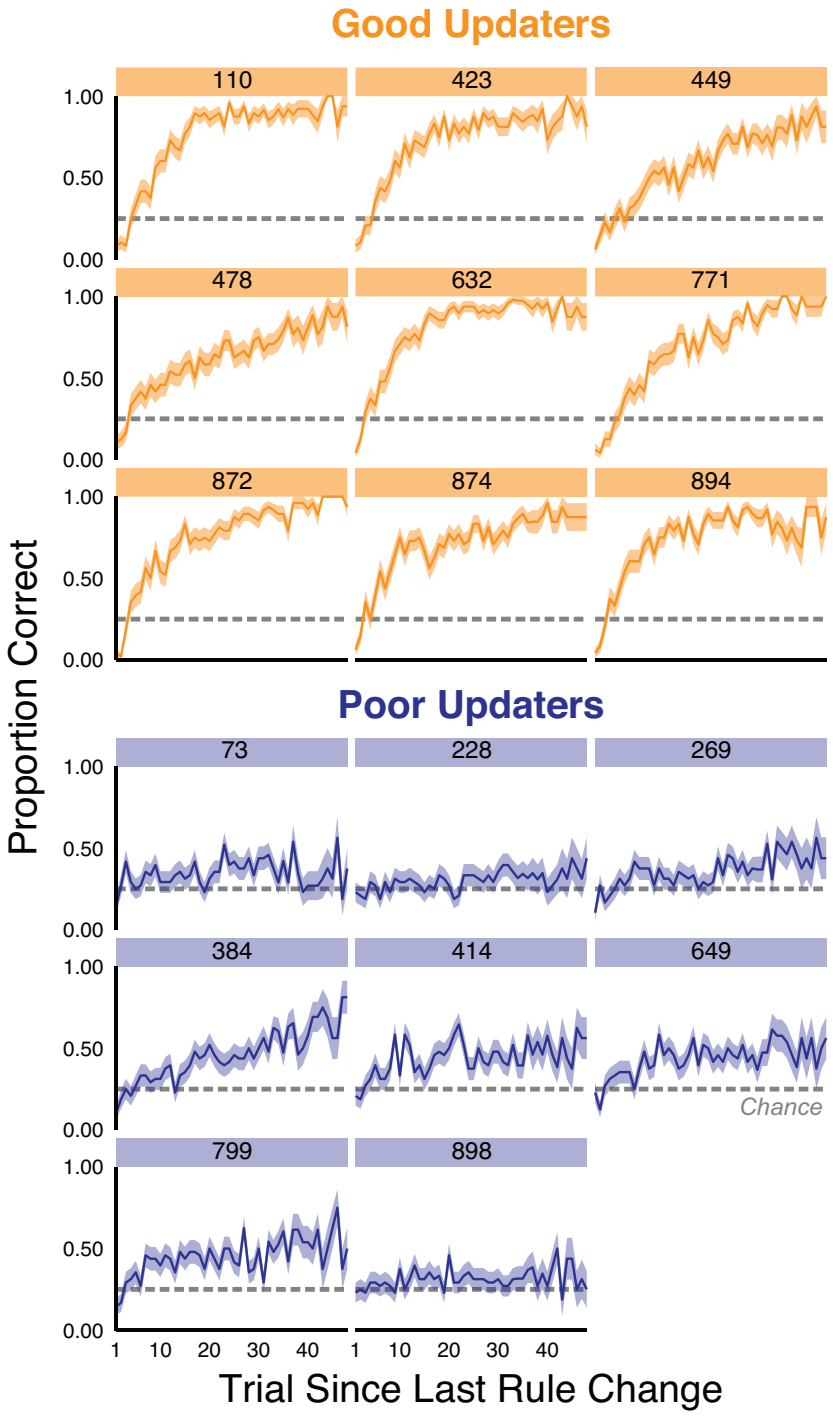

Figure 4.6. Individual patient performance. Correct response rates are displayed for each patient. Subject numbers are indicated at the top of each graph. Patients that performed at or above the level of healthy older adults were classified as Good Updaters (orange) and those that performed below this mark were classified as Poor Updaters (blue). Solid lines represent each participant's mean correct response rate across all session, and shading representing ± one standard error of the mean.

Demographic information of Good and Poor Updaters was compared using independent t-tests. These comparisons revealed a trending but non-significant difference in age (Mean age (95% CI): Good Updaters = 61.6 years (57.8-65.2), Poor Updaters = 70.9 years (61.6-80.2); $t(9.18)=1.838$, $p = .10$), and no differences in years of education (Mean years of education (95% CI): Good Updaters = 13.0 years (11.1-14.9), Poor Updaters = 11.5 years (10.0-13.0); $t(14.46)=1.207$, $p = .245$), MoCA scores (Mean MoCA scores (95% CI): Good Updaters = 24.8 (22.8-26.8), Poor Updaters = 22.5 (18.9-26.1); $t(11.11)=1.085$, $p = .301$), lesion volume (Mean lesion volume (95% CI): Good Updaters = 6,765 voxels (324-13,208), Poor Updaters = 11,894 voxels (4,281-19,506); $t(14.23)=1.008$, $p = .330$), or time post-stroke (Mean months post stroke: Good Updaters = 32.4 months (14.1-50.8), Poor Updaters = 30.1 months (11.5-48.6); $t(14.90)=.179$, $p = .860$).

Consistent with this partition, Good Updaters performed significantly better than Poor Updaters (Mean correct response rate (95% CI): Good Updaters = .70 (.68-.72), Poor Updaters = .38 (.37-.39); $\chi^2(1) = 32.986$, $p < .001$). The performance of the Good Updater group was nominally but not significantly higher than the HC-O group ($\chi^2(1) = 3.643$, $p = .06$), and very similar to the HC-Y participant groups ($\chi^2(1) = .001$, $p = .976$). Poor Updaters performed worse than the healthy control groups (all ps $< .001$). It is worth noting however, that the Poor Updater patient group did perform above chance ($t(7) = 5.365$, $p < .002$), indicating that they were, to some extent, learning and updating to task contingencies (Figure 4.7).

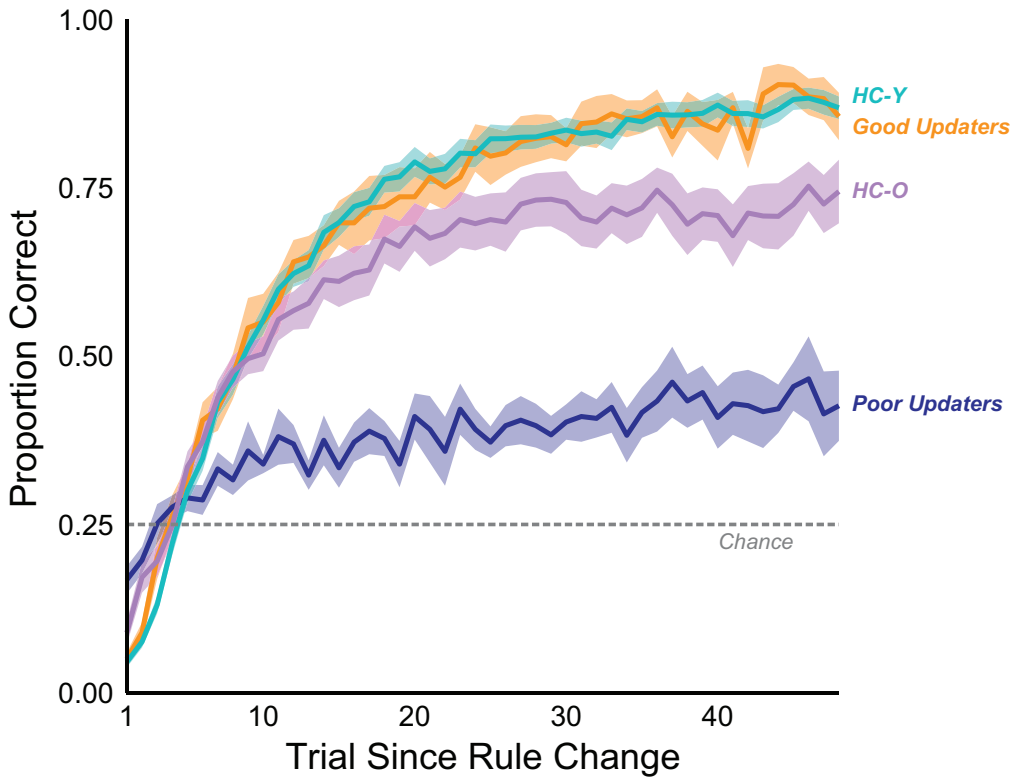

Figure 4.7. PROBE task performance between Good and Poor Updaters. Participant correct response rates were averaged over all task trials since the last rule change and compared between healthy controls and Good and Poor Updaters. Good Updaters (orange) performed as well as healthy younger adults (HC-Y – cyan) and healthy older adults (HC-O - purple). Poor Updaters (blue) performed above chance, but had lower correct response rates than any of the other groups. Solid lines represent group means and shading represents ± one standard error of the mean.

To determine the factors differentiating Poor and Good Updaters, participant susceptibility to 'trap-trials' was analyzed (*i.e.*, responses on the trials in which incorrect responses were marked "correct", and correct responses were marked as "incorrect"). 'Susceptibility' to trap-trials was assessed by measuring the probability of making a correct response *after* having seen a trap-trial – the higher the probability of making a correct response after a trap-trial, the lower the trap-trial susceptibility. After a rule has changed, susceptibility to trap trials should decrease as participants have more experience with the new rule. To measure this, trials since the last rule shift were binned into quartiles of 12 trials, and the proportion of correct responses following trap-trial trials were calculated for each quartile. Susceptibility should be highest in the first quartile, nearest when a switch has occurred, and lowest in the last bin, when participants have learned the new rule.

Performance was examined using a mixed factorial ANOVA, with mean proportion of correct responses *after* a trap trial as the dependent variable, Participant Group (HC-Y, HC-O, Good Updater, Poor Updater) as a between subjects factor, and Trial Bin (First, Second, Third, and Fourth quartile) as a within subjects factor. The omnibus analysis revealed an overall significant main effect for both Trial Bin ($F(3, 195) = 221.70$, $MSE = .01$, $p < .001$), and Participant Group ($F(3, 65) = 24.44$, $MSE = .05$, $p < .001$), along with a Trial Bin x Patient Group interaction ($F(9, 195) = 9.33$, $MSE = .01$, $p < .001$). Separate repeated measures ANOVAs conducted for each participant group revealed a main effect of Trial Bin for HC-Y participants ($F(3, 99) = 189.80$, $MSE = .01$, $p < .001$), HC-O participants ($F(3, 51) = 29.11$, $MSE = .01$, $p < .001$), and Good Updaters

($F(3, 24) = 35.04$, $MSE = .01$, $p < .001$), but no main effect for Poor Updaters ($F(3, 21) = 1.674$, $MSE = .008$, $p = .203$).

One way ANOVAs were next run to compare correct response rates between participant groups at each quartile. Mean correct response rates did not differ between participant groups for the first quartile (Mean Proportion Correct Responses (95% CI): HC-Y = .28 (.27-.30), HC-O = .26 (.25-.27), Good Updaters = .26 (.25-.28), Poor Updaters = .23 (.21-.25); $F(3,65) = 1.309$, $MSE = .01$, $p = .279$). However, correct response rates differed in the second quartile (HC-Y = .66 (.64-.69), HC-O = .52 (.48-.57), Good Updaters = .61 (.56-.66), Poor Updaters = .29 (.26-.32); $F(3,65) = 12.67$, $MSE = .03$, $p < .001$), third quartile (HC-Y = .77 (.75-.79), HC-O = .56 (.50-.60), Good Updaters = .74 (.70-.78), Poor Updaters = .30 (.28-.33); $F(3,65) = 26.31$, $MSE = .02$, $p < .001$), and fourth quartile (HC-Y = .80 (.78-.83), HC-O = .58 (.53-.63), Good Updaters = .70 (.64-.75), Poor Updaters = .32 (.28-.37); $F(3,65) = 20.72$, $MSE = .03$, $p < .001$).

Tukey HSD post-hoc tests were performed between participant groups at the second, third, and fourth quartiles to determine which groups differed in correct response rates. As is evident from Figure 4.8, Poor Updaters had significantly lower correct response rates than any other group by the second quartile (all ps $< .006$), and this difference continued into the third and fourth quartiles (all ps $< .003$). Conversely, correct response rates for Good Updaters did not differ significantly from HC-Y participants in any quartile (all ps $> .347$), did not differ significantly from HC-O participants in the second or fourth quartiles (all ps $> .304$), and had higher correct responses rates than HC-O participants in the third quartile. These results indicate that part of the difficulty for Poor Updaters is correctly identifying trap trials, and ignoring the incongruent feedback.

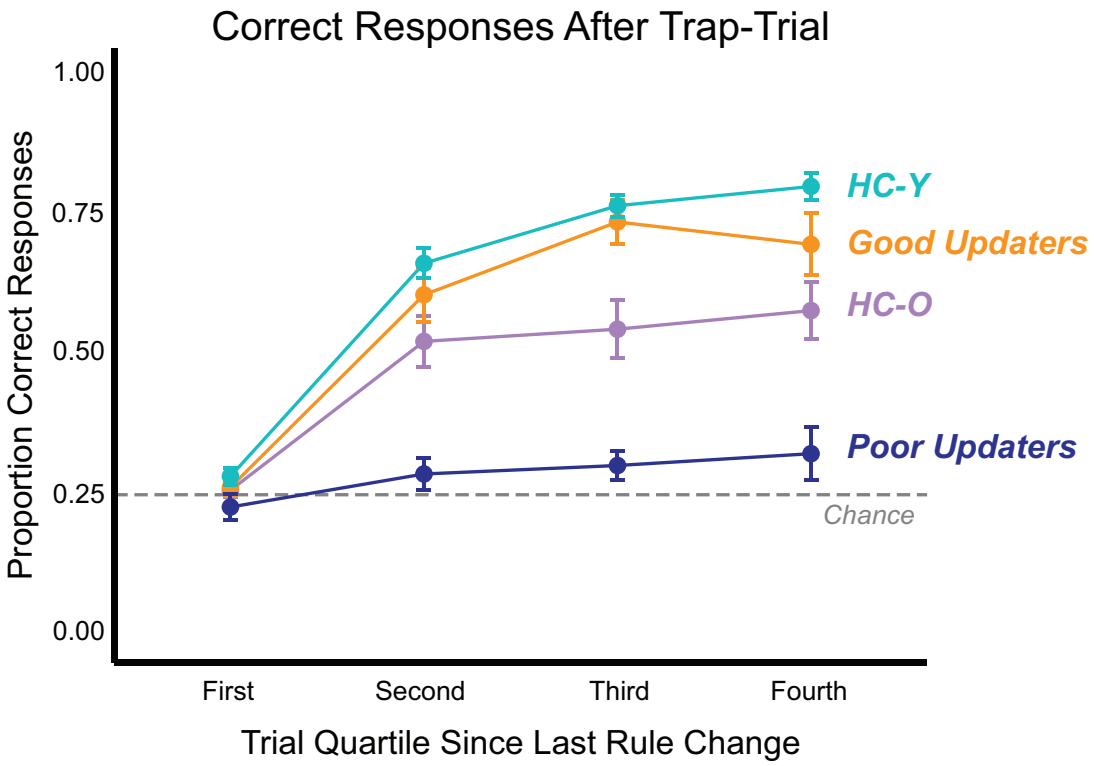

Figure 4.8. Participant correct responses following trap-trials. Participant responses were binned into four quartiles of trials and mean correct response rates were calculated for trials that followed trap-trials. Overall, Poor Updaters (blue) had lower correct response rates following trap-trials in the last three quartiles of trials than healthy younger adults (HC-Y – cyan), healthy older adults (HC-O – purple), and Good Updaters (orange). HC-O participants also had lower correct response rates in the last three quartiles of trials than HC-Y participants. Dots represent group means at each quartile and error bars represent ± one standard error of the mean.

There were also differences in correct response rates between the healthy control groups. HC-Y participants had higher correct response rates than HC-O participants throughout the second, third, and fourth quartiles (all *p*s < .03). This result suggests that some of the updating differences found between HC-O and HC-Y participants could be due to HC-O participants having more difficulty identifying trap-trials than HC-Y participants (Figure 4.8).

Model fits of patient behaviour

Patient performance was next compared by fitting correct response rates using both the PROBE and RL models (see Table 4.3 for mean group fit parameters). Goodness of fit was assessed by computing Akaikes Information Criterion (AIC; Akaike, 1974) and Bayes Information Criterion (BIC; Schwarz, 1978) values using the maximum likelihood estimates obtained from the best fitting sets of parameters for each model (mean values displayed in Table 4.3). Both of these values provide information about the comparative goodness of fit between different candidate models, while correcting for the number of free parameters contained within each model (penalizing models with higher numbers of free parameters). Both of these values are commonly used, with BIC values providing slightly higher penalties for added free parameters (Wagenmakers & Farrell, 2004).

Table 4.3 PROBE and RL Model Parameter Fits.

	HC-Y	HC-O	Good Updaters	Poor Updaters
PROBE parameters				
AIC values	2776 (117)	3378 (224)	2712 (181)	4908 (99)
BIC Values	2815 (117)	3418 (224)	2752 (181)	4947 (99)
Buffer capacity (N)	2.41 (.14)	3.44 (.33)	3.44 (.47)	4.5 (.46)
Volatility (τ)	.12 (.01)	.08 (.01)	.09 (.02)	.07 (.01)
Recollection entropy (η)	.24 (.03)	.45 (.07)	.36 (.11)	.82 (.07)
Prior reliability bias (θ)	.47 (.25)	1.40 (.16)	.35 (.56)	1.81 (.25)
Learning rate (α)	.28 (.02)	.45 (.05)	.36 (.05)	.22 (.07)
Inverse temperature (β)	16 (3)	32 (17)	12 (4)	63 (54)
Noise (ε)	.08 (.01)	.07 (.01)	.04 (.01)	.17 (.03)
RL parameters				
AIC values	3347 (75)	3755 (157)	3342 (104)	4709 (193)
BIC Values	3364 (75)	3773 (157)	3359 (104)	4726 (193)
Learning rate (α)	.81 (.02)	.80 (.05)	.84 (.03)	.39 (.15)
Inverse temperature (β)	4 (.15)	4 (.61)	5312 (5309)	6 (2)
Noise (ε)	.07 (.01)	.08 (.02)	.06 (.02)	.16 (.05)

Note. Numbers represent mean parameter values with the standard error of the mean noted in parentheses. HC-Y = Healthy Control – Young Adults; HC-O = Healthy Control – Older Adults.

To compute model evidence, Akaike weights (AICw) and BIC weights (BICw) were computed for each model fit. These values indicate the relative goodness of fit of a particular model compared to other candidate models (with values tending towards 1 indicating stronger model evidence; Wagenmakers & Farrell, 2004)

When the model evidence was compared between RL fits and PROBE fits in the Good Updaters group, all nine patients were better fit by the PROBE model than the RL model (PROBE AICw and BICw values all > .999; Figure 4.9). Of the eight patients classified as Poor Updaters, five were better fit by the RL model than the PROBE model (patients 73, 228, 384, 414, 898; RL AICw and BICw values all > .999), and one was better fit by the PROBE model (patient 649, PROBE AICw and BICw value > .999). The remaining two patients (269 and 799) had AICw values that favoured the PROBE model (both AICw values = .99), but BICw values that favoured the RL model (RL BICw values: 269 = .632, 799 = .887). For these last two patients, given that BIC values are more conservative than AIC values, this difference in AIC and BIC model evidence suggests that the PROBE model likely provides a better fit, but not much beyond the fits provided by the RL model (Wagenmakers & Farrell, 2004).

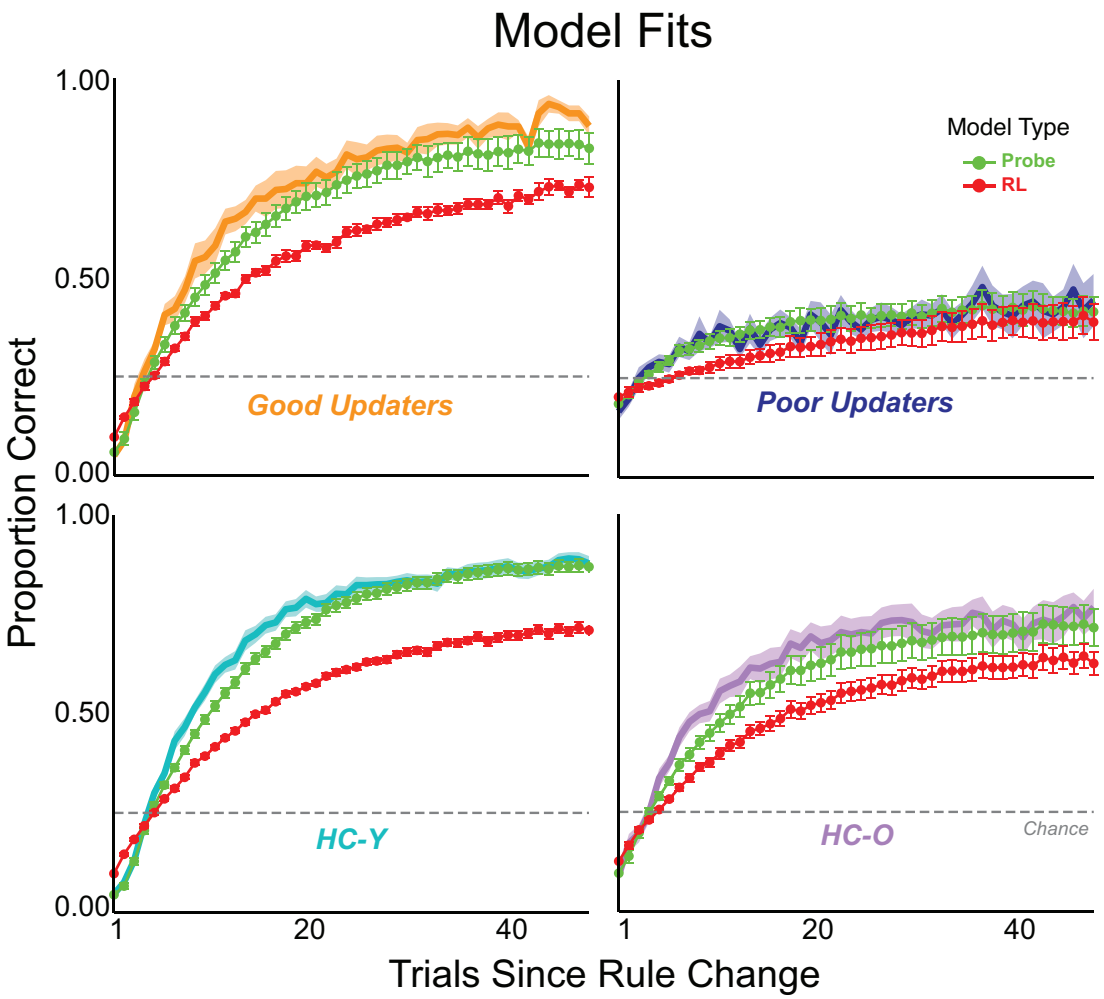

Figure 4.9. PROBE and reinforcement learning (RL) model fits for all participant groups. Participants correct response rates were fit with the PROBE and RL models. Individual participant PROBE and RL parameters were used to run simulations for each participant and the mean simulation results are displayed above. Good Updaters (orange), healthy younger adults (HC-Y – cyan), and healthy older adults (HC-O – purple) were overall better fit by the PROBE model (green). The majority of Poor Updaters (blue) were best fit by the RL model (red). Solid lines and dots represent group means, and shading and error bars represent ± one standard error of the mean.

The same model evidence comparisons were computed for the two healthy control groups. Overall, the PROBE model provided a better fit than the RL model for 30 of the 34 participants in the HC-Y group and 17 of the 18 participants in the HC-O group (all PROBE AICw and BICw values > .999). The remaining four HC-Y and one HC-O participants were better fit by the RL than the PROBE model (all RL AICw and BICw values > .999).

Model parameter differences between participant groups

Participant model parameters were then compared between participant groups using one way ANOVAs and subsequent Tukey HSD post-hoc test. It is important to note that comparing parameters in this way is primarily meaningful between groups that are best fit by the same model. With this in mind, the noise (ε) parameter in both models seemed to distinguish the Poor Updaters from the other participant groups. This parameter represents the proportion of the data that cannot be explained by either model (*i.e.*, responses that are as likely to have been generated by a random process than from the model components). There was a significant difference between groups when comparing PROBE ε values ($F(3,65) = 11.15$, $MSE = .002$, $p < .001$), with Poor Updaters having higher ε values than all other participant groups (all Tukey HSD p values <.001). The RL noise values also differed between groups ($F(3,65) = 3.105$, $MSE = .006$, $p < .04$), with Poor Updater having higher noise values than HC-Y participants ($p < .03$) and Good Updaters ($p < .05$), and nominally but not significantly higher values than HC-O participants ($p = .11$). These results suggest that the learning and updating processes in Poor Updaters were noisier than those in participants from other groups. It is worth

noting, however, that when the AIC values were compared between Poor Updater RL fits, and AIC and BIC values obtained from a model assuming random responses, all patient AICw and BICw values favored the RL model over this random model (all AICw and BICw values > .999). This demonstrates that although noisy, the updating processes used by Poor Updaters still contained some structure.

Some of the PROBE parameter values also differed between the HC-Y and HC-O participants. Learning rate (α) values differed between participant groups ($F(3,65) = 5.821$, $MSE = .03$, $p < .002$), with HC-O participants having higher α values than HC-Y participants ($p < .007$), as did prior reliability bias (η) values ($F(3,65) = 13.99$, $MSE = .06$, $p < .001$), volatility (τ) values ($F(3,65) = 5.132$, $MSE = .002$, $p < .004$), and buffer (N) values ($F(3,65) = 9.112$, $MSE = 1.275$, $p < .001$). When comparing parameter differences between the three participant groups best fit by the PROBE model, the HC-O group had higher mean α, η, and N values than participants in the HC-Y group (all ps < .02), while the HC-Y group had higher mean τ values than participants in the HC-O group ($p < .01$). None of these parameters differed between Good Updaters and either healthy control group (all ps > .08).

Lesion analyses

Damaged brain regions were compared across patients to determine whether any candidate regions were related to the observed performance differences. Since there were no overall updating differences between RBD and LBD patients, lesions were flipped such that the hemisphere with the most significant area of brain damage was compared across subjects (*i.e.*, the right hemisphere regions damaged in RBD patients were

compared to the left hemisphere brain damaged regions in LBD patients). All lesion analyses below were performed on these flipped scans.

Lesion overlay analyses were performed on the lesions traces for Good and Poor Updaters. Among the 8 patients identified as Good Updaters, four had overlapping lesions in parts of the precentral sulcus, inferior frontal gyrus pars opercularis and triangularis, rolandic operculum, and the insula. Among the nine patients identified as Poor Updaters, five had overlapping damage to the inferior frontal gyrus pars opercularis and pars triangularis, rolandic operculum, and insula.

Although the insula was commonly affected in both Good and Poor Updaters, different subregions of the insula were affected between groups. Of the 13 patients with insular damage, nine patients had damage to the *anterior* portion of the insula, while the remaining four had damage in middle or posterior insular regions. Of the eight patients classified as Poor Updaters, seven were identified as having anterior insular damage, while the remaining two patients with anterior insular damage, and all four patients with middle/posterior insular damage, were classified as Good Updaters. It is worth noting that the one Poor Updater that did not present with anterior insular damage (patient 649) was also the only patient from the Poor Updater group with both AICw and BICw favouring the PROBE model over the RL model.

As an additional analysis the flipped lesions were analyzed using voxel-based lesion-symptom mapping (VLSM; Bates et al., 2003). Continuous statistics were computed using non-parametric Brunner-Munzel rank order statistics, implemented in the NPM software package (Rorden et al., 2007). Non-parametric permutation thresholding was applied to correct for multiple comparisons and ensure a family-wise error rate of 5%

(Nichols & Holmes, 2002; Kimberg, Coslett, & Schwartz, 2007). Mean patient correct response rates across all sessions were used as a continuous variable to identify the regions of brain damage that best predict updating deficits on the PROBE task. Given the relatively small sample size, to reduce the number of voxel-wise comparisons and increase statistical power, only voxels damaged in at least 30% of patients (*i.e.*, 5 patients) were used for the analysis. Overall, this analysis revealed a cluster of three adjacent brain regions that were significantly related to mean correct responses (required BM statistic = 3.34, observed BM statistic = 3.54). This cluster included the anterior insula (region size = 59 voxels), and two regions of the inferior frontal gyrus (pars orbitalis, region size = 34 voxels, and pars triangularis, region size = 112 voxels; Figure 4.10).

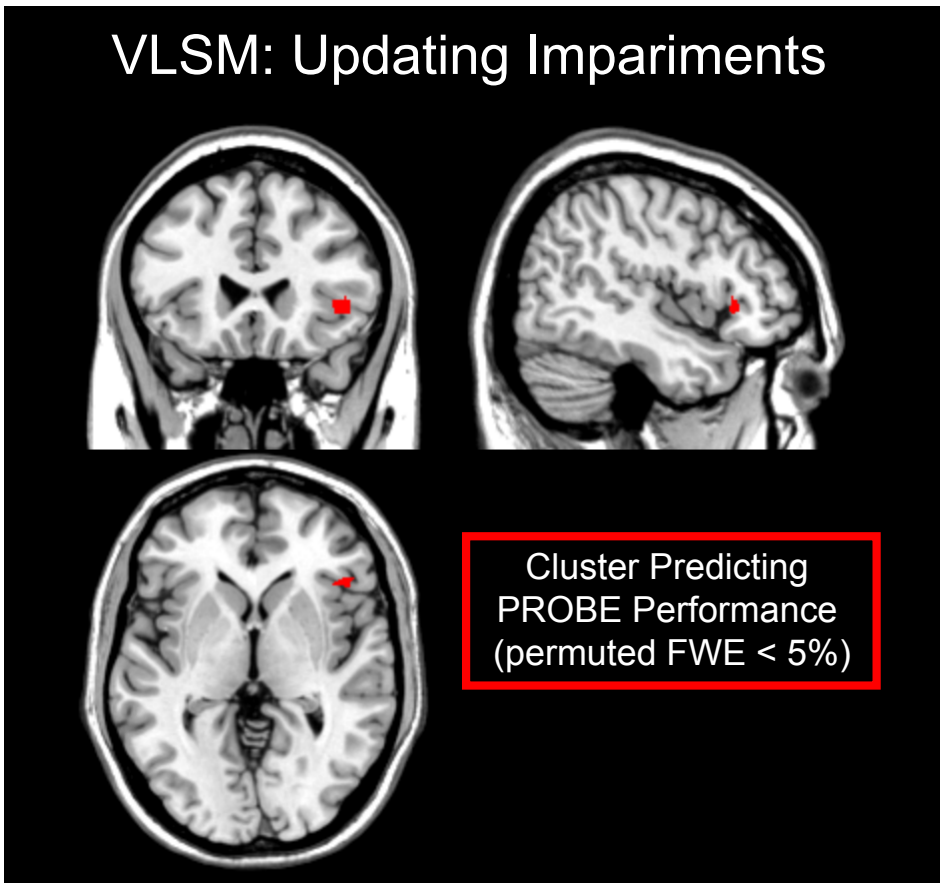

Figure 4.10. Voxel cluster related to PROBE task updating performance. Voxel-based Lesion Symptom Mapping (VLSM) was performed comparing patient flipped lesions to their mean correct response rates on the PROBE task. This analysis revealed a cluster of voxels (red) comprising adjacent regions of the anterior insula and inferior frontal gyrus that significantly predicted deficits in updating performance (with a permuted family-wise error rate < 5%).

Next, additional VLSM analyses were performed using model fit and noise values since these values differed between the Good and Poor Updater groups. Model fit values were computed for each subject as a difference scores between each patient's PROBE AIC and BIC values, and their respective RL AIC and BIC values. It is worth noting that AICs and BICs are both transformations of likelihood values; therefore, although their absolute values may differ, the proportional difference between PROBE and RL AIC values will be the same as the proportional differences between PROBE and RL BIC values. A VLSM analysis of AIC and BIC difference scores did not reveal any brain regions with large enough BM statistics to meet the required permutation threshold (required BM statistic = 3.41). However, the same exact cluster identified for mean correct responses was identified as the region with the highest BM statistic (observed BM statistic = 3.26; uncorrected *p* value = .0006).

A VLSM analysis using patient RL noise parameters as a predictor also failed to find any significant clusters (required BM statistic = 3.30). However, the highest identified BM statistic was found in a small cluster of the anterior insula (region size = 9 voxels; observed BM statistic = 2.92; uncorrected *p* value = .002).

4.4. Experiment 4.1: Discussion

The current study had two goals: 1) to characterize behaviour during updating to identify the exploratory strategies used by RBD and LBD patients, and 2) identify the contributions of damage to specific brain regions to updating performance deficits. Based on previous results, RBD patients were expected to show poorer updating than LBD

patients. As evident in Figure 4.4, overall updating performance did *not* differ between LBD and RBD patients.

Although no overall updating differences were found, LBD patients did perform better in the *recurrent* session than in the *open* session, whereas RBD patients did not. These differences do not seem related to explicit recall abilities: LBD patients could not recall the recurring rules in the post-experimental questionnaire any better than RBD patients. This difference could be attributed to *implicit* learning properties that some researchers have linked to right hemisphere functioning. Wolford and colleagues (2000) found that split-brain patients showed an advantage in statistical learning tasks when information was presented in their left hemifield (*i.e.*, to their right hemisphere), compared to when presented in the right hemifield (*i.e.*, to their left hemisphere; Wolford, Miller, & Gazzaniga, 2000). Damage to the right hemisphere has also been linked to statistical and implicit learning deficits (Shaqiri & Anderson, 2013). Conversely, LBD has in some cases provided *improved* performance on statistical learning tasks. LBD patients in Danckert and colleagues' study (2012) *outperformed* healthy controls on the RPS scissors task, using a more optimal strategy to exploit the statistics afforded by the play environment. Although speculative, the consistent advantage LBD patients demonstrated in the recurrent condition may indicate that implicit learning advantages supported by intact right hemisphere structures may have been at play in their performance.

It is possible that the main hemispheric differences found in previous studies may be due to more posterior brain regions, rather than the frontal regions targeted in the current study. Frontal regions were chosen for the current study because of their presence

in a number of previous updating studies, particularly those interested in the *exploratory* components of updating (*e.g.*, Donoso et al., 2014). Patients in the studies performed by Danckert and colleagues (2012) and by Stöttinger and colleagues (2014) had a less restricted range of damage, many of them being recruited based on the presence or absence of spatial neglect (a clinical syndrome most commonly found with superior temporal/inferior parietal damage; Corbetta, Kincade, Lewis, Snyder, & Sapir, 2005). Additionally, the right temporoparietal junction (TPJ) and the right inferior parietal lobule (IPL) have been linked to the capacities of theory of mind and empathy (Decety & Lamm, 2007), and even understanding jokes (Brownell, Michael, Powelson, & Gardner, 1983) and riddles (Vocat et al., 2012). Some researchers have proposed that these deficits reflect problems of updating (Geng & Vossel, 2013; Filipowicz et al., 2016): theory of mind and empathy require an agent to update their perspective to infer the cognitive or emotional state of others, while jokes and riddles work by setting up an expectation that is then violated with the punch line. Future research examining patient populations with more posterior lesions, in addition to converging methodologies such as fMRI and lesion simulation techniques (*e.g.*, transcranial magnetic stimulation), could help identify whether more posterior regions such as the IPL or TPJ contribute to the hemispheric effects observed in previous studies.

Although there were no clear hemispheric differences in the current study, frontal brain damage clearly has an influence on updating performance. Patient performance collapsed across hemispheres separated into two distinct groups: approximately half of the patients updated at or above the level of healthy older adults (Good Updaters), whereas the other half of patients performed well below this level (Poor Updaters; Figure

4.6). A closer examination of patient performance demonstrated that Poor Updaters were more susceptible to 'trap trials' than Good Updaters, making it more difficult for Poor Updaters to settle on a new mental model after a rule switch. This tendency reflects an updating strategy that makes *incremental* changes to an existing mental model, where feedback is used to update specific stimulus-action mappings. Consistent with this interpretation, the majority of Poor Updaters were better fit by the RL model than the PROBE model.

In contrast to the RL model, the PROBE model assumes that, rather than making incremental changes to an existing model, memory and executive processes arbitrate between different candidate mental models. All of the Good Updaters were best fit by the PROBE model, demonstrating that performance differences between Good and Poor Updaters can be attributed to the efficiency of the exploratory strategies they used to update. Additionally, when examining model fit parameters, Poor Updaters had significantly higher ε noise parameters than did the Good Updaters. This noise parameter measures the proportion of responses that are better explained by random responses than by those predicted by the model. This suggests that, in addition to using differing exploratory strategies, the Good Updaters used their strategies more consistently, whereas Poor Updaters where more prone to make random responses. It is important to note that Poor Updaters were not *only* making random responses – these patients updated at a rate above chance, and were better fit by the RL model than by a random response model.

The final analysis was to examine the differences in damaged brain regions that could contribute to some of the observed updating difficulties. Lesion overlay and VLSM

analyses revealed two regions that were significant predictors of updating behaviour: the anterior insula and the inferior frontal gyrus (IFG). Previous research has implicated both of these regions in updating behaviour, particularly during the exploratory phase.

The anterior insula is often implicated in updating research (McGuire et al., 2014; Stöttinger et al., 2015; Palminteri et al., 2012; Preuschoff, Quartz, & Bossaerts, 2008). Its role has primarily been suggested to represent the *uncertainty* of a given environment (Preuschoff et al., 2008; McGuire et al., 2014). Menon and Uddin (2010) also provide evidence that the anterior insula and anterior cingulate cortex (ACC) form a 'salience' network that coordinates activity of a fronto-parietal 'central executive' network responsible for processing salient, bottom-up sensory information (Menon & Uddin, 2010). These interpretations match results from an fMRI study of perceptual updating discussed earlier (Stöttinger et al., 2015). Participants in this experiment saw images that gradually morphed from one object (*e.g.*, a cat) to a second object (*e.g.*, a swan), indicating via button press the point at which they started seeing the second object. The anterior insula and central executive network were found to be strongly active when participants reported a switch. However, the anterior insula and medial prefrontal cortex (mPFC; including the dorsal ACC) were also found to be active on the image *preceding* the switch (images with high perceptual uncertainty; Stöttinger et al., 2015).

A difficulty in understanding the specific functions of the insula during updating is that it is often co-activated with the ACC/mPFC (Stöttinger et al., 2015; McGuire et al., 2014; Menon & Uddin, 2010). Functional imaging evidence provides insight into the distinct functions these regions subsume during the updating process. In an fMRI study of the frontal regions involved in updating using the PROBE task, the ACC/mPFC was

primarily active when participants detected that their current rule was no longer valid, and entered the exploration period (Donoso et al., 2014). That is, the ACC/mPFC was most active at the points during the task where the rule a participant was using became *unreliable*, and the participant entered an *exploration* phase. These results propose that the primary role for the ACC/mPFC is to *initiate* the exploration process once a mental model is deemed unreliable. In contrast, this same study also found that the anterior insula was active *during* the exploration period, *after* a switch had been detected. This proposes that, in the context of updating, the role of the ACC/mPFC could be to *initiate* the exploratory process, while the anterior insula *sustains* exploration.

This interpretation is consistent with the results of the current study. Eight of the nine Poor Updaters had damage to the anterior insula. These patients did not have difficulty detecting that changes were occurring – if anything, their responses to 'trap-trials' suggest the opposite: they may have been susceptible to thinking that changes were occurring too frequently (an observation that was sometimes informally reported by patients in this group). Instead, as suggested by the differences in noise parameter estimates, patients in this group seemed to have difficulty *sustaining* a reliable updating strategy, sometimes lapsing into periods of random responses. Indeed, although tentative, the VLSM analysis found that the anterior insula was the brain region most strongly linked to higher noise values.

The lateral prefrontal cortex has also been implicated in the exploratory components of updating. Theories on the executive organization of the prefrontal cortex implicate the lateral prefrontal cortex, including regions of the IFG observed in the current study, in the processes required to arbitrate between and settle on a new candidate

mental model (Domenech & Koechlin, 2015; Koechlin & Summerfield, 2007). The VLSM analysis on model fit differences found that damage to the IFG and anterior insula increased the likelihood that patients would be best fit by the RL model over the PROBE model. This finding supports the notion that the IFG is involved in the exploratory processes required to consider alternative models when updating. Functional imaging studies also support this explanation. In an fMRI study of the PROBE task, regions of the IFG were associated with updating events in which a counterfactual rule being considered becomes reliable (Donoso et al., 2014). In an fMRI study of the picture morphing task, the IFG was one of the regions that was strongly active when participants reported switching from an initial perceptual representation to a novel second representation of the ambiguous image (Stöttinger et al., 2015). Both of these activation patterns could be related to participants considering alternate, candidate interpretations to fit the perceptual evidence they observe.

In addition to the patient updating behaviour, some differences were observed between younger and older healthy controls. Older adults did not perform as well as younger adults on the PROBE task, being more susceptible to trap-trials, and showing no benefit for the recurrent session. Although both control groups were best fit by the PROBE model, differences in the model parameters provide some hints as to why these groups differed.

Older adults had higher η prior reliability values, indicating less influence from previously learned strategies. This result likely explains why older adults did not show any benefits of the recurrent session over the open session. Additionally, older adults had higher α learning rates and lower τ volatility values than younger adults. The learning rate

parameter captures the changes made to the mental model driving behaviour (*i.e.*, the actor), while the volatility parameter captures the perceived rate of change *between* different task rules. This suggests that, relative to younger adults, older adults focused more on making changes to their existing mental models, leading to more susceptibility to trap-trials, whereas younger adults were relatively more focused on detecting rule changes. Indeed, recent research has found that the ability to learn the volatility, or 'hazard rate', of an environment declines with age, particularly in highly uncertain environments (Nassar, Bruckner, Gold, Li, Heekeren, & Eppinger, 2016).

Although not the primary goal of this experiment, the results suggest that updating exploration strategies change with age. Future studies specifically designed to measure these differences could provide a better understanding of how learning and adaptive behaviour evolve over time.

In summary, this study examined some of the potential consequences of frontal brain damage on mental model updating. Although none of the predicted hemispheric differences were present in the current sample, behavioural results clearly indicated a bimodal distribution of performances among the patients (Figure 4.6). Some patients managed to update efficiently by choosing an exploratory strategy and using it consistently. Others updated more poorly – their difficulties attributed to problems sustaining exploratory strategies, while also being prone to noisy feedback. Lesion comparisons between these two patient groups suggested that damage to the inferior frontal gyrus and to the anterior insula may have contributed to the observed differences. Although speculative at this point, the inferior frontal gyrus may be involved in arbitrating between distinct strategies when the need for updating arises, while the

anterior insula may be involved in efficient exploration. While more research will be required to confirm these hypotheses, these results provide insights into the specific regions involved in updating.

Chapter 5: General Discussion

Effective behaviour relies on our ability to build and update mental models of our world (Tenebaum et al., 2011; Bach & Dolan, 2012; Johnson-Laird, 2004). The current thesis examined some of the component processes involved in updating, with the aim of refining our understanding of the behavioural and neural mechanisms involved. This final chapter provides a summary of the major findings from this thesis, while highlighting limitations and suggestions for future research.

Chapter 2 examined the challenge of representing participant mental models in a probabilistic learning task. This chapter was focused on testing assumptions about how prior mental models are measured, and how these priors influence learning. Contrary to methods that assume homogeneous priors, participants started the task with a number of idiosyncratic priors (Experiment 2.1). These priors also influenced how effectively participants managed to learn from new events in the environment, highlighting the importance of accurately capturing participant priors.

A limitation to the generalizability of Experiment 2.1 is that participants were only exposed to environments with a known solution (*i.e.*, ball drops in a standard Plinko game should approximate a normal distribution; Galton, 1889). In a different environment, where contingencies are more uncertain (*e.g.*, a Plinko task with a variable start point for the ball, or where the pegs are occluded) it may be possible that priors would be more homogeneous. Although priors may be homogeneous in some task environments, this fact is often assumed rather than tested explicitly (Nassar et al., 2010, 2012; Strange et al., 2005; Mars et al., 2008; McGuire et al., 2014). Indeed, priors can still be found to differ between subjects, regardless about how much or how little

information is provided about task contingencies (Green et al., 2010). Future research interested in measuring probabilistic learning should characterize the types of possible priors participants bring to a task, whether to confirm that they are homogeneous, or capture the different types of priors that are possible in a given environment.

An additional factor that was not considered in Experiment 2.1 was the contribution of a person's *confidence* in their prior. Confidence has been shown to influence decision making, with research generally finding that high mental model confidence makes it less likely for new information to be integrated (Einhorn & Hogarth, 1978; Fischoff, Slovic, & Lichtenstein, 1977; Hollard & Massoni, 2015; Miller, Spengler & Spengler, 2015; Nickerson, 1998). It is possible that although priors could differ between participants, low confidence in these idiosyncratic priors could mitigate the influence of these priors on learning. For example, participants in the Jagged condition started the task with a few interspersed bars – in strictly probabilistic terms, this indicates a prior with *high certainty* about where a ball would fall on future trials given that a few slots are represented with high relative mass. However, in some cases these participants also reported drawing a few bars to get more information from the task, suggesting that they had *low* confidence in their initial estimates. This suggests that to fully capture the influence of priors, it may be important to measure both the type of prior (*e.g.*, in the Plinko task, the probability distribution drawn by a participant), and the degree of confidence in that prior (*e.g.*, a rating of confidence in how likely the prior is to occur). Future studies could, in addition to capturing participant priors, ask participants to report how confident they are in their priors. These results could provide important information

as to the relationship between the priors participants bring to a task, and how their influence is affected by confidence.

Chapter 3 examined how unexpected, *surprising* events influence our ability to detect and update to environmental changes. A number of previous studies propose that surprise and updating should be positively correlated (Nassar et al., 2010; O'Reilly et al., 2013; McGuire et al., 2014). Other research opposes this view, suggesting that highly surprising information is devalued (Nickerson, 1998; DeGardelle & Sumemrfield, 2010). The results from Experiment 3.1 and 3.2 found that in some circumstances, surprise and updating are positively related, while in others, they are negatively correlated.

An extension to this research is to understand the specific situations in which participants devalue surprising information. The nature of the high surprise changes in Experiment 3.2 was that surprising information was intermixed with low surprise events. Under this scenario, it is possible that the weight given to the surprise of an observation is influenced by its proximity in time to low surprising events. Under an efficient coding framework, in order for highly surprising information to be integrated, these events would need to be presented in near succession, with relatively few low surprise events in between (Wei & Stocker, 2015; Summerfield & Tsetsos, 2015). The applications of efficient coding are only starting to be explored in the realms outside of perception (Summerfield & Testsos, 2015), but could provide an explanation for some of the seeming discrepancies between optimal and suboptimal accounts of mental model building and updating. Future computational modeling attempts, particularly those adopting a Bayesian framework, could model these differences by constraining some of the components of their models through the principles of efficient coding.

Chapter 4 examines the strategies used to *explore* alternative mental models once a change has been detected. This was specifically examined in populations with right and left hemisphere brain damage, as previous updating deficits in patients with RBD have been suggested to stem from issues with exploration (Sepavhand et al., 2014; Filipowicz et al., 2016). Although Experiment 4.1 did not reveal any clear updating differences between right and left hemisphere brain damaged patients, patient performance clearly separated into a group of 'Good Updaters' and a group of 'Poor Updaters'. Computational modeling of patient performance determined that the exploratory strategies used differed between these groups, with Good Updaters using more effective exploration strategies that Poor Updaters. Additionally, Good Updaters made more consistent use of their strategy, while Poor Updaters were more prone to committing noisy and inconsistent responses. Patient lesion analysis revealed that the damage to the anterior insula and inferior frontal gyrus (IFG) predicted updating difficulties, suggesting that these regions are involved in selecting and maintaining a participant exploration strategy when updating.

A limitation with Experiment 4.1 is that the results were obtained from a relatively small sample size of patients for a VLSM study. Additionally, although efforts were made to recruit patients with primarily frontal damage, the locations of individual lesions still differed considerably.

A major strength of patient research is that, while imaging methods such as fMRI provide information about the regions and networks *involved* in certain cognitive processes, patient work can provide information about those that are *necessary* (Rorden et al., 2007). To fully understand the role and extent of brain networks involved in any

given process, converging methodologies that include neuroimaging and patient work provide a more complete understanding of the processes being studied. With this in mind, despite the small sample size, the anterior insula and IFG found in Experiment 4.1 are often implicated in neuroimaging studies of updating (McGuire et al., 2014; Stöttinger et al., 2015; Donoso et al., 2014; Domenech & Koechlin, 2015; Summerfield & Koechlin, 2007). This suggests that although the sample of patients was small, the regions found in the study are plausible contributors to some of the observed deficits. Future patient studies could test the proposed roles of these regions more specifically by either testing larger patient samples, or by specifically targeting these regions more directly. A potential future study could be to compare updating performance between a patient group with selective damage to the anterior insula and another with selective damage to the IFG, and model the exploratory strategies used by these groups. The results from Experiment 4.1 predict that patients with IFG damage should be more likely to use simpler reinforcement learning type strategies to explore new mental models, but do so more consistently; conversely, patients with anterior insular damage may use more effective PROBE-like strategies to update, but be more prone to lapses of random responses. A more directed study such as this, along with converging neuroimaging results, could help identify the specific roles these brain regions play in the networks involved in updating.

In summary, the work from this thesis provides important insights into the mechanisms that underlie mental model building and updating. It provides information about the ways in which mental models can be studied and represented, while testing assumptions found in previous research. It also explores how these components are

impacted by brain damage, highlighting the importance of specific regions in the updating process. There is still much work to be done to fully uncover the mechanisms that guide our decisions. However, more refined techniques to describe and quantify decision processes will help bring us closer to understanding the fundamental processes that guide our behaviour.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE transactions on automatic control*, *19*(6), 716-723.

Attneave, F. (1959). *Applications of information theory to psychology: A summary of basic concepts, methods, and results.* New York, NY; Holt.

Bach, D. R., & Dolan, R. J. (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience*, *13*(8), 572-586. doi: 10.1038/nrn3289

Barlow, H. B. (1961). Possible principles underlying the transformations of sensory messages. In W. A. Rosenblith (Ed.), *Sensory Communication* (pp. 217-234). Cambridge, MA; MIT press.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Libear Mixed-Effects Models Using lme4. *Journal of Statistics Software*, *67*(1), 1-48.

Bates, E., Wilson, S. M., Saygin, A. P., Dick, F., Sereno, M. I., Knight, R. T., & Dronkers, N. F. (2003). Voxel-based lesion–symptom mapping. *Nature neuroscience*, *6*(5), 448-450.

Behrens, T.E.J., Woolrich, M. W., Walton, M. E., & Rushworth, M. F. S. (2007). Learning the value of information in an uncertain world. *Nature Neuroscience*, *10*(9), 1214-21. doi:10.1038/nn1954

Bestmann, S., Harrison, L. M., Blankenburg, F., Mars, R. B., Haggard, P., Friston, K. J., & Rothwell, J. C. (2008). Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. *Current Biology*, *18*(10), 775–80. doi:10.1016/j.cub.2008.04.051

Bowers, J. S., & Davis, C. J. (2012a). Bayesian just-so stories in psychology and neuroscience. *Psychological Bulletin*, *138*(3), 389–414. doi:10.1037/a0026450

Bowers, J. S., & Davis, C. J. (2012b). Is that what Bayesians believe? reply to Griffiths, Chater, Norris, and Pouget (2012). *Psychological Bulletin*, *138*(3), 423–6. doi:10.1037/a0027750

Brownell, H. H., Michel, D., Powelson, J., & Gardner, H. (1983). Surprise but not coherence: sensitivity to verbal humor in right-hemisphere patients. *Brain and Language*, *18*(1), 20–27. doi:10.1016/0093-934X(83)90002-0

Chater, N., Tenenbaum, J. B., & Yuille, A. (2006). Probabilistic models of cognition: conceptual foundations. *Trends in Cognitive Sciences*, *10*(7), 287–91. doi:10.1016/j.tics.2006.05.007

Collins, A., & Koechlin, E. (2012). Reasoning, learning, and creativity: frontal lobe function and human decision-making. *PLoS Biology*, *10*(3), e1001293. doi:10.1371/journal.pbio.1001293

Corbetta, M., Kincade, M. J., Lewis, C., Snyder, A. Z., & Sapir, A. (2005). Neural basis and recovery of spatial attention deficits in spatial neglect. *Nature Neuroscience*, *8*(11), 1603–1610. doi:10.1038/nn1574

Costello, F., & Watts, P. (2014). Surprisingly Rational : Probability Theory Plus Noise Explains Biases in Judgment. *Psychological Review, 121*(3), 463–480. doi:10.1037/a0037010

Cox, D. & Snell, E. (Eds.). (1989) *Analysis of binary data* (2nd ed.). New York, NY; Chapman and Hall.

Danckert, J., Stöttinger, E., Quehl, N., & Anderson, B. (2012). Right hemisphere brain damage impairs strategy updating. *Cerebral Cortex*, *22*(12), 2745–2760. doi:10.1093/cercor/bhr351

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876-879.

Dawson, N. V, & Abkes, H. A. L. R. (1987). Systematic Errors in Medical Decision Making. *Journal of General Internal Medicine*, *2*(3), 183–187. doi:10.1007/BF02596149

De Gardelle, V., & Summerfield, C. (2011). Robust averaging during perceptual judgment. *Proceedings of the National Academy of Sciences*, *108*(32), 13341–13346. doi:10.1073/pnas.1104517108

Decety, J., & Lamm, C. (2007). The role of the right temporoparietal junction in social interaction: how low-level computational processes contribute to meta-cognition. *The Neuroscientist*, *13*(6), 580-593.

Domenech, P., & Koechlin, E. (2015). Executive control and decision-making in the prefrontal cortex. *Current opinion in behavioral sciences*, *1*, 101-106.

Donoso, M., Collins, A. G. E., & Koechlin, E. (2014). Human cognition. Foundations of human reasoning in the prefrontal cortex. *Science*, *344*(6191), 1481–6. doi:10.1126/science.1252254

Doya, K., Ishii, S., Pouget, A., & Rao, P.N. (Eds.). (2007). *Bayesian Brain: Probabilistic approaches to neural coding*. Cambridge, MA; MIT press.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. doi:10.1038/415429a

Estes, W. K. (1950). Towards a statistical theory of learning. *Psychological Review*, *57*(2), 94–107. doi:10.1037/h0058559

Filipowicz, A., Anderson, B., & Danckert, J. (2016). Adapting to change: The role of the right hemisphere in mental model building and updating. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *70*(3), 201-218.

Filipowicz, A., Anderson, B., & Danckert, J. (2014). Learning what from where: Effects of spatial regularity on nonspatial sequence learning and updating. *The Quarterly Journal of Experimental Psychology*, *67*(7), 1447–56.

Fischhoff, B., Slovic, P., & Lichtenstein, S. (1977). Knowing with certainty: The appropriateness of extreme confidence. *Journal of Experimental Psychology: Human perception and performance*, *3*(4), 552.

Fiser, J., Berkes, P., Orbán, G., & Lengyel, M. (2010). Statistically optimal perception and learning: from behavior to neural representations. *Trends in Cognitive Sciences*, *14*(3), 119–30. doi:10.1016/j.tics.2010.01.003

Fisk, J. E. (2002). Judgments under uncertainty: Representativeness or potential surprise?. *British Journal of Psychology*, *93*(4), 431-449.

Galton, F. (1889). *Natural inheritance*. London; Macmillan.

Geng, J. J., & Vossel, S. (2013). Re-evaluating the role of TPJ in attentional control: contextual updating?. *Neuroscience & Biobehavioral Reviews*, *37*(10), 2608-2620.

Gigerenzer, G., & Gaissmaier, W. (2011). Heuristic decision making. *Annual Review of Psychology*, *62*, 451–82. doi:10.1146/annurev-psych-120709-145346

Green, C. S., Benson, C., Kersten, D., & Schrater, P. (2010). Alterations in choice behavior by manipulations of world model. *Proceedings of the National Academy of Sciences*, *107*(37), 16401–16406. doi:10.1073/pnas.1001709107

Griffiths, T. L., Chater, N., Norris, D., & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): comment on Bowers and Davis (2012). *Psychological Bulletin*, *138*(3), 415–22. doi:10.1037/a0026884

Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: exploring representations and inductive biases. *Trends in Cognitive Sciences*, *14*(8), 357–64. doi:10.1016/j.tics.2010.05.004

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal predictions in everyday cognition. *Psychological Science*, *17*(9), 767–773. doi:10.1111/j.1467-9280.2006.01780.x

Harrison, L. M., Duggins, A., & Friston, K. J. (2006). Encoding uncertainty in the hippocampus. *Neural Networks*, *19*(5), 535–546. doi:10.1016/j.neunet.2005.11.002.Encoding

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review*, *7*(2), 185–207. doi:10.3758/BF03212979

Hilbert, M. (2012). Toward a synthesis of cognitive biases: how noisy information processing can bias human decision making. *Psychological Bulletin*, *138*(2), 211–37. doi:10.1037/a0025940

Hock, H. S., Bukowski, L., Nichols, D. F., Huisman, A., & Rivera, M. (2005). Dynamical vs. judgmental comparison: hysteresis effects in motion perception. *Spatial Vision*, *18*(3), 317–335. doi:10.1163/1568568054089393

Hogarth, R. M., & Einhorn, H. J. (1992). Order effects in belief updating: The belief-adjustment model. *Cognitive Psychology*, *24*(1), 1-55. doi:10.1016/0010-0285(92)90002-J

Hollard, G., Massoni, S., & Vergnaud, J. C. (2016). In search of good probability assessors: an experimental comparison of elicitation rules for confidence judgments. *Theory and Decision*, *80*(3), 363-387.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*, *15*(1), 1-25.

Kimberg, D. Y., Coslett, H. B., & Schwartz, M. F. (2007). Power in voxel-based lesion-symptom mapping. *Journal of Cognitive Neuroscience*, *19*(7), 1067-1080.

Johnson-Laird, P. N. (2013). Mental models and cognitive change. *Journal of Cognitive Psychology*, *25*(2), 131-138. doi:10.1080/20445911.2012.759935

Johnson-Laird, P. N. (2004). The history of mental models. In K. Manktelow & M. C. Chung (Eds.), *Psychology of reasoning: Theoretical and historical perspectives* (pp. 179–212). Hove, Sussex: Psychology Press.

Jones, M., & Love, B. C. (2011). Bayesian Fundamentalism or Enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, *34*(4), 169–188. doi:10.1017/S0140525X10003134

Jueptner, M., Stephan, K. M., Frith, C. D., Brooks, D. J., Frackowiak, R. S., & Passingham, R. E. (1997). Anatomy of motor learning. I. Frontal cortex and attention to action. *Journal of Neurophysiology*, *77*(3), 1313-1324.

Jueptner, M., Frith, C. D., Brooks, D. J., Frackowiak, R. S. J., & Passingham, R. E. (1997). Anatomy of motor learning. II. Subcortical structures and learning by trial and error. *Journal of Neurophysiology*, *77*(3), 1325-1337.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY; Farrar, Straus and Giroux.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2014). The Goldilocks effect in infant auditory attention. *Child development*, *85*(5), 1795-1804.

Kidd, C., Piantadosi, S. T., & Aslin, R. N. (2012). The Goldilocks effect: Human infants allocate attention to visual sequences that are neither too simple nor too complex. *PloS one*, *7*(5), e36399.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. doi:10.1016/j.tins.2004.10.007

Koechlin, E., & Summerfield, C. (2007). An information theoretical approach to prefrontal executive function. *Trends in cognitive sciences*, *11*(6), 229-235.

Koehler, D. J., & James, G. (2009). Probability matching in choice under uncertainty: Intuition versus deliberation. *Cognition*, *113*(1), 123-127.

Kolling, N., Behrens, T. E., Mars, R. B., & Rushworth, M. F. (2012). Neural mechanisms of foraging. *Science*, *336*(6077), 95-98.

Körding, K. P., & Wolpert, D. M. (2006). Bayesian decision theory in sensorimotor control. *Trends in Cognitive Sciences*, *10*(7), 319–326. doi:10.1016/j.tics.2006.05.003

Lee, N. Y. L., & Johnson-Laird, P. N. (2012). Strategic changes in problem solving. *Journal of Cognitive Psychology*, *25*(2), 1–9. doi:10.1080/20445911.2012.719021

Lewandowsky, S., Griffiths, T. L., & Kalish, M. L. (2009). The Wisdom of Individuals: Exploring People's Knowledge About Everyday Events Using Iterated Learning. *Cognitive Science*, *33*(6), 969–98. doi:10.1111/j.1551-6709.2009.01045.x

Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., & Bestmann, S. (2008). Trial-by-trial fluctuations in the event-related electroencephalogram reflect dynamic changes in the degree of surprise. *The Journal of Neuroscience*, *28*(47), 12539–45. doi:10.1523/JNEUROSCI.2925-08.2008

Mayr, U. (1996). Spatial attention and implicit sequence learning: Evidence for independent learning of spatial and nonspatial sequences. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(2), 350–364. doi:10.1037//0278-7393.22.2.350

McGuire, J. T., & Kable, J. W. (2015). Medial prefrontal cortical activity reflects dynamic re-evaluation during voluntary persistence. *Nature neuroscience*, *18*(5), 760-766.

McGuire, J. T., Nassar, M. R., Gold, J. I., & Kable, J. W. (2014). Functionally Dissociable Influences on Learning Rate in a Dynamic Environment. *Neuron*, *84*(4), 870–881. doi:10.1016/j.neuron.2014.10.013

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, *214*(5), 655-667. doi:10.1007/s00429-010-0262-0

Miller, D. J., Spengler, E. S., & Spengler, P. M. (2015). A meta-analysis of confidence and judgment accuracy in clinical decision making. Journal of Counseling Psychology, 62, 553–567. http://dx.doi.org/10.1037/ cou0000105 Mohr, J. J. (2016)

Nasreddine, Z. S., Phillips, N. A., Bédirian, V., Charbonneau, S., Whitehead, V., Collin, I., ... & Chertkow, H. (2005). The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *Journal of the American Geriatrics Society*, *53*(4), 695-699.

Nassar, M. R., Bruckner, R., Gold, J. I., Li, S. C., Heekeren, H. R., & Eppinger, B. (2016). Age differences in learning emerge from an insufficient representation of uncertainty in older adults. *Nature Communications*, *7*(11609), 1-13.

Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature Neuroscience*, *15*(7), 1040–6. doi:10.1038/nn.3130

Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An approximately Bayesian delta-rule model explains the dynamics of belief updating in a changing environment. *The Journal of Neuroscience*, *30*(37), 12366–12378. doi:10.1523/JNEUROSCI.0822-10.2010

Neal, R. M. (2003). Slice sampling. *Annals of statistics*, *31*(3), 705-741.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, *2*(2), 175–220. doi:10.1037//1089-2680.2.2.175

Nissen, M. J., & Bullemer, P. (1987). Attentional requirements of learning: Evidence from performance measures. *Cognitive Psychology*, *19*(1), 1-32. doi:10.1016/0010-0285(87)90002-8

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, *304*(5669), 452-454.

O'Reilly, J. X., Schüffelgen, U., Cuell, S. F., Behrens, T. E., Mars, R. B., & Rushworth, M. F. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, *110*(38), E3660-E3669. doi: 10.1073/pnas.1305373110

O'Reilly, J. X. (2013). Making predictions in a changing world—inference, uncertainty, and learning. *Frontiers in Neuroscience*, *7*(105), 1-10.

Orbán, G., Fiser, J., Aslin, R. N., & Lengyel, M. (2008). Bayesian learning of visual chunks by human observers. *Proceedings of the National Academy of Sciences*, *105*(7), 2745–50. doi:10.1073/pnas.0708424105

Otto, A. R., Taylor, E. G., & Markman, A. B. (2011). There are at least two kinds of probability matching: evidence from a secondary task. *Cognition*, *118*(2), 274–9. doi:10.1016/j.cognition.2010.11.009

Palminteri, S., Justo, D., Jauffret, C., Pavlicek, B., Dauta, A., Delmaire, C., Delmaire, C., Czernecki, V., Carine, K., Capelle, L., Durr, A., Pessiglione, M. (2012). Critical Roles for Anterior Insula and Dorsal Striatum in Punishment-Based Avoidance Learning. *Neuron*, *76*(5), 998–1009. doi:10.1016/j.neuron.2012.10.017

Park, J., & Sloman, S. A. (2014). Causal explanation in the face of contradiction. *Memory & Cognition*, *42*(5), 806–20. doi:10.3758/s13421-013-0389-3

Patrick, J., & Ahmed, A. (2014). Facilitating representation change in insight problems through training. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *40*(2), 532–43. doi:10.1037/a0034304

Peirce, J. W. (2009). Generating Stimuli for Neuroscience Using PsychoPy. *Frontiers in Neuroinformatics*, *2*, 10. doi:10.3389/neuro.11.010.2008

Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *The Journal of Neuroscience*, *28*(11), 2745–2752. doi:10.1523/JNEUROSCI.4286-07.2008

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/

Ritter, F. E., & Schooler, L. J. (2001). The learning curve. In N. J. Smelser & P. B. Baltes (Eds.), *International Encyclopedia of the Social and Behavioural Sciences*, (pp. 8602–8605). Amsterdam: Pergamon.

Robertson, E. M., Tormos, J. M., Maeda, F., & Pascual-Leone, A. (2001). The Role of the Dorsolateral Prefrontal Cortex during Sequence Learning is Specific for Spatial Information. *Cerebral Cortex*, *11*(7), 628–635. doi:10.1093/cercor/11.7.628

Rodriguez, A. & Laio, A. (2014). Clustering by fast search and find of density peaks. *Science, 344*(6191), 1492-1496.

Rorden, C., Bonilha, L., Fridriksson, J., Bender, B., & Karnath, H. O. (2012). Age-specific CT and MRI templates for spatial normalization. *Neuroimage*, *61*(4), 957-965.

Rorden, C., Karnath, H. O., & Bonilha, L. (2007). Improving lesion-symptom mapping. *Journal of Cognitive Neuroscience*, *19*(7), 1081-1088.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science*, *274*(5294), 1926-1928. doi:10.1126/science.274.5294.1926

Schwarz, G. (1978). Estimating the dimension of a model. *The annals of statistics*, *6*(2), 461-464.

Sepahvand, N. M., Stöttinger, E., Danckert, J., & Anderson, B. (2014). Sequential decisions: a computational comparison of observational and reinforcement accounts. *PloS one*, *9*(4), e94308.

Shafer, G. (1990). Perspectives on the Theory and Practice of Belief Functions. *International Journal of Approximate Reasoning*, *4*(5), 323–362. doi:10.1016/0888-613X(90)90012-Q

Shaqiri, A., & Anderson, B. (2013). Priming and statistical learning in right brain damaged patients. *Neuropsychologia*, *51*(13), 2526–2533. doi:10.1016/j.neuropsychologia.2013.09.024

Shanks, D. R., Tunney, R. J., & Mccarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making, 15*(3), 233–250. doi:10.1002/bdm.413

Shannon, C. E. (1948). A Mathematical Theory of Communication. *The Bell System Technical Journal*, *27*, 379–423.

Stöttinger, E., Filipowicz, A., Valadao, D., Culham, J. C., Goodale, M. A., Anderson, B., & Danckert, J. (2015). A cortical network that marks the moment when conscious representations are updated. *Neuropsychologia*, *79*, 113-122.

Stöttinger, E., Filipowicz, A., Danckert, J., & Anderson, B. (2014). The effects of prior learned strategies on updating an opponent's strategy in the rock, paper, scissors game.*Cognitive Science*, *38*(7), 1482–92. doi:10.1111/cogs.12115

Stöttinger, E., Filipowicz, A., Marandi, E., Quehl, N., Danckert, J., & Anderson, B. (2014). Statistical and perceptual updating: Correlated impairments in right brain injury. *Experimental Brain Research*, *232*(6), 1971–1987. doi:10.1007/s00221-014-3887-z

Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., & Friston, K. J. (2005). Information theory, novelty and hippocampal responses: unpredicted or unpredictable?. *Neural Networks*, *18*(3), 225–30. doi:10.1016/j.neunet.2004.12.004

Summerfield, C., & Tsetsos, K. (2015). Do humans make good decisions?. *Trends in Cognitive Sciences*, *19*(1), 27–34. doi:10.1016/j.tics.2014.11.005

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1, No. 1). Cambridge, MA: MIT press.

Teigen, K. H., & Keren, G. (2003). Surprises : low probabilities or high contrasts?. *Cognition*, *87*(2), 55–71. doi:10.1016/S0

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: statistics, structure, and abstraction. *Science*, *331*(6022), 1279–85. doi:10.1126/science.1192788

Toni, I., Krams, M., Turner, R., & Passingham, R. E. (1998). The Time Course of Changes during Motor Sequence Learning : A Whole-Brain fMRI Study. *NeuroImage*, *8*(1), 50–61.

Turk-Browne, N. B., Jungé, J., & Scholl, B. J. (2005). The automaticity of visual statistical learning. *Journal of Experimental Psychology: General*, *134*(4), 552–64. doi:10.1037/0096-3445.134.4.552

Tversky, A., & Koehler, D. J. (1994). Support theory: A nonextensional representation of subjective probability. *Psychological Review*, *101*(4), 547–567. doi:10.1037//0033-295X.101.4.547

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. *Science*, *211*(4481), 453–458. doi:10.1126/science.7455683

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, *15*(1), 273-289.

Vocat, R., Saj, A., & Vuilleumier, P. (2013). The riddle of anosognosia: does unawareness of hemiplegia involve a failure to update beliefs?. *Cortex*, *49*(7), 1771-1781.

Vulkan, N. (2000). An economist's perspective on probability matching. *Journal of Economic Surveys*, *14*(1), 101–118. doi:10.1111/1467-6419.00106

Wagenmakers, E. J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic bulletin & review*, *11*(1), 192-196.

Wei, X. X., & Stocker, A. A. (2015). A Bayesian observer model constrained by efficient coding can explain'anti-Bayesian'percepts. *Nature Neuroscience*, *18*(10), 1509-1517.

Wilson, R. C., Nassar, M. R., & Gold, J. I. (2010). Bayesian online learning of the hazard rate in change-point problems. *Neural computation*, *22*(9), 2452-2476. doi: 10.1162/NECO_a_00007

Wolford, G., Miller, M. B., & Gazzaniga, M. (2000). The left hemisphere's role in hypothesis formation. *The Journal of Neuroscience*, *20*(6), 1-4. doi:10.1026/0932-4089.51.1.1

Appendix 1

Model descriptions

PROBE model description

The PROBE model was created as a biologically plausible, online algorithm approximating a Dirichlet process mixture. It assumes that on any given trial *t*, mappings between a given stimulus s_t, action a_t, and outcome o_t depend on some hidden state, or 'rule' denoted R_t^*. An environment is assumed to be uncertain, changing, and open ended – the occurrence of hidden rules, therefore, are independent of stimuli and actions, and are presumed by the model to be independent of other hidden rules (*i.e.*, only one hidden rules can apply at any given moment). The goal of the model is to estimate the correct hidden rule by estimating a rule's reliability, and remembering any rule that has been deemed reliable in the past.

At any given point, the PROBE model attempts to identify a rule *R* that corresponds to the current hidden rule. A rule corresponds to a policy of stimulus action outcome expectations that attempt to predict the best action *a* for any stimulus *s*. A rule being used to drive actions is called an *actor*. Rules contains two internal mappings: a selective mapping $Q(s, a)$ and a predictive mapping $\gamma(o, s, a)$.

Selective mappings for a current rule R_c, $Q_c(s, a) = E(r[o]|s, a, R^* = R_c)$, encode the expected reward value $r[o]$ of an outcome *o* given that an action *a* has been made to a stimulus *s*. This mapping determines the policy that guides actions in response to a particular stimulus. This policy orients behaviour towards the most rewarding actions and are learned using reinforcement learning rules outlined below.

An actor's selective mapping $Q_{actor}(s,a)$ computes the behavioural policy $P(a|s_t)$ – *i.e.*, the probability to select an action *a* in response to a stimulus *s* on trial *t* – according to an ε-softmax with an inverse temperature β:

$$P(a \mid s_t) = (1-\varepsilon)\frac{\exp\left(\beta Q_{actor}(s_t,a_t)\right)}{\sum_{a=1\ldots n_a}\exp\left(\beta Q_{actor}(s_t,a_t)\right)} + \frac{\varepsilon}{n_a} \qquad \text{(A.1)}$$

where *n* is the total number of possible actions. After each outcome o_t, the actor's selective mapping $Q_{actor}(s,a)$ is updated based on the outcome's reward $r[o_t]$ through a Rescorla & Wagner reinforcement learning rule (Rescorla & Wagner, 1972):

$$Q^{t+1}_{actor}(s_t,a_t) = \alpha r[o_t] + (1-\alpha)Q^{t}_{actor}(s_t,a_t) \qquad \text{(A.2)}$$

with α representing the learning rate.

Predictive mappings, $\gamma_c(o,s,a) = P(o|s,a,R^* = R_c)$, encode the likelihood that an outcome *o* will be obtained given an action *a* to a stimulus *s* given the rule that is currently being used R_c. Predictive mappings estimate the probability of specific action outcomes independently of outcome valences or reward values. These mappings prove crucial in order to evaluate the uncertainty of external contingencies, and to estimate the volatility associated with external states and action outcomes.

The original model (Collins & Koechlin, 2012) also includes a contextual mapping, which tracks contextual information that could provide information about rule switches. As with Donoso and colleagues (2014), the task used in Experiment 4.1 did not include contextual information; therefore contextual mappings were not considered in the current version of the PROBE model.

The model assumes that a person monitors the reliability of a maximum of *N* potential rules. This number of rules could be considered a kind of "buffer", and is included as a free parameter in model estimation. The notation $\{1,\ldots,N_t\}$ denotes the number of rules being monitored at any given point throughout the task, with the number of monitored rules $R_i(N_t \leq N)$. The absolute reliability $\lambda_i\ (t)$ of rules $R_i \in \{1,\ldots,N_t\}$ is computed as the posterior probability that the hidden rule R_t^* on any trial t matches R_i given all observations, including those seen on preceding trials:

$$\lambda_i(t) = P(R_t^* = R_i \mid s_t, o_{t-1}, past) \qquad \text{(A.3)}$$

where "past" refers to all other observable events occurring in the preceding trials. To calculate changes in reliability, $R_t^* = R_0$ represents an event in which $R_t^* \notin \{1, \ldots, N_t\}$, $\gamma(.,.,.)$ indicates the likelihood of an action given the current event, and $\lambda_i(t)$ represents the posterior probability that this event occurs given all observations. With this notation, the updating rule for absolute reliabilities $\lambda_i(t), i \in \{0,1, \ldots, N_t\}$ is calculated using Bayesian calculus:

$$\lambda_i(t+1) = \frac{(1-\tau)\mu_i(t) + \tau \sum_{j \in \{0,1,\ldots,N_t\}} \mu_j(t)}{Z_t^{\lambda}} \quad \text{with } \mu_i(t) = \frac{\gamma_i(o_t, s_t, a_t)\lambda_i(t)}{Z_t^{\mu}} \tag{A.4}$$

Both Z_t^{λ} and Z_t^{μ} are normalization factors, and τ represents the perceived volatility, that is, the probability that a hidden state can change between an outcome o_t and the next stimulus s_{t+1}.

Based on the equations above, a crucial term required to infer a rule's absolute reliability is $\gamma_0(.,.,.)$ and is important in order to learn the mappings of monitored strategies. It can formally be written as follows:

$$\gamma_0(o,s,a) = P(o \mid s,a,R_t^* \notin \{1,\ldots,N_t\} \tag{A.5}$$

It is important to note that regardless of the stimuli and actions, all of the possible outcomes that could be produced by a current set of monitored rules are equally likely to occur when the current hidden state R_t^* is unknown – *i.e.*, when $R_t^* \notin \{1, \ldots, N_t\}$. Therefore, $\gamma_0(o, s, a)$ is represented as a constant γ_0 calculated as the equal probability of an action outcome produced by the entire set of monitored rules. It is also worth noting that the value $\lambda_0(t)$ does not need to be explicitly computed given that $\lambda_0(t) + \sum_{i \in \{1,\ldots,N_t\}} \lambda_i\,(t) = 1$.

A key feature of the PROBE model is to infer the absolute reliability of a limited set of rules. Each monitored rule $i \in \{1, \ldots, N_t\}$ is therefore seen as either being reliable (more likely to be matching the current hidden rule, or $\lambda_i(t) > 1 - \lambda_i(t)$), or unreliable (a situation in which $\lambda_i(t) < 1 - \lambda_i(t)$). If there is a reliable monitored strategy, all other monitored rules are by definition *unreliable* given that $\sum_{i \in \{1,\ldots,N_t\}} \lambda_t\,(t) < 1$. A reliable rule is identified as an *actor*, meaning that this rule becomes the unique action selector (and learner of external contingencies) at that particular moment. When the model settles on a reliable actor, it is said to be in *exploitation* mode. Conversely, when all monitored rules become unreliable, the model switches to an *exploration* mode. During exploration, a new, provisional actor R_p is created by marginalizing over the range of rules stored in long term memory. The provisional actor's mappings M_p (which include selective and predictive mappings Q_p and γ_p respectively) are computed as follows:

$$M_p = \eta U + (1-\eta)\sum_m M_m \qquad \text{(A.6)}$$

where the index *m* runs over the entire set of rules stored in long-term memory, and M_m denotes each rule *m*'s internal mappings. The *U* in this equation is the uniform density over actions meant to represent recollection entropy, and η a recollection scaling parameter that determines how much the model relies on mappings from long-term memory.

Once created, the provisional actor R_p is endowed with a prior reliability λ_{prior} that is computed as minimizing the reliability information from monitored strategies. As

demonstrated by Collins and Koechlin (2012), the reliability λ_{prior} ranges between $1/(N_t + 1)$ and 1/3. Thus, the provisional actor is initially unreliable and *probed* to try to learn external contingencies. As this provisional actor learns, it may become reliable, while other monitored rules become unreliable. In these circumstances, the provisional actor is *confirmed*, its mappings are stored in long-term memory, the model exits the exploration state, and enters an exploitation state with the provisional actor as its new reliable rule. In cases where the *N* buffer capacity is reached, the least recent rule is discarded from the buffer, and replaced with the previously reliable rule.

It is also possible for a monitored rule in the buffer, other than the provisional actor, to become reliable. In these circumstances, the provisional actor is *rejected*, the exploration period terminates, and the reliable monitored rule becomes the new actor.

Overall, the PROBE model contains 7 free parameters that can vary between subjects:

Buffer capacity (N). An integer value that corresponds to the number of rules, including the actor, that can concurrently be monitored. It is worth noting that the probe actors are added to this buffer during periods of exploration.

Perceived volatility (τ). A parameter ranging between 0 and 1 that represents the likelihood that a hidden state will change at any given point throughout the task.

Recollection entropy (η). A continuous parameter ranging between 0 and 1 that determines how much a probe actor's mappings rely on mappings stored in long-term memory (with values closer to 1 indicated less influence from long-term memory).

Prior reliability bias (θ). Participants may be more or less inclined to confirm rather than reject newly created provisional actors. As such, the prior reliability λ_{prior}

could be biased. To account for this possibility, a biased prior reliability is calculated as $\lambda_{biased\ prior} = \theta \times \frac{1}{2} + (1-\theta)\lambda_{prior}$. The higher the value of $\lambda_{biased\ prior}$, the more participants are biased towards confirming rather than rejecting newly created actors. The θ parameter scales the deviation of a provisional actor's prior reliability from an uninformative prior to the reliability threshold of 0.5 required to confirm an actor.

Learning rate (α). Standard reinforcement learning rate parameter that can range between 0 and 1, and determines the weight given to an outcome.

Inverse temperature (β). Standard continuous reinforcement learning parameter that scales the 'greediness' of an action selection. Higher values of β indicate higher biases towards actions that maximize reward output.

Noise (ε). Standard continuous reinforcement learning parameter that can vary between 0 and 1, and accounts for lapses in subject responses. The larger this parameter, the more lapses unaccounted for by the model.

Reinforcement Learning model description

The reinforcement learning model (RL) fit to participant responses in Experiment 4.1 uses past outcomes to build a selective mapping $Q(s,a)$ to represent stimulus-response associations. These mappings are built using the same reinforcement learning rules used in equations A.1 and A.2. As such this RL model contains three free parameters: α (learning rate), β (temperature), and ε (noise).