Belief updating and information seeking in decision making under uncertainty

Daniel Bennett

Student ID: 518253
ORCID: 0000-0001-6608-9026

Submitted in total fulfilment of the requirements of the degree of Doctor of Philosophy
March, 2017

Melbourne School of Psychological Sciences
The University of Melbourne
Abstract

Adaptive goal-directed behaviour depends on a well-calibrated internal model of the environment. In particular, learning and decision making is most efficient when there is a good match between an agent’s model of the environment and true environmental contingencies. Accordingly, recent theories in cognitive neuroscience have proposed that a primary computational goal of the central nervous system in humans and other animals is the construction and maintenance of these internal models (also termed beliefs). The present thesis aimed to investigate the implications of this belief-centric perspective for the study of decision making under uncertainty, using computational modelling of behaviour in concert with analysis of human electroencephalography (EEG) data. Across one behavioural and three EEG studies, this thesis addressed two related questions concerning participants’ updating of beliefs. Study 1 and Study 2 sought to characterise the neural and cognitive processes by which participants incorporated new information into their beliefs (belief updating). Study 3 and Study 4 sought to investigate the nature of participants’ preferences for acquiring such information, and the neural substrates of this preference (information seeking).

In Study 1, 18 healthy young adult participants completed a perceptual learning task—in which participants learned to identify the target contrast of a greyscale checkerboard stimulus using monetary feedback—while EEG was recorded. Formal comparison of computational cognitive models revealed that participants’ behaviour in this task was better explained by a model implementing Bayesian belief updating than by a model implementing a simpler win-stay lose-shift (WSLS) heuristic. Belief variables computed from this Bayesian model were then used as predictors in a single-trial regression analysis of event-related potential (ERP) data. This analysis revealed that the amplitude of the frontocentral P3 component of the ERP was positively associated with belief update size, and that the amplitude of the stimulus-preceding negativity component was negatively associated with belief uncertainty. These results provided evidence that belief update size and belief uncertainty had distinct neural signatures that could be tracked in single trials in specific ERP components. The results further suggested that the cognitive mechanisms underlying belief updating in this task could be described well within a Bayesian framework.

Study 2 investigated the effect of task motivation on participants’ use of Bayesian versus heuristic task strategies, and upon the neural substrates of belief updating. Using a modified variant of the perceptual learning task from Study 1, Study 2 presented feedback to participants in the form of either monetary reward, or as affectively neutral instructional
directives. Using model-based clustering based upon formal comparison of computational cognitive models, Study 2 identified two distinct participant subgroups. The first subgroup used Bayesian inference in the monetary condition, but switched to a heuristic strategy in the instructive feedback condition; by contrast, the second subgroup always used the Bayesian inference strategy, regardless of feedback condition. It was found that only the strategy-switching subgroup showed worse performance for instructive than for monetary feedback, whereas the Bayesian subgroup did not. This pattern of performance was reflected by similar differences between subgroups in neural encoding of feedback in two components of the event-related potential: the P3, and the late positive potential. These findings suggested that selection of Bayesian versus heuristic strategies in perceptual learning may depend critically on participants’ motivational state, and that individual differences in strategy-switching may underlie group-level differences in neural encoding of feedback.

Study 3 investigated information seeking behaviour in two experiments, each with 40 healthy participants. In these experiments, participants completed a novel information-seeking task in which they could choose to pay a monetary cost to receive advance information about the outcome of a monetary lottery. Importantly, acquiring information did not alter lottery outcome probabilities. In the first experiment, it was found that participants were willing to incur considerable monetary costs to acquire payoff-irrelevant information about the lottery outcome. This behaviour was well explained by a computational cognitive model in which information preference resulted from aversion to temporally prolonged uncertainty. In the second experiment, it was found that preference for information was modulated by the rate of information delivery. These results strongly suggest that humans assign an intrinsic value to information, in a manner inconsistent with normative accounts of decision making under uncertainty.

Finally, Study 4 investigated the neural substrates of the intrinsic valuation of information, using EEG data were recorded from 22 participants performing the information seeking task developed in Study 3. Behavioural results replicated the findings of Study 3. Analysis of ERPs elicited by informative cues revealed that the feedback-related negativity, an ERP component linked to reward processing, independently encoded both a reward prediction error and an information prediction error. These findings are consistent with the hypothesis that information seeking results from processing of information within neural reward circuits, and provide further evidence for an intrinsic valuation of information.
Overall, these findings demonstrate that a belief-centric research perspective has substantial explanatory power for behavioural and neural data in human decision making under uncertainty.
Declaration

This is to certify that

(1) This thesis comprises only my original work toward the degree of Doctor of Philosophy, except where otherwise indicated in the preface,

(2) Due acknowledgement has been made in the text to all other material used,

(3) This thesis is fewer than 100 000 words in length, exclusive of tables, figures, references, footnotes, and appendices.

…………………………

Daniel Bennett
Preface

For each study reported herein, experimental data was collected by the PhD candidate with assistance from members of the Decision Neuroscience Laboratory at The University of Melbourne. Each empirical manuscript was drafted entirely by the PhD candidate, and was then revised and edited with the assistance of the manuscript’s named co-authors. All co-authors have agreed to the use of these manuscripts in this thesis, and have provided signed copies of the co-author authorisation form.
Acknowledgements

I thank all members past and present of the Decision Neuroscience Laboratory for their friendship and support. I would particularly like to thank Maja Brydevall, Karen Sasmita, and Hayley Warren for their assistance with data collection for the experiments reported in this thesis. I would also like to express my appreciation for the support given to me by current and former staff and students of the School of Psychological Sciences, especially Olivia Carter, Simon Lilburn, Dan Little, and David Sewell. I thank my supervisors for their sound guidance and counsel throughout my studies; I am confident that I will continue to rely on these in the future. I would also like to express my deepest appreciation and admiration to Dale Laverty and Scott Hammerton, of Newcastle, NSW.

Finally—and most importantly—I thank my partner Clair, without whom the last four years would have been neither possible nor worthwhile.
Publications, conference abstracts, and awards during candidature

Peer-reviewed journal articles


Peer-reviewed conference abstracts

Bennett, D., Sasmita, K., Murawski, C., & Bode, S. (2016) Bayes if it pays: Switches between Bayesian and heuristic learning strategies are encoded in the event-related potential. *Australasian Experimental Psychology Conference*. University of Melbourne: Melbourne, Australia. *(Oral presentation)*


**Grants and awards**

2016

**Special Career Development Support Scheme** Melbourne School of Psychological Sciences, The University of Melbourne

**Runner-up, Mendelsohn Student Award** The University of Melbourne

2015

**Postgraduate Student Presentation Award** Melbourne School of Psychological Sciences, The University of Melbourne

2014

**Melbourne Abroad Travelling Scholarship** The University of Melbourne

2013

**Australian Postgraduate Award** Australian Government, Department of Education and Training
# Table of Contents

Abstract ii  
Declaration v  
Preface vii  
Acknowledgements ix  
Publications, conference abstracts, and awards  
  Peer-reviewed journal articles xi  
  Conference abstracts xi  
  Grants and awards xii  
Table of Contents xiii  
List of Figures 1  
Abbreviations 3  
Thesis Presentation 4  
1. Foundations of the Present Inquiry 6  
2. Review of Belief Updating Literature 10  
  2.1. Bayesian belief updating 12  
    2.1.1. Bayesian belief updating in cognitive psychology 13  
    2.1.2. Bayesian belief updating in economics 16  
    2.1.3. Bayesian belief updating in cognitive neuroscience 17  
      2.1.3.1. fMRI studies of Bayesian belief updating 18  
      2.1.3.2. EEG studies of Bayesian belief updating 20  
  2.2. Non-Bayesian belief updating 22  
    2.2.1. Conservatism 23  
    2.2.2. Order effects 25  
    2.2.3. Asymmetric belief updating 27  
    2.2.4. Confirmation bias 30  
    2.2.5. Heuristics 31  
  2.3. Rationale for research agenda 34  
3. Review of Information Seeking Literature 38  
  3.1. Overview and definitions 38  
  3.2. Theories of information valuation 41  
    3.2.1. Theories positing an instrumental value of information 42
3.2.2. Theories positing an intrinsic value of information

3.3. *Empirical studies of information seeking and information valuation*

3.3.1. Instrumental information

3.3.2. Non-instrumental information

3.4. *Rationale for research agenda*

4. Rationale for Methodological Approach

4.1. *Electroencephalography and event-related potentials*

4.2. *General rationale for behavioural tasks*

4.3. *Overview of behavioural tasks and computational models*

4.3.1. Study 1: Single-trial event-related potential correlates of belief updating

4.3.2. Study 2: Electrophysiological indices reflect switches between Bayesian and heuristic strategies in perceptual learning

4.3.3. Study 3: Intrinsic valuation of information in decision making under uncertainty

4.3.4. Study 4: The feedback-related negativity encodes an information prediction error in decisions to seek non-instrumental information

4.4. *Summary*

5. Study 1: Single-trial event-related potential correlates of belief updating

5.1. *Aims of Study 1*

5.2. *Manuscript*

5.3. *Summary*

6. Study 2: Electrophysiological indices reflect switches between Bayesian and heuristic strategies in perceptual learning

6.1. *Aims of Study 2*

6.2. *Manuscript*

6.3. *Summary*

7. Study 3: Intrinsic valuation of information in decision making under uncertainty

7.1. *Aims of Study 3*

7.2. *Manuscript*

7.3. *Summary*
8. Study 4: The feedback-related negativity encodes an information prediction error in decisions to seek non-instrumental information

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.1. Introduction</td>
<td>165</td>
</tr>
<tr>
<td>8.2. Method</td>
<td>167</td>
</tr>
<tr>
<td>8.2.1. Participants</td>
<td>167</td>
</tr>
<tr>
<td>8.2.2. Materials and procedure</td>
<td>167</td>
</tr>
<tr>
<td>8.2.3. Data analysis</td>
<td>168</td>
</tr>
<tr>
<td>8.3. Results</td>
<td>174</td>
</tr>
<tr>
<td>8.3.1. Behavioural results</td>
<td>174</td>
</tr>
<tr>
<td>8.3.2. ERP results</td>
<td>175</td>
</tr>
<tr>
<td>8.4. Discussion</td>
<td>182</td>
</tr>
</tbody>
</table>

9. General Discussion

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1. Review of key findings</td>
<td>187</td>
</tr>
<tr>
<td>9.2. Implications of research findings</td>
<td>191</td>
</tr>
<tr>
<td>9.2.1. Belief updating</td>
<td>191</td>
</tr>
<tr>
<td>9.2.2. Information seeking</td>
<td>195</td>
</tr>
<tr>
<td>9.3. Limitations</td>
<td>200</td>
</tr>
<tr>
<td>9.4. Future research directions</td>
<td>204</td>
</tr>
<tr>
<td>9.5. Conclusion</td>
<td>206</td>
</tr>
</tbody>
</table>

10. References | 208 |

Appendices

- Appendix A: The neural mechanisms of Bayesian belief updating | 235 |
- Appendix B: Systematic overestimation of reflection impulsivity in the information sampling task | 239 |
List of Figures

Chapter 4

Figure 4.1. Schematic of perceptual learning task.

Figure 4.2. Schematic overview of the computational cognitive models tested in Study 2.

Figure 4.3. Schematic of non-instrumental information seeking task.

Chapter 5

Figure 1. Task schematic.

Figure 2. Mean accuracy as a function of within-block trial number across participants.

Figure 3. Computational belief variables as a function of trial number.

Figure 4. P3 analysis.

Figure 5. Stimulus-preceding negativity analysis.

Chapter 6

Figure 1. Trial schematic.

Figure 2. Behavioural results.

Figure 3. Event-related potential grand average waveforms.

Figure 4. Event-related potential scalp maps.

Figure 5. Event-related potential component amplitudes as a function of feedback condition and behavioural subgroup.

Chapter 7

Figure 1. Task schematic.

Figure 2. Behavioural results for Experiment 1.

Figure 3. Model fit.

Figure 4. Individual-participant model fits.

Figure 5. Block-wise behavioural results for Experiment 1.

Figure 6. Behavioural results for Experiment 2.
Figure 7. Trial structure represented as a Markov Decision Process.

Figure 8. Latent mixture model of choice probability.

Chapter 8

Figure 8.1. Schematic overview of task events.

Figure 8.2. Mean proportion of informative stimulus choices as a function of information cost.

Figure 8.3. Grand average waveform for reward prediction errors at electrode Cz.

Figure 8.4. Scalp topography of mean difference between negative and positive reward prediction errors during the FRN analysis window.

Figure 8.5. Grand average waveform for information prediction errors at electrode Cz.

Figure 8.6. Scalp topography of mean difference between negative and positive information prediction errors during the FRN analysis window.

Figure 8.7. Grand average waveforms at electrode Cz for contrasts reported in Sections 8.3.2.1 to 8.3.2.5.
Abbreviations

**ACC:** Anterior cingulate cortex  
**BOLD:** Blood oxygen level-dependent  
**CB:** Consistent Bayesian  
**EEG:** Electroencephalography  
**ERP:** Event-related potential  
**EVPI:** Expected value of perfect information  
**EVSI:** Expected value of sample information  
**fMRI:** Functional magnetic resonance imaging  
**FRN:** Feedback-related negativity  
**ICS:** Incentive-compatible switching  
**IFG:** Inferior frontal gyrus  
**IPE:** Information prediction error  
**LPP:** Late positive potential  
**RPE:** Reward prediction error  
**SPN:** Stimulus-preceding negativity  
**UP:** Uncertainty penalty  
**WSLS:** Win-stay lose-shift
Thesis Presentation

This thesis is submitted as a series of three empirical articles (Chapters 5 to 7), one previously unpublished empirical chapter (Chapter 8), and five chapters comprising, respectively, prefatory material (Chapter 1), a review of previous literature (Chapters 2 and 3), a general methodological rationale (Chapter 4), and a general discussion of thesis results (Chapter 9). This presentation is in accordance with the guidelines for submission of a thesis with publication provided by the School of Graduate Research at The University of Melbourne, as well as advice communicated by the Chair of Examiners of the School of Psychological Sciences at The University of Melbourne.

Although they form separate parts of an overarching research investigation, as detailed in Chapter 1, each of the three empirical articles presented as Chapters 5 to 7 has been written to be read as a stand-alone investigation. As such, in-text references for these three chapters are followed by separate reference lists at the end of each chapter, and references in these chapters are not reproduced in the final reference list unless cited elsewhere in the thesis. At the beginning of each of these chapters, prefatory material has been included to contextualize the chapter with respect to other chapters, and to the overall research aims of this thesis. Chapters 5 and 7 have been accepted for publication in peer-reviewed academic journals, and are therefore presented in the format in which they appeared in these journals. Chapter 6 is an empirical chapter currently under review at Biological Psychology, and is therefore presented according to the formatting conventions of that journal.

Figure numbering conventions in this thesis are as follows: figures in Chapters 5, 6, and 7 are numbered sequentially according to the numbering conventions of the respective journals. In all other chapters, figures are named according to the convention Figure X.Y, where X denotes the chapter number within which a figure appears and Y denotes the number of the figure within that chapter (for instance, Figure 8.2 denotes the second figure in Chapter 8).
1. Foundations of the Present Inquiry

Adaptive goal-directed behaviour is dependent upon a well-calibrated internal model of the environment (Berkes, Orbán, Lengyel, & Fiser, 2011; R. S. Sutton & Barto, 1998; Tolman, 1948; Zhang, 1997). In particular, decision making under uncertainty is most efficient when there is a close correspondence between an agent’s internal model of its environment and true environmental contingencies (Conant & Ashby, 1970). An illustration of this point is given by human pedestrians, who are able to safely cross busy roads because they possess an internal model of this environment which specifies that cars stop at red lights and drive through green lights. By comparison, animals with a less well-developed internal model of this environment have great difficulty in crossing the same road safely. Similarly, animals’ learning about environmental contingencies is significantly faster, and therefore more effective, when they possess a good internal model of the structural relationships between different features of the environment. A foraging animal, for instance, will learn the availability of food from different sources more rapidly if it has an internal model of the covariance between different food sources (e.g., that visually similar trees are likely to produce fruit at similar times) than if it does not have such a model and attempts to learn about each food source independently (Eldar, Rutledge, Dolan, & Niv, 2016).

A prominent area of recent inquiry in cognitive neuroscience has been to understand the role played in judgment and decision making by these internal models of the environment, which can also be termed beliefs. Within this broader belief-centric research program, a number of distinct research questions have been addressed. These include the structure and function of humans’ internal models of the environment (Chater & Oaksford, 2008; Fiorillo, 2008; Friston, 2010), the circumstances under which behaviour relies upon internal models rather than upon simpler model-free habits (Daw, Niv, & Dayan, 2005; Gläscher, Daw, Dayan, & O’Doherty, 2010), the neural processes by which internal models of the environment are updated (termed belief updating; Kolossa, Kopp, & Fingscheidt, 2015; Nassar, Wilson, Heasly, & Gold, 2010; O’Reilly et al., 2013), and information seeking, concerning the value placed by humans and other animals upon information that can be used to improve the accuracy of internal models of the environment (Bromberg-Martin & Hikosaka, 2009, 2011; Vasconcelos, Monteiro, & Kacelnik, 2015; Zentall & Stagner, 2011). The latter two topics—belief updating and information seeking—were the focus of the research project presented in
this thesis. Specifically, this thesis aimed to elucidate the neural substrates of belief updating and information seeking, and to use behavioural and neural evidence to compare different theories regarding the cognitive mechanisms underlying these processes in humans. As such, this thesis was primarily concerned with the overarching question of how information is used by the human brain to update internal models of the environment. This question was addressed by investigating two distinct but related questions related to the updating of internal models: first, by what process information is incorporated into these internal models (belief updating) and, second, what neural substrates were associated with human preferences for acquiring such information (information seeking). As such, although belief updating and information seeking are addressed separately within this thesis, the two topics were considered in the context of, and as different perspectives upon, the same underlying belief-centric research program.

In the domain of belief updating, the present thesis addressed several important open questions. The first of these open questions was with what neural process or processes belief updating is implemented in the human brain. Although it has been proposed that the updating of internal models of the environment may present a unifying computational account of the central nervous system (Fiorillo, 2008), the neural correlates of belief updating are yet to be elucidated precisely. In the past five years, a growing body of research has investigated this question using neuroimaging techniques such as functional magnetic resonance imaging (fMRI; e.g., O’Reilly et al., 2013) and electroencephalography (EEG; e.g., Kolossa et al., 2015). One candidate neural index which has been proposed by EEG studies of belief updating is the P3 component of the event-related potential (ERP), the amplitude of which has been theorised to index the size of updates to participants’ beliefs under a Bayesian statistical framework (Kopp, 2008). Study 1 of this thesis tested whether this was the case using data from healthy young adult participants performing a simple perceptual learning task.

The second open question in the domain of human belief updating concerned the nature of the computations (Bayesian versus non-Bayesian) by which beliefs are updated. The belief-centric research program described above has often been linked to the ‘Bayesian brain’ hypothesis (Knill & Pouget, 2004). One strong formulation of this hypothesis states that the statistical computations underlying the generation and maintenance of internal models of the environment in the human brain take place
according to Bayes’ Rule (Friston, 2010; Friston et al., 2016). However, it is important to note that adopting a belief-centric research frame does not necessarily require adopting this strong version of this hypothesis. There also exist non-Bayesian theories of human belief updating (e.g., El-Gamal & Grether, 1995; Hogarth & Einhorn, 1992; see Chapter 2 for a summary of this literature), and one of the aims of this thesis was to use behavioural and neural data from human participants to help adjudicate between Bayesian and non-Bayesian accounts of belief updating. This question is addressed by Study 2 of this thesis. Study 2 used a similar perceptual learning task to Study 1, with the addition of a feedback incentive manipulation, to assess two questions: first, whether the presence of rewarding incentives for task performance affected participants’ use of Bayesian versus non-Bayesian task strategies, and, second, whether Bayesian and non-Bayesian task strategies were differentially reflected in the ERP components implicated in belief updating by the results of Study 1.

The second research topic of this thesis was information seeking. Whereas the research into belief updating in Study 1 and Study 2 outlined above investigated the neural processes underlying the incorporation of new information into individuals’ internal models, research into information seeking considered attitudes toward information itself. Within this domain of research, a key question has been whether human decision makers assign only an instrumental value to information, or whether they act as though information also has some intrinsic value (cf. Grant, Kajii, & Polak, 1998; Hirshleifer & Riley, 1979; D. B. Lawrence, 1999). In this context, instrumental valuation refers to a preference for acquiring information only when that information can be used to acquire other tangible rewards, such as food or money. By contrast, theories of intrinsic valuation of information posit that the acquisition of information may be itself rewarding, even where that information cannot necessarily be used to do anything (e.g., Grant et al., 1998)

Instrumental and intrinsic theories of information valuation make qualitatively similar behavioural and neural predictions in most natural decision problems. This is because the information that is available to decision makers in such everyday settings is instrumentally useful, or at least potentially so. For instance, seeking information about weather conditions by checking a forecast allows individuals to make a better-informed decision about whether or not to pack an umbrella for their commute. Where these two theories differ is with respect to their predictions regarding the valuation of non-instrumental information, defined as information which is not of any use for
acquiring tangible rewards such as food or money (Lanzetta & Driscoll, 1966)
. Instrumental value theories predict that, since non-instrumental information by
definition possesses no instrumental value, decision makers should not display a
preference for acquiring such information. By contrast, intrinsic value theories predict
that individuals ought to display a preference for acquiring non-instrumental
information, since information is hypothesised to possess a value independent of its
instrumental use. This prediction was tested across two separate behavioural
experiments by Study 3 of the present thesis.

Finally, the neural substrates of information valuation remain an open area of
research interest. One prominent theory, the ‘common currency’ hypothesis of
Bromberg-Martin and Hikosaka (2011) has proposed that the intrinsic value of
information is derived from a representation of information within canonical neural
reward-processing circuits within the basal ganglia and the dopaminergic midbrain. As
such, information and reward are posited to possess a ‘common currency’ in terms of
neural representation, which may explain why information is encoded in these reward-
sensitive brain areas in animals such as monkeys and pigeons which display intrinsic
valuation of information (Blanchard, Hayden, & Bromberg-Martin, 2015; Zentall &
Stagner, 2011). Study 4 of the present thesis tested the predictions of this common
coding hypothesis in human participants using EEG. Specifically, this study tested
whether ERP components typically associated with the processing of rewarding stimuli
also encoded the information value of non-informative stimuli.
2. Review of Belief Updating Literature

One of the two primary research topics of this thesis is the neural and cognitive mechanisms of belief updating. Before reviewing past literature on this topic it is necessary to give working definitions both of belief and of belief updating.

This thesis follows recent literature in cognitive neuroscience (e.g., Huber, Klucharev, & Rieskamp, 2015; O'Reilly et al., 2013; Vossel, Mathys, Stephan, & Friston, 2015) in defining a belief as an individual’s probabilistic cognitive representation of some aspect of the environment, where larger probabilities denote greater subjective plausibility of a particular state or event (de Finetti, 1974; Jaynes, 2003; Savage, 1954). For example, one might summarise an individual’s beliefs regarding the possibility of rain in the next week by assigning a probability of 2/3 to the event ‘rain’ and a probability of 1/3 to the event ‘no rain’. Representing beliefs by probabilities allows for transparent interpretation of the relative plausibility of different events; for instance, in the example above it is sensible to state that the individual considers it twice as likely that it will rain than that it will not rain. This probabilistic definition of belief can be considered a quantifiable counterpart of historical conceptualisations of belief in psychology, such as William James’ definition of a belief as “the psychic attitude in which our mind stands towards [a] proposition taken as a whole” (1890, p. 287). Adopting a probabilistic definition of beliefs is equivalent to assuming that the plausibility of different propositions can be summarised by a probability distribution.

Several technical points regarding this definition of belief warrant further discussion. First, it is important to note that representing degrees of belief in terms of probabilities is in keeping with a Bayesian rather than a frequentist interpretation of probability (as discussed in, e.g., Cosmides & Tooby, 1996; Wagenmakers, Lee, Lodewyckx, & Iverson, 2008). In addition, there are several colloquial definitions of the term ‘belief’ that are excluded from the definition given above, and which will not be the subject of any further discussion in this thesis. Specifically, the definition above concerns itself with beliefs which are evidential rather than existential (cf. James, 1890, pp. 289-290). Evidential beliefs are those which concern the plausibility of different propositions on the basis of empirical evidence (e.g., the belief that it will rain tomorrow, or the belief that the toss of a coin will come up heads); excluded from this definition are existential beliefs in the goodness or truth of some concept (e.g., belief
in democracy, belief in God). This is equivalent to the distinction proposed by Price (1965) between ‘belief-that’ (evidential beliefs) and ‘belief-in’ (existential beliefs). This thesis is solely concerned with the former.

Having thus defined belief, it is also necessary to provide a working definition of belief updating. This thesis adopts a definition of belief updating as the sequential revision of beliefs on the basis of new information. Since we have defined beliefs above as probabilistic, belief updating can therefore be considered as fundamentally a form of probabilistic inference. I therefore follow N. H. Anderson (1981) in considering belief updating a dynamic, sequential process:

“Information is received a piece at a time and integrated into a continuously evolving impression. Each such impression, be it of a theoretical issue, another person, or a social organization, grows and changes over the course of time. At any point in time, therefore, the current impression looks both forward and back. In one perspective, the current impression is the cumulated resultant of all past information. In the other perspective, it is the initial impression into which future information will be integrated.” (N. H. Anderson, 1981, p. 144)

The empirical investigations of belief updating reviewed in this chapter will therefore be concerned primarily with the sequential updating of beliefs in response to multiple new pieces of information presented one after another.

Belief updating is a ubiquitous feature of human cognition, fundamental to cognitive processes including but not limited to judgment, decision making, and causal inference (Hogarth & Einhorn, 1992). However, there exists a prominent theoretical divide in research assessing how human participants perform belief updating. This divide concerns the question of whether or not human belief updating can be well described as Bayesian probabilistic inference. This distinction is of great theoretical importance because Bayes’ theorem (Bayes, 1763; Laplace, 1812) describes a mathematically optimal way of integrating new information into uncertain prior beliefs; as such, proposing that participants follow Bayes’ theorem in belief updating is equivalent to asserting that human belief updating closely follows normative standards. This contention has proved controversial. As such, this literature review will distinguish research adopting a Bayesian perspective on human belief updating (discussed in Section 2.1) from research seeking to characterise the ways in which human belief
updating deviates from normative Bayesian standards (henceforth referred to as non-Bayesian belief updating, and discussed in Section 2.2).

2.1 Bayesian belief updating

The statistically optimal way of updating one’s beliefs given new information is specified by Bayes’ theorem (Bayes, 1763; Laplace, 1812). In the terminology of Bayesian statistics, this process is one in which prior beliefs regarding the probability of some hypothesis are transformed after the observation of new information into posterior beliefs; this process is crucially dependent upon the likelihood of observed data given the hypothesis. Bayes’ theorem can be formally stated as follows:

\[
P(H|D) = \frac{P(H) \times P(D|H)}{P(D)}\]

This equation specifies how beliefs regarding some hypothesis \(H\) should be updated upon observing some new data \(D\). The left hand side of the equation, \(P(H|D)\) denotes the posterior probability of the hypothesis given the data, and is calculated by multiplying prior belief in the hypothesis, \(P(H)\), by the likelihood of the data given that the hypothesis is true, \(P(D|H)\), and dividing by the probability of the data \(P(D)\). For the studies reviewed in this thesis, \(P(D)\) is generally treated as a normalisation factor, since it is constant with respect to the hypothesis under consideration. For this reason, Bayes’ theorem can also be rewritten to specify that posterior beliefs are proportional to the product of prior and likelihood:

\[
P(H|D) \propto P(H) \times P(D|H)\]

Patterns of belief updating which are consistent with this framework are termed Bayesian (or, frequently, ‘Bayes-optimal’).

It is important to note that, although the probability calculus of Bayesian belief updating can be expressed simply, and is easily implemented in several lines of computer code, in naturalistic environments Bayesian belief updating can become extremely computationally complex. The reason for this is that, although the application of Bayesian belief updating to a single hypothesis is simple, in the presence of multiple hypotheses Bayesian belief updating requires integration over every possible hypothesis. For naturally occurring inference problems (e.g., learning a
language, predicting the movements of the stock market) there may be an extremely large, or even infinite, number of such hypotheses. As such, full Bayesian inference for naturalistic tasks can become extremely complex, or even computationally intractable (Cooper, 1990).

The assumption that human belief updating is Bayesian underpins a number of recent theories of judgment and inference in cognitive psychology and cognitive neuroscience (e.g., Courville, Daw, & Touretzky, 2006; Griffiths & Tenenbaum, 2006) as well as theories of rational choice in economics (see Camerer & Loewenstein, 2004 for review). I review theories and findings concerning Bayesian belief updating in these three disciplines in separate sections below, before turning to the broader question of non-Bayesian belief updating in Section 2.2.

2.1.1 Bayesian belief updating in cognitive psychology

Research into Bayesian belief updating in cognitive psychology has focused on the question of whether human belief updating can be described accurately as a form of Bayesian inference. In recent years, research in cognitive psychology has studied Bayesian belief updating within a broader research program alternately termed rational analysis (J. R. Anderson, 1990; Chater & Oaksford, 1998), or Bayesian cognitive science (Chater & Oaksford, 2008). This approach has its roots in early cognitive research comparing human probabilistic inference to normative standards from statistical decision theory (e.g., Peterson & Beach, 1967; Wallsten, 1972). However, recent research in Bayesian cognitive science is distinguished from these earlier studies in that it considers human cognition from a teleological perspective. That is, this research program assesses the extent to which cognition can be considered an optimal response to the information-processing demands of the natural environment (J. R. Anderson, 1991b). This approach has been used to investigate a range of cognitive phenomena, including categorisation (J. R. Anderson, 1991a; Tenenbaum & Griffiths, 2001), memory (J. R. Anderson & Schooler, 1991) reasoning (Oaksford & Chater, 1994, 1996), and language learning (Xu & Tenenbaum, 2007).

For reasons described in Section 2.1, Bayesian cognitive science has naturally tended to adopt a Bayesian perspective upon the question of human belief updating. As such, the focus of this research program has been upon characterising ways in which human belief updating and probabilistic inference resemble Bayesian norms. For
instance, an influential study by Griffiths and Tenenbaum (2006) provided evidence for the Bayes-optimality of simple predictions made by participants on the basis of minimal information (e.g. “If you made a surprise visit to a friend, and found that they had been watching a movie for 30 minutes, what would you predict for the length of the movie?”; Griffiths & Tenenbaum, 2006, p. 770). A similar conclusion was reached in a study by Cosmides and Tooby (1996), in which the authors presented participants with a number of classical probabilistic inference questions (e.g., given the base rate of a disease and the sensitivity of a diagnostic test for that disease, what is the probability that a person who tests positive actually has the disease?). The authors then demonstrated that, when information was presented to participants in a format compatible with frequentist analysis, participants’ responses were consistent with posterior probabilities generated by applying Bayes’ theorem. Notably, this result was inconsistent with earlier ‘heuristics and biases’ research suggesting that participants display systematic deviations from optimality when performing probabilistic inference tasks (see Section 2.2 for a more detailed review of this literature). Cosmides and Tooby (1996) therefore concluded that human belief updating is closer to Bayes-optimality than had been previously supposed.

Results consistent with this conclusion have also been reported based on behavioural performance on other cognitive tasks. A study by Wallsten and Sapp (1977), for instance, reported that participants’ subjective probability estimates in a hypothesis comparison task were closely approximated by Bayesian posterior distributions in a sequential probabilistic inference task. More recently, studies by Meyniel, Schlunegger, and Dehaene (2015) and E. C. Yu and Lagnado (2012) also reported that participants’ subjective probabilities in sequential inference tasks closely matched posterior probabilities of normative Bayesian belief updating. Importantly, however, the latter study added a caveat to this conclusion. While concluding that behaviour was consistent with Bayesian belief updating, E. C. Yu and Lagnado (2012) nevertheless highlighted two limitations of Bayesian theories for understanding behaviour in their task (in which participants observed play on virtual poker machines before estimating the machines’ payoff probabilities). The first limitation noted by E. C. Yu and Lagnado (2012) was that Bayesian theories “do not specify the underlying cognitive process” governing participants’ behaviour (E. C. Yu & Lagnado, 2012, p. 7). That is, although participants’ responses were consistent with Bayesian updating, this data did not license any inferences about whether participants actually implemented
Bayesian inference, or whether they instead used a less complex strategy which simply *approximated* Bayesian inference in a particular task environment. Secondly, the authors also noted that Bayesian models of this task could in fact account for *any* payoff probability a participant might conceivably report, simply by inferring that participants were basing their belief updates upon misinformed prior beliefs regarding the noise in the poker machines’ outcome-generating process. This led E. C. Yu & Lagnado (2012) to discount the plausibility of purely Bayesian theories for this task, since a theory which can equally account for any possible observed data is in principle unfalsifiable.

This latter point echoes general criticisms of Bayesian cognitive science that have been raised in a number of recent commentaries (e.g., Bowers & Davis, 2012; M. Jones & Love, 2011; Marcus & Davis, 2013). These commentaries have advanced a number of practical and theoretical criticisms of Bayesian cognitive science, two of which are of particular relevance to Bayesian belief updating: first, the structural flexibility of Bayesian theories of belief updating, and second, issues of cognitive task selection. The former point is similar to that described by E. C. Yu and Lagnado (2012), and notes that many Bayesian models possess arbitrary flexibility, such that they can provide post-hoc explanations for any pattern of belief updating through an arbitrary specification of the prior (Bowers & Davis, 2012; Daunizeau et al., 2010; M. Jones & Love, 2011), or likelihood function (Daunizeau et al., 2010). As such, even when Bayesian belief updating models fit data well, the models’ flexibility means that it may not be warranted to conclude that participants were truly updating beliefs in a Bayes-optimal manner. The second issue raised by these commentaries—task selection—is discussed in detail by Marcus and Davis (2013). These authors noted that, among studies providing evidence for the Bayes-optimality of human cognition, researchers tended to base their conclusions upon data from a select group of tasks, while ignoring other classes of cognitive tasks in which human performance has been shown to deviate from optimality. For instance, whereas Griffiths and Tenenbaum (2006) reported that participants’ probabilistic inferences closely matched Bayesian norms, studies using different probabilistic inference tasks have reached the opposite conclusion (e.g., the gambler's fallacy; Tversky & Kahneman, 1974; the conjunction fallacy; Tversky & Kahneman, 1983). Of course, the mere presence of inconsistent evidence does not invalidate the claims of Bayesian cognitive science, since it may be that participants perform Bayesian belief updating only for relatively simple problems. However, Bayesian belief updating models are often tailored to particular task environments, and
do not provide principled explanations for why participants might be Bayes-optimal in one task, but not in another (Marcus & Davis, 2013). It is therefore important, in assessing the question of Bayesian belief updating, to consider the extent to which any given findings may be specific to a particular task context. Non-Bayesian belief updating effects in cognitive psychology are therefore considered separately in Section 2.2.

2.1.2 Bayesian belief updating in economics

As well as cognitive psychology, belief updating has also been a topic of research in the economic disciplines of decision theory, management theory, and microeconomics. Since an important aim of research in these fields is to identify the best ways for ideal agents to make decisions in the presence of uncertainty (see, e.g., Clemen, 1996), the issue of how agents ought to update uncertain beliefs on the basis of new evidence is of primary concern. In economics, the majority of the belief updating research is often interpreted as taking a prescriptive view upon belief updating. In other words, this research has sought to identify a set of principles which agents ought to follow when updating beliefs in order to maximise their expected future utility (Blume & Easley, 1995; Marimon, 1997; Raiffa & Schlaifer, 1961). This is in contrast to the research in psychology summarised above, which can be more accurately thought of as aiming to provide a good descriptive account of how humans update their beliefs. Given the prescriptive focus of much economic research, and given that the mathematically optimal way of updating beliefs is given by Bayes’ theorem, standard economic models of belief updating have generally adopted a Bayesian framework (for a detailed review see Zellner, 1988). Important exceptions to this generalisation are discussed in Section 2.2.

It is possible to trace the concern with belief updating in economics to the statistical principles of subjective probability outlined by Savage (1954), and to the subsequent influence of this work on decision theory (see Kiefer & Nyarko, 1995 for review). Savage (1954) laid out a statistical framework for decisions under risk, and described a number of conditions under which an agent could be said to hold rational preferences over different outcome states of the environment, given a particular set of subjective beliefs about the likelihood of moving into each of those states. Although this axiomatic model was not primarily cast in terms of Bayesian inference in its
original formulation, it has subsequently been shown that the subjectivist statistical principles laid out by Savage (1954) necessarily entail Bayesian belief updating, such that an agent following these axioms would always update beliefs in a Bayes-optimal fashion (Cogley & Sargent, 2008; Kiefer & Nyarko, 1995). Such Bayes-optimal belief updating is therefore also economically rational; that is, given a particular prior belief and conditional upon a particular sequence of observations, an agent which updates its beliefs according to Bayes’ theorem will on average derive at least as much profit, and typically more profit, than an agent following any other belief updating scheme.

However, it should be noted that although Bayesian perspectives of belief updating account for much of standard economic theory regarding decision making under uncertainty, significant portions of the economics literature have also considered non-Bayesian perspectives. In particular, behavioural economics has sought to extend the standard framework to describe human participants’ deviations from Bayes-optimality (e.g., Gigerenzer & Goldstein, 1996; Kahneman & Tversky, 1972). This issue is reviewed further in Section 2.2 below.

2.1.3 Bayesian belief updating in cognitive neuroscience

Recent belief updating research in cognitive neuroscience has tended to accept the assumption that Bayesian inference provides a good description of belief updating. Then, premised upon this assumption, cognitive neuroscience research has investigated the question of how Bayesian belief updating is implemented in the human brain. Alternatives to the assumption that human belief updating is Bayes-optimal have only rarely been considered within this literature (e.g. Achtziger, Alós-Ferrer, Hügelschäfer, & Steinhauser, 2014; Sharot, Korn, & Dolan, 2011; Vossel et al., 2015; this literature is discussed further in Section 2.2). On the basis of this assumption, research has then sought to map brain regions involved in belief updating using neuroimaging techniques (e.g. Behrens, Woolrich, Walton, & Rushworth, 2007; Iglesias et al., 2013; McGuire, Nassar, Gold, & Kable, 2014; O'Reilly et al., 2013), to characterise the spatial and temporal dynamics of these regions during belief updates (Nassar et al., 2012; Vossel et al., 2015), and to describe neural signatures of belief updating which may be taken as an index of belief updates even in the absence of subjective report (Kolossa, 2016; Kolossa et al., 2015; Kopp, 2008). The two main research techniques which have been used to investigate these questions are functional magnetic resonance imaging (fMRI),
and electroencephalography (EEG). Results from these two neuroimaging techniques are discussed separately below.

2.1.3.1 fMRI studies of Bayesian belief updating

Using fMRI, a neuroimaging technique which uses blood-oxygen-level dependent (BOLD) magnetic signal strength as a proxy measure of neural activation, researchers have identified a number of brain regions associated with Bayesian belief updating. However, the identity of these brain regions, and their patterns of activation, depended upon the belief updating task being performed by participants, and also upon how the concept of belief updating was operationalised when analysing fMRI data. It is therefore important to consider differences between tasks used in different studies when interpreting results. Overall, three prefrontal and fronto-temporal brain regions have been strongly linked to the implementation of belief updating: the anterior cingulate cortex (ACC), bilateral inferior frontal gyri (IFG), and the anterior insular cortex (Behrens et al., 2007; d'Acremont, Schultz, & Bossaerts, 2013; Huber et al., 2015; Iglesias et al., 2013; McGuire et al., 2014; O'Reilly et al., 2013; Preuschoff, Quartz, & Bossaerts, 2008; Vossel et al., 2015).

Of these brain regions, activation in the ACC has been shown to encode the magnitude of belief updates in participants performing a simple saccade-based probabilistic learning task (O'Reilly et al., 2013). In the context of this simple saccade response task, O'Reilly et al. (2013) operationalised belief updating by defining the magnitude of a belief update as the dissimilarity of prior and posterior distributions, measured by the distributions’ Kullback-Leibler divergence (see Cover & Thomas, 1991). This can be used as a measure of belief updating because larger belief updates are associated with greater dissimilarity between prior and posterior belief distributions, and therefore greater Kullback-Leibler divergence. O'Reilly et al. (2013) found that ACC showed greater activation when participants made larger updates to their beliefs, and less activation for smaller belief updates. The encoding of this belief updating signal in ACC was dissociable from the encoding of a separate but mathematically related quantity, stimulus-bound surprise, which O'Reilly et al. (2013) found to be reflected in neural activation within posterior parietal cortex. Interestingly, ACC activation has also been shown to index environmental volatility in a dynamic and uncertain environment (Behrens et al., 2007). This finding also provides indirect
evidence for the role of the ACC in belief updating: when learning from sequential noisy observations in a volatile environment, the effect of a new observation on beliefs ought to be inversely proportional to the degree of environmental volatility (Dayan, Kakade, & Montague, 2000). Behrens et al. (2007) interpreted their results as suggesting that the ACC may be involved in weighting new information to determine the degree to which prior beliefs are updated, consistent with the findings of O'Reilly et al. (2013). This may also be consistent with the putative role of the ACC in conflict detection (Carter & Van Veen, 2007), since belief updating is likely to involve the resolution of conflict between competing hypotheses regarding the causal structure of the environment. In support of the conclusions of O’Reilly et al. (2013), a subsequent study using a distinct Bayesian model which permitted the disentanglement of uncertainty-based from surprise-based belief updating also linked anterior prefrontal activation general—and ACC activation specifically—with uncertainty-based belief updating (McGuire et al., 2014).

A second brain region which has been implicated in Bayesian belief updating is bilateral IFG. Rather than encoding the magnitude of belief updates directly, it has been shown using fMRI that neural activation in these regions is sensitive to the violation of outcome expectancies, which may be a cue that beliefs ought to be updated (d’Acremont et al., 2013). Using a Bayesian belief updating model termed the Hierarchical Gaussian Filter (HGF; Mathys, Daunizeau, Friston, & Stephan, 2011; Mathys et al., 2014), Iglesias et al. (2013) demonstrated that activation in bilateral IFG correlated with a cue-outcome association prediction error, weighted by the uncertainty of beliefs. A similar result was also shown by Huber et al. (2015), who found that activation in IFG and anterior insula reflected belief updating only for belief updates driven by privately available information, as distinct from belief updates driven by publically available social information. The co-activation of IFG and anterior insula in this study is consistent with previous regions showing that these two brain regions often show coordinated patterns of activation (Craig, 2009); as such, it is possible that these two regions may constitute a single functional unit in belief updating. The anterior insular cortex has also been shown to encode a ‘risk prediction error’ when participants update their beliefs about the likelihood of future reward (Preuschoff et al., 2008). This finding also suggests a role for anterior insula in belief updating, but given the substantial differences in how belief updating is defined in different studies, it remains to be determined what this role may be.
Indeed, the degree of inconsistency in fMRI results between different studies of belief updating is likely to be related to the wide range of different behavioural tasks used to assess belief updating, as well as these studies’ differing operationalisations of belief updating. It follows that that there is likely to be no single brain region which is solely responsible for implementing Bayesian belief updating; rather, belief updating appears to depend on the coordinated activity of a diverse network of brain regions (Vossel et al., 2015). Consistent with this conclusion, several recent studies have provided evidence that belief updating depends on the neurotransmitters norepinephrine (Nassar et al., 2012) and acetylcholine (Vossel et al., 2014), both of which are present throughout cortex and subcortex, and which are not localised to any of the particular brain regions discussed above.

In support of a distributed-network account of belief updating, Nassar et al. (2012) provided evidence for a relationship between pupil diameter and belief updating. The authors of this study found that transient fluctuations in participants’ pupil diameter reflected the volatility of new information, and baseline pupil diameter reflected the uncertainty of participants’ beliefs. These observed relationships provide information regarding the neural mechanisms of belief updating, since tonic and phasic changes in pupil diameter are thought to be related to norepinephrinergic neural activity in the locus coeruleus, a brainstem nucleus (Aston-Jones & Cohen, 2005). Moreover, the observed relationship was not epiphenomenal, but displayed causal properties: an unrelated experimental manipulation of participants’ pupil size also had a systematic effect on patterns of belief updating (Nassar et al., 2012). With respect to acetylcholine, it has been shown by Vossel et al. (2014) that that administration of the acetylcholine agonist galantamine affected participants’ belief updating (Vossel et al., 2014), in line with the hypothesis that acetylcholine levels may encode uncertainty regarding cue-outcome contingencies (A. Yu & Dayan, 2005). Specifically, using the HGF to estimate participants’ latent beliefs, Vossel et al. (2014) found that increases in cortical acetylcholine levels were associated with faster belief updating regarding the outcome contingencies of simple visual cues.

2.1.3.2 EEG studies of Bayesian belief updating

Past research has typically used EEG to investigate Bayesian belief updating by assessing the covariance of belief updating variables with the amplitude of different
components of the event-related potential (ERP; see Luck, 2005) evoked by the presentation of informative stimuli. ERP analysis methodology involves calculating an average voltage waveform associated with an event of interest; it is then possible to interpret characteristics of this waveform—such as the amplitude or latency of different peaks—with reference to underlying cognitive processes of interest. Two ERP components in particular have been linked, either directly or indirectly, to belief updating: the P3 (S. Sutton, Tueting, Zubin, & John, 1967), and the stimulus-preceding negativity (SPN; Brunia, 1988). These two components are discussed separately below.

The P3 is a positive ERP component whose amplitude has been linked to a large number of cognitive processes, including processing of stimulus probability (Mars et al., 2008; S. Sutton et al., 1967), stimulus novelty (Friedman, Cycowicz, & Gaeta, 2001; S. Sutton, Braren, Zubin, & John, 1965) stimulus value (Begleiter, Porjesz, Chou, & Aunon, 1983; Sato et al., 2005) emotional salience of stimuli (Johnston, Miller, & Burleson, 1986), and status of stimuli as targets or distractors (Squires, Squires, & Hillyard, 1975). It is useful to distinguish between two topographically distinct subcomponents of the P3: the fronto-centrally distributed P3a, and the centro-parietally distributed P3b (see Polich, 2007 for review). In recent years, several different theoretical perspectives have each suggested that, given the cognitive variables encoded in the P3 (particularly the P3a), this component may represent the output of neural processes implementing Bayesian belief updating (Donchin & Coles, 1988; Kopp, 2008; Mars et al., 2008). This is because most of the manipulations affecting P3 amplitude summarised above affect the degree to which participants must revise their beliefs regarding cue-outcome contingencies or other probabilistic features of the task environment. Moreover, given that one cortical region proposed as a generator for the P3a component is the ACC (Volpe et al., 2007), a role for the P3a in belief updating is consistent with evidence from fMRI studies reviewed above that the ACC is critically involved in belief updating (Behrens et al., 2007; O'Reilly et al., 2013). Recently, several sets of results have provided empirical support for this proposal, finding that P3a amplitude encoded Bayesian belief update size in a simple probabilistic inference task (Kolossa, 2016; Kolossa et al., 2015). However, it remains to be determined whether this is also true in more naturalistic task settings.

The second ERP component which may be linked with Bayesian belief updating is the SPN, a slow negative deflection associated with the anticipation of upcoming stimuli (Brunia, 1988). Increased SPN amplitudes are associated with stimuli which
reinforce past responses, both via delivery of reward (Masaki, Yamazaki, & Hackley, 2010), and via delivery of instructive feedback (Damen & Brunia, 1994; Morís, Luque, & Rodriguez-Fornells, 2013). Moreover, SPN amplitude also varies as a function of the expected information content, as well as the predictability, of reinforcing stimuli (Catena et al., 2012; Kotani et al., 2003). As a result, it is possible that SPN amplitude may be reconceptualised as reflecting neural computations regarding belief uncertainty prior to belief updates, though this proposition has not heretofore been tested explicitly.

2.2 Non-Bayesian belief updating

Implicitly or explicitly, the literature reviewed in the preceding section describes human belief updating as a Bayesian process. As such, these theories and results share a common assumption that, given a generative model of the environment, new information is integrated with prior beliefs according to Bayes’ theorem. However, there is substantial empirical evidence to suggest that in many task environments, participants’ belief updating deviates significantly and systematically from Bayesian norms. The following sections review these non-Bayesian belief updating effects, with a focus upon characterising the circumstances in which these effects have been elicited. Five distinct non-Bayesian belief updating effects will be reviewed: conservatism, order effects, asymmetry, conflict with heuristics, and confirmation bias.

Parenthetically, although the focus of this section is upon empirical results concerning non-Bayesian belief updating in psychology, neuroscience, and economics, it is important to note that non-Bayesian belief updating has also been the focus of theoretical analysis in economics. This research has led to a number of formal economic models of non-Bayesian belief updating, including those which permit agents to choose prior beliefs as they please (Akerlof & Dickens, 1982; Brunnermeier & Parker, 2005), those which amend Bayes’ theorem by allowing favourable and unfavourable information to be treated differently (Compte & Postlewaite, 2004; Rabin & Schrag, 1999), and those which allow for non-Bayesian subjectivity both in selection of prior beliefs and in the manner in which prior beliefs are updated (Epstein, 2006; Epstein, Noor, & Sandroni, 2008). Although of relevance to some of the specific non-Bayesian belief updating effects discussed below (particularly asymmetry and confirmation bias), the present literature review will not analyse these models in depth. This is
because unlike the Bayesian models described in the previous section, these economic models are likely to be of limited general relevance to experiments in psychology and neuroscience since, as noted by Mobius, Niederle, Niehaus, and Rosenblat (2014; p. 2), each non-Bayesian model is “too ad-hoc and restrictive to explain non-Bayesian inference in general”. As such, this chapter will first discuss four known violations of Bayes-optimal belief updating (Sections 2.2.1 to 2.2.4) before discussing an alternative non-Bayesian framework based on the use of heuristics (Section 2.2.5).

2.2.1 Conservatism

Among the core findings of the cognitive revolution in psychology was that, contrary to Bayesian theories, human belief updating is generally conservative (Beach, 1966; Edwards, 1968; Peterson & Miller, 1965; Peterson, Schneider, & Miller, 1965; Phillips & Edwards, 1964, 1966; Phillips, Hays, & Edwards, 1966; Slovic & Lichtenstein, 1971). That is, upon the presentation of new information, it was observed that participants generally updated their beliefs in the correct direction, but not to the extent prescribed by Bayes’ theorem. In the commonly used ‘book bag’ task of Phillips and Edwards (1966), for instance, participants were informed that poker chips would be drawn from one of two ‘book bags’, each containing chips of two colours in different known ratios (e.g., seven red chips and three blue chips in Bag One versus three red and seven blue chips in Bag Two). On each trial, one of the two bags was selected randomly by the experimenter, and participants observed a number of draws with replacement of chips from this chosen bag. After each new chip was shown, participants were asked to report their updated beliefs regarding which of the two bags the experimenter was drawing from. The conservatism effect was that participants’ probabilistic belief updates were generally much smaller than prescribed by Bayes’ theorem, resulting in overall under-confidence. The practical effect of this conservatism was that participants’ probability estimates fell closer to 0.5 than did the ‘correct’ Bayesian probability. For a given pattern of evidence, in other words, participants displayed less certainty regarding the identity of the chosen bag than was statistically warranted by their observations. Although one possibility was that conservatism might simply have resulted from the artificial task demands of the book bag task, subsequent research showed that conservative belief updating was also observed in more complex
and naturalistic tasks (Alker & Hermann, 1971; Phillips et al., 1966) and in applied domains such as accounting (e.g., Eger & Dickhaut, 1982).

Although the cognitive and neural processes responsible for conservative belief updating remain unclear, early research implicated participants’ calculation of data likelihoods as one plausible cognitive mechanism. Prescriptively, Bayes-optimal belief updates should be proportional to the product of prior belief in each hypothesis, denoted $P(H)$, and the likelihood of the observed data given each hypothesis, $P(D|H)$. For a particular prior, therefore, conservatism might result either from erroneous calculation of data likelihood, or a mechanistic failure to correctly integrate prior and likelihood, or both of the above. A study by Peterson, DuCharme, and Edwards (1968) suggested that the former explanation—erroneous calculations of the likelihood of the data under different hypotheses—provided a good explanation for participants’ conservative belief updating. In this study, in addition to eliciting the strength of participants’ beliefs after new information was observed, Peterson et al. (1968) also elicited participants’ estimates of $P(D|H)$, the likelihood of the observed data given each hypothesis. It was found that, like overall belief updates, estimates of $P(D|H)$ were markedly conservative, and that much of the overall conservatism of belief updating could be explained by assuming that participants were updating their beliefs using these erroneous likelihoods rather than the ‘correct’ likelihood values. In line with this explanation, a more recent study by Corner, Harris, and Hahn (2010) has suggested that external factors may influence the perceived reliability of data sources. These differences in perceived reliability may result in participant scepticism regarding the observed data and, therefore, in poorly specified individual estimates of $P(D|H)$ and conservative belief updating.

Individual differences in conservative belief updating appear to be somewhat related to general cognitive ability. A recent study by Oechssler, Roider, and Schmitz (2009) reported that participants who scored highly on a task assessing general cognitive reflection displayed less conservatism than participants who scored low on general cognitive reflection. Interestingly, conservatism in belief updating may also be less pronounced in some psychiatric conditions than in the general population: individuals with a diagnosis of delusional schizophrenia displayed less conservative patterns of belief updating than healthy controls (Huq, Garety, & Hemsley, 1988). In this case, however, the authors urged caution in the interpretation of results, suggesting that this finding may have been the result of a general bias towards overconfidence in
this disorder ‘cancelling out’ the conservatism effect. Separately, research in behavioural economics has suggested that conservatism may at least partly be an artefact of the procedure by which participants’ beliefs are elicited. When participants’ beliefs were elicited using an incentive-compatible procedure (a Becker-DeGroot-Marshak auction; Becker, DeGroot, & Marschak, 1964), participants’ belief updates were less conservative than in standard tasks using simpler measures of belief strength such as subjective probability report (Holt & Smith, 2009). Finally, it has also been found that conservatism in belief updating increases with increasing task complexity and perceived importance of outcomes (Alker & Hermann, 1971), suggesting that stronger task motivation and higher cognitive load may also exacerbate conservative belief updating.

Although conservative belief updating has not been a common topic of recent research in cognitive neuroscience, several studies have found that including a conservatism-like parameter improved the goodness-of-fit of probabilistic learning models to participant choice data (Kolossa, 2016; Kolossa et al., 2015; McGuire et al., 2014; Nassar et al., 2010). This can be implemented either by assuming that participants extract less than the full likelihood information from each new piece of data (e.g., Nassar et al., 2010), or by assuming that participants re-weight probabilities in belief updates to place more weight on extreme-probability events and less weight on moderate-probability events (e.g., Kolossa et al., 2015).

2.2.2 **Order effects**

One assumption of Bayesian belief updating is that sequential belief updates are dependent only upon the information provided by stimuli, and not upon the order in which this information is presented. To illustrate this, consider two pieces of information: \(A\) and \(B\). According to Bayesian accounts of belief updating, participants’ beliefs after observing these two stimuli should be identical whether they are presented in the order \(AB\) or the order \(BA\). However, numerous studies have shown that this is not the case. Instead, participants’ behaviour has overwhelmingly displayed pronounced order effects in belief updating (see Hogarth & Einhorn, 1992 for a comprehensive review). In general, research has identified two forms of order effects in sequential belief updating: primacy effects and recency effects. Primacy effects refer
to the finding that people often update beliefs more in response to information early in a sequence than to information late in a sequence (N. H. Anderson, 1973; N. H. Anderson & Barrios, 1961; Asch, 1946; Benassi, 1982; Curley, Young, Kingry, & Yates, 1988; Dale, 1968; Mantonakis, Rodero, Lesschaeve, & Hastie, 2009; McAndrew, 1981; Roby, 1967; Tesser, 1968; Yates & Curley, 1986). By contrast, recency effects are the converse of primacy effects, and refer to cases in which greater updating weight is placed upon more recent items (N. H. Anderson, 1959; Furnham, 1986; Hands & Avons, 2001; Levin & Schmidt, 1970; Luchins, 1958; Pitz & Reinhold, 1968; Shanteau, 1970). Both forms of order effect have been observed in similar task contexts, including probabilistic judgment tasks (Dale, 1968; Pitz & Reinhold, 1968), judgments of character (Asch, 1946; Luchins, 1958), and attributions of performance to individuals (Allen & Feldman, 1974; E. E. Jones, Rock, Shaver, Goethals, & Ward, 1968). In some cases, both primacy and recency effects have been observed within the same study (see, e.g., E. E. Jones et al., 1968; Mantonakis et al., 2009; Stewart, 1965). In judgments of character, for instance, Asch (1946) observed that participants’ ratings of hypothetical individuals’ personalities were affected more by items earlier in a list than those late in a list, whereas the results of Luchins (1958) suggested that the converse was true.

Although the presence of both primacy and recency effects might appear paradoxical, the model of belief updating proposed by Hogarth and Einhorn (1992) demonstrates that it is possible to account for both types of effects simultaneously using an overall framework of belief ‘anchoring and adjustment’ (cf. Tversky & Kahneman, 1974). Synthesising previous literature, Hogarth and Einhorn (1992) concluded that the relative preponderance of primacy and recency effects in different studies depended on a number of experimental factors, most notably the manner in which participants’ beliefs were elicited (either at the end of a sequence of stimuli, or step-by-step after each new stimulus), and whether the information-bearing stimuli were simple or complex (for example, simple single-number stimuli versus complex 600-word messages). The authors concluded, first, that step-by-step belief elicitation tended to be associated with recency effects, whereas end-of-sequence belief elicitation was associated with primacy effects, and, second, that simpler tasks tended to induce primacy effects, whereas more complex tasks tended to induce recency effects. The model of Hogarth and Einhorn (1992) accounted for these effects by assuming that participants ‘anchored’ their beliefs at a current state of knowledge, and ‘adjusted’ (that
is, updated) their beliefs each time a response was made. Although initially only applicable to situations in which different pieces of evidence contradicted one another, a subsequent study by Tubbs, Gaeth, Levin, and Van Osdol (1993) extended this model to account for recency effects for consistent pieces of evidence. Tubbs et al. (1993) showed that similar qualitative patterns to the model of Hogarth and Einhorn (1992) could be predicted for consistent pieces of evidence if there were a sufficiently large difference in the degree of belief updating associated with different stimuli.

Finally, a recent study by Jasper, Kunzler, Prichard, and Christman (2014) identified an interesting pattern of individual differences in order effects in belief updating. These authors reported an interaction between the strength of order effects across individuals and the strength of handedness, with participants of mixed handedness (who reported using their non-dominant hand for many everyday activities) exhibiting a stronger recency effect for two pieces of inconsistent information than those of strong handedness (who reported using their dominant hand for most everyday activities). Although Jasper et al. (2014) proposed that this might be the result of a hemispheric asymmetry in belief updating between participants of strong handedness and participants of mixed handedness, this proposal has not yet received support from any neuroscientific research.

2.2.3 Asymmetric belief updating

One frequently studied non-Bayesian belief updating effect, termed asymmetric belief updating (Sharot & Garrett, 2016) or the ‘good news-bad news effect’ (Eil & Rao, 2011), is that people tend to make larger updates to their beliefs following favourable information than following unfavourable information. For instance, participants given feedback from a sham IQ test that indicates that they are more intelligent than previously estimated tend to make larger updates to their beliefs than participants who receive feedback that they are less intelligent than previously estimated (Eil & Rao, 2011; Mobius et al., 2014). This pattern of belief updating is non-Bayesian, since Bayesian updating is necessarily symmetrical across information of different valences. It has been proposed that asymmetric belief updating may underlie such general effects as the bias toward optimistic self-appraisals in the general population (Lench & Ditto, 2008; Lund, 1925; Weinstein, 1980, 1987), and the self-serving causal attribution bias (Bradley, 1978; but see also D. T. Miller & Ross, 1975).
Asymmetric belief updating has typically been investigated using variants of the ‘belief formation updating task’ (Chowdhury, Sharot, Wolfe, Düzel, & Dolan, 2014; Garrett & Sharot, 2014; Garrett et al., 2014; Korn, Sharot, Walter, Heekeren, & Dolan, 2014; Kuzmanovic, Jefferson, & Vogeley, 2015, 2016; Moutsiana, Charpentier, Garrett, Cohen, & Sharot, 2015; Moutsiana et al., 2013; Sharot, Guitart-Masip, Korn, Chowdhury, & Dolan, 2012; Sharot, Kanai, et al., 2012; Sharot et al., 2011). In this task, participants first estimate their likelihood of experiencing a number of negative life events (e.g., burglary, diagnosis of Alzheimer’s disease), and are then given feedback regarding the true average probability of the event across the population. Trials are divided into a ‘good news’ condition (comprising trials in which the participant overestimated their likelihood of experiencing the adverse event) and a ‘bad news’ condition (in which the participant underestimated the adverse event’s likelihood). Crucially, participants then report a second set of estimated likelihoods for the same events, with the difference between the first and the second estimate taken as to indicate the size of the belief update. A robust finding in this task is that participants update their beliefs more in the good news condition than in the bad news condition, indicative of asymmetric belief updating (Sharot et al., 2011).

The belief formation updating task results have also revealed several interesting patterns of asymmetric belief updating in developmental and clinical samples. Moutsiana et al. (2013) assessed asymmetric belief updating in nine- to 26-year-olds and found that children exhibited more asymmetric belief updating than adolescents and young adults. This effect was driven by smaller belief updates in response to unfavourable information for younger children, whereas the degree of belief updating in response to favourable information remained relatively constant across different age groups. At the other end of the developmental spectrum, Chowdhury et al. (2014) found that older adults (mean age 66 years) displayed a stronger asymmetric belief updating effect than younger adults (mean age 22 years), with older adults making both larger belief updates in response to favourable information and smaller belief updates in response to unfavourable information than younger adults. It has also been shown, moreover, that unlike healthy adults, adults with a diagnosis of major depressive disorder do not display an asymmetric belief updating (Garrett et al., 2014; Korn et al., 2014). On the basis of this result, Korn et al. (2014) proposed that the tendency toward asymmetric belief updating observed in the general population may, in fact, be a protective factor for individuals’ mental health.
Neuroimaging studies have implicated a diverse set of brain regions in asymmetric belief updating, including ventromedial prefrontal cortex, inferior frontal gyrus, and inferior parietal lobule (Garrett et al., 2014; Kuzmanovic et al., 2016; Sharot et al., 2011), as well as connectivity within a left fronto-subcortical network including inferior frontal gyrus, insular cortex, putamen, and pallidum (Moutsiana et al., 2015). Of these regions, the left interior frontal gyrus appears to be of particular importance, since disruption of this region using repetitive transcranial magnetic stimulation has been shown to dispel participants’ asymmetry in belief updating by increasing the extent to which participants incorporated unfavourable information into their beliefs (Sharot, Kanai, et al., 2012). The authors interpreted this as reflecting the role of the interior frontal gyrus in general inhibitory processing. This is also consistent with the evidence from neuroimaging studies reviewed in Section 2.1.1.1 which implicates IFG in Bayesian belief updating, and suggests that IFG activation may reflect the belief updating process more generally, for both symmetric and asymmetric belief updates.

In addition, complementary evidence from psychopharmacological studies has provided evidence that levels of the neuromodulator dopamine and the neuropeptide oxytocin may play a role in asymmetric belief updating. Sharot, Guitart-Masip, et al. (2012) demonstrated that administration of the dopamine precursor dihydroxy-L-phenylalanine (L-DOPA) exacerbated asymmetric belief updating in healthy adult participants relative to participants given the serotonin reuptake inhibitory citalopram. This effect was driven by a particularly pronounced reduction in belief updating among participants who received L-DOPA. Similarly, a recent study by Ma et al. (2016) reported increased belief updating in response to favourable news and decreased belief updating in response to unfavourable news following intranasal administration of oxytocin.

Several recent studies have suggested that asymmetric belief updating may be specifically related to beliefs about the self, and that this effect may be less robust or even reversed in other domains. Garrett and Sharot (2014) reported that no updating asymmetry was observed in a task where participants made estimates of the likelihood of adverse events in the general population, rather than for themselves personally. Kuzmanovic et al. (2015, 2016) replicated this finding, and also found that the self-specific belief updating asymmetry was only present in participants high in trait optimism. This self-specific effect is consistent with the framework proposed by Köszegi (2006), within which an optimistic belief updating asymmetry is proposed to
result from participants’ desire to maintain positive beliefs about their own abilities (termed ‘ego utility’). Convergent evidence from the behavioural economics literature also suggests that the asymmetric belief updating effect may be specific to beliefs regarding the self. Ertac (2011) found that belief updating following performance feedback was only asymmetric for feedback regarding one’s own performance, and not for the performance of others. Intriguingly, Ertac (2011) also found that participants displayed a pessimistic rather than an optimistic belief updating asymmetry, contrary to the results of studies using the belief formation updating task of Sharot et al. (2011). Coutts (2016) also reported a pessimistic belief updating bias in several tasks assessing simple probability estimations, suggesting that the valence of the belief updating asymmetry may depend on task-specific demands.

Interestingly, a recent commentary by Shah, Harris, Bird, Catmur, and Hahn (2016) has also suggested that the tendency toward optimistic belief updating reported by Sharot et al. (2011) may be an artefact of the particular statistical properties of the belief formation updating task. As such, caution is warranted in interpreting behavioural and neuroimaging findings derived from this task.

2.2.4 Confirmation Bias

The cognitive literature also describes one additional form of non-Bayesian inference relevant to belief updating: confirmation bias (Nickerson, 1998; Wason, 1960). Confirmation bias in belief updating refers to the finding that participants tend to interpret new information in a manner which supports their pre-existing beliefs (Kunda, 1990). For instance, Gilovich (1983) found that participants who had placed monetary bets on the outcomes of football games tended to accept wins at face value, but to discount and explain away losses. Similarly, Baron (1995) asked participants to make notes about the morality of early abortion before evaluating notes made by other hypothetical students, and found that participants tended to rate arguments which supported their own beliefs as stronger than arguments which did not.

However, interpreting the literature on confirmation bias with respect to belief updating is made difficult by several factors. Primary among these is that the term confirmation bias in fact refers to a number of distinct phenomena, only some of which are related to belief updating (see Hahn & Harris, 2014 for review). Indeed, Nickerson defines confirmation bias as “a generic concept that subsumes several more specific
ideas that connote the inappropriate bolstering of hypotheses or beliefs whose truth is in question” (1998; p. 175). For instance, although the term confirmation bias has sometimes been used to refer to the interpretation of new information to confirm pre-existing beliefs, as described above, one prominent line of research instead defines confirmation bias as a tendency to seek out confirmatory information (Cosmides, 1989; Wason, 1960). Studies adopting this definition of confirmation bias tend to leave unaddressed the mechanistic question of how beliefs are updated, and it is therefore difficult to interpret the findings of these studies in terms of belief updating. Furthermore, an additional difficulty for interpreting confirmation bias in terms of belief updating is that, even in those cases theoretically relevant to belief updating, the tasks used to investigate confirmation bias do not involve sequential belief revision. Instead, tasks used in this literature tend to involve giving subjective ratings of argument strength (Baron, 1995), to measure the weight placed on information in terms of the time spent talking about it (Gilovich, 1983), or to assess hypothetical willingness-to-pay for virtual stock options (Kuhn & Knutson, 2011). As such, beyond noting a general tendency for participants to place greater weight on information consistent with their prior beliefs, it is difficult to draw any quantifiable or substantive predictions regarding sequential belief updating from the confirmation bias literature.

2.2.5 Heuristics

Finally, in addition to the effects described in previous sections of conservatism, information order, asymmetry, and confirmation bias, a more general class of non-Bayesian belief updating effects derives from the interference of judgment heuristics. Heuristics are defined as “simple principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations” (Tversky & Kahneman, 1974, p. 1124), and are thought to provide simple ‘rule of thumb’ strategies which are computationally simple, but which avoid many of the cognitive demands of formal probabilistic inference (Goldstein & Gigerenzer, 2002; Kahneman & Tversky, 1972; Tversky & Kahneman, 1973, 1974).

The key insight for understanding the effects of heuristics on belief updating is drawn from theories of bounded rationality (see, e.g., Conlisk, 1996; Simon, 1956). Such theories seek to ground general principles of cognition in physiological facts concerning the resource limitations of the human brain:
“Because of the limits on their computing speeds and power, intelligent systems must use approximate methods to handle most tasks. Their rationality is bounded.” (Simon, 1990, p. 6)

One consequence of bounded rationality in judgment and decision making generally, and belief updating specifically, is that cognitive resources should be allocated so as to optimise decisions about behaviourally relevant outcomes, while minimising expenditure of cognitive resources on irrelevant or inconsequential tasks (Pitz & Sachs, 1984; Simon, 1956). By producing reasonably accurate choices while consuming relatively few cognitive resources, heuristics provide be a valuable tool to overcome cognitive resource constraints (Conlisk, 1996; Gigerenzer & Goldstein, 1996). Numerous different heuristics have been identified, including availability (Tversky & Kahneman, 1973), adjustment-and-anchoring (Tversky & Kahneman, 1974), recognition (Goldstein & Gigerenzer, 2002) and the take-the-best heuristic (Gigerenzer & Goldstein, 1996). Of particular importance for this review, two heuristics have been identified which interfere with human belief updating. These are the representativeness heuristic (Kahneman & Tversky, 1972) and the win-stay lose-shift heuristic (Robbins, 1952).

The representativeness heuristic dictates that the subjective likelihood of a sample is driven largely by the resemblance between the sample and the population from which it is drawn. Kahneman and Tversky (1972, pp. 447-448) illustrated this heuristic using a book-bag task similar to that devised by Phillips and Edwards (1966) to test conservatism. For instance, given two samples, one with five red chips and one black chip, and another with 15 red chips and 11 black chips, most participants reported stronger belief in a red majority given the first sample than given the second sample. However, for a binomial probabilistic inference task, the probability that one colour is in the majority depends only on the difference between the numbers of chips of each colour, and not on the relative proportions of each colour. As such, contrary to participants’ beliefs, both samples gave the same objective probability that red chips are in the majority in the book bag. This appears counter-intuitive, since in the first case there are five times as many red chips and black chips, whereas in the second case this ratio is approximately 1.36. Kahneman and Tversky identified this as an example of the representativeness heuristic, since participants made inferences on the basis of the resemblance between samples and the population, rather than the truly pertinent
statistical features of the data. The representativeness heuristic has been shown to interfere with belief updating even in simple contexts such as book-bag or urn-ball tasks (Achtziger et al., 2014; El-Gamal & Grether, 1995; Grether, 1980, 1992). For instance, Grether (1980) demonstrated that, consistent with the representativeness heuristic, when updating beliefs participants tended to over-weight the information provided by samples which resembled populations from which they might have been drawn. This result was replicated by Grether (1992), who also showed that the effect of representativeness was relatively invariant to the presence of monetary incentives. These results appear to be inconsistent not only with Bayesian belief updating, but also with the conservatism effect first described by Edwards (1968) and Phillips and Edwards (1966). However, the posited effects of representativeness and conservatism are not necessarily contradictory: a subsequent study by El-Gamal and Grether (1995) demonstrated that participants’ belief updates were best explained by a model in which both a representativeness effect and conservatism affected participants’ belief updates. As such, the effects of representativeness on belief updating appear to be complementary, rather than contradictory, to those of conservatism. Interestingly, the effect of the representativeness heuristic on belief updating is related to a proposal by Griffin and Tversky (1992) that participants place greater weight on the strength of evidence (defined as the proportion of evidence in favour of each hypothesis) than the weight of evidence (sample size). Kvam and Pleskac (2016) provided empirical evidence for this hypothesis from response time data, and concluded that “the accumulation of evidence for making judgments and decisions is not a true Bayesian updating process” (p. 177).

The second heuristic which has been identified as interfering with belief updating is the win-stay lose-shift heuristic (Robbins, 1952). This heuristic occurs in contexts in which participants make choices which receive reward or reinforcement, and refers to the finding that, rather than performing Bayesian belief updating, many participants simply re-select choice options which have previously led to reinforcement, or shift to a new option if a previous choice is not reinforced. Previous research has shown that, even in studies where optimal Bayesian models fit group-level data well, a substantial proportion of participants nevertheless made choices better explained by a win-stay lose-shift heuristic (Steyvers, Lee, & Wagenmakers, 2009). For this reason, one topic of recent research has been to investigate the ways in which participants manage a trade-off between Bayesian belief updating and a win-stay lose-
shift heuristic based on repeating choices which have previously led to reinforcement (Achtziger & Alós-Ferrer, 2013; Achtziger, Alós-Ferrer, Hügelschäfer, & Steinhauser, 2015; Charness, Karni, & Levin, 2007; Charness & Levin, 2005). A consistent finding of this research is that the win-stay lose-shift heuristic disrupts belief updating when the tendency to repeat a reinforced choice conflicts with the demands of Bayesian belief updating. For instance, one study by Achtziger et al. (2015) modified a classical book bag-style task such that, although participants were still required to guess which of two bags chips were being drawn from, the payoffs associated with the different bags were different. This created a situation in which Bayesian belief updating, which ought to be insensitive to the consequences of belief updates, could conflict with the win-stay lose-shift heuristic, which suggested that participants should update beliefs in line with past reinforcement. The authors observed that participants did indeed update beliefs according to this heuristic, resulting in more erroneous choices, and that this effect was more pronounced when incentives were higher.

Finally, a recent study by d'Acremont and Bossaerts (2016) also showed that, in the presence of leptokurtotic feedback noise (similar to Gaussian noise, but with a higher peak and broader tails), participants did not behave in accordance with a Bayes-optimal belief updating strategy. Instead, participants’ behaviour was better explained by a form of model-free reinforcement learning termed Contrarian Reinforcement Learning. This model assumed that participants repeated actions that had led to previous reinforcement, similar to the WSLS heuristic, but also assumed that participants adjusted their belief updating according to the degree of autocorrelation between sequential prediction errors. This result provides further evidence that non-Bayesian belief updating may occur following interference from competing heuristic strategies.

2.3 Rationale for Research Agenda

In this section, I will summarise the literature presented in the preceding sections of Chapter 2, in order to identify several important unresolved research questions concerning the neural and cognitive mechanisms of human belief updating.

The belief updating literature reviewed above draws a critical distinction between Bayesian and non-Bayesian belief updating. Bayesian belief updating is defined as updating that takes place according to rules of probabilistic inference set out
in Bayes’ Rule, and has been the dominant approach to the recent study of belief updating in economics and cognitive neuroscience. By contrast, non-Bayesian belief updating is defined as belief updating which takes place in a fashion inconsistent with Bayes’ Rule. Although non-Bayesian approaches to belief updating have historically been more strongly associated with research in cognitive psychology, there has also been a growing focus upon non-Bayesian belief updating in recent cognitive neuroscience research. This tension between Bayesian and non-Bayesian accounts of belief updating is central to the present thesis. Specifically, since the majority of research to date has adopted either a Bayesian or a non-Bayesian perspective on belief updating, the circumstances under which participants might select between Bayesian and non-Bayesian task strategies remains an open question. In addition, the neural substrates of belief updating in general, and of Bayesian versus non-Bayesian updating specifically, remain relatively poorly understood. The present thesis sought to investigate these questions.

The behavioural evidence in favour of Bayesian belief updating reviewed in this chapter consists of studies showing that, first, in some circumstances participants’ probabilistic inference closely follows rules of Bayesian inference, and, second, that computational models implementing Bayesian inference have been found to provide a good account of human behaviour in simple judgment and decision-making tasks. By contrast, evidence for non-Bayesian belief updating comprises documentation of a variety of empirical effects inconsistent with Bayesian inference. Four specific non-Bayesian belief updating effects were reviewed in this chapter: conservatism, order effects, asymmetric belief updating, and confirmation bias. In addition, a distinct non-Bayesian perspective on belief updating reviewed in this section was participants’ use of heuristic strategies in preference to Bayes-optimal belief updating. The use of heuristic strategies will be the primary applied focus of the present thesis, since this effect is most amenable to investigation using incentivised learning and decision-making tasks. A more general rationale for the use of such tasks to estimate participants’ beliefs in the present thesis is set out in Chapter 4.

According to one common framework for understanding heuristics, the use of non-Bayesian heuristic strategies—as well as the interference between these heuristics and Bayesian inference—can be accounted for by general principles of resource-rationality in cognition. This principle holds that agents ought to balance their use of different cognitive strategies according to the computational costs associated with
different behavioural strategies. This allows agents to optimise decisions about behaviourally relevant outcomes, while minimising expenditure of cognitive resources on irrelevant or inconsequential tasks. By producing reasonably accurate choices while consuming relatively few cognitive resources, heuristics such as win-stay lose-shift provide be a valuable tool to overcome cognitive resource constraints. An open research question is therefore the circumstances under which this effect occurs.

As well as this behavioural question, the neural substrates of human belief updating remain to be precisely characterised. Recent research has begun to investigate this question using neuroimaging techniques such as fMRI and EEG. Studies reviewed in this chapter using fMRI have suggested that a set of frontal cortical brain regions are likely to be involved in belief updating, specifically including the anterior cingulate cortex and bilateral inferior frontal gyri. In parallel, EEG research has implicated several event-related potential components in Bayesian belief updating, including both the P3 and the SPN. Moreover, convergent evidence has suggested the possibility that findings in these two neuroimaging modalities may reflect the same underlying belief updating mechanisms. In order to better understand the neural correlates of belief updating, therefore, Study 1 of the present thesis investigated the neural correlates of belief updating in a perceptual learning task using a single-trial regression approach to analysis of event-related potentials. This analysis was motivated by the limitations of a standard condition-based averaging approach for the analysis of EEG data when studying belief updating, and is introduced and detailed further in Chapter 4. Study 1 specifically sought to test the hypothesis that the P3 component of the event-related potential encoded the magnitude of Bayesian belief updates, as proposed by Kopp (2008). In addition, Study 1 performed an exploratory analysis to investigate the relationship between the SPN component of the event-related potential and belief uncertainty. In order to address the question of Bayesian versus non-Bayesian belief updating, Study 1 formulated two computational models implementing both a Bayesian belief updating strategy and a non-Bayesian heuristic task strategy, and formally compared these models’ goodness of fit to participants’ choice data.

Study 2 sought to follow up on the results of Study 1 by further investigating the conditions under which participants used Bayesian versus non-Bayesian belief updating strategies. It follows from the theoretical material set out in Section 2.2.5 that a key factor underlying the use of heuristics rather than Bayesian belief updating is motivation: since Bayesian inference is complex and computationally costly, the
selection of a Bayesian belief updating strategy is likely to depend on participants’ task motivation. The presence of rewarding incentives for good performance, for instance, may motivate participants to use computationally demanding task strategies involving Bayesian belief updating. On the other hand, in the absence of strong task motivation, it might be expected that individuals would demonstrate an increased reliance on non-Bayesian heuristic strategies. Study 2 investigated this question by introducing a feedback incentive manipulation (monetary versus purely instructive feedback) for the perceptual learning task developed within Study 1. Study 2 then assessed the effect of this manipulation both on participants’ behaviour, and upon the neural substrates of belief updating as manifested in EEG data. This question was addressed using model-based cluster analysis, an analytic technique which involves the use of computational cognitive models to subset participants into behaviourally distinct subgroups, and which is introduced further in Chapter 4 of the present thesis.
3. Review of Information Seeking Literature

Alongside belief updating, the second area of research interest for this thesis is information seeking. Both research topics can be conceptualised in terms of a belief-centric account of human judgment and decision making. However, whereas studies of belief updating investigate the cognitive processes by which agents use information to update their beliefs, the literature on information seeking is concerned with decision-makers’ attitudes toward information itself.

3.1 Overview and definitions

The attitudes of human decision-makers toward informative stimuli have been the subject of research in cognitive psychology since the cognitive revolution of the 1950s and 1960s (Edwards, 1962; Tversky & Edwards, 1966). Indeed, it has been argued that the cognitive revolution itself was touched off in part by psychologists’ attempts to apply the recently developed principles of information theory (Shannon, 1948; Shannon & Weaver, 1949) to questions regarding human behaviour (Barsalou, 1992; G. A. Miller, 2003). Although it is generally accepted that information theory does not provide a satisfactory general theory of cognition (see Luce, 2003), a key insight of this research program has informed much of subsequent cognitive psychology and cognitive neuroscience. This insight is that the human mind can be usefully conceptualised as an information-processing machine, and that formal mathematical tools such as those provided by information theory may have great utility in quantifying the functions carried out by the mind (Sternberg & Sternberg, 2012).

Moreover, it is also highly pertinent that information is necessarily defined with reference to the beliefs of some observer; this is true both in the colloquial definition of information as *processed data*, and in the formal definition from information theory (provided below). As such, there is a clear link between the information seeking literature reviewed in this chapter and the belief updating literature reviewed previously. Studies of belief updating consider how agents’ probabilistic models of the environment are revised on the basis of new information. However, in naturalistic environments this information is rarely acquired passively; instead it is often necessary for agents to actively acquire information with which to update their beliefs (a lost tourist, for instance, may ask a passer-by for directions). The information seeking
literature is concerned with how and why agents go about acquiring this new information.

This thesis adopts the formal mathematical definition of information provided by information theory (Shannon, 1948). This is distinct in important respects from the colloquial definition of information as meaning any form of processed data (as in “information technology”, for instance). In information theory, the information content of a stimulus is defined by the effect of that stimulus on the uncertainty of an observer’s beliefs. Formally, this follows from a quantification of uncertainty as entropy, a concept borrowed by analogy from thermodynamics. Formally, the entropy of a probability mass function $P(X)$ is denoted $H(X)$, and can be calculated as follows:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$$

Where $n$ refers to the number of possible discrete values that the function $P(X)$ can take. The entropy of a probability distribution is maximised when all outcome values are equally likely, representing maximal uncertainty (for instance, a uniform distribution over possible outcome values for a bounded random variable, or a success probability of 0.5 for a binomial random variable). The information content of a stimulus can then be formally defined as the mutual information $I$ of prior and posterior beliefs, quantified as the reduction in the entropy of a probabilistic belief distribution from prior beliefs to posterior beliefs upon observation of the stimulus:

$$I = H(X)_{\text{posterior}} - H(X)_{\text{prior}}$$

Under this definition, positive values of $I$ are possessed by stimuli which increase overall confidence (that is, reduce the entropy of beliefs), whereas stimuli which reduce overall confidence (increase the entropy of beliefs) are associated with negative values of $I$. An important consequence of this specification is that information is inherently defined with respect to the uncertainty of beliefs, such that the concepts of uncertainty and of information are not mathematically separable. It should also be noted that the definition of uncertainty adopted in this thesis refers to the entropy of participants’ beliefs. This is distinct from the economic concept of Knightian uncertainty, which is defined as immeasurable and irreducible, and which occurs when possible future
outcomes are unknown or to which it is impossible to assign a probability (Knight, 1921).

Adopting this definition has had important implications for recent research into human attitudes toward information. Since information is only uniquely defined with respect to some observer’s uncertainty, information in the formal sense of the term is generally understood subjectively, and is intimately linked to the beliefs of the observer. The exception to this generalisation is in the case of the application of the equations above to probabilities defined by objective frequency counts; for instance, Shannon (1951) calculated the entropy of printed English based on the relative frequency with which different letters follow one another across the entire language. However, even in this case it might be argued that the objective frequency counts correspond to the letter transition probabilities learned by native speakers of English. In studies of judgment and decision making, information has generally been defined as the reduction of uncertainty in an agent’s internal model of behaviourally relevant aspects of the environment. So, for instance, information may take the form of a predictive signal about the likelihood of an upcoming reward (e.g., Bromberg-Martin & Hikosaka, 2009; Preuschoff et al., 2008), or feedback enabling a participant to reduce their uncertainty about the reward contingencies associated with different actions (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Speekenbrink & Konstantinidis, 2015).

Information is vitally important in making decisions: well-informed decisions are associated with better overall outcomes on average than uninformed decisions. This can be rephrased in terms of beliefs to state that the closer the match between an agent’s beliefs regarding the environment and the true state of the environment, the more likely that decisions based on the agent’s beliefs will result in desired outcomes. In a sense, this principle can be considered an example of the good regulator theorem from control theory, which states that any good regulator of a system must be a model of that system (Conant & Ashby, 1970). The value of information is also clearly demonstrated in applied domains; for instance, well-informed animals make more effectively foraging decisions (Piorli & Card, 1999; Reebs, 2000), and well-informed investors are better able to maximise their monetary profits (Hirshleifer & Riley, 1979).

The practical utility of information is encapsulated by theories seeking to specify when and how agents ought to seek out information. A normative interpretation of economic decision analysis, for instance, prescribes that agents should seek information when its resolution of uncertainty will increase overall expected payoffs.
(Clemen, 1996; Howard, 1966; Raiffa & Schlaifer, 1961). In cognitive neuroscience, recent theories of active inference have proposed that, as well as hedonic reward value, stimuli should be assigned an epistemic value based on their information content (Friston et al., 2015; Friston et al., 2013). The reason for this is that, given the complexity of the natural environment, information is often useful in the long term even when its short-term utility is unclear. However, in spite of the apparent agreement between theories that information possesses value, a key matter of disagreement between theories is what kind of value information decision-makers assign to information, and what the cognitive source of this value might be. The next section provides an in-depth examination of the assumptions and behavioural consequences of different theoretical conceptualisations of the value of information.

3.2. Theories of information valuation

A key theoretical distinction can be drawn in information seeking research between the instrumental and the intrinsic value of information. Theories espousing instrumental valuation of information, such as decision analysis, hold that the value of information ought to be determined solely on the basis of the effect of possessing information on future payoffs. By contrast, theories such as active inference hold that as well as instrumental value, information ought to be treated as though it were of intrinsic value. In this dichotomy, instrumental value is defined as the usefulness of information with respect to acquiring direct, tangible payoffs such as food or money (Lanzetta & Driscoll, 1966). By contrast, intrinsic valuation of information involves attitudes toward the resolution of uncertainty itself, independent of whether resolving uncertainty is associated with better or worse overall outcomes (Grant et al., 1998; Grant, Kajii, & Polak, 2000; Kreps & Porteus, 1978). In part, this theoretical discrepancy may be attributable to the differing aims of different theories of information valuation. A key divide is between prescriptive theories of information valuation, which seek to lay out principles guiding how agents ought to assign a value to information, and descriptive theories, which seek to account for actual observed patterns of behaviour (in economics, a comparable distinction is sometimes drawn between normative and positive theories; see Loasby, 1976). There is a general trend for prescriptive theories to adopt instrumental accounts of information valuation, and for descriptive theories to adopt intrinsic perspective. This distinction is not complete,
however; for instance, it can be argued that the theory of active inference (Friston et al., 2015; Friston et al., 2013) is a prescriptive theory which posits an intrinsic value of information. Moreover, it is important to make clear that the distinction between intrinsic and instrumental valuation of information for economic theories represents a claim about the components of outcomes which are considered valuable, and not necessarily a claim about the nature of utility itself. As such, from the perspective of economics the distinction between intrinsic and instrumental valuation of information would be seen to be procedural, rather than structural.

More generally, descriptive theories of information valuation have been motivated by a desire to account for empirical findings which are inconsistent with normative prescriptions. Before such findings are discussed in detail, however I will first review and provide a detailed overview of different instrumental and intrinsic theories of information valuation.

3.2.1 Theories positing an instrumental value of information

This section discusses theories positing a solely instrumental value of information. Such theories have been generally associated with normative approaches in the disciplines of economics, finance, and stochastic control theory. Although a full account of all theories of instrumental information valuation in these fields is beyond the scope of this review (see Hirshleifer & Riley, 1979; D. B. Lawrence, 1999 for comprehensive surveys of theories of information valuation in these fields), it is nevertheless instructive to identify two influential perspectives espousing instrumental valuation of information. These are statistical decision analysis (Clemen, 1996; Howard, 1966; LaValle, 1978; Raiffa & Schlaifer, 1961) and information valuation in the presence of attitudes toward risk (D. B. Lawrence, 1999, ch. 7).

Statistical decision analysis (Clemen, 1996) is an area of economic theory which uses quantitative analysis of stochastic processes to facilitate good decision making under uncertainty. Decision analysis aims to define metrics which economic agents can use to help aid their decision making; as such, decision analysis can be considered a prescriptive theory of decision making. This is a broad area of research, including many other topics in addition to the valuation of information. Nevertheless, since one frequently useful action in the presence of uncertainty is to gather information, decision
analysis has sought to identify and characterise the circumstances under which information should be sought out.

In defining the (instrumental) value of information, decision analysis defines two metrics: the expected value of perfect information (EVPI) and the expected value of sample information (EVSI). The EVPI is defined as the increase in a rational agent’s expected monetary payoffs if the agent were to have perfect information about what the future state of the world is to be (Hammitt & Shlyakhter, 1999; Raiffa & Schlaifer, 1961). For instance, the EVPI of knowing the winning numbers of a one-thousand-dollar lottery is one thousand dollars, since an agent who knew the winning numbers in advance would be able to select these numbers, and thereby win the lottery. In applied settings, the EVPI is rarely directly relevant, since perfect information about the future state of the world is extremely difficult to obtain. However, the EVPI does provide an upper bound on the (instrumental) value of information in decision analysis. Notably, this value is solely instrumental, since the value of information is only defined in terms of the change in expected payoffs which is associated with the possession of that information. Decision analysis therefore also uses a more realistic measure, the EVSI, for cases in which perfect information is not available. The EVSI is calculated by integrating over all possible outcomes of an information-seeking action, and is equal to the probability-weighted sum of the change in payoffs which would result from making the best possible decision following each possible outcome of the informative test. Decision analysis then prescribes that information should be acquired if the cost of obtaining information is less than the EVSI (Stigler, 1961).

There are several notable shortcomings to the use of EVSI as a measure of information value. First, although easily calculable for simple examples, such as a test with only two outcomes, the calculation of the EVSI may be computationally intractable in cases where there are many possible outcomes of an informative test. Secondly, and most notably, the EVSI is calculated according to a probability-weighted sum over possible test outcomes, where probability weights are derived from the agent’s subjective prior probabilities of each possible outcome of the test. As such the EVSI is not an objective quantity, but is belief-dependent. This belief-dependence of the EVSI means that the usefulness of the EVSI as a measure is dependent on having well-calibrated prior beliefs. In the case of poorly calibrated prior beliefs, the EVSI may be systematically biased: for instance, Hammitt and Shlyakhter (1999) detail how
overconfident prior beliefs may result in an estimate of the EVSI which underestimates the true instrumental value of information.

Finally, the quantities of EVPI and EVSI as defined by decision analysis assume a linear payoff function for both gains and losses. That is, these quantities assume, first, that the utility of a monetary gain increases linearly with monetary value and, second, that gains and losses are weighted equally by agents. Economic research into risk aversion and prospect theory has demonstrated that both of these assumptions are likely to be untenable for describing the preferences of human decision makers (Kahneman & Tversky, 1979; Von Neumann & Morgenstern, 1944). As a result, the calculations of EVPI and EVSI as defined by classical statistical decision analysis (Howard, 1966; Raiffa & Schlaifer, 1961) fail to take into account important features of human preferences regarding risk and loss. D. B. Lawrence (1999) provides a generalised decision-theoretic account of the instrumental value of information in the presence of loss aversion and risk attitudes. In the case of risk aversion, for instance, information may possess a value to a risk-averse agent in excess of its EVSI simply because acquiring information results in reduced variability of outcomes, and therefore reduced risk (LaValle, 1978). However, as noted by D. B. Lawrence (1999), risk aversion can only account for this information premium when it is possible for an agent to make a decision on the basis of the information thus acquired. As demonstrated by the literature reviewed in Section 3.3.2, behavioural evidence suggests that human decision makers have preferences even with respect to information that cannot be used for a decision.

3.2.2 Theories positing an intrinsic value of information

In contrast to instrumental theories of information valuation, intrinsic valuation theories across neuroscience, psychology and economics posit that agents have preferences with respect to information itself (or should have such a preference, in the case of prescriptive theories). This intrinsic valuation is independent of the instrumental utility of the information for acquiring tangible rewards such as food, water, or monetary reinforcement.

Economic theories espousing intrinsic valuation of information are built upon the foundation of the instrumental theories of information valuation summarised in the previous section. Intrinsic theories expand upon this foundation by proposing that agents have additional preferences with respect to aspects of decision options other than
solely mean and variance of payoffs. This may include a preference for early resolution of uncertainty, as posited by Kreps and Porteus (1978), an intrinsic preference for information (Grant et al., 1998, 2000), or anticipatory preferences with respect to future events (Caplin & Leahy, 2001). The precise source of information’s posited intrinsic value differs between theories; however, the crucial feature of each theory is that the resulting preference for information is necessarily non-instrumental, since information’s value derives from the resolution of uncertainty itself, independent of the payoffs which result from the resolution of uncertainty.

Kreps and Porteus (1978) posit that agents may prefer to resolve their uncertainty about future states of the world earlier rather than later, as a result of an axiomatic specification of utility termed recursive expected utility (see also Epstein & Zin, 1989). Such a preference for early resolution of uncertainty induces an intrinsic valuation with respect to information because of the mathematical relationship between information and uncertainty described in Section 3.1: agents who wish to resolve their uncertainty early will naturally place a positive value on informative stimuli, which resolve uncertainty. A similar intuition is formalised in the theoretical framework proposed by Grant et al. (1998), which posits an intrinsic preference for information itself. Other theorists have speculated that intrinsic valuation of information might each result from an aversion to uncertainty itself (Caplin & Leahy, 2001; Epstein, 2008). This would ground axiomatic economic theories of information valuation in a psychologically plausible mechanism: if a decision-maker found uncertainty itself to be aversive, information would provide negative reinforcement by relieving the disutility associated with being in an uncertain state (cf. Herrnstein & Hineline, 1966).

Similarly, physiological principles are at the heart of active inference theory (Friston et al., 2016; Friston et al., 2015; Friston et al., 2013), which also posits an intrinsic value of information. In contrast to the axiomatic economic theories described above, however, the intrinsic value of information in active inference is derived from an application of the “free energy principle” (Friston, 2009, 2010; Friston & Kiebel, 2009; Friston, Kilner, & Harrison, 2006) to decision making under uncertainty. Active inference is a Bayesian theory of central nervous system function (cf. Knill & Pouget, 2004) in which cognition is considered an approximate Bayesian inference scheme based upon the minimisation of an information-theoretic quantity termed free energy. This statistically complex quantity can, in this context, be thought of as a measure of the discrepancy between an agent’s preferred environmental outcomes and its long-
term beliefs about the likely causes of those outcomes, given a particular model of how environmental outcomes are generated. This latter point is particularly important since it implies that all belief updating performed by agents is Bayes-optimal, subject to a particular set of prior beliefs (Brown, 1981). As detailed in Section 2.1.2, this assumption has been contested in other contexts.

Friston et al. (2015) details two ways in which agents behaving according to theories of active inference can minimise free energy: first, by maximising the probability of occurrence of preferred outcomes, and, second, by minimising the expected uncertainty of future beliefs (cf. Schwartenbeck et al., 2015). In practical terms, these two methods of minimising free energy theoretically respectively correspond to extrinsic reward (seeking tangible rewards such as food or money) and epistemic reward, or the resolution of environmental uncertainty. This epistemic reward is therefore a form of intrinsic valuation of information. In this sense, active inference theory is a prescriptive theory of cognition, since it proposes that in the presence of uncertainty regarding the causes of observations in its environment, an agent will derive long-term value from reducing its uncertainty regarding those causes. Put another way: since the long-term stability of biological agents is dependent upon a good match between an agent’s beliefs and the true generative model of the environment, it is adaptive for agents to seek out information with which to update their beliefs, even if the tangible gains associated with this belief updating are not immediately clear. As such, active inference entails intrinsic valuation of information.

Finally, although the intrinsic valuation theories discussed heretofore have generally focused upon an appetitive value of information, some theories have also emphasised the possibility that agents may exhibit an aversion to information, even when there is no cost associated with gathering that information (Andries & Haddad, 2015; Barigozzi & Levaggi, 2010; Lazrak, 2004). For instance, it is commonly observed that some individuals avoid medical tests which would reveal whether they have a serious medical condition such as Huntington’s Disease (Sweeny, Melnyk, Miller, & Shepperd, 2010), in spite of the fact that having this information is associated with better long-term outcomes. As such, aversion to information is necessarily a form of (negative) intrinsic valuation of information. To demonstrate this, consider that instrumental valuation of information involves agents who, given a choice between more information and less, either prefer more information to less (if the information is instrumentally useful), or are indifferent between more and less information (if the
information is non-instrumental). Under this framework, there is no mechanism by which agents can prefer less information to more; indeed, Blackwell’s theorem (Blackwell, 1953) proves that an agent that seeks to maximise expected future reward must in all circumstances weakly prefer more information to less. An agent who prefers a less informative signal to a more informative signal must therefore be responding to some non-instrumental feature of the information, and can therefore be characterised as exhibiting an intrinsic valuation of information.

It is interesting to note that aversion to information can be accounted for by economic theories positing intrinsic valuation of information such as that of Kreps and Porteus (1978), or that of Grant et al. (1998). These theories can account for information aversion by means of an inversion of the standard formulation: a preference for late resolution of uncertainty or an intrinsic aversion to information, respectively. By contrast, it is not immediately clear how aversion to information can be explained within active inference theory. In addition, it is important to note that the aversion to information described by Sweeny et al. (2010) has only been observed with respect to information about unpleasant future consequences, and might therefore be related to a prospective aversive response to the potential for this negative outcome to occur. Because of this potential confound, all studies of information valuation in the present study used a task in which participants were offered early information about either positive or neutral valenced outcomes, but not outcomes of a negative valence.

3.3 Empirical studies of information seeking and information valuation

The next section of this literature review summarises empirical findings regarding information seeking and information valuation within psychology, neuroscience and behavioural economics. In doing so, it is crucial to distinguish between two different types of information: instrumental information and non-instrumental information (Lanzetta & Driscoll, 1966; Wolosin & Wolosin, 1967). These two types of information are strongly related to the two theories introduced in Section 3.2, of instrumental and intrinsic valuation of information respectively. Instrumental information is information which can be used (or can potentially be used) to aid decision making. According to this definition, therefore, the large majority of all information available in natural settings is instrumental information. By contrast, non-instrumental information, defined as information which cannot be used to aid any decision or condition any behaviour,
occurs relatively infrequently in naturalistic environments. Nevertheless, non-instrumental information is of particular importance for assessing intrinsic valuation of information in experimental settings. This is because if information is non-instrumental, any preference observed with respect to an informative stimulus must be the result of intrinsic valuation of information.

I review psychological and neuroscientific findings pertaining to these two different types of information separately below.

3.3.1 Instrumental information

One of the primary goals of research into the valuation of instrumental information is to determine whether the value placed upon such information is commensurate with the true statistical value of the information. To do this, a number of different tasks have been designed in which participants choose between either gathering or forgoing information which is useful for a future decision. It is then possible to identify the factors which affect participants’ willingness to seek information, as well as to assess whether patterns of information gathering meet standards of optimality. One common approach has been to place participants in a situation where acquiring information is associated with a cost, and then quantify participants’ willingness-to-pay for information. Two primary tasks have been used in this research program: the “observe or bet” task (Navarro, Newell, & Schulze, 2016; Rakow, Newell, & Zougkou, 2010; Tversky & Edwards, 1966), and optional stopping tasks (Averbeck, 2015; Busemeyer & Rapoport, 1988; Edwards, 1965; Juni, Gureckis, & Maloney, 2016; Rapoport & Burkheimer, 1971; Rapoport & Tversky, 1970).

Observe or bet tasks are prediction tasks in which the act of information acquisition is explicitly separated from, and traded off against, the prediction of future outcomes. By observing participants’ relative preference for acquiring information versus predicting outcomes, it is possible to derive a proxy measure of participants’ preference for instrumental information. For instance, in the version of the observe-or-bet task introduced by Tversky and Edwards (1966), participants were presented with a ‘machine’ with a left light globe and right light globe, only one of which could be illuminated at any time. Participants were informed that their task was to predict which light globe would be illuminated on each trial, and were presented with two choice options on each trial: they could either place a bet as to which of the two lights would
be illuminated, or they could observe which light was illuminated without placing a bet (the information-seeking choice). Crucially, if participants chose to place a bet, they would earn money in the case of a correct prediction but would not actually observe the outcome itself. This design feature explicitly decorrelated the instrumental information value and the reward value of feedback, since on each trial it was possible to have either information or monetary reward, but not both. Tversky and Edwards (1966) found two primary behavioural effects using this task: first, that participants over-sampled information prior to betting for the first time and, second, that participants who were informed that illumination probabilities might change over time chose to observe significantly more often than participants who were informed that illumination probabilities would remain stationary. The former result can be attributed to an excessively high information value relative to the optimal policy. The latter result was interpreted as reflecting the fact that information is of greater instrumental value in relatively more complex environments than in a less complex environment such as the case of stationary illumination probability. More recently, using the same task, Rakow et al. (2010) found that participants who displayed a valuation of information closer to the optimal value tended to be those with a higher working memory capacity. This suggests that the excessive instrumental valuation of information first reported by Tversky and Edwards (1966) may have been associated with a decreased ability to incorporate previous information into beliefs among some participants. More recently, Navarro et al. (2016) found that behaviour on this task was closer to the predictions of an optimal strategy than suggested by Tversky and Edwards (1966), and suggested that performance on the task was crucially dependent on the ability to learn the best task strategy over multiple iterations of the task. Navarro et al. (2016) found that, when participants were permitted to perform the observe-or-bet task multiple times, participants increasingly adopted the optimal task strategy, and therefore the optimal valuation of information, with greater experience on the task.

The second type of task used to assess the trade-off between information and reward is the optional stopping task. Like the observe-or-bet task, optional stopping tasks offer participants a choice between seeking instrumental information and seeking monetary reward. However, whereas the cost of information in observe-or-bet tasks is an opportunity cost, in optional stopping tasks the cost of acquiring information is made explicit, and participants may then choose how many times they pay this cost prior to making a decision. Research using optional stopping tasks has produced results
somewhat at odds with research using observe-or-bet tasks, in spite of the tasks’ putative design similarities. For instance, in contrast to the excessive valuation of instrumental information reported in an observe-or-bet task by Tversky and Edwards (1966), results from early research with optional stopping tasks suggested that, where participants’ strategies deviated from optimality, this was because participants assigned a smaller value to instrumental information than was optimal (Rapoport & Tversky, 1970). Moreover, a descriptive modelling study by Busemeyer and Rapoport (1988) found that participants’ information-seeking behaviour on an optional stopping task was best explained by a ‘myopic decision rule’ model. In this model, participants sample information in a short-sighted fashion, until whereby participants continue to purchase information until the expected loss of their decision after $n$ pieces of information is less than or equal to the sum of the costs of purchasing $n + 1$ pieces of information. Such a strategy is short-sighted because it involves looking only one observation ahead at each point in time, rather than optimising behaviour with respect to all possible future observations. More recently, Juni et al. (2016) used a modified optional stopping task in which participants attempted to estimate the mean of a bivariate Gaussian distribution on the basis of a number of sequential observations, and found no systematic deviations from optimality, with over-sampling occurring in some conditions and under-sampling occurring in different conditions. The authors reported that deviations from optimality were more closely related to the variability of observed samples, with participants sensitive to this sample dispersion even though the optimal model was not.

Another form of optional stopping task frequently used to assess preferences for acquiring instrumental information is the Information Sampling Task (IST) of Clark, Robbins, Ersche, and Sahakian (2006). In contrast to the optional stopping and observe-or-bet tasks described above, which have generally assessed healthy participant samples, the IST has generally been used in clinical samples to investigate a psychological construct termed reflection impulsivity. Reflection impulsivity is defined as the tendency to gather and evaluate information before making a decision (Kagan, 1966); this construct is therefore highly relevant to the question of how information is valued. In the IST, participants are presented with an array of 25 closed boxes on a $5 \times 5$ grid. When opened, each box can be one of two colours. Participants’ task is to predict which of the two colours will be in the majority after all boxes in the grid are opened.
Crucially, participants are permitted to sample information by opening boxes prior to making a prediction; the IST’s two primary outcome measures are (a) the number of boxes opened by the participant prior to making a decision, and (b) the probability that the participant’s choice was correct, based on the colours of the boxes opened at the time of response, denoted $P(\text{correct})$. Of the two measures, the $P(\text{correct})$ measure is thought to provide the more sensitive measure of participants’ reflection impulsivity, since a particular number of boxes opened may be associated with quite different levels of uncertainty depending on the colours of the open boxes. For instance, ten open boxes might consist of nine blue boxes and one yellow box, or five blue boxes and five yellow boxes; these two scenarios have the same number of open boxes but very different values of $P(\text{correct})$. The majority of published studies using the IST report $P(\text{correct})$ as a primary outcome measure (e.g. Clark, Roiser, Robbins, & Sahakian, 2008; Crockett, Clark, Smillie, & Robbins, 2012; Passetti, Clark, Mehta, Joyce, & King, 2008; Solowij et al., 2012).

In spite of the fact that the IST can be considered a form of optional stopping task, and can therefore be used to measure preference for information, most research using the IST has not addressed the question of information valuation directly. Instead, the focus of this research has been upon assessing deficits in reflection impulsivity in different clinical populations and under different psychopharmacological manipulations. Disorders of reflection impulsivity have been reported in clinical conditions including substance use disorder (Clark et al., 2006), Alzheimer’s Disease (Zamarian, Benke, Brand, Djamshidian, & Delazer, 2015), major depressive disorder (Tavares et al., 2007), obsessive-compulsive disorder (Chamberlain, Fineberg, et al., 2007), and problem gambling (A. J. Lawrence, Luty, Bogdan, Sahakian, & Clark, 2009). In addition, the IST has been used alongside other neurocognitive measures to assess the psychological effects of pharmacological agents such as buspirone (Chamberlain, Müller, et al., 2007), methylphenidate (DeVito et al., 2009) and modafinil (Scoriels, Barnett, Soma, Sahakian, & Jones, 2012).

Finally, the valuation of instrumental information has also been assessed indirectly by a body of research in neuroscience and psychology concerning the exploration/exploitation dilemma (J. D. Cohen, McClure, & Yu, 2007). This dilemma concerns the fact that, in uncertain and dynamic natural environments, biological agents face a complex trade-off between investigating their environment, so as to learn its structure and be able to make better decisions in future (exploration), and choosing the
best available choice option on the basis of their current knowledge (exploitation). The exploration-exploitation dilemma can therefore be conceptualised as a trade-off between information-seeking (exploratory) and reward-seeking (exploitative) actions, similar to observe-or-bet and optional stopping tasks. Unlike the research reviewed above, however, research into the exploration-exploitation dilemma has tended to assess behaviour using variants of a behavioural paradigm termed the $k$-armed bandit task. In this task, participant choose on each trial between $k$ different choice options, each associated with a different payout amount or payout probability. In the classical version of this task formalised by Gittins (1979), the payout probability of each choice option can change over time, but only when the option is chosen. In another common variant of this paradigm termed the restless $k$-armed bandit task, payout amounts change each trial, even for options that are not selected (Daw et al., 2006; Jepma & Nieuwenhuis, 2011; Speekenbrink & Konstantinidis, 2015). Such tasks present participants with a tension between exploratory actions (choosing different options to determining which provides the greatest payout at any point in time) and exploitative actions (choosing options with little uncertainty and known payouts).

From the perspective of instrumental versus intrinsic valuation of information, a key question addressed by exploration-exploitation research is whether human participants behave as though they are only interested in maximising their overall expected reward, or whether they also seek information for its own sake, independent of the tangible value associated with this knowledge. In a restless $k$-armed bandit task, this question can be answered by investigating whether participants act solely to maximise their payouts (corresponding to solely instrumental valuation of information), or whether they also behave as though they derive some value from reducing their uncertainty about different choice options (corresponding to intrinsic valuation of information). This literature provides somewhat inconsistent findings with respect to this question: whereas Daw et al. (2006) found no evidence that preference for different choice options was affected by an ‘uncertainty bonus’ proportional to options’ uncertainty, Speekenbrink and Konstantinidis (2015) reported that computational models including an uncertainty bonus substantially outperformed models without an uncertainty bonus. Although both of these studies used restless $k$-armed bandit tasks, it is possible that subtle differences in the statistical process generating payouts between studies may be responsible for these discrepant results. Specifically, Daw et al. (2006) generated choice payouts according to a random walk.
function with a drift term that decreased the value of high-paying options and increased the value of low-paying options over time. By contrast, Speekenbrink and Konstandinidis (2015) used a simpler Gaussian random walk process, with no drift term, such that low-paying options did not naturally increase in value over time. This difference may have resulted in different conclusions regarding participants’ preferences with respect to information: since Daw et al. (2006) assumed that participants knew that low-paying options would appreciate over time, an additional uncertainty bonus may have had little explanatory value. This discrepancy demonstrates the importance of carefully considering how assumptions about participants’ beliefs might affect calculations of the value of information. This point is also underscored in active inference theory (see Section 3.2) and by critiques of Bayesian models of belief updating (see Section 2.1.2).

Neuroimaging and psychopharmacological studies have also used k-armed bandit tasks to identify the neural processes which facilitate exploratory behaviour. The study by Daw and colleagues (2006) discussed above also measured brain activation with fMRI while participants were performing the bandit task. The authors found that, relative to exploitative choices, exploratory choices were associated with increased BOLD signal in bilateral frontopolar cortex. The authors hypothesised that, given the involvement of this region in cognitive and behavioural control (Koechlin, Basso, Pietrini, Panzer, & Grafman, 1999; E. K. Miller & Cohen, 2001; Ramnani & Owen, 2004), this activation was likely to be reflective of a role for top-down cognitive control in disengaging from an exploitative task strategy to engage in exploratory behaviour. Other research in cognitive neuroscience has pointed toward the neurotransmitter norepinephrine (NE) as crucially involved in managing the exploration-exploitation trade-off in humans. Aston-Jones and Cohen (2005) proposed an ‘adaptive gain’ theory of norepinephrinergic function in decision making, whereby NE neurons in the locus coeruleus (LC) promote task engagement by switching between two modes of firing: phasic (high firing rates for brief periods) and tonic (lower but more sustained neural firing). The adaptive gain theory posits that phasic LC-NE activity is involved in optimisation of task performance by maintaining attention and focus, and tonic LC-NE is associated with strategic disengagement and mind-wandering. In the context of the exploration-exploitation dilemma, therefore, phasic LC-NE activity is posited to map onto exploitative behaviour, whereas the disengagement associated with tonic LC-NE activity is theorised by Aston-Jones and Cohen (2005) to trigger more stochastic and
exploratory choice behaviour. McClure, Gilzenrat, and Cohen (2006) further proposed that this LC-NE system might interact with dopaminergic activity, such that dopaminergic activity controls learning via its effects on synaptic plasticity, whereas LC-NE firing provides higher-level adaptive control of behaviour in a manner consistent with the adaptive gain theory. Empirical support for this model was provided by Jepma and Nieuwenhuis (2011), who used pupillometry during a k-armed bandit task to demonstrate that pupil diameter, a putative index of LC-NE firing rates, tracked shifts between exploratory and exploitative choice modes.

3.3.2 Non-instrumental information

Along with the research summarised in the previous section into the valuation of instrumental information, a somewhat smaller body of research has investigated how humans and other animals assign value to non-instrumental information. To review, non-instrumental information is information upon which no behaviour can be usefully conditioned. As such, non-instrumental information can be thought of as having no practical use with respect to outcomes other than the resolution of uncertainty itself, since acquiring such information does not result in any change with respect to tangible expected outcomes. For instance, one example of non-instrumental information might be the winning numbers of a lottery for which one cannot purchase a ticket: such information is non-instrumental because no practical change in behaviour can possibly result from possessing this information.

A number of studies in psychology and cognitive neuroscience have investigated demand for non-instrumental information in human participants. In general, such research involves presenting participants with real or hypothetical scenarios in which non-instrumental information is available upon request, and then assessing the strength of participants’ preferences for this information. For example, a classic study by Lanzetta and Driscoll (1966) gave participants a task in which they could receive either ten cents reward, a mild but aversive electrical shock to the calf, or nothing on each trial. When participants were given the option to find out in advance which of these outcomes would take place, participants consistently expressed a preference for acquiring this information, despite the fact that the information did not alter the outcome in any way, and was therefore non-instrumental. Interestingly, participants displayed strong individual differences in the proportion of trials in which information
was sought (range: 5.56 to 98.89 per cent). However, overall preference for information did not vary depending on the exact configuration of possible outcomes available on each trial, with the authors reporting no significant differences in information preference between a reward/no reward condition, a shock/no shock condition, and a reward/shock condition (Lanzetta and Driscoll, 1966). This suggests that the observed behaviour related to participants’ uncertainty itself, rather than to the outcomes about which participants were uncertain. Similar results were reported in a recent study by Pierson and Goodman (2014), who used a survey measure to investigate participants’ preference for acquiring non-instrumental information. The authors assessed the effect of varying outcome probability on participants’ preference for acquiring non-instrumental information. Interestingly, the authors found that preference for non-instrumental information increased as the probability of a positive outcome increased if information was free, whereas preference for information was strongest for an intermediate probability of winning (50 per cent) for information with a non-zero cost. Similarly, it has also recently been shown that preference for non-instrumental information increases with increasing duration of uncertainty (Iigaya, Story, Kurth-Nelson, Dolan, & Dayan, 2016).

As well as these studies with human participants, the non-instrumental information valuation research also includes a number of studies conducted with non-human primates and other animals as subjects. This stands in contrast to the instrumental information valuation studies reviewed above, which almost entirely comprised experiments with human participants.

The animal research literature has investigated the valuation of non-instrumental information using a specific variant of the task first tested in humans by Lanzetta and Driscoll (1966). In this non-instrumental information sampling task, as adapted by Spetch, Belke, Barnet, Dunn, and Pierce (1990), the animal is placed in an environment in which a primary reward—such as water, fruit juice, or grain pellets—is delivered probabilistically. On each trial, the animal may either receive or not receive the reward after some delay period; typically, there is an equal probability of receiving and of not receiving the reward. Then, in the delay prior to the delivery of this reward, the animal may view either an informative or a non-informative stimulus. In birds, this action is performed by pecking a response key (Gipson, Alessandri, Miller, & Zentall, 2009; Spetch et al., 1990; Vasconcelos et al., 2015; Zentall & Stagner, 2011); in macaque monkeys, this choice is made via visual saccade (Blanchard et al., 2015; Bromberg-
In this task context, it has been observed that pigeons, starlings and macaque monkeys each display a clear preference for observing the informative stimulus over the non-informative stimulus despite the fact that this information is non-instrumental (Bromberg-Martin & Hikosaka, 2009, 2011; Vasconcelos et al., 2015; Zentall & Stagner, 2011). Moreover, when a cost is placed on the informative stimulus and animals are given the option of sacrificing part of the reward in exchange for early information, both birds and monkeys display a willingness to do so. This suggests that the value of the non-instrumental information is not only metaphorical, but that animals directly ascribe a value to information and are willing to sacrifice reward to acquire it. The results summarised from previous studies above about intrinsic valuation of non-instrumental information by humans predict that human participants should display a similar willingness to pay for non-instrumental information (Niv & Chan, 2011). However, this question has not been assessed to date.

Moreover, single-neuron recordings from monkeys performing this task have provided evidence that non-instrumental information may share neural processing substrates with primary reward in non-human primates. Bromberg-Martin and Hikosaka (2009) found that midbrain dopamine neurons typically associated with the processing of primary rewards such as food and water, also encoded the expected value of information during the delay prior to the receipt of reward. This encoding covaried across different individuals proportional to the strength of their preference for information. Moreover, a subsequent study showed that another vital component of information, an information prediction error (IPE), was encoded within the same lateral habenula neurons known also to encode reward prediction errors (RPE; Hong & Hikosaka, 2008; Matsumoto & Hikosaka, 2007). Finally, a recent study showed that during a non-instrumental information seeking task, monkey orbitofrontal cortex represented both the primary reward value and the information value of predictive cues (Blanchard et al., 2015). This was interpreted as a reward value for information consistent with the putative role of the orbitofrontal cortex in representing reinforcing features of stimuli in reward-based behaviour and learning (Rushworth, Noonan, Boorman, Walton, & Behrens, 2011). On the basis of this set of findings, it has been suggested that information and reward may be represented in common neural structures, and that this *common currency* for information and reward may be the physiological source of the reward value of information in non-human primates (Bromberg-Martin & Hikosaka, 2011).
The common currency hypothesis predicts that processing of non-instrumental information should share neural substrates with reward processing in human participants; however, to date this prediction has not been tested explicitly. Nevertheless, circumstantial evidence in favour of the common currency hypothesis was provided by an fMRI study by Kang et al. (2009). In this study, the authors assessed the desirability of information using a paradigm in which participants were presented with simple trivia questions (e.g., “what musical instrument was invented to sound like a human voice singing?”), and asked to rate their curiosity about the answer as well as their confidence that they knew the correct answer. Behavioural results indicated that participants expressed the greatest levels of curiosity for intermediate levels of confidence (i.e. neither complete certainty nor complete uncertainty), and that the presentation of the answers to the trivia questions was associated with increased BOLD signal in the ventral striatum. Since this brain region is a key reward-processing area of the basal ganglia (Pagnoni, Zink, Montague, & Berns, 2002), this finding is consistent with the common coding hypothesis of information. However, owing to the relatively imprecise operationalisation of information by Kang et al. (2009), further research into this question in human participants is warranted.

Finally, a parallel approach to the question of non-instrumental information has investigated the effect of such information on participants’ decision making, rather than assessing demand for information directly. For instance, Tversky and Shafir (1992) provided evidence for such an effect in the context of disjunctive choice. In disjunctive choice scenarios, participants are given the option of acquiring information that is irrelevant to a decision at hand: for instance, in one question participants were asked whether they would prefer to find out whether they had passed or failed a university exam before making a decision about whether to go on a Christmas holiday. In such scenarios, Tversky and Shafir (1992) found that participants behaved differently in the subsequent decision depending on whether they had first acquired the non-instrumental information about the irrelevant outcome. Shafir and Tversky (1992) subsequently replicated this finding in the context of a one-off prisoner’s dilemma game, in which participants could either cooperate or not cooperate with a second player to achieve mutually desirable outcomes. The authors showed that behaviour on this task differed depending on whether participants had first acquired non-instrumental information. Such findings are inconsistent with strictly instrumental valuation of information, since under solely instrumental valuation of information, preferences over choice options
should not be affected by information about an irrelevant outcome (the ‘sure thing principle’ of Savage (1954)). Bastardi and Shafir (1998) proposed that these results are consistent with participants’ preference for acquiring non-instrumental information not because of its instrumental effects (which were nil), but because acquiring information gave participants greater confidence in their decision at the point of making that decision. A similar conclusion was reached by Eliaz and Schotter (2010) on the basis of results from an economic decision task showing that participants expressed a willingness to pay for non-instrumental information only when it would give them more confidence in a decision they had previously made. The authors interpreted results as suggesting that participants were ‘paying for confidence’, since they observed that participants did not acquire information when it was just information and did not pertain to a decision they had made. This stands in contrast to the results of Lanzetta and Driscoll (1966), who found that attitudes were related not to a desire for confidence, but to a desire for information itself.

Convergent evidence for valuation of non-instrumental information (independent of confidence) was also provided by Perone and Baron (1980). In this experiment, the authors used a simple behavioural conditioning paradigm in which participants could pull on a plunger; this action was paired with monetary reward if the plunger pull followed a predictive stimulus (the deflection of a needle on a dial). Crucially, participants were also given the opportunity to observe a discriminative stimulus (a diode whose colour predicted the likelihood of an upcoming deflection). Choosing to observe this discriminative stimulus was termed ‘observing behaviour’, and it was found that the delivery of predictive information in the discriminative stimulus reinforced participants’ observing behaviour, such that participants often sought to view the discriminative stimulus even when there was no practical benefit for them in doing so. Unlike comparable results in pigeons (Auge, 1974), for human participants this reinforcement occurred even when the information provided indicated that the delivery of monetary reward was unlikely. These results are contrary to the predictions of purely Pavlovian reinforcement theory, but are in line with an information hypothesis stating that “stimuli correlated with non-reinforcement or other aversive events are reinforcing when they provide information about those events” (Perone & Baron, 1980, p. 259). From the perspective of behavioural theory, this result therefore suggests that non-instrumental information is of inherent reinforcing value, contrary to classical
decision theory but in line with the preference for non-instrumental information observed in the remainder of the literature summarised in this section.

3.4 Rationale for Research Agenda

This section will synthesise the literature discussed in the preceding sections of Chapter 3, in order to identify several important unresolved research questions concerning the neural and cognitive processes underlying human information seeking in decision making under uncertainty.

This chapter has reviewed theoretical and empirical literature concerning human information-seeking behaviours, and has identified a key issue in this research field as the nature of information’s value in decision making under uncertainty. A key theoretical distinction in this regard is between theories positing solely an instrumental value of information and theories positing an additional intrinsic value of information. According to commonly used definitions of these terms, instrumental valuation of information implies that information is assigned a value solely with respect to its instrumental properties; that is, as a means to an end, such that information is valuable to the extent that it can be used to acquire other tangible rewards, such as food or money. By contrast, theories positing an intrinsic value of information hold that, in addition to an instrumental value of information, humans assign (or should assign) a positive value to information that reduces uncertainty about the environment, even when this reduction of uncertainty does not aid in planning of any future behaviour.

Instrumental and intrinsic theories of information valuation therefore make differing predictions with respect to decision makers’ attitudes concerning information that cannot be used to aid decision making (termed non-instrumental information). Under an instrumental valuation framework, if no decision can be usefully conditioned upon a piece of information then the information need not be acquired. Intrinsic valuation of information, by contrast, predicts that decision makers should express a preference for acquiring non-instrumental information as a consequence of the intrinsic value assigned to this information.

Empirically, past studies of information seeking with human participants can be divided into those which study participants’ preference for acquiring information that is useful for a decision (instrumental information) and those which study participants’ preferences for acquiring non-instrumental information. The difference between these
two forms of information is crucial for distinguishing between instrumental and intrinsic theories of information valuation because intrinsic theories, but not instrumental theories, predict that decision makers will have preferences with respect to non-instrumental information. Non-instrumental information seeking tasks therefore present a means by which to test whether instrumental or intrinsic theories of information valuation provide a better account of human information-seeking.

Intrinsic valuation of information would predict that decision makers should be willing to pay a monetary cost to acquire early but non-instrumental information about whether they would win or lose a lottery. Animals other than humans including monkeys, pigeons, and starlings, display this behaviour, but it is an open question whether human participants also do so. Study 3 of the present thesis sought to address this gap in the literature, and therefore to determine whether instrumental or intrinsic valuation of information provided a better account of human information-seeking behaviour. To do so, Study 3 developed a novel adaptation of a task from the animal literature to assess human participants’ willingness to pay for non-instrumental information.

A second prominent topic of recent information seeking research has been the neural substrates of information seeking in decision making under uncertainty. The use of information seeking tasks with non-instrumental information is considered to be advantageous because this design permits the processing of information to be dissociated from the processing of the outcomes to which the information pertains. Accordingly, recent animal literature has primarily used a methodology involving single-neuron recordings from non-human primates completing non-instrumental information seeking tasks. In human participants, however, the neural substrates of information seeking have typically been investigated using instrumental information-seeking tasks, as for example in the context of the exploration-exploitation dilemma. As such, the neural substrates of information seeking have not previously been investigated in human participants using non-instrumental information seeking tasks equivalent to state-of-the-art research in the animal literature.

Findings from studies of information seeking in monkeys have suggested that non-instrumental information is encoded within brain regions in the dopaminergic midbrain typically associated with processing of primary reward, such as the ventral tegmental area, substantia nigra pars compacta, and lateral habenula. These findings have led to the development of a ‘common currency’ hypothesis of information seeking
by Bromberg-Martin and Hikosaka (2011). This hypothesis proposes that a preference for acquiring information in non-human primates results from a common neural representation between information and reward. In humans, this hypothesis predicts that the neural substrates of (non-instrumental) information processing will be similar to the neural substrates of reward processing, indicative of the common processing architecture proposed by Bromberg-Martin & Hikosaka (2011) to underlie intrinsic valuation of information. Study 4 investigated this question using EEG data recorded from human participants completing the non-instrumental information seeking task previously developed by Study 3.
4. Rationale for Methodological Approach

This chapter sets out the rationale for the methodological approach employed in the present thesis, and reviews the quantitative tools used to draw conclusions regarding the neural substrates of belief updating and information seeking. The chapter is organised as follows: the neurophysiological underpinnings of EEG signals are first reviewed, as well as the event-related potential methodology that is commonly used to analyse EEG data. I will then describe the theoretical rationale for the use of incentive-compatible behavioural tasks in belief-centric research such as the present thesis. Finally, I will provide an overview of the tasks and models used in each of the four experiments reported in this thesis, detailing how these general methodological considerations were applied in this thesis.

Overall, the present thesis adopted an approach based in the principles of model-based cognitive neuroscience. This research program aims to find common ground between the disciplines of mathematical psychology and cognitive neuroscience (Forstmann & Wagenmakers, 2015). Research in this field makes use of quantitative tools for modelling behavioural data developed within mathematical psychology—as well as other disciplines such as machine learning and decision theory—and uses these models to assist in the analysis of neural data collected with neuroimaging techniques such as fMRI (e.g., Forstmann et al., 2008) or EEG (e.g., Boehm, van Maanen, Forstmann, & van Rijn, 2014). EEG will be the neuroimaging technique employed in the present thesis, and the neurophysiological principles underlying this technique are therefore introduced in Section 4.1 below.

A model-based approach to the analysis of the EEG data collected in this thesis was necessitated by this thesis’s research focus on participants’ beliefs. Beliefs are by definition latent constructs, and must be estimated from behaviour rather than observed directly, as discussed in more detail in Section 4.2. This estimation was achieved using computational modelling of participants’ belief states, with the resulting estimates of belief states used to inform EEG analyses. The specific models used in the present thesis were drawn from the Bayesian modelling and the reinforcement learning literatures (cf. Dayan & Niv, 2008; O'Reilly & Mars, 2015), and are introduced in Section 4.3.

4.1 Electroencephalography and event-related potentials
EEG is a non-invasive neuroimaging technique used to record the electrophysiological activity of the brain from electrodes on the surface of the scalp. In the typical 64-electrode recording configuration, as was used for the collection of the EEG data reported in Chapters 5, 6, and 8 of the present thesis, the electrodes are located across the scalp according to a standardised placement map, the International 10-20 system (Klem, Luders, Jasper, & Elger, 1999). The most significant physiological contributor to the electrical activity of the brain recorded at these scalp electrodes is synaptic (Olejniczak, 2006). Specifically, the electrical activity recorded at the scalp reflects current sources derived from the sum of inhibitory post-synaptic potentials (IPSPs) and excitatory post-synaptic potentials (EPSPs) in the cerebral cortex (Speckmann & Elger, 2005). These post-synaptic potentials are generated by the movement of ions across dendritic membranes following polarisation (for IPSPs) or depolarisation (for EPSPs) of neurons. As such, EEG data do not reflect neuronal firing rates directly, but rather summed inhibitory and excitatory synaptic currents. However, since these post-synaptic potentials themselves give rise to neuronal depolarisation, and therefore to the generation of action potentials, electrical activity as recorded by EEG provides a gross index of neuronal activation within superficial regions of cortex.

EEG data has good temporal resolution, but poor spatial resolution (Niedermeyer & Lopes da Silva, 2005). The good temporal resolution of EEG results from the nature of the physiological signal that is recorded. Since the underlying signal recorded by EEG—voltage fluctuations at the surface of the scalp—is analogue, the signal can be digitised and recorded at arbitrarily high sampling rates. The EEG data recorded in the present thesis, for instance, were recorded at a sampling rate of 512Hz, corresponding to approximately one voltage recording every two milliseconds. However, this good temporal resolution is offset by the relatively poor spatial resolution of EEG. For a given pattern of observed EEG activity, it is not possible to determine precisely what pattern of neural activity has produced this signal—although there do exist some statistical toolboxes which attempt to overcome this ‘inverse problem’, such as sLORETA (Pascual-Marqui, 2002). There are several biophysical reasons for this poor spatial resolution. The first of these concerns the distance of scalp electrodes from the sources of the electrical activity in the brain. This is problematic because electrical field strength decreases proportionate to squared radial distance from the source, meaning that the EEG signal from subcortical regions attenuates rapidly and is extremely difficult to detect by electrodes located at the scalp. The second problem is
that the electrical fields generated by the brain are affected by volume conduction, meaning that intervening tissue types such as cerebro-spinal fluid and the skull have a ‘smearing’ effect on the recorded EEG signal (Burle et al., 2015; Srinivasan, Nunez, Tucker, Silberstein, & Cadusch, 1996). Moreover, EEG signals recorded at the scalp necessarily reflect the summed activity of very large numbers of individual neurons; as such, since neurons with opposite spatial orientations produce electrical fields which cancel one another, the EEG recorded at the scalp tends to be produced by neurons which are spatially aligned with one another (Lopes da Silva & van Rotterdam, 2005). As a result, the electrical activity recorded by EEG electrodes largely represents the activity of sources such as cortical pyramidal neurons, which tend to be aligned in parallel with one another and perpendicular to the scalp. In addition, stronger electrical activity is detected in the EEG from cortical gyri than from sulci, since grey matter in sulci are further from the scalp, and tend to be spatially opposed by grey matter on the opposite wall of the sulcus. The recorded activity of EEG therefore disproportionately represents certain cortical sources, meaning that EEG data do not provide a complete index of whole-brain activity. In addition, even for activity which reaches the scalp and is detected by EEG electrodes, it has been shown that it is mathematically impossible to solve the inverse problem by identifying a unique cortical source which is the generator for this data (Lopes da Silva & van Rotterdam, 2005). This presents a major limitation to the interpretability of EEG data with respect to its spatial origin, since a given signal could be associated with any number of possible brain regions. The limitations described above have informed the development of analysis techniques for EEG data such as event-related potential (ERP) analysis, which uses cross-trial and cross-participant averaging to decompose the EEG signal into interpretable components. The present thesis adopts this methodology, in concert with computational cognitive models of behaviour.

The typical methodology for ERP analysis is as follows: an event of interest is identified, such as the presentation of a visual stimulus or the execution of a motor response. This event is typically repeated a number of times across different trials of a cognitive task (typically at least ten trials per condition, but often more; see Luck, 2005). The goal of ERP analysis is to identify an average potential associated with this event by extracting ‘epochs’ of a constant length (e.g., from 100 milliseconds before to 1000 milliseconds after the event of interest), and averaging across different epochs of the same event to identify an average waveform. This event-related potential is said to
be ‘time-locked’ to the event of interest, and the shape of the waveform can be interpreted with respect to underlying cognitive processes. Typically, task events will be grouped into different conditions (e.g., the presentation of a common target versus an uncommon distractor; Polich, 2012), with the aim of identifying components of the ERP waveform which differ between different conditions. This averaging approach assumes that variability between different epochs can be treated as zero-mean noise; as will be discussed in Section 4.3.1, this assumption may be restrictive and can be overcome by analysing single-trial ERP data rather than averaged ERP data. Likewise, this averaging approach also assumes that variability between individual participants can be treated as a random variable and not modelled directly; as will be discussed in Section 4.3.2, this assumption ignores the possibility that qualitatively different participant subgroups may be present in the data.

Components of the ERP are typically named according to their polarity (positive versus negative) and the position of their peak in the ERP waveform (either ordinal or time-latency). This leads to components such as the N1 (first negative peak), P3 (third positive peak), or P300 (positive peak occurring at approximately 300ms post-stimulus, and synonymous with P3). However, other components are often named according to their functional role (e.g., the feedback-related negativity, FRN; the stimulus-preceding negativity, SPN). Over decades of research into the functional significance of ERPs, different components have been linked to different cognitive processes. As a result, ERP amplitude differences between different task conditions can be readily interpreted with respect to differences in underlying cognitive processes.

4.2 General rationale for behavioural tasks

A crucial methodological issue for the belief-centric research methodology adopted by the present thesis was that individuals’ beliefs are by definition latent constructs. That is, the beliefs of participants completing a cognitive task cannot be observed directly, only inferred by observing behaviour. The strength of the inferences drawn in the present thesis was therefore dependent upon a well-specified quantitative model of the correspondence between participants’ behaviour and their latent beliefs.

Depending on the outcome measure used, an individual’s actions may give either direct or indirect access to the state of that individual’s beliefs. To illustrate this, consider that there are a number of different measures which can be used to infer an
individual’s beliefs regarding the likelihood of rain on a given day. A direct measure of this belief would be to explicitly ask the individual to assign a probability between 0 and 1 corresponding to their belief regarding the subjective likelihood of rain. A second, more indirect option would be to observe the individual’s behaviour, and then to infer the state of the beliefs which motivated this behaviour: for instance, it is reasonable to infer that an individual who takes an umbrella to work in the morning has assigned a large subjective probability to the possibility of rain. Although the direct approach to belief elicitation intuitively appears more informative than the indirect method, this is not necessarily the case for the experimental tasks typically used in cognitive psychology and cognitive neuroscience. There are several reasons for this, most notably the issue of incentive compatibility of behaviour, and the limited probabilistic sophistication of many human participants.

The issue of incentive compatibility refers to the fact that, in many experimental cognitive tasks, participants have little or no incentive to give an accurate report of their beliefs if asked directly. This is especially true in the case of studies collecting data from undergraduate students who have participated in exchange for course credit. By contrast, it is relatively straightforward to collect an incentive-compatible indirect measure of participants’ beliefs, such as actions taken in a learning task with real monetary rewards as performance outcomes. Using incentive-compatible rewards as outcomes means that participants are motivated to behave in a way that allows for more accurate (indirect) estimation of the state of their latent beliefs. Indeed, previous research has demonstrated that task performance is much less variable, and therefore provides a more accurate estimate of participants’ latent preferences and beliefs, when incentive-compatible measures are used (Holt & Smith, 2009). Incentive-compatible methods for eliciting direct probability estimates of beliefs have been developed in the economics literature, such as the Becker-DeGroot-Marschak auction (Becker et al., 1964). Although these measures represent an improvement over their unincentivised counterparts, they nonetheless depend upon several auxiliary assumptions, namely that the participant seeks to maximise their expected utility (Hossain & Okui, 2013), and that the participant maps the subjective plausibilities of events into numerical probabilities in a linear fashion. For reasons detailed below, the latter assumption may not hold within untrained participant samples.

A second potential confound for direct probabilistic belief elicitation, and therefore a second argument in favour of indirect assessment of participants’ beliefs
using incentivised tasks, concerns the issue of probabilistic sophistication. Many expected-utility theories of choice assume probabilistic sophistication (e.g., Machina & Schmeidler, 1995; Sarin & Wakker, 2000), specifically that decision makers’ subjective probabilities across different conjunct sets of events are complete (sum to one) and are linearly weighted, meaning that increases of a given size in reported probabilities correspond to equivalent increases in subjective probability (for example, a change from $p = .01$ to $p = .02$ is of the same subjective magnitude as a change from $p = .97$ to $p = .98$). In fact, however, both of these assumptions are regularly violated in empirical data: Ellsberg (1961) demonstrates that participants can, in some cases, act as though they possess subjective probabilities which sum to greater than one, and advances in prospect theory Tversky and Kahneman (1992) have demonstrated that participants display nonlinear subjective probability weightings, with an overestimation of the frequency of rare events ($p < .1$) and an underestimation of the frequency of frequent events ($p > .9$). As a result, it is not straightforward to derive valid estimates of belief strength from subjective probability estimates. This presents a methodological confound for the direct elicitation of participants’ beliefs.

For these reasons, therefore, the present thesis estimated participants’ beliefs indirectly, based on behaviour in simple choice and learning tasks. In this respect the thesis is consistent with a large body of recent research in cognitive neuroscience, which has similarly estimated latent states of beliefs from choices in simple tasks rather than eliciting probabilistic beliefs directly (e.g., Daunizeau et al., 2010; Mars et al., 2008; O’Reilly et al., 2013; Vossel et al., 2015). Further details of the tasks and computational models used are provided separately for each experiment below.

4.3 Overview of behavioural tasks and computational models

4.3.1 Study 1: Single-trial event-related potential correlates of belief updating

Study 1 of this thesis sought to identify the neural substrates of belief updating using EEG. Specifically, we sought to test the hypothesis that the amplitude of the P3 component of the event-related potential encoded the magnitude of participants’ updates to their beliefs (Kopp, 2008). A secondary aim of this study was to investigate how the SPN component of the event-related encoded belief uncertainty.
This study elicited participants’ beliefs using an incentive-compatible perceptual learning task (see Figure 4.1), consistent with the rationale set out in Section 4.2. This task required participants to learn the association between different contrasts of a visually presented checkerboard stimulus and monetary reward. This mapping was constant within each block, but differed between blocks.

Figure 4.1. Schematic of perceptual learning task. Following a self-paced button press, a checkerboard stimulus was presented whose contrast changed linearly. The participant could at any time select the contrast displayed on screen by pressing a button with the right index finger. The trial continued until a button was pressed or until stimulus duration exceeded 30 s. Following the participant’s choice, the selected contrast remained on screen for 2 s, after which time the monetary reward associated with the chosen contrast was displayed for 2.5 s. In the event that no button was pressed within 30 s, feedback was a message reminding the participant of the task instructions.

During each block, participants performed a number of consecutive trials in which they aimed to choose the contrast associated with the maximum reward (target contrast). After choosing a contrast, participants received the reward associated with the chosen contrast. Crucially, the amount of reward which participants received for a given contrast was determined by the proximity of the chosen contrast to the maximally rewarding target contrast. Concretely, reward was assigned as a function of the difference between the chosen and target contrasts, and reward per trial was in the range 0-25 cents (rounded to the nearest integer value). The mapping was a symmetrical triangular function with a centre of zero percent contrast difference, a half-width of 15 percent contrast difference, and a height of 25 cents.
As such, received reward was maximal when the participant responded at the target contrast, and decreased monotonically with increasing difference of chosen contrast from the target. Importantly, reward was zero for responses at greater than 15 percent distance. By choosing different contrasts and obtaining associated rewards over a number of trials, participants were able learn the target contrast and thereby maximise their winnings. Prior to testing, participants received training to instruct them in the shape of the reward function, and were informed that each block would have a different target in the range 10-100 percent. Participants completed 15 blocks of the task in total.

Given this perceptual learning task, we formulated a computational model that provided a trial-by-trial estimate of participants’ beliefs regarding the identity of the target contrast. This model was necessary in order to estimate the magnitude of updates to participants’ beliefs, and specifically to quantify the magnitude of the belief update following the presentation of each piece of monetary performance feedback. The magnitude of a belief update was defined as the dissimilarity of prior and posterior beliefs; this measure provides an index of the amount of information provided by a piece of feedback since more informative feedback is associated with greater dissimilarity between prior and posterior beliefs, and therefore with larger belief updates.

The Bayesian updating model applied to this task assumed that participants had an internalised representation of the response mapping function, and that learning occurred in a Bayes-optimal fashion, subject to perceptual uncertainty. This latter uncertainty parameterisation was motivated by the fact that feedback was presented to participants along with a visual representation of the chosen contrast. As such, participants did not know exactly which contrast they had chosen, and there was therefore an inherent degree of uncertainty concerning which contrast the observed feedback pertained to. For instance, a participant choosing at 60 percent contrast and observing 20 cents feedback would not necessarily know whether this feedback was associated with 58 versus 60 versus 62 percent contrast. To capture this uncertainty, the Bayesian updating model used in Study 1 assumed that participants performed Bayes-optimal belief updating subject to a variable amount of zero-mean Gaussian noise affecting the perceived chosen contrast. A free parameter $\sigma$, which corresponded to the standard deviation of this Gaussian noise, was fit to participants’ behaviour. In addition, the model also assumed that different participants could differ in perceptual uncertainty.
as a result of individual differences in visual acuity. The perceptual uncertainty parameter \( \sigma \) was therefore permitted to vary between different participants. Belief update magnitudes were calculated separately for each participant at the estimated value of the \( \sigma \) parameter. Further technical details concerning the model are provided in Chapter 5.

In addition, in order to investigate the issue of Bayesian versus non-Bayesian belief updating strategies, we also formulated a non-Bayesian model of performance on the perceptual learning task. This allowed us to investigate whether participants were indeed behaving in a manner consistent with Bayesian belief updating, and were not adopting any form of non-Bayesian heuristic strategy (see Section 2.2.5). Specifically, we implemented a model which assumed that participants were performing the task using a Win-Stay Lose-Shift (WSLS; Robbins, 1952) heuristic strategy rather than a Bayesian updating strategy. A WSLS strategy predicts that, rather than maintaining a full model-based representation of an outcome space, participants simply repeat choices which have previously been associated with any reward, and switch randomly to a new choice if the previous choice did not receive reinforcement. For the perceptual learning task employed in Study 1, in which participants could either receive some or no monetary reinforcement on each trial, the WSLS model was implemented such that participants were assumed to attempt to reproduce the previous choice if this previous choice had resulted in any non-zero monetary reward amount. On the other hand, if the previous choice had resulted in zero monetary reward, participants were assumed to randomly choose a new contrast on the subsequent trial. Using formal model comparison based on the Bayesian Information Criterion (BIC) statistic, it was possible to determine whether the Bayesian updating model or the WSLS heuristic model provided a better account of data, both for individual participants and across the entire participant sample.

Finally, in order to link belief updating with the amplitude of the P3 and SPN components of the event-related potential, it was necessary to expand upon traditional forms of ERP analysis. As discussed in Section 4.1 above, traditional ERP analysis extracts a measure of the amplitude of a particular component by averaging across all epochs in which the event of interest is present. This statistical approach implicitly assumes that any variability in the amplitude of the ERP component between epochs can be considered noise. Although this method is very suitable for the comparison of
distinct experimental conditions (e.g., targets versus distractors), a condition-based averaging approach is not suitable for assessing the relationship between the amplitude of an ERP component and a continuous independent variable, such as belief update magnitude. One solution to this issue would be to discretise the continuous variable, using a method such as a median split. However, such an approach would result in a degree of arbitrariness concerning the number of conditions into which to discretise the variable, as well as a loss in statistical power (McClelland, Lynch, Irwin, Spiller, & Fitzsimons, 2015). As such, in order to investigate the relationship between continuous predictor variables and ERP amplitudes, Study 1 of the present thesis adopted a single-trial regression approach to ERP analysis (Boehm et al., 2014; Kolossa et al., 2015; Mars et al., 2008). In this analysis, dependent variables were the amplitudes of each ERP component, and the independent variables were model-derived trial-by-trial estimates of the magnitude of belief updates (for the P3 analysis) and belief uncertainty (for the SPN analysis). This regression analysis was conducted separately for each participant, and the resulting \( \beta \) coefficients subjected to a \( t \)-test against zero across participants, in order to make population-level inferences concerning the encoding of particular belief variables in particular ERP components.

4.3.2 Study 2: Electrophysiological indices reflect switches between Bayesian and heuristic strategies in perceptual learning

Study 2 of the present thesis was designed to investigate several questions related to the neural substrates of belief updating left unresolved by Study 1. In particular, this study investigated the factors underlying participants’ use of Bayesian belief updating versus non-Bayesian heuristics (such as WSLS) in the perceptual learning task developed by Study 1. This allowed for identification of patterns of neural data which differentiated between use of Bayesian belief updating and the use of a non-Bayesian heuristic task strategy.

A key factor likely to contribute to the selection between Bayesian and heuristic strategies is task motivation. As discussed in Section 2.2.5, Bayesian belief updating is computationally demanding compared to the use of simpler heuristics. As such, participants’ willingness to expend cognitive resources in the use of a Bayesian belief updating strategy is likely to depend upon their motivation to complete the task. Study 2 assessed the performance of a new participant sample on a modified version of the
perceptual learning task developed in Study 1, once again while EEG data were recorded. Whereas in Study 1 this perceptual learning task was completed with solely monetary feedback, Study 2 administered this task with two separate feedback conditions: a monetary condition identical to that in Study 1, and an instructive condition which provided feedback on participants’ performance in the form of instructional directives. These two feedback conditions were completed by all participants in a repeated-measures design. By altering the monetary incentive for participants to expend cognitive resources in performing this task, the use of these two feedback conditions was expected to manipulate participants’ task motivation, and therefore to impact upon the selection of Bayesian versus heuristic task strategies. The two feedback conditions were matched in terms of the information content of feedback, by precisely matching the shape of the mathematical function which mapped the distance of participants’ choices from the target into feedback values (a step-wise triangular function, as described in Section 4.3.1). Rather than receiving monetary reward between 1 and 25 cents, for responses within 15 percent contrast distance of the target, in the instructive feedback condition participants were informed of the exact distance of their chosen contrast from the target (e.g., ‘Your choice was 6.4 percent from the target’). For responses at greater than 15 percent distance from the target, which received 0 cents reward in the monetary condition, in the instructive condition participants were simply told that their response was “too far” from the target. This ensured that, in spite of the categorical differences in the modality of the feedback, the information content of the two feedback conditions was identical. Participants completed 14 blocks of the task in total, with seven blocks each of the instructive and monetary feedback conditions, with presentation order counterbalanced across participants.

As in Study 1, Study 2 investigated belief updating with the aid of computational cognitive models, and made use of the two basic models which were developed for the perceptual learning task in Study 1: the Bayesian belief updating model, and the WSLS model. In Study 2, however, rather than using these models to quantify belief updating variables for use in ERP analysis, these two models were used to investigate the effect of the feedback manipulation on participants’ use of Bayesian versus heuristic task strategies. This was done by considering that participants might switch between strategies corresponding to the two models according to the two feedback conditions of the task, as detailed in Figure 4.2 below. For instance, a participant might adopt a
Bayesian strategy in the monetary feedback condition but switch to a heuristic strategy in the instructive feedback condition, where there was less incentive to perform accurately (the Incentive-Compatible Switching model in Figure 4.2). Alternatively, a participant might adopt a Bayesian strategy in both conditions (the Consistent Bayesian model in Figure 4.2). Further technical details regarding model fitting procedures are provided in Chapter 6.

Figure 4.2. Schematic overview of the computational cognitive models tested in Study 2. Models were defined according to the factorial interaction of two possible behavioural models (Bayesian Updating versus Heuristic) with two feedback conditions (Instructive and Monetary). This resulted in four possible models: two in which participants adopted the same strategy in both feedback conditions (unshaded boxes) and two in which participants switched strategies between feedback conditions (shaded boxes).

For each participant, models were fit using maximum-likelihood estimation, and formal model comparison using the BIC statistic was used to determine which of the four models provided the best explanation of that participant’s behaviour. These individual-participant model fits were then used to perform a model-based cluster
analysis (Fraley & Raftery, 1998), in which each participant was assigned to one of
four subgroups depending on which of the four computational models provided the best
fit to their choice behaviour. These subgroups were then compared to one another with
respect to both their performance accuracy in different feedback conditions, as well as
to the patterns of neural encoding of feedback in the different conditions, as reflected
in the event-related potential. Both behavioural and neural data were assessed using
analyses of variance (ANOVAs), with feedback condition (monetary versus
instructive) as a within-participants factor and model subgroup as a between-
participants factor. In this analysis, the interaction term between feedback condition
and model subgroup provided an empirical test of whether patterns of differences in
behavioural and ERP data between the two feedback conditions differed as a function
of the model subgroup to which participants were assigned.

The use of model-based cluster analysis to investigate differences in ERP data
between distinct subgroups of participants overcomes an important potential limitation
of ERP analysis, namely its assumption of the homogeneity of an effect across all
participants in a sample. Absent a between-participants grouping factor such as
subgroup membership, standard within-participants ANOVAs that are typically used to
analyse ERP data in classical ERP research (e.g., M. X. Cohen, Elger, & Ranganath,
2007; Hajcak, Moser, Holroyd, & Simons, 2007; Luck, 2005) data residuals are
assumed to be normally distributed across participants. This assumption results from
the use of a Gaussian error term in the general linear model, and is equivalent to
assuming that the deviance of individual participants from the marginal group mean
can be treated as (Gaussian) noise. This assumption is likely to be met if the effect of
interest is present in approximately equal magnitude across the participant sample.
However, the assumption is very likely to be violated if the participant sample consists
of distinct subgroups of participants, such as the different model subgroups in Study 2.
Using a model-based cluster analysis allows these different subgroups to be identified
in a principled manner, and for subgroup membership to be used as a between-
participants clustering variable. Moreover, the use of formal computational model
comparison to inform clustering also overcomes some of the limitations of model-free
cluster analyses (e.g., k-means clustering). Simpler clustering methods such as these
have been criticised for assigning participants to clusters in a way which solely seeks
to minimise within-cluster variance, and which may not represent meaningful
qualitative differences between different participants (Meehl, 1992).
4.3.3 Study 3: Intrinsic valuation of information in decision making under uncertainty

Study 3 of the present thesis focused upon the question of information seeking, rather than belief updating as in Studies 1 and 2. Study 3 sought to develop a novel behavioural task to investigate participants’ preferences for acquiring non-instrumental information, and comprised two separate behavioural experiments investigating participants’ willingness to pay for non-instrumental information (defined as information upon which no decision could be usefully conditioned, as per Section 3.3). The first experiment in Study 3 investigated the effect of information cost on participants’ preference for acquiring information, and the second experiment investigated the effects of both the cost of information and the rate at which it was delivered upon willingness to pay for information. In both experiments, computational modelling was used to formalise and test predictions of different theories of information valuation.

The non-instrumental information seeking task used in Study 3 was developed by adapting extant tasks from the animal literature for use with human participants. These animal tasks were developed for use in macaque monkeys (Blanchard et al., 2015; Bromberg-Martin & Hikosaka, 2009, 2011), as well as birds including pigeons (Gipson et al., 2009) and starlings (Vasconcelos et al., 2015). The underlying structure of each of these tasks is the same: on each trial, a lottery is played out in which the animal can receive either a large or a small reward with equal probability. These rewards are primary reinforcers of motivational significance to the animal, such as food pellets for birds, or water for monkeys. The outcome of this lottery is revealed only after a delay period; during this delay period the animal is given the choice of viewing either an informative stimulus, which accurately predicts the outcome of the lottery, or a non-informative stimulus, which is perceptually equivalent to the informative stimulus but which has no predictive validity for the outcome of the lottery.

The present thesis made several modifications to this basic task in order to adapt it for testing with human participants (see Figure 4.3). In order to elicit incentive-compatible behaviour from human participants, we replaced the large and small primary rewards used in animal research with a lottery between 20 cents and 0 cents monetary reward. In addition, the nature of the informative stimulus was altered: rather than delivering all information about the lottery outcome using a single stimulus, the present thesis used a task in which information was delivered piece-by-piece. Specifically, information was delivered in the form of an array of five cards, each of
which could be either red or black with equal probability. Cards’ colours were initially hidden, and were uncovered one-at-a-time at a constant interval during the delay period prior to the receipt of the lottery outcome. Participants were informed that, if they chose to observe the informative stimulus, a majority of black cards predicted a win outcome, whereas a majority of red cards predicted a loss outcome. The non-informative stimulus also consisted of a visually identical array of red and black cards, but participants were informed that the relative proportions of red and black cards in the non-informative stimulus held no predictive validity for the outcome of the lottery. As such, since informative and non-informative stimuli were perceptually identical, any preference for observing the informative stimulus rather than the non-informative stimulus could be interpreted in terms of participants’ preference for acquiring non-instrumental information. The rationale for using a piecewise informative stimulus rather than a one-off signal, as in the animal literature, was two-fold. The first rationale was that this design allowed for manipulation of the rate at which information was delivered in the second experiment of Study 3, simply by manipulating the interval at which cards were uncovered. Second, delivering information piecewise across five different cards introduced variability into the amount of information delivered by each individual card; this was a necessary consideration for the analysis of the neural encoding of information using EEG in Study 4 (discussed further in Section 4.3.4 below).

**Figure 4.3.** Schematic of non-instrumental information seeking task. Participants first received information regarding the identity and cost of the informative stimulus (both counterbalanced across trials), and then made a choice using left and right arrow keys within 2 seconds (left/right mapping of A and B counterbalanced across trials). Participants were then presented for 2 seconds with a choice information screen,
following which cards from the chosen stimulus were revealed sequentially at a constant rate of 3 seconds per card (18 seconds total delay). Participants were informed that all outcomes were predetermined, and that choice of stimulus was unrelated to win probability. If the participant failed to respond during the choice window, the non-informative stimulus was shown and no reward was subsequently delivered.

The final modification which was made to the information seeking task in Study 3 was the introduction of a cost manipulation, whereby the cost of observing the informative stimulus varied across different trials. This allowed us to assess whether participants were willing to sacrifice money in exchange for early but non-instrumental information. This cost manipulation was implemented by subtracting the cost of the informative stimulus from participants’ winnings only if they chose to view the informative stimulus prior to a win outcome; no cost was deducted if they observed the informative stimulus but did not win the lottery. This ensured that participants could not lose money by observing the informative stimulus, and thereby ensured that choices were not confounded by loss aversion. The same four cost conditions were used across all experiments in Study 3 and 4: 0 cents (i.e. free information), 1 cent, 3 cents, and 5 cents.

Finally, Study 3 used computational modelling of behaviour to instantiate the predictions of different theories of information value, rather than to quantify belief variables or to cluster different participants into subgroups, as in Studies 1 and 2 respectively. Specifically, Study 3 sought to determine whether instrumental or intrinsic theories of information valuation provided a better account of participants’ choice data. Crucially, as detailed in Section 3.2, these two theories make differing predictions regarding participants’ attitudes toward the non-instrumental information available in Study 3. Intrinsic theories of information valuation predict that participants should place a positive value on early resolution of uncertainty about the lottery outcome, and should be willing to pay to observe the non-informative stimulus. By contrast, instrumental theories of information valuation predict that participants should be indifferent toward the information provided by the informative stimulus, since the information imparted by this stimulus cannot be used to improve tangible future outcomes. Moreover, under instrumental valuation, participants should never be willing to pay for non-instrumental information. In order to test these predictions, Study 3
formulated two novel computational models instantiating the assumptions underlying each of these theories. These models were an Expected Value of Information model, which assumed solely instrumental valuation of information, and an Uncertainty Penalty model, which assumed intrinsic valuation of information derived from an aversion to temporally prolonged uncertainty (Bromberg-Martin & Hikosaka, 2009; Epstein, 2008). Formal comparison of these two models allowed the study to identify which model provided the best account of data (both at the aggregate level and for individual participants), and therefore to identify whether participants’ choice behaviour was better explained by instrumental versus intrinsic valuation of information. Further details concerning the mathematical specification of these models are provided in Chapter 7.

4.3.4 Study 4: The feedback-related negativity encodes an information prediction error in decisions to seek non-instrumental information

The goal of Study 4 was to investigate the neural substrates of information seeking, using EEG data from participants completing the behavioural task developed in Study 3. To do this, Study 4 investigated how the feedback-related negativity, an event-related potential component typically associated with the processing of rewarding stimuli, encoded the piece-by-piece delivery of information in the informative stimulus of this task. This design demonstrates an additional advantage of modifying the information seeking task originally used in animal research (e.g., Bromberg-Martin & Hikosaka, 2011) to allow for piece-wise delivery of information, as described in Section 4.3.3. The use of red and black cards with known occurrence probabilities to deliver information meant that it was mathematically possible to quantify the probability of a win outcome after each successive card, and to therefore to quantify the extent to which individual cards altered uncertainty regarding the lottery outcome. Using these measures, it was then possible to derive measures of the amount of information delivered by each card using tools from information theory (Shannon, 1948). As set out in Section 3.1, information was quantified as the reduction of entropy in beliefs as a result of observing each card stimulus.

Two important quantities were calculated for each card: a reward prediction error and an information prediction error. Since the win and loss outcomes did not change within a trial, the reward prediction error was mathematically determined by
the difference between the prior and posterior probabilities of winning the lottery; as such, black cards in the informative stimulus were associated with positive reward prediction errors, and red cards were associated with negative reward prediction errors (cf. Schultz, Dayan, & Montague, 1997). Information prediction errors, by contrast, were defined as the difference between the expected and the actual amount of information imparted by each card. Defined by analogy with reward prediction errors, positive information prediction errors were associated with stimuli which imparted more information than expected, and negative information prediction errors were associated with stimuli which impart less information than expected. This means that positive information prediction errors resulted from stimuli which increased participants’ certainty—whether or winning or of losing, and vice versa for negative information prediction errors.

Having calculated these quantities, Study 4 investigated how positive and negative reward prediction errors and positive and negative information prediction errors were encoded in the feedback-related negativity component of the event-related potential. According to the influential hypothesis of Holroyd and Coles (2002), this component reflects a reward prediction error resulting from the disinhibition of neurons in the anterior cingulate cortex by mesencephalic dopamine neurons. This analysis involved a straightforward condition-based ERP analysis (Luck, 2005), and did not necessitate an advanced analysis of ERPs using single-trial regression or model-based cluster analysis, as in Studies 1 and 2.

4.4 Summary

Over four studies, the present thesis investigated the cognitive processes and neural substrates underlying belief updating and information seeking in human judgment and decision making. Using a combination of computational modelling of behaviour and analysis of human EEG data, Studies 1 and 2 investigated belief updating, and Studies 3 and 4 investigated information seeking. Specifically, Study 1 aimed to identify the neural substrates of belief updating in a perceptual learning task, and Study 2 built on these findings by further investigating the circumstances under which human participants employ Bayesian versus non-Bayesian belief updating, and the event-related potential correlates of these different strategies. Study 3 then developed a novel behavioural task to investigate willingness to pay for non-instrumental information
among human participants. Finally, Study 4 investigated the neural substrates of information valuation using EEG data collected from participants completing this information-seeking task.
5. Study 1: Single-trial event-related potential correlates of belief updating

This chapter presents a study investigating the neural correlates of Bayesian belief updating, using EEG data recorded from human participants completing a simple perceptual learning task. This chapter was published in the journal *eNeuro* (Bennett, Murawski, & Bode, 2015), and will be presented in this format.

5.1 Aims of Study 1

The aims of Study 1 were twofold. The experiment reported in this study sought, first, to investigate the event-related potential correlates of belief updating and, second, to compare Bayesian and non-Bayesian models of belief updating for a perceptual learning task with monetary feedback.

This study sought to investigate the neural correlates of belief updating using a single-trial analysis of ERPs evoked by the presentation of monetary feedback. Specifically, this study developed a Bayesian belief updating model to estimate participants’ latent belief states, and used the resulting estimates to quantify two belief variables of interest: update size and uncertainty. These variables were then used to investigate the relationship between belief update size and single-trial amplitudes of the P3 component of the ERP, and between belief uncertainty and single-trial amplitudes of the stimulus-preceding negativity (SPN) component of the ERP.

A secondary aim of this study was to compare a Bayesian belief updating model with a non-Bayesian heuristic model, and thereby to determine which model provided a better account of participants’ choice data. Previous research has found that, even in tasks for which a Bayesian model provided a good overall account of choice data (e.g., a multi-armed bandit task; Steyvers, Wagenmakers & Lee, 2009), many individual participants’ choices were better explained by a win-stay lose-shift heuristic than by the optimal Bayesian strategy. Study 1 sought to determine whether this was also the case in simple perceptual learning.

5.2 Manuscript
Single-Trial Event-Related Potential Correlates of Belief Updating\textsuperscript{1,2,3}

\textsuperscript{1}Daniel Bennett,\textsuperscript{1,2} Carsten Murawski,\textsuperscript{2} and Stefan Bode\textsuperscript{1}

Abstract
Belief updating—the process by which an agent alters an internal model of its environment—is a core function of the CNS. Recent theory has proposed broad principles by which belief updating might operate, but more precise details of its implementation in the human brain remain unclear. In order to address this question, we studied how two components of the human event-related potential encoded different aspects of belief updating. Participants completed a novel perceptual learning task while electroencephalography was recorded. Participants learned the mapping between the contrast of a dynamic visual stimulus and a monetary reward and updated their beliefs about a target contrast on each trial. A Bayesian computational model was formulated to estimate belief states at each trial and was used to quantify the following two variables: belief update size and belief uncertainty. Robust single-trial regression was used to assess how these model-derived variables were related to the amplitudes of the P3 and the stimulus-preceding negativity (SPN), respectively. Results showed a positive relationship between belief update size and P3 amplitude at one fronto-central electrode, and a negative relationship between SPN amplitude and belief uncertainty at a left central and a right parietal electrode. These results provide evidence that belief update size and belief uncertainty have distinct neural signatures that can be tracked in single trials in specific ERP components. This, in turn, provides evidence that the cognitive mechanisms underlying belief updating in humans can be described well within a Bayesian framework.

Significance Statement
Recent theories propose that a central function of the brain is belief updating, the process by which internal models of the environment are revised. However, despite strong implications for cognition, the neural correlates of belief updating remain poorly understood. This study combined computational modeling with analysis of the event-related potential (ERP) to investigate neural signals, which systematically reflect belief updating in each trial. We found that two ERP components, P3 and stimulus-preceding negativity, respectively encoded belief update size and belief uncertainty. Our results shed light on the implementation of belief updating in the brain, and further demonstrate that computational modeling of cognition in ERP research can account for variability in neural signals, which has often been dismissed as noise.

Introduction
In an uncertain and dynamically changing world, survival depends upon having accurate beliefs about the environment. The more accurately an agent’s beliefs predict environmental contingencies such as threats from
predators or the availability of food, the more effectively the agent can plan its actions (Gläscher et al., 2010; Wunderlich et al., 2012). In particular, where environmental contingencies are unknown or nonstationary, an agent should constantly update beliefs in order to produce adaptive behavior (Behrens et al., 2007). Belief updating has generally been studied within a Bayesian framework (Nassar et al., 2010; Stern et al., 2010), wherein beliefs are described by probability distributions over possible states of the world. Bayesian belief updating is captured by the transformation of prior beliefs into posterior beliefs after new information is observed (Knill and Pouget, 2004; Courville et al., 2006).

Recent theories propose that belief updating may be a general principle underlying neural functioning, not merely an adaptive feature of cognition (Fiorillo, 2008, 2012; Friston, 2010). This hypothesis has strong implications for the understanding of human cognition (Bubic et al., 2010; Schwartenbeck et al., 2013). However, while general computational principles of belief updating are well understood, details of the mechanisms by which belief updating is performed in the human brain remain unclear. In addition, some recent research has suggested that the ability of decision makers to update beliefs in a Bayesian fashion may depend on the complexity of the decision situation and on the availability of heuristic alternatives to Bayesian updating (Achtziger et al., 2014, 2015). The present study addressed these questions by comparing Bayesian and heuristic accounts of belief updating, and by assessing how Bayesian belief updating was associated with two event-related potential (ERP) components typically linked with prediction and learning: the P3 and the stimulus-preceding negativity (SPN).

These components are implicated in belief updating by their association with learning and prediction. The P3 is a positive ERP component, the amplitude of which indexes the information content or surprise of an eliciting stimulus (Sutton et al., 1967; Mars et al., 2008). Under the context-updating hypothesis, P3 amplitude is thought to reflect the updating of internal schemata representing stimulus context (Donchin and Coles, 1988). These functions are broadly compatible with belief updating in the Bayesian sense of the term (Kopp, 2008). Furthermore, Mars et al. (2008) hypothesized that a fronto-central subcomponent of the P3 (the P3a; Polich, 2007) encodes belief update size. The present study explicitly tested this hypothesis.

The SPN is a negative-going slow wave elicited by stimulus anticipation (Brunia, 1988). SPN amplitude increases prior to stimuli delivering response reinforcement, both for reward (Masaki et al., 2010) and for instructive feedback (Moris et al., 2013), and covaries with the predictability and expected information of feedback (Kotani et al., 2003; Catena et al., 2012). The present study investigated whether SPN amplitude was related to belief uncertainty prior to updating.

We recorded the electroencephalogram (EEG) from participants performing a perceptual learning task with monetary feedback and used a Bayesian framework to estimate participants’ beliefs at each trial. Model-derived variables related to belief updating were then used to regress single-trial variations in ERP components (Bénar et al., 2007; Mars et al., 2008; van Maanen et al., 2011; Ostwald et al., 2012; Lieder et al., 2013; Kolossa et al., 2015).

Materials and Methods

Participants

Participants were 18 right-handed individuals with normal or corrected-to-normal visual acuity. Human subjects were recruited from among the staff and students of The University of Melbourne. The exclusion criterion was a medical history of any neurological disorder, including migraine and epilepsy. Informed consent was acquired from all participants in accordance with the Declaration of Helsinki, and approval was obtained from The University of Melbourne Human Research Ethics Committee.

One participant was excluded from analysis because of poor EEG signal quality. A second participant was excluded from analysis after a postexperiment debriefing revealed inadequate task understanding. For two other participants, computer error resulted in incomplete acquisition of EEG data. For these participants, behavioral analyses are reported only for task blocks in which complete EEG data were available (6 and 7 of 15 blocks, respectively). Final analyses were performed on data acquired from 16 participants (mean age, 22.63 years; age range, 18-29 years; 6 females).

In order to incentivize task performance, participants received monetary compensation for participation that was proportional to task winnings. Actual remuneration values were within the range of AUD $20-30 (mean remuneration, AUD $25.89; SD, AUD $4.36).

Behavioral paradigm

Participants performed a novel perceptual learning task while EEG data were recorded. The task required participants to learn an arbitrary mapping between the contrast of a stimulus and monetary reward. This mapping was constant within each block, but differed between blocks. During each block, participants performed a number of consecutive trials in which they aimed to choose the contrast associated with the maximum reward (target contrast). The stimulus was a grayscale checkerboard stimulus (Fig. 1A), which was presented on each trial for a duration of up to 30 s. During this time, the contrast of the checkerboard changed linearly (Fig. 1B), and the participant could at any time choose the contrast displayed on screen by pressing a button with the right index finger.

Acknowledgments: We thank Birte Forstmann, Robert Hester, Daniel Little, Bowen Fung, and Christina van Heer for comments on the manuscript; David Sewell and Jutta Stahl for general feedback; Damien Crone and Daniel Rosenblatt for assistance with data collection; and Hayley McFadyen for assistance with EEG preprocessing.

Correspondence should be addressed to Daniel Bennett, Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia. E-mail: daniel.bennett@unimelb.edu.au. DOI: http://dx.doi.org/10.1523/ENEURO.0076-15.2015

Copyright © 2015 Bennett et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International, which permits unrestricted use, distribution, and reproduction in any medium provided that the original work is properly attributed.
After choosing a contrast, participants received the reward associated with the chosen contrast. Crucially, the amount of reward that participants received for a given contrast was determined by the proximity of the chosen contrast to the maximally rewarding target contrast. Concretely, reward was assigned as a function of the difference between the chosen and target contrasts, and reward per trial was in the range 0–25 cents (rounded to the nearest integer value). The mapping (Fig. 1C) was a symmetrical triangular function with a center of 0% contrast difference, a half-width of 15% contrast difference, and a height of 25 cents. As such, the received reward was maximal when the participant responded at the target contrast and decreased linearly with increasing difference of chosen contrast from the target. The reward was 0 for responses at >15% distance. Feedback received was rounded to the nearest whole-cent value.

After choosing a contrast, participants received the reward associated with the chosen contrast. Crucially, the amount of reward that participants received for a given contrast was determined by the proximity of the chosen contrast to the maximally rewarding target contrast. Concretely, reward was assigned as a function of the difference between the chosen and target contrasts, and reward per trial was in the range 0–25 cents (rounded to the nearest integer value). The mapping (Fig. 1C) was a symmetrical triangular function with a center of 0% contrast difference, a half-width of 15% contrast difference, and a height of 25 cents. As such, the received reward was maximal when the participant responded at the target contrast, and decreased monotonically with increasing difference of chosen contrast from the target. The reward was 0 for responses at >15% distance. This relationship is formally expressed in Equation (1):

$$R(t, x_t) = \begin{cases} 
25 - \frac{(5\ | r_t - x_t |)}{3}, & |r_t - x_t | < 15 \\
0, & |r_t - x_t | \geq 15
\end{cases}$$

where $t$ is the trial number, $r_t$ is the target contrast on trial $t$, and $x_t$ is the participant’s chosen contrast on trial $t$. 

**Figure 1.** A, Following a self-paced button press, a checkerboard stimulus was presented whose contrast changed linearly. The participant could at any time select the contrast displayed on screen by pressing a button with the right index finger. The trial continued until a button was pressed or until stimulus duration exceeded 30 s. Following the participant’s choice, the selected contrast remained on screen for 2 s, after which time the monetary reward associated with the chosen contrast was displayed for 2.5 s. In the event that no button was pressed within 30 s, feedback was a message reminding the participant of the task instructions. B, Two demonstrative examples of stimulus contrast as a function of elapsed time. Example trial 1 (blue) has an initial contrast of 63%, is initially increasing, and has a half-cycle period of 9 s. Example trial 2 (red) has an initial contrast of 39%, is initially decreasing, and has a half-cycle period of 6 s. The checkerboard stimulus phase reversed at a rate of 12 Hz. C, Functional mapping between the contrast difference from target and monetary reward. The mapping was a symmetrical triangular function with a center of 0% contrast difference, a half-width of 15% contrast difference, and a height of 25 cents. As such, the received reward was maximal when the participant responded at the target contrast and decreased linearly with increasing difference of chosen contrast from the target. The reward was 0 for responses at >15% distance. Feedback received was rounded to the nearest whole-cent value.
By choosing different contrasts and obtaining associated rewards over a number of trials, participants were able learn the target contrast and thereby maximize their winnings. One important feature of the task was that participants were never informed of the exact contrast value they had chosen. As a result, there remained at all times a degree of uncertainty concerning contrast to which the observed feedback pertained.

The initial contrast and initial direction of contrast change were randomly determined on each trial using a Matlab random number generator with unique seeds for each participant. The half-cycle period, defined as the time required for the contrast of the checkerboard to change from one extreme to the other, was likewise randomly selected as 6, 7, 8, or 9 s on each trial in order to nullify the potential confound of learning based on temporal cues. The checkerboard phase reversed at a rate of 12 Hz, giving it a flickering appearance.

Prior to testing, participants received training to instruct them in the shape of the reward function and were informed that each block would have a different target in the range of 10-100%. Participants then completed 15 blocks of the task, each with a different target contrast, over approximately 60 min. Each block continued until the cumulative checkerboard presentation duration for the block exceeded 3 min, or until 25 trials were completed, whichever occurred sooner. As a result, the number of trials per block varied (mean, 18.46 trials; SD, 3.68 trials). This ensured that participants could not rush through the task, and that it was not possible to trade off experiment duration against monetary winnings. Finally, target contrasts were assigned subject to the constraint that the reward available for lowest and highest contrasts must be 0. In practice, because of the width of the reward distribution (Fig. 1), this meant that target contrasts were assigned on the interval [25, 85] rather than the interval [10, 100]. This ensured that the total reward available in each block was equivalent, and that feedback was always equally interpretable. Participants were not informed of this manipulation.

Stimuli were presented using a Sony Trinitron G420 CRT monitor at a framerate of 120 Hz. During task performance, participants were seated comfortably in a darkened room, using a chin rest at a distance of 77 cm from the screen. Checkerboard stimuli were 560 × 560 pixels in size, measuring 19.5 × 19.5 cm on the screen and subtending a visual angle of 14.43° by 14.43°. Responses were recorded using a five-button Cedrus Response Box.

EEG data acquisition

The electroencephalogram was recorded from 64 Ag/AgCl active scalp electrodes located according to the International 10-20 system. Electrodes interfaced with a BioSemi ActiveTwo system running ActiView acquisition software, and used an implicit reference during recording. Data were linearly detrended and re-referenced off-line to an average of mastoid electrodes. The vertical and horizontal EOGs were recorded from infraorbital electrodes that were horizontally adjacent to the left eye. The EEG was recorded at a sampling rate of 512 Hz. Using a linear finite impulse response filter, data were high-pass filtered at 0.1 Hz, low-pass filtered at 70 Hz, and notch filtered at 50 Hz to remove background electrical noise. Data were analyzed in epochs consisting of data from 1500 ms before to 1500 ms after the presentation of monetary feedback.

During preprocessing, data were first manually screened to exclude epochs contaminated by skin potential or muscle artifacts. Poor-quality data channels were then identified visually and corrected using the spline interpolation routine as implemented by the EEGLAB processing toolbox (Delorme and Makeig, 2004). An independent-components analysis, as implemented in the EEGLAB toolbox, was performed on the resulting dataset to identify and remove components related to eye movements and eye-blink artifacts. A final impartial artifact screening procedure was performed to exclude from analysis all epochs in which maximum/minimum amplitudes exceeded ±500 µV. Finally, a standard current source density (CSD) analysis was conducted on epoched EEG data for each of the 64 electrode sites using the CSD toolbox (version 1.1; Kayser and Tenke, 2006). This analysis calculates the spatial second derivative of voltage distribution over the scalp, and is a commonly applied procedure in the P3 and SPN literature (Gaeta et al., 2003; Catena et al., 2012). Spatial filters, such as CSD, are recommended for single-trial EEG analysis because of their ability to extract estimates of activity that are unique to each electrode, which increases the signal-to-noise ratio of individual trial CSD-ERPs, thereby augmenting the statistical power of analysis (Blankertz et al., 2008).

Single-trial CSD-ERP calculation

Single-trial P3 amplitudes were calculated at the following four electrodes typically investigated in condition-based P3 ERP research: FCz, Cz, CPz, and Pz (Mecklinger and Ullsperger, 1993; Troche et al., 2009). These electrodes were chosen to allow investigation of the effects of belief update on the topographically distinct P3a (fronto-central) and P3b (parietal) subcomponents of the P3 (for review of P3 subcomponents, see Polich, 2007).

For each electrode, P3 amplitude was calculated as the maximum voltage in the window from 300 to 450 ms after feedback presentation. This window was chosen according to a consensus estimate of latency of the peak of the P3 (Polich, 2007) and accounted for trial-to-trial variability in P3 peak latency. Voltages at each electrode were baseline corrected to the mean voltage within the period from 0 to 200 ms prefeedback.

Single-trial SPN amplitudes were calculated at 10 electrodes typically investigated in condition-based SPN ERP studies: F3, F4, C3, C4, T7, T8, P3, P4, O1, and O2 (Kotani et al., 2003). This allowed the investigation of the relationship between belief uncertainty and SPN amplitude at bilateral frontal, central, temporal, parietal, and occipital electrodes. For each electrode, SPN amplitude was calculated as the mean voltage in the window from 0 to 500 ms prior to the presentation of feedback. This window was longer than that used in some previous studies (Ko-
but this was considered necessary to stabilize the measurement volatility associated with the calculation of SPN amplitudes in single trials rather than from averaged waveforms. Voltages were baseline corrected at each electrode to the mean voltage within the period from 1300 to 1500 ms prefeedback.

Overview of behavioral models
We estimated two competing behavioral models: an unbiased updating model and a win-stay lose-shift (WSLS) heuristic model. The updating model assumed that participants maintained a belief distribution over the entire range of possible contrasts and updated this distribution as feedback provided new information on each trial. By contrast, the WSLS model assumed that, rather than maintaining a full belief distribution across contrasts, choices exhibited a one-trial memory such that participants tried to repeat the choice of the previous trial if it had resulted in any reward, and shifted randomly to a new contrast otherwise. Both models are formally specified below.

Parameters were estimated for each participant with maximum likelihood estimation using the interior point algorithm as implemented in MATLAB (MathWorks). Standard statistical model comparison tools were used to identify which model provided the best account of observed choices. The best-fitting model from this comparison was used in subsequent analyses of ERP results.

Unbiased updating model
For the unbiased updating model, a variant of a Bayesian grid estimator (Moravec, 1988) was used to obtain estimates of participants’ belief uncertainty and belief update size on each trial. In general terms, the model made a probabilistic estimate on each trial of participants’ beliefs regarding the target contrast. These estimates could then be used to quantify (1) the degree of belief uncertainty in any given trial and (2) how beliefs changed from trial to trial as new feedback information was received.

Structurally, the model describes participants’ prior beliefs at each trial t by a probability mass function (PMF) \( \theta_t \) over a contrast space divided into J discrete bins 1, 2, 3, . . . , J, such that the value of the PMF at each bin j, \( \theta_t(j) \), represented the subjective probability that the target contrast \( r_t \) fell within bin j on trial t. Bins had a width of 0.61% contrast, which was chosen as the largest value sufficient to resolve different monetary feedback values. As a result, the belief distribution contained \( J = 148 \) contrast bins on the interval [10, 100]. At the beginning of each block, this distribution was initialized according to a discrete uniform distribution, reflecting participants’ a priori uncertainty regarding the target contrast. Use of an uninformative starting prior is consistent with the modeling protocols of similar studies (Mars et al., 2008; Ostwald et al., 2012). Except for transitions between one block and the next, beliefs were considered to be updated sequentially, such that the posterior distribution of trial t was the prior distribution for trial \( t + 1 \).

For each trial t, participants observed the feedback \( f_t \) after the choice of contrast bin \( x_t \), which was determined according to the feedback mapping function \( R \) specified by Equation (1). Upon receipt of monetary feedback, the prior \( \theta_t \) was updated for each contrast bin j according to Bayes’ rule, as follows:

\[
\theta_{t+1}(j) = \frac{\theta_t(j) P(r_t | x_t, j) \varepsilon_j}{P(r_t | x_t)}
\]

The left-hand side of Equation (2) is the value of the posterior belief distribution for bin j, calculated by multiplying the participant’s prior belief that the target contrast fell within bin j, \( \theta_t(j) \), by the likelihood of observing the choice/feedback pair if the target were in bin j, \( P(r_t | x_t, j) \varepsilon_j \), and dividing by the marginal likelihood of the update \( P(r_t | x_t) \).

Importantly, in the task used in the present study, participants did not possess perfect knowledge of which contrast they had chosen (e.g., if the true value of a participant’s chosen contrast was 50%, the participant might know only that he or she had chosen some contrast between 40% and 60%). To account for this response uncertainty, the likelihood \( P(r_t | x_t, j) \varepsilon_j \) in Equation (2) was expressed as a probability-weighted sum over all contrasts the participant might have believed he or she had chosen. As such, the likelihood was considered not at a single contrast value but over the set of all candidate contrast bins \( J^* \), \( J^* = J \), as follows:

\[
P(r_t | x_t, J^*) = \sum_j \{ P(r_t | x_t, j) \varepsilon_j \}
\]

For each candidate contrast \( j^* \) in the set \( J^* \), the probability \( P(r_t | x_t, j^*) \) was equal to 1 if it was logically possible under the task feedback mapping for the target contrast r to belong to bin j if feedback \( f_t \) was observed after a choice of contrast \( x_t \), and was 0 otherwise. That is:

\[
P(r_t | x_t, J^*) = \begin{cases} 1, R(r_t, x_t) = f_t, \\ 0, R(r_t, x_t) \neq f_t. \end{cases}
\]

Each candidate contrast likelihood was then weighted by the subjective probability \( P(r_t | x_t, j^*) \) that the chosen contrast \( x_t \) was equal to the candidate contrast \( j^* \). This subjective probability reflects participants’ response uncertainty and was calculated as the function \( G_{\alpha, \sigma} \), a 0 mean Gaussian function of the contrast difference between the true chosen contrast \( x_t \) and the candidate contrast \( j^* \), as follows:

\[
P(x_t = j^*) = G_{\alpha, \sigma}(x_t, j^*, \sigma) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left[ -\frac{(x_t - j^*)^2}{2\sigma^2} \right]
\]

The SD \( \sigma \) of the distribution function reflects the degree of response uncertainty, such that greater values of \( \sigma \) result in more weight being given to candidate contrasts at a greater distance from the true chosen contrast. In the case of 0 response uncertainty, Equation (5) reduces to a Dirac \( \delta \) function. Given Equations (3) and (5), Equation (2) can be rewritten:
For an intuitive understanding of this model parameterization, consider the case of a participant who has perfect knowledge of exactly which contrast he or she has chosen. In this case, \( \sigma = 0 \) and \( \Pr(x_i = x_{i-1}) \) is equal to 1 where \( x_i = x_r \), and 0 elsewhere. In this case, the likelihood in Equation (3) is calculated exclusively on the basis of the true chosen contrast, and the participant is able to make very precise inferences from the observed feedback. In the present study, it was considered highly unlikely that participants had perfect knowledge of their chosen contrast. By allowing \( \sigma \) to vary, the model allows that participants consider a range of alternative hypotheses concerning the chosen contrast when updating their beliefs. The parameter \( \sigma \) was permitted to vary between participants when fitting the unbiased updating model.

To implement this model, we made the further assumption that participants’ choices were determined by beliefs, such that contrast bins with a higher probability of containing the target contrast had a higher probability of being chosen, subject to the response uncertainty during choice. Formally, the PMF for contrast choices over the set of contrast bins \( J \) was determined by convolving the prior belief distribution \( \theta_0 \) by the response uncertainty function \( G_J \) over the set of contrast bins \( J \), as follows:

\[
\Pr(x_i) = \frac{(\theta_0^* G_J)[J]}{k}
\]  

(7)

where \( k \) is a normalization constant ensuring that \( \Sigma \Pr(x_i) = 1 \).

As an illustration of how this model operates, we can assess the effects on belief of receiving feedback of \( f_t = 20 \) cents after a choice of \( x_i = 50\% \) contrast on trial 1 (\( t = 1 \)). For the sake of simplicity, rather than enumerating effects across the entire belief distribution, we consider the effects of observing this feedback on one contrast bin of the belief distribution centered around 60.2% contrast (\( j = 83 \)). Since we are considering the first trial of a block, prior belief probability for this contrast \( \theta_1 (83) = 1 / J = 0.007 \). If we assume that the perceptual uncertainty parameter \( \sigma \) is equal to 15, then, by Equation (3), the likelihood \( \Pr(20c, 50\% | x_i \in 60.2\%) \) is equal to 0.026. In order to calculate the posterior probability, we multiply the likelihood 0.026 by the prior belief probability 0.007 and divide by the marginal likelihood to normalize, giving \( \theta_2 (83) = 0.013 \). By calculating the ratio of posterior and prior, we observe that the participant’s subjective belief that the target contrast falls within this bin has nearly doubled in strength as a result of the information provided by feedback: \( \theta_2(j) / \theta_1(j) = 0.013 / 0.007 = 1.86 \).

Win-stay lose-switch heuristic model

Unlike the unbiased updating model, the WSLS model does not assume that participants maintain a belief distribution over the entire range of contrasts. Instead, this model predicted that participants’ behavior on a given trial was a function of whether or not they had received reinforcement on the preceding trial (Robbins, 1952). Specifically, the model assumed that participants attempted to repeat the contrast choice of the previous trial if they had received any monetary reward on the previous trial (win), subject to response uncertainty, or shifted randomly to a new contrast if they had not received monetary reward (loss) or at the start of a new block. This gives the following choice probability function:

\[
\Pr(x_i \in j) = \begin{cases} 
\delta(j - x_{i-1}) G_J[j] & , f_{t-1} > 0 \\
\frac{k}{j} & , otherwise
\end{cases}
\]  

(8)

where \( k \) is a normalization constant. Equation (8) implements the win case with the convolution of the 0 mean Gaussian response uncertainty function given in Equation (5) with the Dirac delta function \( \delta \), which is equal to 1 at the contrast bin chosen in the previous trial contrast and 0 elsewhere. This allows for the WSLS model to account for response uncertainty in a fashion similar to that of the unbiased updating model, thereby ensuring that predicted choice probabilities are comparable across the two models.

Calculation of belief updating variables

For the unbiased updating model, which assumed that participants updated a belief distribution across all contrasts, estimations of subjective belief distributions could be used to calculate the following three variables of interest on each trial: belief uncertainty prior to the receipt of feedback; postfeedback belief update size; and postfeedback surprise (Mars et al., 2008).

Belief uncertainty was calculated as Shannon entropy (Shannon, 1948) over contrast bins of the prior distribution, as follows:

\[
H(\theta_0) = - \sum_j \theta_0(j) \log_2 \theta_0(j)
\]  

(9)

Shannon entropy was used as an uncertainty metric because the entropy \( H \) of a probability distribution represents the degree of uncertainty coded by that set of probabilities. The entropy of a distribution is equal to 0 only in the case of complete certainty, when all probabilities but one are 0. Conversely, the entropy of a distribution is maximal when all probabilities have an equal value, as in a uniform distribution. In the present study, therefore, higher entropy values of the belief distribution reflected greater levels of belief uncertainty.

Belief update size was calculated as the mutual information of prior and feedback. This quantity represents the degree to which uncertainty is resolved in the transformation from prior to posterior probabilities, and corresponds to the information content \( I \) of feedback: the more informative feedback is, the greater the reduction in uncertainty from prior to posterior beliefs. Accordingly, belief update size was calculated as the difference in entropy between prior and posterior beliefs, as follows:
Table 1. Summary of statistical analyses

<table>
<thead>
<tr>
<th>Data structure</th>
<th>Type of test</th>
<th>Observed power</th>
</tr>
</thead>
<tbody>
<tr>
<td>a Normally distributed</td>
<td>Single-sample t test</td>
<td>1.0</td>
</tr>
<tr>
<td>b Model likelihoods</td>
<td>BIC</td>
<td>Not applicable</td>
</tr>
<tr>
<td>c Normally distributed</td>
<td>Single-sample t test</td>
<td>1.0</td>
</tr>
<tr>
<td>d Normally distributed</td>
<td>Single-sample t test</td>
<td>0.54</td>
</tr>
<tr>
<td>e Normally distributed</td>
<td>Single-sample t test</td>
<td>0.65</td>
</tr>
<tr>
<td>f Normally distributed</td>
<td>Single-sample t test</td>
<td>0.06</td>
</tr>
<tr>
<td>g Normally distributed</td>
<td>Single-sample t test</td>
<td>1.0</td>
</tr>
<tr>
<td>h Normally distributed</td>
<td>Pearson correlation</td>
<td>0.99</td>
</tr>
<tr>
<td>i Normally distributed</td>
<td>Single-sample t test</td>
<td>0.95</td>
</tr>
<tr>
<td>j Normally distributed</td>
<td>Single-sample t test</td>
<td>1.0</td>
</tr>
<tr>
<td>k Normally distributed</td>
<td>Repeated-measures ANOVA</td>
<td>0.77</td>
</tr>
<tr>
<td>l Normally distributed</td>
<td>Repeated-measures ANOVA</td>
<td>0.08</td>
</tr>
<tr>
<td>m Normally distributed</td>
<td>Repeated-measures ANOVA</td>
<td>0.13</td>
</tr>
<tr>
<td>n Normally distributed</td>
<td>Single-sample t test</td>
<td>0.31</td>
</tr>
<tr>
<td>o Normally distributed</td>
<td>Single-sample t test</td>
<td>0.97</td>
</tr>
<tr>
<td>p Normally distributed</td>
<td>Single-sample t test</td>
<td>0.98</td>
</tr>
<tr>
<td>q Normally distributed</td>
<td>Single-sample t test</td>
<td>1.0</td>
</tr>
</tbody>
</table>

This value was calculated for each trial and provided a model-based estimate of the degree to which feedback was used by participants to update their beliefs regarding the location of the target contrast in contrast space. Larger values of \( I \) indicate greater resolution of uncertainty, and therefore larger belief updates.

In addition, we note that in the literature, belief update size is sometimes also measured by a metric termed Bayesian surprise (Baldi and Itti, 2010; Ostwald et al., 2012), which can be calculated as the Kullback–Leibler divergence of prior and posterior. In order to allow comparison between the present study and previous research, Bayesian surprise, denoted \( I_{KL} \) (Kullback–Leibler divergence), was also calculated as an alternative measure of belief update size, as follows:

\[
I_{KL}(\theta; x, f) = H(\theta) - H(\theta | x, f) = H(\theta) - H(\theta | f) \tag{10}
\]

Finally, we calculated feedback surprise \( S \), a measure of the improbability of observing a particular feedback value given a certain contrast choice under certain beliefs (Shannon, 1948). Formally, this was computed as the negative logarithm of the probability of observing a certain feedback value \( f \), given the prefeedback belief distribution \( \theta \), and the chosen contrast value \( x \):

\[
S(\theta, f, x) = -\log Pr(f | x, \theta) \tag{12}
\]

It has previously been shown that surprise was encoded in the amplitude of the P3 at parietal electrodes in a serial reaction time task (Mars et al., 2008), and this quantity was therefore calculated in order to allow us to dissociate any observed effects of belief updating from effects of surprise. Importantly, while there is a superficial conceptual resemblance between belief update size and surprise, the two quantities are mathematically distinct (Baldi and Itti, 2010). Feedback surprise relates to the probability of occurrence of a particular feedback value; it is calculated as a function of the prior predictive distribution over possible observations. By contrast, belief updating relates to the degree to which feedback causes beliefs to be modified, and is calculated as a function of the prior and posterior distributions over parameters. Moreover, it has been shown that the two quantities have distinct neural substrates, with belief updating encoded in anterior cingulate cortex (ACC) and surprise encoded in posterior parietal cortex (O’Reilly et al., 2013). Furthermore, from a statistical perspective, an important difference between surprise and belief updating is that belief updating is calculated as the distance measure between prior and posterior belief distributions, whereas surprise is calculated only at a single point in the prior distribution.

Single-trial regression analysis of belief updating

Robust single-trial multiple regression analyses were used to investigate (1) the effect of feedback reward, feedback surprise, and belief update size on the amplitude of the post-feedback P3 component; and (2) the effect of belief uncertainty on the amplitude of the pre-feedback SPN. To account for individual variability in the amplitude of ERP components, both P3 and SPN amplitudes were normalized on an individual-participant level prior to regression analysis. To account for heteroscedasticity in the relationship between model-derived belief variables and single-trial ERP amplitude estimates, robust (weighted least squares) linear regression analyses were used. For all ERP analyses, regressions were run separately for each participant at each electrode, and resulting \( \beta \) coefficients were subjected to Bonferroni-corrected single-sample \( t \) tests in order to determine whether the effect of each predictor significantly different from 0 across participants.

Results

Behavioral task

Table 1 presents an overview of all statistical analyses reported. Across participants, responses became more
precise with increasing within-block trial number (mean $\beta = -0.65$, $t_{(19)} = -9.66$, $p = 0.00000008^a$), indicating acceptable task performance (Fig. 2). The mean absolute difference between the chosen contrast and the target contrast in the final trial of blocks was 9.24% (SD, 8.48%). This demonstrates that, while participants achieved proficiency on the task, their performance did not reach an absolute ceiling before block termination.

Model comparison
We used standard model comparison techniques in order to determine which of the two computational models described above provided the best account of participants’ choices. Table 2 presents Bayesian information criterion (BIC) values for the unbiased updating and WSLS models. Use of BIC allows us to identify models that account for data in a parsimonious way by balancing measures of parsimony (number of parameters) against measures of goodness-of-fit (log likelihood).

It can be seen that the unbiased updating model provided the best overall account of participants’ choices$^b$. This model assumed that participants maintained a complete belief distribution over the contrast space and that belief updates were unbiased by the direction of contrast movement at the time of choice. Furthermore, examination of model fits for individual participants using participant-specific BIC values revealed that the unbiased updating model provided the best account of choices for a clear majority of participants (Table 2, N best fit column).

As a result, all ERP analyses made use of belief variables calculated from the unbiased updating model.

Computational model
Across participants, pretrial belief uncertainty, as quantified by the unbiased updating model, was found to significantly predict choice accuracy on the upcoming trial (mean $\beta = 5.71$, $t_{(19)} = 11.74$, $p = 0.000000006^c$). Moreover, model-estimated belief uncertainty predicted choice accuracy even after accounting for the effects of the following three linear and nonlinear trial number regressors: a linear term, a quadratic term, and a cubic term. In this analysis, we found significant effects for the quadratic trial number term (mean $\beta = 0.12$, $t_{(19)} = 2.15$, $p = 0.048^d$) and the cubic trial number term (mean $\beta = -0.004$, $t_{(19)} = -2.44$, $p = 0.03^e$), but not for the linear effect of trial number (mean $\beta = -0.28$, $t_{(19)} = -0.42$, $p = 0.68$).

However, even when accounting for these effects of trial number, the linear relationship between model-estimated belief uncertainty and choice accuracy was still strong (mean $\beta = 9.75$, $t_{(19)} = 6.68$, $p = 0.0000007^f$). This result indicates that belief uncertainty was predictive of choice accuracy even when linear and nonlinear trial-by-trial learning effects were accounted for, suggesting that the task model fit the data well and validating the use of variables derived from this model in single-trial regression analyses. Figure 3 presents descriptive statistics for each of the calculated belief variables as a function of trial number.

In the task model, participants’ response uncertainty was captured by the parameter $\sigma$, the SD of the Gaussian noise affecting the marginal likelihood of belief updates. Across participants, estimates of $\sigma$ had a mean value of 12.99 (SD, 4.42), and fit values of $\sigma$ were positively correlated with participants’ overall task performance, as measured by the average deviation between chosen and target contrasts ($r_{(19)} = 0.86$, $p = 0.000002^g$). Individual differences in $\sigma$ were therefore behaviorally relevant, such that individuals with less response uncertainty tended to respond closer to the target contrast on average. This further validates our use of the Bayesian grid estimator to represent participants’ beliefs.

Single-trial regression analysis
P3
Single-trial regression analysis found a positive effect of belief update size on P3 amplitude at electrode Cz, CPz, or Pz, and no significant effect of reward magnitude or feedback surprise on P3 amplitude at any electrode. This indicates that

![Figure 2](image-url) Mean accuracy as a function of within-block trial number across participants. Accuracy is presented as the absolute difference of chosen and target contrasts, where lower numbers indicate better task performance. Error bars represent the SEM. Note that the number of trials per block varied across blocks and participants, as a result some participants did not complete >19 trials in any block. This confound limited the interpretability of accuracy data for trial numbers >20, and the final data point of the series therefore represents mean accuracy across trials 19–25 for each participant.

![Figure 3](image-url) Figure 3 presents descriptive statistics for each of the calculated belief variables as a function of trial number.

![Figure 4](image-url) Single-trial regression analysis for P3. There was no effect of belief update size on amplitude at electrodes Cz, CPz, or Pz, and no significant effect of reward magnitude or feedback surprise on P3 amplitude at any electrode. This indicates that

### Table 2. Summary of behavioral model fits for 4417 choices by 16 participants

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters per participant</th>
<th>Parameters</th>
<th>Belief distribution</th>
<th>Log-likelihood</th>
<th>BIC</th>
<th>N best fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unbiased updating</td>
<td>1</td>
<td>$\sigma$</td>
<td>Yes</td>
<td>$-20190$</td>
<td>40515</td>
<td>11</td>
</tr>
<tr>
<td>Win-stay/lose-shift</td>
<td>1</td>
<td>$\sigma$</td>
<td>No</td>
<td>$-20950$</td>
<td>40834</td>
<td>5</td>
</tr>
</tbody>
</table>

---

$^a$ Single-trial regression analysis on P3 amplitude at electrode FCz (mean $\beta = 0.27$, $t_{(19)} = 3.33$, $p = 0.005^i$, Bonferroni corrected; Fig. 4). There was no effect of belief update size on P3 amplitude at electrodes Cz, CPz, or Pz, and no significant effect of reward magnitude or feedback surprise on P3 amplitude at any electrode. This indicates that
single-trial amplitudes of the fronto-central P3a directly indexed model-derived measures of belief update size. Figure 4B displays the average voltage scalp distribution, and Figure 4C illustrates the difference map for large and small belief updates during the P3 time window. Table 3 displays a correlation matrix of the predictor variables included in the P3 regression analysis. Note that P3 regression analyses included either I or I_B as measures of belief update size, but never both.

As illustrated by Figure 3C, there was a significant tendency for belief update size I to reduce as the trial number increased (mean Spearman correlation across participants = −0.67, t_{115} = −17.91, p = 2 × 10^{-11}). As a result, we considered the possibility that the single-trial relationship between P3 amplitude and belief update size might have been confounded by an incidental effect of trial number on P3 amplitude. In order to address this possibility, we ran a control analysis in which trials were partitioned according to both trial number and belief update size. In this analysis, each trial was designated as an “early-,” “middle-,” or “late stage” trial, corresponding respectively to trial numbers 1–5, 6–10, and 11–15. Trials were also designated as either “small” or “large” belief updates according to a median split separately for each participant. We then used 3 × 2 repeated-measures ANOVA to assess separately the effects of trial number (early, middle, late) and belief update size (small, large) on mean P3 amplitudes at electrode FCz. Consistent with the single-trial regression results presented above, ANOVA results indicated a significant main effect of update size, F_{(1,15)} = 8.40, p = 0.01, with large belief updates (mean = 0.053 μV/cm², SD = 0.017 μV/cm²) associated with significantly larger P3 amplitudes than small belief updates (mean = 0.049 μV/cm², SD = 0.016 μV/cm²). There was no main effect of trial number on P3 amplitude (F_{(2,14)} = 0.25, p = 0.78), and no interaction between belief update size and trial number (F_{(2,14)} = 0.63, p = 0.55). These results support the contention that fronto-central P3 amplitude indexed belief update size, and suggest that this effect was not confounded by any incidental effects of trial number.

Interestingly, there was no relationship between belief update size and P3 amplitude at any electrode when belief update size was calculated as Bayesian surprise I_B rather than mutual information I (mean β = 0.20, t_{115} = 1.54, p = 0.14). This appears to suggest that the observed effects are specific to the mutual information formulation of belief update size. Note that regression analyses were each run with either I or I_B as measures of belief update size, never both.

Across participants, the mean P3 peak latency at electrode FCz was 338.4 ms (SD = 5.29 ms). There were no effects of reward, belief update size, or surprise on P3 peak latency at any electrode assessed.

**SNP**

Single-trial regression analysis found a small but significant negative effect of belief uncertainty (formally, the entropy of the belief distribution approximated by a Bayesian grid estimator) on SPN amplitude at electrodes C3 (mean β = −0.06, t_{115} = 3.56, p = 0.003, Bonferroni corrected; illustrated in Fig. 5) and P4 (mean β = −0.05,
This result indicates that higher levels of belief uncertainty were associated with smaller SPN components. That is, the more certain participants were regarding the location of the target contrast in contrast space, the greater the amplitude of the SPN evoked in anticipation of feedback stimuli. A median split was used to divide trials into two bins for each participant, corresponding to large and small belief updates according to model-derived estimates. This median split was for display purposes only and was not used in the main regression analysis, which was based on single-trial amplitudes.

Discussion
This study combined single-trial analysis of ERPs with computational modeling of belief. Our results showed that two mathematically distinct belief variables—update size and belief update size (KL)—were associated with SPN amplitudes. A strong negative correlation between belief uncertainty and trial number (mean Spearman correlation $r = -0.94$, $t(15) = -115.20$, $p = 2 \times 10^{-23}$, Fig. 3A), as would be expected in a task in which participants learned incrementally from each trial. The strength of this relationship precluded a factorial control analysis to dissociate the effects of belief uncertainty and trial number on SPN amplitudes.

Table 3. Correlation matrix for predictors in P3 regression analysis

<table>
<thead>
<tr>
<th></th>
<th>Reward</th>
<th>Belief update size ($I$)</th>
<th>Belief update size ($I_{KL}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reward</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Belief update size ($I$)</td>
<td>0.22 (0.19)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Belief update size ($I_{KL}$)</td>
<td>$-0.24$ (0.12)</td>
<td>0.64 (0.16)</td>
<td>1</td>
</tr>
<tr>
<td>Surprise</td>
<td>0.45 (0.21)</td>
<td>0.22 (0.12)</td>
<td>0.05 (.14)</td>
</tr>
</tbody>
</table>

Data are presented as mean Spearman coefficient across participants (SD).
and uncertainty—were encoded in distinct ERP components in a perceptual learning task. The combination of methods that we used linked the fine-grained information contained in single-trial EEG data with model-based estimates of participants’ latent beliefs, which would have been inaccessible to explicit testing. Our results suggest that trial-by-trial variations in the P3 and SPN reflect fundamental and distinct neural processes by which beliefs regarding the structure of the environment change over time.

Participants performed a simple perceptual learning task in which they learned a functional mapping between stimulus contrast and monetary reward. The task was both naturalistic and challenging: even with extensive practice, participants’ performances did not reach a ceiling, suggesting that participants continued to update beliefs throughout the experiment. We used a probabilistic model, termed the unbiased updating model, to infer participants’ beliefs at each trial from their choice history and found that model-based estimates of belief uncertainty predicted future choices well. The unbiased updating model gave better predictions of behavior than a competing model assuming a win-stay/lose-switch choice process in which participants chose on the basis of reward received on the previous trial rather than updating a full belief distribution.

We used the unbiased updating model to quantify the following three latent belief variables: belief uncertainty, belief update size, and surprise (Mars et al., 2008; Baldi and Itti, 2010; O’Reilly et al., 2013). We then investigated how the estimates of belief of the model update size and belief uncertainty were encoded in the P3 and SPN components of the ERP, respectively.

At the fronto-central midline electrode FCz, we found a significant positive relationship between postfeedback belief update size and single-trial P3 amplitude. This indicates that larger P3 amplitudes were observed in trials where feedback caused larger belief updates. Variability in single-trial P3 amplitude was best explained by regression using a model-derived estimate of belief update size and could not be accounted for by alternative regressors such as reward amount or feedback surprise. This is consistent with the hypothesis that P3 amplitude reflects a Bayesian belief-updating mechanism (Kopp, 2008; Mars Fig. 5. Stimulus-preceding negativity analysis. A, Median split waveforms for 0–1500 ms prior to the visual presentation of feedback. The SPN regression analysis window from 0 to 500 ms preceding feedback is indicated by the gray bar. ERP waveforms were low-pass filtered at 30 Hz for display purposes only. B, Mean voltage topography during SPN analysis window from 0 to 500 ms prior to visual presentation of feedback (time = 0). C, Topography of the mean voltage difference between high and low uncertainty trials across participants during the SPN analysis window. A median split was used to divide trials into two bins for each participant, corresponding to high and low belief uncertainty according to model-derived estimates. This median split was for display purposes only and was not used in the main regression analysis, which was based on single-trial amplitudes.
This theory attributes variability in P3 amplitude to the engagement of cognitive processes for revising internal models of the environment and predicts that larger updates to beliefs will be associated with larger P3 amplitude. Our study, using a single-trial regression approach, allowed for a direct test of this hypothesis, and our results provide broad support for the theory. In addition, we note that the observed association between belief update size and P3 amplitude disappeared when Bayesian surprise, rather than mutual information, was used as a measure of belief update size. The reason for this discrepancy is unclear, but may be related to differences in statistical power associated with the different temporal dynamics of the two measures (Table 1, Fig. 3). Other metrics, including a free-energy theoretical quantity termed model adjustment, have also been used in the ERP literature (Lieder et al., 2013). Future research should seek to provide a unifying account of belief updating by investigating circumstances under which these different metrics make differing cognitive and behavioral predictions.

The significant single-trial relationship between belief update size and P3 amplitude was restricted to a fronto-central midline electrode, with no evidence for a comparable effect at centro-parietal midline electrodes. This partition corresponds to a distinction drawn between the following two subcomponents of the P3: the fronto-central P3a and the centro-parietal P3b (Polich, 2007). In the present study, the P3a, but not the P3b, was an index of belief update size. In this regard, it is of particular interest that a previous study by Mars et al. (2008) found that feedback surprise, but not belief update size, was encoded in the P3b subcomponent, leading the authors to speculate that the P3a component may encode update size but not surprise. This proposal received empirical support from our findings. The observed results are broadly consistent with recent research investigating Bayesian single-trial properties of the P3 in a prediction task without reinforcement (Kolossa et al., 2015). Furthermore, the dissociation between frontal encoding of belief update size and parietal encoding of surprise is consistent with evidence from functional magnetic resonance imaging research. O’Reilly et al. (2013) measured brain activity during a saccadic eye movement task, and found that, whereas belief update size was encoded in ACC, surprise was encoded in posterior parietal cortex. Convergent methodologies, therefore, have shown that belief update size is encoded in both ACC and in the fronto-central P3a component of the ERP. Since the ACC has been proposed as a possible source of the P3a (Volpe et al., 2007), these results may be manifestations of the same underlying process. However, we note that since we did not use a standard P3a paradigm with novel nontarget distractors, it is possible that the P3a component encoding belief update size in the present study might also simply be labeled an anterior P3. To date, this nomenclature remains ambiguous (Luck, 2005; Polich, 2007).

A link between the P3 and belief updating has the potential to unify a number of disparate experimental findings. Larger P3 potentials are elicited by infrequent stimuli (Sutton et al., 1965), novel stimuli (Friedman et al., 2001), and stimuli imparting information (Sutton et al., 1967). Since these manipulations each vary the extent to which participants must revise an internal model of the environment, belief updating might be considered a general principle linking each of these observations. Moreover, the Bayesian perspective is broadly compatible with context-updating theory, which proposes that P3 amplitude reflects the revision of schemata concerning stimulus context (Donchin and Coles, 1988). Prior beliefs in the Bayesian sense are conceptual cognates of context schemata, and belief updating equivalent to schema revision. Of course, a Bayesian framework cannot account for all manipulations that affect P3 amplitude (Kopp, 2008). Other important manipulations include effects of stimulus value (Begleiter et al., 1983; Sato et al., 2005) emotional salience (Johnston et al., 1986), and target/nontarget status (for review, see Squires et al., 1975). The triarchic model of Johnson (1986) suggests that both the transmission of information (analogous to the effect of a Bayesian belief update) and stimulus meaning contribute to the amplitude of the P3. Since stimulus meaning was not manipulated in the present study, we are unable to assess how its effects might have interacted with observed effects of belief updating. Integrating these manipulations is a task for future research.

The present study also observed a significant negative relationship between belief uncertainty and prefeedback SPN amplitude. At electrodes C3 and P4, larger SPN components were observed in trials in which participants’ beliefs were more certain. The SPN has previously been linked to the anticipation of feedback that provides response reinforcement (Damen and Brunia, 1994). The left central electrode C3 was situated over primary motor cortical areas responsible for the right index finger button press that indicated participants’ choices. The observed association between uncertainty and SPN amplitude at C3 may therefore reflect motor learning, since preparatory neural activity in motor cortex is known to be associated with rapid visuomotor learning (Muellbacher et al., 2001; Paz et al., 2003). Likewise, encoding of belief uncertainty at electrode P4 may reflect anticipatory prefeedback processing, which is consistent with previous studies showing involvement of parietal SPN in reward processing (Kotani et al., 2003). However, we note that, whereas the present study found a negative association between SPN amplitude and uncertainty, one recent study (Catena et al., 2012) found a positive effect at frontal electrodes. Of course, it is problematic to compare frontal with central and parietal SPN, since different regions are likely to be recruited in different cognitive processes. Nevertheless, an important difference between the present study and that of Catena et al. (2012) pertains to the operationalization of uncertainty. We used a task in which uncertainty was reducible: with practice, participants could become more certain about the contrast–reward mapping. By contrast, Catena et al. (2012) tested irreducible uncertainty by varying cue–outcome association strength. The resultant use of different cognitive processes may explain the discrepancy between electrophysiological findings. Further-
more, we note that the SPN is generally elicited only during the period prior to the occurrence of a stimulus. As such, our finding that SPN amplitude indexes uncertainty is specific to the case of temporal anticipation and does not necessarily fully define a general principle of the neural encoding of uncertainty. Future research should seek to determine how belief uncertainty is encoded when there is not a well defined future time at which uncertainty will be resolved.

In the P3 analysis, an additional factorial control analysis demonstrated that single-trial regression results were unlikely to have been affected by the possible confound of trial number. In the SPN analysis, by contrast, a relationship between trial number and belief uncertainty was an inherent feature of the learning task used in the present study, it was not possible to rule out a possible mediating effect of trial number on the relationship between SPN amplitude and belief uncertainty. Further research is required to determine whether the relationship between SPN amplitude and belief uncertainty holds even when uncertainty is not monotonically decreasing as a function of trial number.

In the present study, our intention was not to give a complete overview of the ERP correlates of feedback processing, but rather to investigate the role in belief updating of two particular ERP components (the P3 and SPN) that have been implicated in belief updating by past research. Indeed, the general neural response to feedback is likely to recruit many processes other than just those associated with the P3 and SPN, and research using different experimental tasks from the present study has identified other ERP components involved in learning from feedback. In particular, a large body of research suggests the importance of the feedback-related negativity (FRN; Eriksen and Stoffregen, 1973), the feedback potential (FP; Donchin et al., 1981). This component has been strongly linked to the evaluation of feedback outcomes (Yeung and Sanfey, 2004; Achtziger et al., 2015) and has been theorized to index the magnitude of a reward prediction error associated with reinforcement learning (Holroyd and Coles, 2002). Given this theory, in the present study we would have expected the FRN to encode not the size of a belief update, or the uncertainty of beliefs per se, but the valence of feedback outcomes relative to participants’ expectations. This is conceptually a separate aspect of learning from the model-based definition of belief updating used in the present study. Furthermore, a recent review (Lutf, 2014) noted that it is problematic to investigate the FRN in tasks such as that used in the present study, in which reward and performance feedback are delivered contemporaneously. Since the task used in the present study was not optimized for the investigation of the FRN component, we chose to exclude the FRN from our model-based single-trial regression analysis. Future research should investigate the interaction of the FRN with the ERP components identified in the present study by making use of a belief-updating task in which reward and performance feedback are orthogonal.

Finally, we note that while we assessed belief updating within a Bayesian framework, there is evidence that humans also perform non-Bayesian belief updating in some circumstances (Hogarth and Einhorn, 1992; Stern et al., 2010). We do not make the strong claim that all neural computations underlying perceptual learning take place according to Bayesian principles; instead, it is likely that the ability of decision makers to make use of Bayesian updating is constrained by the complexity of the decision situation and by the availability of heuristic alternatives to Bayesian updating (Achtziger et al., 2014, 2015). However, the results of the present study show that a Bayesian updating model outperformed a non-Bayesian heuristic model for a relatively simple perceptual learning task. Non-Bayesian belief updating may have distinct ERP correlates in more complex environments, as suggested by Achtziger et al. (2014, 2015), and further research is required to reconcile these perspectives.

In summary, the present study provides evidence that single-trial EEG data can be used to track the evolution of latent states of belief in humans. Our results build an empirical bridge between general theories of belief updating in cognition and a long tradition of research into the functional significance of ERPs. More broadly, our findings are a novel demonstration of the value and viability of computational cognitive modeling in EEG research.

References


September/October 2015, 2(5) e0076-15.2015 eNeuro.sf.nog


5.3 Summary

In sum, Study 1 found that single-trial amplitudes of two ERP components were associated with two distinct belief variables. Specifically, this study found that single-trial P3 amplitude was positively associated with belief update size, and that single-trial SPN amplitude was negatively associated with belief uncertainty. In addition, the results of Study 1 also revealed that a Bayesian belief updating model provided a better overall account of participants’ choice behaviour than a WSLS heuristic model. Interestingly, however, there was considerable heterogeneity in models fits across participants, such that a minority of participants (five of sixteen) produced behavioural data more consistent with the WSLS heuristic than with Bayesian belief updating.

Several additional questions concerning the neural substrates of belief updating were raised by these findings. First, Study 1 did not provide a mechanistic account of how the P3 component was implicated in Bayesian belief updating. Indeed, an open question is whether the observed relationship between P3 amplitude and belief update size was functional or epiphenomenal. The results of Study 1 are consistent with the hypothesis that anterior P3 amplitude indexes the output of a neural process which actively implements Bayesian belief updating (Kolossa et al., 2015; Mars et al., 2008); however, it is also possible that the observed association may have resulted from a relationship between P3 amplitude and other neurocognitive processes that simply happened to coincide with belief updating in Study 1. One such potential explanation may be reward valence processing (Johnson, 1986), since the structure of the perceptual learning task was such that more rewarding pieces of reinforcement tended to provide more information about the target contrast than zero-reward feedback. In order to provide more compelling support for the Bayesian belief updating model, it would be necessary to demonstrate that the involvement of the P3 in belief updating was independent of the processing of the rewarding aspects of monetary feedback.

Finally, it is interesting to note that a secondary finding of Study 1 was that a number of participants behaved more in accordance with a WSLS heuristic than with Bayesian belief updating. The results of Study 1 alone do not permit speculation as to why this might be, though the theory of bounded rationality suggests that the use of resource-cheap heuristics such as WSLS over computationally expensive Bayesian strategies may be driven by cognitive resource limitations (Gigerenzer & Goldstein, 1996). As such, the presence of monetary rewards for performance in the perceptual
learning task used in Study 1 may have been a factor which motivated the majority of participants’ use of Bayesian belief updating. This explanation would predict that a variant on this perceptual learning task with non-monetary performance feedback would be associated with lower overall task motivation, and therefore with a greater proportion of participants adopting a WSLS strategy.
6. Study 2: Electrophysiological indices reflect switches between Bayesian and heuristic strategies in perceptual learning

This chapter presents an EEG study investigating the effect of task motivation on belief updating among participants completing a variant of the perceptual learning task introduced in Study 1. A model-based cluster analysis (Fraley & Raftery, 1998) was used to investigate differences in behavioural and neural data as a function of whether participants consistently adopted a Bayesian belief updating task strategy, or switched between Bayesian and heuristic task strategies depending on the motivational value of feedback. This chapter is a manuscript under review at the journal *Neuropsychologia*, and will be presented in this format.

6.1 Aims of Study 2

Study 2 sought to investigate several issues concerning the neural substrates of belief updating left unresolved by Study 1. Specifically, this study aimed to assess the effect of a feedback reward manipulation on participants’ use of Bayesian versus heuristic task strategies. As well as addressing this behavioural research question, the feedback manipulation also allowed Study 2 to investigate whether the neural substrates of belief updating identified in Study 1 were specific to use of a Bayesian belief updating strategy, or whether they were also implicated when participants adopted a WSLS heuristic strategy.

To investigate these questions, Study 2 introduced a within-participants feedback manipulation to the perceptual learning task developed by Study 1. For this feedback manipulation, participants completed the perceptual learning task in one of two feedback conditions: monetary, or instructive. The monetary feedback condition was identical to the design of Study 1, in which participants received larger monetary rewards for responses closer to the target contrast. By contrast, the novel instructive feedback condition developed for Study 2 delivered performance feedback to participants in the form of an instructional directive (e.g., ‘Your choice was 12 per cent from the target’). Crucially, the instructive feedback condition was designed such that the functional form of the feedback mapping was equivalent to the monetary feedback condition; as such, the information content of feedback was held constant across conditions. This permitted differences in performance between conditions to be interpreted with respect to participants’ task motivation, rather than to incidental
factors. All participants completed both the instructive and the monetary feedback conditions, with condition order counterbalanced across participants.

To investigate the neural substrates of Bayesian strategies versus heuristic strategies, Study 2 used a model-based cluster analysis based on the two computational cognitive models introduced in Study 1. This analysis classified participants into subgroups depending on patterns of switching between Bayesian and heuristic strategies according to the motivational value of feedback. Then, using model subgroup membership as a between-participants grouping variable in subsequent ERP analyses, Study 2 assessed the effects of Bayesian versus heuristic strategies on the neural encoding of feedback in different ERP components. This allowed Study 2 to test the functional role played by the P3 component in Bayesian belief updating, and thereby to address an open research question from Study 1.

6.2 Manuscript
ABSTRACT

Given finite cognitive resources, agents should allocate these to maximise desirable outcomes while minimising cognitive effort. This trade-off has often been studied as a competition between Bayesian inference and ‘fast-and-frugal’ heuristic strategies. An important open question in this regard is whether utilisation of Bayesian inference is dependent upon motivational state, and how this is reflected in the brain. We recorded electroencephalography from 23 participants performing a perceptual learning task with both monetary and a non-monetary instructive feedback conditions. Using model-based cluster analysis, we found that only participants who switched between a Bayesian and a heuristic strategy showed worse performance for instructive than monetary feedback, whereas participants who consistently employed Bayesian inference showed equivalent performance. This pattern was mirrored by differences in neural encoding of feedback in two event-related potential components: the P3, and the late positive potential. Findings suggest that Bayesian inference in perceptual learning critically depends on motivational state.

Keywords: event-related potential, Bayesian brain, model-based cognitive neuroscience, P3, feedback-related negativity, late positive potential
1. INTRODUCTION

Humans possess finite cognitive resources. In judgment and decision making these resources should be allocated so as to optimise decisions about behaviourally relevant outcomes, while minimising expenditure of cognitive resources on irrelevant or inconsequential tasks (Pitz and Sachs, 1984; Simon, 1956). Cognitive resource constraints are thought to provide a principled explanation for the finding that in many tasks humans rely on simple heuristics rather than adopting superior but more computationally demanding task strategies (Boureau and Dayan, 2011; Goldstein and Gigerenzer, 2002; Tversky and Kahneman, 1973, 1974). By producing reasonably accurate choices while consuming relatively few cognitive resources, heuristics can be a valuable tool to overcome cognitive resource constraints (Conlisk, 1996; Gigerenzer and Goldstein, 1996). However, the cognitive and neurophysiological factors affecting the use of optimal strategies versus resource-cheap heuristics remain unclear. In particular, one open question is how the presence of reward and motivational state might affect strategy choice (Achtziger et al., 2015; Charness and Levin, 2005). This question is of particular importance given ongoing debate in educational psychology and personnel economics regarding the efficacy of using rewards to incentivise performance (Hidi, 2015; Lazear, 2000).

In reinforcement learning research, one way that the trade-off between optimal and heuristic strategies has been conceptualised is as a competition between Bayesian inference, which is statistically ideal but computationally demanding, and a win-stay lose-shift (WSLS) heuristic (Bennett et al., 2015; Charness and Levin, 2005; Steyvers et al., 2009). Whereas Bayesian inference involves repeatedly revising a complete prior belief distribution, the WSLS heuristic assumes that participants simply select choice options which have previously led to reinforcement, or shift to a new option if a previous choice is not reinforced (Robbins, 1952). Although Bayesian models often
fit learning behaviour well overall, their goodness-of-fit deteriorates sharply as the
cognitive demands of Bayesian inference increase (Payzan-LeNestour and Bossaerts,
2011), or when a WSLS heuristic conflicts with Bayesian inference (Achtziger et al.,
2015; Charness and Levin, 2005). Moreover, even in studies where Bayesian models
fit group-level data well, a substantial proportion of participants nevertheless made
choices better explained by a WSLS heuristic (Bennett et al., 2015; Steyvers et al.,
2009). Crucially, it remains to be determined how these findings can be reconciled
within Bayesian theories of cognition such as the ‘Bayesian brain’ hypothesis (Knill
and Pouget, 2004). Such theories, which include both weak and strong variants, claim
that much of human behaviour can be explained as a form of Bayesian inference (e.g.
Chater and Oaksford, 2008; Friston et al., 2015).

One possible explanation for these findings is that motivational factors may
affect the use of Bayesian versus heuristic strategies. Given the greater cognitive
demands of Bayesian inference compared with simple heuristics, participants may
require more task motivation to employ a Bayesian strategy. As a result, individual
differences in motivational state are likely to result in the utilisation of different task
strategies and the recruitment of different neural processes by different participants.
One corollary of this explanation is that, since feedback delivered in the form of
monetary reward is thought to enhance participants’ cognitive control (Fröber and
Dreisbach, 2014; Jimura et al., 2010), providing rewarding performance feedback
should incentivise the use of Bayesian inference strategies in complex task
environments.

In the present study, we investigated the effect of motivational state on the use
of Bayesian versus heuristic strategies in a perceptual learning task with graded
feedback. In this task, feedback was delivered in the form of either monetary reward
or instructional directives. Importantly, feedback values were constrained such that
the exact amount of information in feedback was identical across the two feedback conditions. Then, to assess the question of strategy selection in a principled way, we formulated competing computational cognitive models implementing Bayesian, heuristic, or mixed strategies, and compared the predictions of these models with behavioural data. In order to elucidate the neural mechanisms underlying selection of Bayesian versus heuristic strategies, we investigated the effect of feedback condition on three event-related potential (ERP) components associated with learning and/or reward processing: the P3 (Polich, 2007), the feedback-related negativity (FRN; Yeung and Sanfey, 2004), and the late positive potential (LPP; Ito et al., 1998). Finally, using a model-based cluster analysis, we investigated how neural encoding of feedback differed between participant subgroups employing different task strategies.

2. MATERIALS AND METHODS

2.1 Participants

Twenty-three participants were recruited from among students of the University of Melbourne, Australia (mean age = 23.40; age range 19-31; 17 female, six male). Participants were right-handed and had normal or corrected-to-normal visual acuity. Exclusion criterion was a medical history of any neurological disorder. Informed consent was acquired from all participants in accordance with the Declaration of Helsinki, and approval was obtained from the University of Melbourne Human Research Ethics Committee (ID 1339694). Participants received monetary compensation for participation (mean = $25.24; SD = 4.05) that was proportional to task winnings in the monetary feedback condition only (see below). For all participants, total remuneration value was within the range AUD $20-30.

Four participants were excluded from analysis of EEG data: one because of an excessive number of artefacts (more than 80 percent of trials affected), one because of
a failure of the eyeblink artefact removal routine, and two additional participants because of computer error during EEG acquisition. Final EEG analyses were therefore performed on data provided by 19 participants (mean age = 23.75; age range 19-31; 13 female, six male).

2.2 Behavioural task

While EEG was recorded, participants performed a perceptual learning task modified from a previous study by Bennett et al. (2015). This task required participants to use visually presented feedback in sequential trials to learn the target contrast of a greyscale checkerboard stimulus (see Figure 1). The present study employed a novel variant of this task in which feedback regarding the target contrast could be either monetary (as in the original paradigm) or instructive.
Figure 1. (A) Trial schematic. Following a self-paced button press, a checkerboard stimulus was presented whose contrast changed linearly. The participant could at any time select the contrast displayed on screen by pressing a button with the right index finger. The trial continued until a button was pressed, or until stimulus duration exceeded 30 seconds. Following the participant’s choice, the selected contrast remained on screen for two seconds, after which time the monetary or instructive feedback associated with the chosen contrast was displayed for 2.5 seconds. In the event that no button was pressed within 30 seconds, feedback was a message reminding the participant of the task instructions. (B) Feedback mapping for monetary feedback condition. The mapping was a symmetrical triangular function with a centre of zero percent contrast difference, a half-width of 15 percent contrast difference, and a height of 25 cents. As such, received reward was maximal when the participant responded at the target contrast, and decreased with increasing difference of chosen contrast from the target. Reward was zero for responses at greater than 15 percent distance. Feedback received was rounded to the nearest whole-cent value. (C) Feedback mapping for the instructive feedback condition. For responses at less than 15 percent difference from the target, participants were informed of the difference between the chosen
contrast and the target (rounded to the nearest of 49 equally spaced values, in order to match precisely the step size of the monetary condition’s feedback mapping). For responses at greater than 15 percent difference from the target, participants were informed only that their response was “too far” from the target (equivalent to the monetary condition’s zero cent feedback).

In each trial, the checkerboard was presented for up to 30 seconds (see Figure 1A), during which time its contrast changed linearly (alternately increasing and decreasing, changing direction at upper/lower contrast bounds). Initial contrast, initial direction of contrast change (increasing/decreasing), and rate of change were randomised on each trial using the same parameters as in Bennett et al. (2015). At any time during stimulus presentation, the participant could choose the contrast displayed on screen by pressing a button with the right index finger. After a delay in which the chosen contrast remained on screen, participants received feedback regarding their chosen contrast. In the monetary condition, this feedback was presented in the form of monetary reward (e.g. “You won 15 cents”) according to a triangular function $M$ of the distance between the chosen and the target contrast (see Figure 1B). Responses closer to the target contrast earned proportionally more (up to a maximum of 25 cents per trial, rounded to the nearest integer), and participants received zero reward for responses at greater than 15 percent distance from the target:

$$M(r_t, x_t) = \begin{cases} 25 - \frac{5|r_t - x_t|}{3}, & |r_t - x_t| \leq 15 \\ 0, & |r_t - x_t| > 15 \end{cases}$$

where $t$ is the trial number, $r_t$ is the target contrast on trial $t$, and $x_t$ is the participant’s chosen contrast on trial $t$. Double bars denote rounding to the nearest integer.

By contrast, feedback in the instructive condition took the form of an explicit instructional directive informing the participant of the distance between their chosen
contrast and the target (e.g. “You were 11.25% away from the target”; see Figure 1C). For responses at greater than 15 percent distance from the target, participants were informed only that their response was ‘too far’ from the target. As such, in the instructive feedback condition the function reward mapping function $M$ from Equation (1) was replaced with the instruction mapping function $I$:

$$I(r_t, x_t) = \begin{cases} |r_t - x_t|, & |r_t - x_t| \leq 15 \\ "Too far", & |r_t - x_t| > 15 \end{cases}$$

(2)

Crucially, in order to ensure strict equivalence in feedback information between instructive and monetary feedback, instructive feedback values were constrained to follow an equivalent functional form to monetary feedback (compare Figures 1B and 1C). This was done by rounding instructive feedback values to the nearest value in the set \{0, 0.625, 1.25, 1.875, … 15\}. For any given sequence of choices, therefore, feedback in the two conditions provided identical information regarding the target contrast. Consequently, any differences in task performance between instructive and monetary conditions cannot be attributed to differences in the information content of feedback.

Prior to the task, participants were trained in interpretation of feedback in both feedback conditions, and testing commenced only when satisfactory levels of task understanding were displayed. Participants then completed 14 blocks of the task in total over approximately 50 minutes. Each block had a different target contrast, selected pseudo-randomly from the interval [25%, 85%]. Monetary and instructive conditions were presented in seven consecutive blocks each, with condition order counterbalanced across participants. Each block continued until cumulative checkerboard presentation duration for the block exceeded three minutes, or until 25
trials were completed, whichever occurred sooner. As a result, the number of trials per block varied, ensuring that participants could not rush through the task in an attempt to trade off experiment duration against monetary winnings. Upon receiving feedback, participants were not informed of the exact numerical contrast level of their choice; instead, the checkerboard remained on-screen at the chosen contrast while feedback was presented. As a result, learning was necessarily affected by perceptual uncertainty regarding the identity of the chosen contrast.

Stimuli were presented using a Sony Trinitron G420 CRT monitor at a framerate of 120 Hz. During task performance, participants were seated comfortably in a darkened room, using a chin rest at a distance of 77 cm from the screen. Checkerboard stimuli were 560 × 560 pixels in size, measuring 19.5 × 19.5 cm on the screen and subtending a visual angle of 14.43 × 14.43°. Responses were recorded using a five-button Cedrus Response Box. All other task parameters were identical to those employed by Bennett et al. (2015), with the exception that the checkerboard in the present task did not phase-reverse, and therefore had a smoothly changing (rather than a flickering) appearance.

2.3 Computational models of behaviour

We tested four competing computational cognitive models of the task by generating all possible configurations of Bayesian and heuristic strategies according to feedback condition. This model configuration permitted us to test formally several possible ways in which participants might have switched, or not have switched, between Bayesian and heuristic strategies as a function of feedback condition (see below for formal specifications and choice rules for the Bayesian and heuristic strategies).

The first model, termed the ‘Consistent Bayesian’ (CB) model, assumed that participants employed a Bayesian inference strategy irrespective of the feedback
condition. Similarly, the second model, the ‘Consistent Heuristic’ (CH) model, assumed that participants employed a WSLS heuristic strategy in both the monetary and the instructive feedback condition. A third possibility was that participants might employ a Bayesian inference strategy when monetary feedback was provided, but switch to a WSLS heuristic strategy in blocks with instructive feedback. Such a model implies the use of a more computationally demanding strategy in the presence of reward and a simpler heuristic strategy in the absence of reward, and was therefore termed the ‘Incentive-Compatible Switching’ (ICS) model. Finally, for the fourth model we also considered the possibility that participants might employ a WSLS heuristic strategy in monetary feedback blocks, and a Bayesian inference strategy in instructive feedback blocks. This model was termed the ‘Incentive-Incompatible Switching’ (IIS) model, and can be thought of as corresponding to the idea of “choking under pressure” (see also Achtziger et al., 2015; Baumeister, 1984, p. 610), in which the presence of monetary incentives produces a decrement in performance.

Each of these four models was parameterised by the perceptual uncertainty parameter $\sigma$ (see below); this parameter was permitted to vary between participants, but was fixed to take the same value across different feedback conditions for each participant, since neither perceptual stimuli nor observation environment varied across feedback conditions. Constraining the $\sigma$ parameter across feedback conditions was done for identification purposes, to ensure that differences in task performance between feedback conditions could not simply be accounted for by changes in a perceptual uncertainty parameter. One prominent criticism of Bayesian models in psychology and neuroscience is that flexible parameterisations of Bayesian models permit qualitatively and quantitatively distinct patterns of behaviour to be described within an identical model architecture (Bowers and Davis, 2012). The constraints on $\sigma$ applied in the present study ensured that the goodness-of-fit of each of the four
models described below gave an unbiased estimate of the likelihood of each behavioural strategy. For each model, \( \sigma \) was estimated using maximum likelihood estimation as implemented in the MATLAB Optimization Toolbox (The Mathworks, Natick, MA) and choice likelihood functions as specified by Bennett et al. (2015). Model fitting was performed with multiple different initial parameter values for each participant to ensure that identified values of sigma corresponded to global rather than local optima.

The relative performance of these four models also informed subsequent ERP analyses. By comparing model fits for individual participants we derived an index of which model strategy provided the best explanation of each participant’s data. Groups of participants who all used the same strategy were then clustered into subgroups for subsequent ERP analyses. Differences in neural encoding of feedback between subgroups were, therefore, interpretable with respect to the different behavioural strategies employed by different participants. This model-based clustering analysis is a principled alternative to more traditional model-free cluster analysis algorithms, which have been criticised for clustering data in order to maximise intra-cluster homogeneity in a way that may not result in psychologically meaningful differences between participant subgroups (Fraley and Raftery, 1998; Meehl, 1992). By contrast, a model-based clustering procedure allows for principled segregation of participants into subgroups representing different computational models of task performance. As such, different subgroups necessarily correspond to meaningful and distinct psychological constructs. In the present study, this approach meant that differences between subgroups in neural encoding of feedback could be readily interpreted with reference to the employment of different behavioural strategies. This approach can be considered a classification-based counterpart of Bayesian model selection (Marković
et al., 2015; Stephan et al., 2009), which also takes into account within-participants variability in the likelihood of different computational models.

It should also be noted that the model of Bayesian inference implemented in the present study is only one among many models capable of approximating Bayesian updating for feedback-based learning under uncertainty (see e.g. Behrens et al., 2007; Mathys et al., 2011; Nassar et al., 2010). The statistical mechanics of Bayesian updating are not unique to the model used in the present study; as such, this model should be considered as representing a specific implementation of broad computational principles which are applicable to a large number of distinct model architectures.

2.3.1 Bayesian Strategy

To model choices under the Bayesian inference strategy, we estimated beliefs using a Bayesian grid estimator (Moravec, 1988) as described and implemented for the perceptual learning task used in the present study by Bennett et al. (2015). This estimator calculated a probabilistic estimate of participants’ beliefs regarding the level of the target contrast in each trial, and used this belief distribution to estimate choice likelihoods. Formally, beliefs were described by a probability mass function $\theta$ over a contrast space discretised into $J$ equally sized bins, where the value of the function $\theta$ at each bin represented the participant’s subjective probability that the target contrast (denoted $r_t$) fell within bin $j$ on trial $t$. On each trial $t$, participants observed the feedback $f_t$ after the choice of contrast bin $x_t$, determined according to the monetary and instructive feedback mapping functions $M$ and $I$ as specified by Equations (1) and (2), respectively. Belief estimates were initialised in each block as a discrete uniform distribution, representing participants’ a priori uncertainty regarding the target contrast level. This belief distribution was then updated sequentially according to
Bayes’ Rule as feedback was received, such that the posterior distribution of trial $t$ formed the prior distribution for trial $t + 1$:

$$
\theta_{t+1}(j) = \frac{\theta_t(j)Pr(f_t, x_t | r_t \in j)}{Pr(f_t, x_t)}
$$

(3)

The left-hand side of Equation 3 represents the posterior belief distribution for contrast bin $j$ following trial $t$, and is calculated by multiplying the participant’s prior belief that the target contrast fell within bin $j$, $\theta_t(j)$ by the likelihood of observing the choice/feedback pair $(f_t, x_t)$ if the target were in bin $j$, $Pr(f_t, x_t | r_t \in j)$, and dividing by the marginal likelihood of the update, $Pr(f_t, x_t)$.

As described above, variability in task performance between participants was captured by the perceptual uncertainty parameter $\sigma$. Formally, $\sigma$ represents the standard deviation of the Gaussian noise affecting belief updates after feedback receipt, such that larger values of $\sigma$ indicate a greater degree of noise in the updating process, and therefore more imprecise belief updates. Since participants were not informed of exactly what contrast they had chosen, but had to estimate this chosen contrast from the visual display, this perceptual uncertainty therefore also results in a Gaussian prior over chosen contrast. For a complete discussion of the mathematical role of $\sigma$ in the Bayesian updating model see Bennett et al. (2015)

To estimate choice likelihood, this model used a probability of maximum utility choice rule (cf. Speekenbrink and Konstantinidis, 2015), whereby contrast bins with a higher probability of containing the target contrast had a proportionally higher probability of being chosen, subject to response uncertainty during choice:

$$
Pr(x_t) = \frac{(\theta * G_\theta)[J]}{k}
$$

(4)

As such, on each trial the choice likelihood probability mass function was determined by convolving the prior belief distribution $\theta$ by the uncertainty function $G_\theta$ over the set of contrast bins $J$, where $k$ is a normalisation constant and square
parentheses denote the domain of convolution. Intuitively, this response model implies that response probabilities are derived by the addition of Gaussian noise to the target contrast distribution $J$. The uncertainty function $G_0$ was a zero-mean Gaussian function of the contrast difference between the true chosen contrast $x_t$ and each bin $x_j$ of the distribution $\theta$. This function was also parameterised by $\sigma$:

$$G_0(x_t, x_j, \sigma) \equiv \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x_t - x_j)^2}{2\sigma^2}}$$  (5)

### 2.3.2 WSLS Heuristic Strategy

In contrast to the Bayesian inference strategy, the WSLS heuristic strategy did not assume that participants attempted to infer the location of the target contrast. Instead, behaviour under the WSLS heuristic was assumed to be driven by a one-trial memory, such that participants’ behaviour on a given trial was a function of whether or not they had received reinforcement on the preceding trial (Robbins, 1952). Specifically, the model assumed that participants attempted to repeat the choice made on the previous trial if this choice had resulted in any level of reinforcement (defined as any monetary reward amount in the monetary feedback condition, and as any numerical instruction amount in the instructive feedback condition). The WSLS model assumed that participants shifted randomly to a new contrast if they received no reinforcement on trial $t - 1$, or at the start of a new block. This gives the following choice probability function:

$$Pr(x_t) = \begin{cases} \frac{\delta(j - x_{t-1}) \ast G_0[\{J\}]}{k}, & f_{t-1} > 0 \\ \frac{1}{j}, & \text{otherwise} \end{cases}$$  (6)

where $G_0$ is defined as per Equation 5, $J$ is the number of bins in the belief distribution, $\delta$ is the Dirac delta function, and $k$ is a normalisation constant. See
Bennett et al. (2015) for a full formal characterisation of the WSLS heuristic as implemented for the perceptual learning task used in the present study.

Note that perceptual uncertainty in the WSLS choice rule was implemented in an identical manner to the Bayesian inference choice rule, using the same zero-mean Gaussian function. This allowed for an identical specification of perceptual uncertainty across all four models, thereby ensuring that predicted choice probabilities were directly comparable between models.

2.4 EEG data acquisition

The electroencephalogram was recorded from 35 Ag/AgCl active scalp electrodes (Fp2, AF7, AF3, AFz, AF4, AF8, F5, F1, Fz, F2, F4, F6, FC1, FCz, FC4, FC6, C5, C3, Cz, C4, CP5, CP3, CP1, CPz, CP6, P5, P1, Pz, P4, P6, POz, PO8, O1, Oz, Iz in the International 10-20 System). Electrodes interfaced with a BioSemi ActiveTwo 64-channel system running ActiView acquisition software, and used an implicit reference during recording. Due to technical problems with electrode hardware, not all 64 channels could be recorded for all participants. Therefore, based on previous (Bennett et al., 2015) and planned analyses, data was acquired from prespecified channels of interest, including all fronto-central and centro-parietal midline electrodes. All electrode channels included in subsequent event-related potential (ERP) analyses were recorded without issue for all participants, and data quality was not compromised. Data were linearly detrended and re-referenced offline to an average of left and right mastoid electrodes. The vertical and horizontal electrooculogram (EOG) were recorded from electrodes infraorbital and horizontally adjacent to the left eye. EEG was recorded at a sampling rate of 512 Hz.

During preprocessing, data were first manually screened to exclude epochs contaminated by skin potential or muscle artefacts. Using a linear FIR filter, data were
then highpass filtered at 0.1 Hz, lowpass filtered at 70Hz, and notch filtered at 50Hz to remove background electrical noise. Epochs were generated consisting of data from 1500 milliseconds before to 1500 milliseconds after feedback presentation. An independent components analysis (ICA) as implemented in the EEGLAB toolbox (Delorme and Makeig, 2004) was performed on the resulting dataset to identify and remove components related to eye movements and eye-blink artefacts. Finally, an automatic artefact screening procedure excluded all epochs from analysis in which maximum/minimum amplitudes exceeded ±200 µV.

2.5 ERP analysis

We assessed three ERP components: the P3, the feedback-related negativity (FRN), and the late positive potential (LPP). Component amplitudes were calculated using estimation routines implemented in the ERPlab plugin (Lopez-Calderon and Luck, 2014), time-locked to feedback presentation on each trial and baseline-corrected from 0 to 500 milliseconds pre-feedback.

P3 amplitude was calculated as the largest positive peak in the window from 250-550ms post-feedback at the frontocentral and centroparietal midline electrodes AFz, Fz, FCz, Cz, and CPz (Bennett et al., 2015). This time window allowed us to estimate peak amplitude within a symmetrical window about the peak of the P3 as identified in grand average waveforms, and was consistent with consensus estimates of P3 analysis windows in past research (see, e.g., Kolossa et al., 2015; Mars et al., 2008; Polich, 2007). Past research suggests that P3 amplitude is an index of individuals’ revision of probabilistic beliefs (Kolossa et al., 2015; Kopp, 2008), and we therefore investigated whether P3 amplitude varied between Bayesian belief updating and heuristic strategies.
At the same midline electrodes as the P3 analysis, FRN amplitude was calculated as the peak-to-peak distance between the most negative peak in the window from 200 to 550 milliseconds and the immediately preceding positive peak. A peak-to-peak measure of the FRN was used rather than a mean amplitude measure to ensure that estimates of FRN and P3 amplitude were statistically independent of one another, and the time window for identification of the FRN peak was selected based upon consensus in previous literature (see, e.g., Achtziger et al., 2015; Frank et al., 2005; Yeung and Sanfey, 2004). FRN amplitude was investigated because of its importance as an index of outcome evaluation in reinforcement learning and feedback processing (Frank et al., 2005; Holroyd and Coles, 2002; Yeung and Sanfey, 2004).

Finally, LPP amplitude was calculated as the mean voltage within the window from 550 to 900 milliseconds post-feedback at the centro-parietal midline electrodes Cz, CPz, and Pz (Hajcak et al., 2009; Ito et al., 1998). This time window was chosen both to accord with previous literature (e.g. Keil et al., 2002), and to ensure that P3 and LPP analysis windows did not overlap. In research studying the processing of emotional stimuli, LPP amplitude is thought to differentially encode positive and neutrally valenced stimuli (Keil et al., 2002; Schupp et al., 2000); we therefore sought to investigate whether LPP amplitude differed between monetary and instructive feedback conditions.

ERP analyses investigated the neural correlates of differential processing of monetary and instructive feedback. Where formal comparison of computational cognitive models indicated the presence of participant subgroups using distinct performance strategies, we investigated interactions between model-derived participant subgroups and feedback condition. This allowed us to identify electrophysiological indices associated with the use of different behavioural strategies in different subgroups.
3. RESULTS

3.1 Behavioural results

Participants completed a variable number of trials per block (mean = 17.57; SD = 2.70). A paired-samples $t$-test indicated that the average number of trials completed per block did not differ between instructive and monetary conditions ($t(22) = 0.33, p = .74$).

Behavioural performance was quantified by choice error, defined as the absolute contrast difference between the chosen contrast and the target contrast on each trial. We investigated differences in choice error as a function of trial number and feedback condition using linear mixed-effects analysis with feedback condition and trial number as fixed effects. Results indicated a significant main effect of trial number ($F(24,60.95) = 13.09, p = 4 \times 10^{-16}$), with performance improving over time within each block (see Figure 2A), and a significant main effect of feedback condition ($F(1,9.29) = 20.07, p = .001$), driven by better overall performance in the monetary than the instructive feedback condition. In addition, we observed a significant interaction between feedback condition and trial number ($F(24, 60.95) = 2.89, p = .0004$). This effect is likely to have been driven by greater differences between monetary and instructive feedback conditions in mid- and late-block trials, rather than in block-initial trials. To test this contention, we divided blocks into three sections: initial (trials 1-5), middle (trials 6-10) and late (trials 11-15), and used paired-sample $t$-tests to assess whether monetary and instructive conditions were associated with different performance in each section. Consistent with the conclusion above, we found better overall performance in the monetary than the instructive block for middle trials ($t(20) = -3.40, p = .003$) and late trials ($t(20) = -3.08, p = .006$), but not for initial trials ($t(20) = -1.87, p = .08$). Such a pattern stands to reason, since participants began
each block with no *a priori* knowledge regarding the target contrast, and were as likely to make a correct as an incorrect initial guess regardless of feedback condition.

![Figure 2](image)

**Figure 2.** (A) Overall task performance among all participants (*n* = 23). All participants’ mean choice error (measured as absolute difference between chosen and target contrasts) as a function of feedback condition (red: monetary feedback; black: instructive feedback) and trial number. Only trials 1-15 are shown, since data was not available for all participants for trials 16 – 25. Error bars represent the standard error of the mean. (B) Overall behavioural results among the subgroup of participants whose choices were best fit by the Consistent Bayesian (CB) model in a two-model comparison (*n* = 13). (C) Overall behavioural results among the subgroup of participants whose choices were best fit by the Incentive-Compatible Switching (ICS) model in a two-model comparison (*n* = 10).

### 3.2 Computational model results

Using standard model comparison techniques, we next determined which of the computational models defined above provided the best account of choices across participants. Table 1 presents Bayesian Information Criterion (BIC) values for each of the four models. Results showed that, as in a previous study using this perceptual learning paradigm, the CB model, which assumed that participants adopted a Bayesian inference strategy in both feedback conditions, provided the best fit to data.
across all participants (Bennett et al., 2015). However, further examination of model fits for individual participants using participant-specific BIC values (see ‘n best fit’ columns) revealed that there was considerable variability in the best-fitting model across participants. Indeed, in spite of providing the best overall fit to data, the CB model was the best-fitting behavioural model for fewer than half \((n = 11)\) of all participants considered separately. This strongly suggests the presence of inter-individual heterogeneity in task strategy. Closer inspection of individual model fits revealed that the second-best-fitting model overall, the ICS model, was the best-fitting model for approximately an additional third of participants. This model assumed that participants switched between a Bayesian inference strategy in the monetary feedback condition and a WSLS heuristic strategy in the instructive feedback condition.
Table 1. Summary of behavioural model fits.

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameters per participant</th>
<th>-LL</th>
<th>BIC</th>
<th>Participants best fit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Four-model comparison</td>
</tr>
<tr>
<td>CB</td>
<td>1</td>
<td>26094.18</td>
<td>52387.07</td>
<td>11</td>
</tr>
<tr>
<td>ICS</td>
<td>1</td>
<td>26178.47</td>
<td>52555.66</td>
<td>7</td>
</tr>
<tr>
<td>IIS</td>
<td>1</td>
<td>26254.21</td>
<td>52707.13</td>
<td>2</td>
</tr>
<tr>
<td>CH</td>
<td>1</td>
<td>26344.96</td>
<td>52888.63</td>
<td>3</td>
</tr>
</tbody>
</table>

BIC: Bayesian Information Criterion; -LL: Negative log-likelihood; CB: Consistent Bayesian model; ICS: Incentive-Compatible Switching model; IIS: Incentive-Incompatible Switching model; CH: Consistent Heuristic model. Overall fit statistics are based on 5651 choices by 23 participants in total.

As such, more than three-quarters of participants were best-fit by either the CB or the ICS model. Furthermore, for four of the five participants best fit by one of the other two models (IIS or CH), the second-best-fitting model was either the CB or the ICS model, such that the CB and ICS models were together either the first- or second-choice model for 22 of the 23 participants. Given the marked superiority of the CB and ICS models, therefore, we divided participants into two subgroups using a two-model comparison of the CB and ICS models (Fraley and Raftery, 1998). This model-based cluster analysis produced a two-model solution in which the principle of parsimony was balanced against the evident heterogeneity in individual task strategies detailed in Table 1.

We termed the two approximately evenly sized participant subgroups resulting from this two-model comparison the ‘CB’ ($n = 13$) and ‘ICS’ ($n = 10$) subgroups respectively. Subgroup membership was included as a between-subjects grouping variable in all subsequent behavioural and ERP analyses, in order to investigate
whether patterns of learning and neural responses to feedback differed between participant subgroups. Furthermore, since the order in which feedback conditions were presented was counterbalanced across participants, we performed an additional control analysis to ensure that the two behavioural subgroups identified by model comparison were not merely a reflection of between-participants differences in condition order. A chi-square test of independence revealed no relationship between condition order and model subgroup ($\chi^2(1) = 0.73, p = .39$); as such, there was no evidence to suggest that the classification of participants into behavioural subgroups was related to the order in which participants completed the instructive and monetary feedback conditions.

Next, we sought to determine whether behavioural strategy (as indexed by subgroup membership) was associated with different levels of overall task performance. We used a $2 \times 2$ repeated-measures ANOVA with within-groups factor of condition (instructive, monetary) and between-groups factor of model subgroup (CB, ICS), and mean choice accuracy across all trials as a dependent variable. We found a significant effect of feedback condition on accuracy ($F(1,21) = 20.88, p = .00017, \eta_p^2 = .50$), indicating better overall performance in the monetary condition than the instructive condition. Crucially, we also found a significant interaction between reward condition and model subgroup ($F(1,21) = 8.13, p = .01, \eta_p^2 = .28$). Follow-up paired-samples $t$-tests revealed that this interaction was driven by a significantly better overall performance in the monetary than the instructive condition in the ICS subgroup ($t(9) = -4.01, p = .003$; see Figure 2C), but not in the CB subgroup, ($t(12) = -1.66, p = .12$; see Figure 2B). In addition, we observed a non-significant trend toward better performance overall among the CB subgroup than among the ICS subgroup ($F(1,21) = 3.71, p = .07$).
3.3 ERP results

We next investigated whether any of the three identified ERP components displayed an interaction between feedback condition and behavioural subgroup that might account for the analogous interaction observed in behavioural data (see Figure 3). This analysis allowed us to identify electrophysiological indices which were associated with the differential relative performance between monetary and instructive feedback in participants who consistently adopted a Bayesian inference strategy (the CB model subgroup), compared with participants who switched between a Bayesian inference strategy and a heuristic strategy in different feedback conditions (the ICS subgroup). Scalp maps for the P3 and LPP analysis windows are presented in Figure 4.

![Graphs showing ERP results](image)

**Figure 3.** Grand average feedback-locked ERP waveforms at electrode Fz, grouped by feedback condition (red: monetary feedback; black: instructive feedback). Time 0 denotes the presentation of performance feedback. (A) Grand average waveforms across all participants. (B) Grand average waveforms across the subgroup of participants whose choices were best fit
by the Consistent Bayesian (CB) model in a two-model comparison ($n = 10$). (C) Grand average waveforms across the subgroup of participants whose choices were best fit by the Incentive-Compatible Switching (ICS) model in a two-model comparison ($n = 9$).

**Figure 4.** Scalp map of mean voltage difference between monetary and instructive feedback (in µV) for different model subgroups and analysis windows. Group-level voltage differences were largely driven by differences among participants in the Incentive-Compatible Switching (ICS) model subgroup, whereas voltage differences were small among participants in the Consistent Bayesian (CB) subgroup. Top row (A-C): P3 analysis window (250-550ms post-feedback). Bottom row (D-F): LPP analysis window (550-900ms post-feedback). Left column (A and D): Mean voltage difference pooled across all participants. Centre column (B and E): Mean voltage difference pooled across participants in the CB model subgroup. Right column (C and F): Mean voltage difference across participants in the ICS subgroup. For all scalp maps, voltages at missing electrodes have been reconstructed using spline interpolation for display purposes only.

### 3.3.1 P3

A $5 \times 2 \times 2$ repeated-measures ANOVA with within-groups factors of electrode (AFz, Fz, FCz, Cz, CPz) and feedback condition (instructive, monetary) and between-groups factor of model subgroup (CB, ICS) revealed a significant main effect of electrode on P3 amplitude ($F(1.50, 25.56) = 13.94, p = .0002, \eta_p^2 = .45$), as well as a significant interaction between feedback condition and model subgroup on P3 amplitude ($F(1,17)$
Follow-up paired-samples $t$-tests marginalised across electrodes revealed that this interaction was driven by a significantly larger P3 amplitudes for monetary than instructive feedback in the ICS subgroup ($t(8) = -2.88, p = .02$), but not in the CB subgroup ($t(9) = .29, p = .78$). This indicates that P3 amplitudes differed between feedback conditions solely for participants who switched between a Bayesian strategy in the monetary feedback condition and a heuristic strategy in the instructive feedback condition; by contrast, there was no difference in P3 amplitudes for participants who employed a Bayesian strategy in both feedback conditions.

Finally, we also observed a non-significant trend toward a main effect of feedback condition on P3 amplitude ($F(1,17) = 3.32, p = .09$). No other main effects or interactions were significant, all $p > .10$.

### 3.3.2 FRN

A $5 \times 2 \times 2$ repeated-measures ANOVA with within-groups factors of electrode (AFz, Fz, FCz, Cz, CPz), condition (instructive, monetary) and between-groups factor of model subgroup (CB, ICS) revealed a significant main effect of feedback condition on FRN amplitude ($F(1,17) = 5.92, p = .03, \eta_p^2 = .26$), with larger FRNs elicited by instructive than rewarding feedback. We also observed a non-significant trend toward an effect of model subgroup on FRN amplitude ($F(1,17) = 4.20, p = 0.06$), with numerically larger FRN amplitudes for the ICS than the CB subgroup (see Figure 5B). There was no interaction between feedback condition and model subgroup, and no other main effects or interactions were significant, all $p > .10$.

### 3.3.3 LPP
A 3x2x2 repeated measures ANOVA with within-groups factors of electrode (CPz, Pz, Cz) and feedback condition (instructive, monetary) and between-groups factor of model subgroup (CB, ICS) revealed a significant main effect of feedback condition on LPP amplitude ($F(1,17) = 21.38$, $p = .0002$, $\eta_p^2 = .56$), as well as a significant interaction between feedback condition and model subgroup ($F(1,17) = 4.85$, $p = .04$, $\eta_p^2 = .22$; see Figure 5C). Follow-up paired-samples $t$-tests marginalised across electrodes indicated that this interaction was driven by a significantly larger LPP amplitudes for monetary than instructive feedback in the ICS subgroup ($t(8) = -8.25$, $p = .00002$), but not for the CB subgroup ($t(9) = -1.39$, $p = .20$). This indicates that LPP amplitudes differed between feedback conditions only among participants who switched between a Bayesian strategy in the monetary feedback condition and a heuristic strategy in the instructive feedback condition; by contrast, there was no difference in LPP amplitudes for participants who employed a Bayesian strategy in both feedback conditions.

No other main effects or interactions were significant, all $p > .10$. 

![Graph A) P3](image1)

![Graph B) FRN](image2)

![Graph C) LPP](image3)
Figure 5. ERP component amplitudes as a function of feedback condition and behavioural subgroup (red: monetary feedback; black: instructive feedback). Error bars represent the standard error of the mean. * indicates $p < .05$. (A) P3 amplitude marginalised across electrodes AFz, Fz, FCz, Cz, and CPz. We observed a significant interaction between feedback condition and model subgroup, such that monetary feedback was associated with larger P3 components than instructive feedback for the ICS subgroup, but not for the CB subgroup. (B) FRN amplitude marginalised across electrodes AFz, Fz, FCz, Cz, and CPz. FRNs were significantly larger in the instructive feedback condition than the monetary feedback condition across both model subgroups. (C) LPP amplitude marginalised across electrodes Cz, CP, and Pz. We observed a significant interaction between feedback condition and model subgroup, such that monetary feedback was associated with larger LPP components than instructive feedback for the ICS subgroup, but not for the CB subgroup.

4. DISCUSSION

The present study assessed the effect of performance incentives on use of Bayesian versus heuristic strategies in a perceptual learning task. We found that, at a group level, participants’ choices were more accurate when feedback was delivered in the form of monetary reinforcement than when it was delivered as instructive directives. Similarly, group-level results suggested differences between monetary and instructive feedback conditions in neural encoding of feedback in three ERP components: the P3, FRN, and the LPP. Critically, however, subsequent analyses informed by computational model comparison revealed that group-level behavioural and neural differences were, in fact, driven almost entirely by participants who switched between a Bayesian and a heuristic strategy according to feedback condition. In participants who consistently applied a Bayesian strategy in both feedback conditions, we observed no behavioural differences between monetary and instructive feedback.

Using a model-based cluster analysis based on formal comparison of computational cognitive models, we identified two distinct participant subgroups: a Consistent Bayesian (CB) subgroup, and an Incentive-Compatible Switching (ICS) subgroup. These two subgroups were associated with two models corresponding to qualitatively distinct behavioural strategies. The CB subgroup comprised participants best fit by a model assuming a Bayesian inference strategy in both the monetary and
the instructive feedback condition. Conversely, participants in the ICS subgroup made choices more consistent with strategic switching between Bayesian inference in the monetary feedback condition and a WSLS heuristic strategy in the instructive feedback condition. This behaviour was consistent with incentive-compatible deployment of cognitive resources, at the cost of poorer performance in the instructive feedback condition. By contrast, participants in the CB subgroup used a Bayesian inference strategy in both feedback conditions, including when there was no monetary reward at stake. Behavioural performance for the CB subgroup therefore did not differ between feedback conditions, and was consistently of a high standard overall. A control analysis revealed that model subgroup was unrelated to the order of feedback conditions, thereby ruling out a purely temporal switching effect.

Alternatively, another possible explanation for the observed behavioural data is that participants may indeed have switched between heuristic and Bayesian inference strategies, but may have done so within individual blocks, rather than between different feedback conditions. For instance, participants may have employed a heuristic strategy in early trials of a block, when they had received relatively little information with which to construct a model of the contrast-reward space. Then, in later trials, once they had acquired sufficient information to constrain a model of this space, participants may have switched to a Bayesian inference strategy. Like the analytic approach employed in this study, this explanation explains differences in performance in terms of strategy-switching. Future research should seek to determine whether within-block strategy-switching, as well as between-block strategy switching, is affected by the presence of monetary incentives.

ERP analyses revealed that, like behaviour, neural encoding of feedback also differed between CB and ICS subgroups. We assessed the effect of feedback condition and participant subgroup on the P3, FRN, and LPP: three ERP components
associated with learning and processing of rewarding stimuli (Achtziger et al., 2015; Bennett et al., 2015; Frank et al., 2005; Hajcak et al., 2009; Keil et al., 2002; Polich, 2007; Yeung and Sanfey, 2004). This analysis showed an interaction of feedback condition and participant subgroup for the amplitudes of two components: the P3 and the LPP. This interaction was driven by differences between feedback conditions in the ICS subgroup only, such that monetary feedback elicited larger P3 and LPPs than instructive feedback. This was not the case for participants in the CB subgroup, who showed P3 and LPP components of similar amplitude in both feedback conditions, without any reduction for instructive feedback. Since only the ICS subgroup was associated with strategy-switching, this implicates the P3 and LPP as components which differentially encoded feedback depending on whether participants employed a Bayesian or a heuristic strategy. We also found a significant main effect of feedback condition for the FRN, indicating that across both participant subgroups, FRN amplitudes were larger for instructive than monetary feedback.

This differential neural encoding of feedback affords insight into the nature of feedback processing in Bayesian and heuristic strategies. In particular, the P3 has been linked in past research to the process of Bayesian belief updating (Bennett et al., 2015; Kolossa et al., 2015). It has been proposed that P3 amplitude indexes the magnitude of belief updates, possibly reflecting the deployment of working memory in the revision of prior beliefs (Kopp, 2008). Differences in feedback encoding between CB and ICS subgroups might therefore be interpreted as reflecting the differential engagement of a belief updating mechanism, since only Bayesian inference involves updating a full belief distribution. This is also in line with the proposal by Kok (1997) that P3 amplitude may reflect general cognitive effort, since Bayesian belief updating requires a greater expenditure of cognitive resources than a simple win-stay lose-switch heuristic.
Differential encoding of feedback in the LPP, by contrast, may reflect sensitivity to the reward valence of feedback. In tasks assessing encoding of affective stimuli, LPP amplitude has been associated with the affective salience of stimuli, such that both positively and negatively valenced stimuli elicited larger LPPs than neutral stimuli (Keil et al., 2002; Schupp et al., 2000). As such, one possible interpretation of LPP encoding differences in the present study is that the strategy-switching ICS group, but not the CB subgroup, perceived a difference in the emotional valence of monetary and instructive feedback. This may reflect a greater degree of reward sensitivity in the ICS subgroup than the CB subgroup, since it has previously been shown that reward processing may recruit different neural regions according to participants’ reward sensitivity (Fröber and Dreisbach, 2014; Jimura et al., 2010).

Finally, we observed an overall effect of feedback condition on FRN amplitude, with a larger FRN for instructive compared to monetary feedback, but found that this effect did not interact with participant subgroup. This finding is in line with the hypothesis that FRN amplitude reflects a relatively automatic binary evaluation of stimulus valence (Yeung and Sanfey, 2004), and may provide an electrophysiological index of affective components of feedback processing (Wiswede et al., 2009). Our findings suggest that feedback value evaluation, as indexed by the FRN, was likely to have been equivalent in extent across all participants, independent of differences between participants in Bayesian versus heuristic task strategies. The smaller FRN elicited by monetary feedback in the present study may therefore reflect the greater overall hedonic value of monetary feedback relative to instructive feedback.

An alternative explanation of the observed behavioural data is that participants may have adopted the same strategy in both feedback conditions, but with varying levels of perceptual uncertainty. This is consistent with the recent proposal...
that the presence of reward may have a noise-reduction effect in cognitive and motor control (Manohar et al., 2015). We were unable to explicitly test this hypothesis in the present study, since allowing the perceptual uncertainty parameter \( \sigma \) to vary between feedback conditions would have compromised the identifiability and interpretability of the different computational models. However, we consider that a strategy-switching account provides a more comprehensive overall account of both behavioural and ERP data, given that the observed ERP effects were specific to components associated with belief updating and reward valence.

More broadly, it is important to note that the method of Bayesian model selection employed by the present study identifies which of a given set of computational models provides the most parsimonious account of behavioural data. Notably, therefore, this method does not provide information as to whether the best-fitting model within this set is also the best out of any possible model that might have been considered. For a task such as the perceptual learning task used in the present study, the space of possible models that might have been fit to the data is extremely large, and it was beyond the scope of the present study to exhaustively compare the fit of all possible learning models to participants’ behaviour. Rather, the goal of model comparison in the present study was to assess the relative performance of two particular task strategies—Bayesian inference and a WSLS heuristic—that have been found to provide a good account of behaviour in our perceptual learning task by previous research (Bennett et al., 2015). The relative performance of different models was then used as a tool to make inferences regarding the effect of the feedback incentive manipulation on both behaviour and neural encoding of feedback.

One potential limitation of the present study is that, although monetary and instructive feedback conditions were identical with respect to the functional form and information content of feedback, the direction of the feedback scale differed between
conditions. Specifically, good task performance was indicated by larger numbers (denoting larger rewards) in the monetary feedback condition, but by smaller numbers (denoting smaller choice-target differences) in the instructive feedback condition. This could present a possible confound to interpretation of ERP results, since large and small numbers are thought to have somewhat distinct neural processing substrates (Hyde and Spelke, 2009). However, we believe that this alternative explanation of results is relatively unlikely, since we observed ERP differences between feedback conditions only in a subset of participants whose behavioural data were best fit by the ICS model. If the overall group differences in P3 and LPP components had been driven by numerosity effects, we would have expected that participants in the CB subgroup would also have shown effects of feedback condition on P3 and LPP amplitude. Since this was not the case, we believe that a simple numerosity effect cannot capture the pattern of individual differences in ERPs and behaviour that we observed. Future research could further investigate this question by transposing both monetary feedback into the loss domain, and thereby equating numerosity across feedback conditions. However, it is likely that the behavioural and neural results of such a study will not be directly comparable to the results of the present study, given the well-documented differences in learning from gains versus learning from losses (e.g. Seymour et al., 2007).

Behavioural and neural differences between CB and ICS subgroups can also be interpreted in terms of intrinsic versus extrinsic motivation (Deci, 1971; Ryan and Deci, 2000). Intrinsic motivation refers to a willingness to perform well because of internal factors, such as taking pride in performance; by contrast, extrinsic motivation refers to performance motivated predominantly to acquire desirable outcomes. The results described above might therefore reflect a difference between intrinsic motivation among CB participants and extrinsic motivation in ICS participants
(Daniel and Pollmann, 2010). Alternatively, subgroups could also reflect differences in participants’ general cognitive resources, since participants with fewer overall cognitive resources might strategically optimise their behaviour with respect to monetary reward by adopting an ICS strategy. Consistent with this account, it has previously been shown that use of an optimal strategy in a bandit task was related to participants’ general intelligence, with more intelligent participants more likely to employ an optimal task strategy (Steyvers et al., 2009).

More broadly, our results have bearing on the hypothesis that Bayesian inference represents a unifying principle of neural computation (the 'Bayesian brain' hypothesis; Knill and Pouget, 2004). This hypothesis has been applied successfully to domains including sensory coding and motor planning (Körding and Wolpert, 2004; Yuille and Kersten, 2006). However, one issue with applying Bayesian inference to higher-level judgement and decision making is that Bayesian inference is resource-intensive, and can be computationally intractable for many real-world tasks (Payzan-LeNestour and Bossaerts, 2011). Indeed, a wealth of evidence demonstrates that in many decision settings, humans fail to employ Bayesian strategies (e.g. Cassey et al., 2016; Gigerenzer and Goldstein, 1996). Moreover, even in cases where Bayesian inference is tenable, many individuals instead rely on heuristic strategies (Bennett et al., 2015; Steyvers et al., 2009). Such evidence appears to challenge the suitability of purely Bayesian models to judgement and decision making. However, the results of the present study show that this impasse might be resolved by considering Bayesian models within a resource-rational framework that also takes cognitive resource limitations into account. Such an approach has been termed procedural rationality (Simon, 1976), or Type II rationality (Good, 1983). Our findings suggest that participants may select among Bayesian and heuristic behavioural strategies according to both associated outcomes and each strategy’s processing costs (cf.
Ortega and Braun, 2013). In computational terms, these processing costs comprise both the computational expense of computing action policies, and the difficulty of learning (computational complexity versus sample complexity; see Boureau and Dayan, 2011). This situates Bayesian models of cognition within an ecologically valid framework in which inference is constrained by the cognitive resource limitations of the human brain. From this perspective, we might conclude that ICS participants found the marginal value of employing a Bayesian over a heuristic strategy to be outweighed by the cognitive costs of Bayesian inference in the instructive feedback condition.

In summary, using a model-based clustering analysis, we identified distinct subgroups of participants who appeared to use different combinations of Bayesian and heuristic strategies in a perceptual learning task. Incentive-compatible switching between Bayesian and heuristic strategies was associated with differences in performance between feedback conditions, as well as pronounced amplitude differences in ERP components linked to belief updating and affective salience processing. Overall, results suggest that motivational state may critically affect the use of Bayesian versus heuristic task strategies. This demonstrates the importance of embedding Bayesian models of cognition within a framework constrained by the cognitive resource limitations of biological agents. In addition, results suggest that individual differences in motivational state and reward sensitivity mediate the effect of incentives on task performance; as such, a one-size-fits-all approach to performance incentivisation in educational psychology or personnel economics is likely to be an oversimplification.
REFERENCES


Hyde, D.C., Spelke, E.S., 2009. All numbers are not equal: an electrophysiological investigation of small and large number representations. Journal of Cognitive Neuroscience 21, 1039-1053.


Using a model-based cluster analysis, Study 2 found that overall patterns of behaviour were well explained by the presence of two distinct subgroups in the data: a ‘Consistent Bayesian’ subgroup, and an ‘Incentive-Compatible Switching’ subgroup. Of these, the Consistent Bayesian subgroup comprised participants whose behavioural data was best fit by a model assuming use of a Bayesian belief updating strategy in both monetary and instructive feedback conditions. By contrast, the Incentive-Compatible Switching subgroup included participants whose behaviour was best explained by the use of a Bayesian strategy in the monetary feedback condition and a WSLS heuristic strategy in the instructive feedback condition.

Crucially, subsequent analyses found that subgroup membership significantly interacted with the effect of feedback condition for both behavioural and ERP data. This interaction was such that differences in overall performance accuracy differed between feedback conditions only for participants in the ICS subgroup, but not for participants in the CB subgroup. Similarly, the ERP interaction effects were such that the amplitudes of the P3 and LPP components of the event-related potential differed between feedback conditions only for participants in the ICS subgroup, who switched strategies between conditions. The amplitude of the FRN component of the ERP also differed between monetary and instructive feedback, but this main effect did not interact with participant subgroup, suggesting that the FRN may have reflected processes independent of the strategy employed by participants.

These findings shed further light on the event-related potential correlates of belief updating identified in Study 1. As discussed in Section 5.3, an open question following Study 1 was whether the association between P3 amplitude and belief update size was functional or epiphenomenal. The results of Study 2 provide evidence in favour of the former hypothesis, since P3 amplitude did not differ between monetary and instructive feedback conditions for all participants, but only for those who switched between a Bayesian strategy and a heuristic strategy. By contrast, there was no difference between feedback conditions in P3 amplitude for participants who adopted a consistent Bayesian strategy. This suggests that variability in P3 amplitude was closely linked with the task strategy employed by participants, and therefore provides further evidence that P3 amplitude may directly reflect the size of participants’
revisions of internal models of the environment, consistent with recent theory (Kolossa et al., 2015; Kopp, 2008; Mars et al., 2008).

The results of Study 2 are also noteworthy within the context of the belief-centric research program set out in Chapter 1. One goal of this research program has been to resolve studies of human belief updating with the proposal that a primary function of the central nervous system is the construction and maintenance of internal models of the environment. This approach leads to a novel perspective upon the distinction between Bayesian and non-Bayesian belief updating. From the perspective of the belief-centric research program (Fiorillo, 2008; Friston et al., 2015), it is natural that Bayesian belief updating should not necessarily be applied to the updating of all internal models of the environment. Rather, given the high computational demands associated with Bayesian inference, an adaptive approach is for agents to use Bayesian mechanisms for updating beliefs when there exist good incentives for doing so. Indeed, literature studying bounded rationality and the use of heuristics has frequently demonstrated that it is computationally infeasible for organisms to maintain models of all environmental contingencies, and that it may be adaptive to employ simpler heuristics when task motivation is low (Gigerenzer & Goldstein, 1996).

A second related aspect of this belief-centric framework concerns the value placed by humans upon information which can be used to update internal models. Study 1 and Study 2 have investigated the neural correlates of the use of this information, and the Bayes-optimality of these processes. However, an important related issue is the dilemma faced by agents who can choose to seek out information with which to update beliefs. Under which circumstances such information should be gathered is a core question for the information seeking literature, and one which will be addressed in Study 3 and Study 4.
7. Study 3: Intrinsic valuation of information in decision making under uncertainty

This chapter presents the results of two behavioural experiments developing and validating a novel non-instrumental information seeking task for human participants. This task was adapted from existing information-seeking tasks in the animal literature (e.g., Bromberg-Martin & Hikosaka, 2009; Zentall & Stagner, 2011), and was used to investigate willingness to pay for non-instrumental information among healthy young adult human participants. This study was published in the journal *PLoS Computational Biology* (Bennett, Bode, Brydevall, Warren, & Murawski, 2016), and will be presented in this format.

7.1 Aims of Study 3

From a methodological standpoint, the primary goal of Study 3 was to develop a behavioural task with which to assess the valuation of non-instrumental information in human participants. As discussed in Section 4.3.3, this involved the adaptation and modification of similar tasks from the animal literature. Study 3 then sought to use this task to investigate preferences for acquiring non-instrumental information in healthy human participants. In two separate experiments, Study 3 investigated the effects of varying the cost of information and the rate of information delivery on participants’ valuation of this information. These experiments tested the specific hypothesis that, like other species of animals, human participants would be willing to sacrifice future reward in exchange for early but non-instrumental information, as well as investigating whether this valuation changed according to the rate of information delivery.

In addition, Study 3 also sought to develop a computational cognitive model of behaviour on this task. To this end, two different models of information valuation, implementing instrumental valuation of information and intrinsic valuation of information respectively, were developed and compared. One of these models, the ‘Uncertainty Penalty’ model, which incorporated an intrinsic value of information, tested whether individual differences in valuation of non-instrumental information could be accounted for by individual differences in aversion to uncertainty, as posited by past research (Bromberg-Martin & Hikosaka, 2011; Epstein, 2008).

7.2 Manuscript
RESEARCH ARTICLE

Intrinsic Valuation of Information in Decision Making under Uncertainty

Daniel Bennett¹,²*, Stefan Bode¹, Maja Brydevall¹, Hayley Warren¹,², Carsten Murawski²

¹ Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Victoria, Australia,
² Department of Finance, The University of Melbourne, Parkville, Victoria, Australia

* daniel.bennett@unimelb.edu.au

Abstract

In a dynamic world, an accurate model of the environment is vital for survival, and agents ought regularly to seek out new information with which to update their world models. This aspect of behaviour is not captured well by classical theories of decision making, and the cognitive mechanisms of information seeking are poorly understood. In particular, it is not known whether information is valued only for its instrumental use, or whether humans also assign it a non-instrumental intrinsic value. To address this question, the present study assessed preference for non-instrumental information among 80 healthy participants in two experiments. Participants performed a novel information preference task in which they could choose to pay a monetary cost to receive advance information about the outcome of a monetary lottery. Importantly, acquiring information did not alter lottery outcome probabilities. We found that participants were willing to incur considerable monetary costs to acquire payoff-irrelevant information about the lottery outcome. This behaviour was well explained by a computational cognitive model in which information preference resulted from aversion to temporally prolonged uncertainty. These results strongly suggest that humans assign an intrinsic value to information in a manner inconsistent with normative accounts of decision making under uncertainty. This intrinsic value may be associated with adaptive behaviour in real-world environments by producing a bias towards exploratory and information-seeking behaviour.

Author Summary

Acquiring information about the external world is vital for planning and decision making. However, recent research has shown that some animals choose to acquire information at considerable cost even when the information is of no practical benefit, a counter-intuitive behavior associated with suboptimal outcomes. In this study, we demonstrate that humans also engage in this suboptimal behavior by forfeiting future monetary reward in exchange for early but unusable information about future outcomes. Our results suggest that participants attach a value to information beyond its purely instrumental value. This preference for information may help account for apparent anomalies in human choice behavior such
as compulsive checking behaviors in obsessive-compulsive disorder, and excessive and wasteful use of uninformative laboratory testing in hospitals.

Introduction

In many decision situations, agents possess only incomplete information about decision outcomes, and may choose to seek out further information before choosing a course of action [1, 2]. For instance, a surgeon considering whether to operate on a tumor might first request a biopsy to determine whether the tumor is malignant or benign. Despite being a key feature of choice problems in natural settings, information seeking is not considered within many standard accounts of decision making under risk and uncertainty [3–5]. Moreover, it has been shown that some animals choose to seek information even when that information cannot be used to improve future outcomes [6–8]. This behaviour, which is suboptimal from the perspective of expected reward maximization, suggests that biological agents may attach a value to information which is not solely defined in terms of tangible future outcomes [9].

Historically, many theories of information valuation have adopted an instrumental framework, in which the value of information is calculated solely in terms of expected instrumental benefit [10–12]. These theories predict that a decision-maker should seek information only if the information is expected to impart a tangible benefit in excess of its cost [11]. For instance, a clairvoyant charging $100 to reveal whether stock prices will rise or fall should only be consulted if a payoff greater than $100 is expected to result from using this information. Instrumental valuation of information is normatively optimal, in the sense that it maximises expected monetary reward. However, one strong prediction of instrumental valuation is that information of no instrumental use for acquiring payoffs (henceforth termed non-instrumental information) should not affect choice behaviour. As a result, instrumental valuation of information cannot easily explain curiosity-driven or purely exploratory behaviours [13, 14].

An alternative proposal is that biological agents may attach an intrinsic value to information, such that information about relevant future outcomes is valued for its own sake, independent of direct, tangible payoffs [15]. Similarly, economic decision theory has posited that humans might possess a preference for early resolution of uncertainty which would result in intrinsic value of information [16–18], and recent theories of active inference propose that choice behaviour can be explained by sensitivity to information gain as well as to extrinsic reward [19]. In support of intrinsic valuation of information, human participants have been shown to prefer early to late information about receiving an unavoidable electric shock [20], and to be conditioned by non-instrumental information in a behavioural conditioning paradigm [21]. Moreover, neural data from humans and non-human primates have shown that non-instrumental information is encoded using similar mechanisms, and within similar circuits, to primary and monetary reward [9, 22–24]. These findings are consistent with the hypothesis that biological agents assign an intrinsic reward value to non-instrumental information about future outcomes using a coding scheme commensurate with primary and monetary reward.

One limitation of previous empirical work assessing preference for information in humans is that information available in decision-making tasks is usually of instrumental benefit to participants, such that it is difficult to dissociate the intrinsic value of information from its instrumental value [25, 26]. To address this issue, the present study adapted a task from the animal literature which allowed preferences for non-instrumental information to be elicited in a well-controlled and incentive-compatible manner [22]. Using this task, we sought to test one
counterintuitive prediction of intrinsic valuation of information: that, like starlings and pigeons, human participants would trade off information against extrinsic reward by sacrificing part of an uncertain future reward in exchange for early but non-instrumental information about reward likelihood [27].

Furthermore, among theories positing an intrinsic value of information, the source of this value is often unspecified. For instance, the Kreps-Porteus model in economic decision theory predicts a preference for early resolution of uncertainty from a particular axiomatic formulation of utility, but does not specify a cognitive mechanism which might drive this preference [16]. One proposal is that preference for non-instrumental information might result from an aversion to temporally prolonged uncertainty, such that agents may seek information in order to obtain relief from uncertainty [13, 28–30]. We therefore tested a novel computational cognitive model, which assumed that inter-individual variability in the intrinsic value of information resulted from stable trait-level individual differences in degree of aversion to uncertainty, against a standard expected reward maximization model, which assumed that information was assigned solely instrumental value [11]. Finally, in order to determine whether the duration of uncertainty affected participants’ preference for information, we also conducted an additional experiment in the rate at which non-instrumental information was delivered was experimentally manipulated.

Results
To titrate preferences for non-instrumental information in human participants, we developed a novel variant of an experimental task used in animal research [7, 8, 22]. In each trial of this task, a lottery was played out in which participants could either win (receiving 20 cents) or lose (receiving 0 cents), with equal probability (see Fig 1a). Participants were asked to express their preference for observing one of two stimuli (termed ‘Set A’ and ‘Set B’) in the delay period prior to the presentation of the lottery outcome. Both stimuli took the form of five-slot arrays of red and black cards, with card colours initially hidden and then revealed one-by-one at a constant rate (see Fig 1b). One of the two stimuli (the ‘informative stimulus’) imparted information regarding the lottery outcome: a majority of black cards indicated that the participant would win, whereas a majority of red cards indicated a loss. By contrast, the other stimulus (the ‘non-informative stimulus’) consisted of five black and red cards whose colours were determined pseudo-randomly, and which therefore imparted no information about the lottery outcome. Informative and non-informative stimuli were therefore perceptually equivalent, but only the informative stimulus imparted information regarding the lottery outcome. Crucially, the information gained by observing the informative stimulus was non-instrumental, since it affected only the participant’s certainty regarding the lottery outcome, not the probabilities of the lottery itself.

To assess participants’ willingness to sacrifice monetary reward for non-instrumental information, on each trial a monetary cost was associated with the informative stimulus. Four cost conditions were assessed: 0 cents (free information), 1 cent, 3 cents, and 5 cents. If participants chose the informative stimulus, this cost was deducted from lottery winnings in the case of a win outcome, but not in the case of a loss outcome (see Fig 1a). Participants could not lose money by observing the informative stimulus, ensuring that information preference was not confounded by loss aversion. Since lottery probabilities were unaffected by participants’ choice of stimulus, expected reward was greater for the non-informative stimulus than the informative stimulus in all non-zero cost conditions. In the zero-cost condition, the expected monetary reward of informative and non-informative stimuli was equal.
We conducted two separate experiments using this paradigm. In Experiment 1, only information cost varied between trials. In Experiment 2, both information cost and information rate (the speed at which cards were revealed) differed between trials. Preference for non-instrumental information

Experiment 1 assessed participants’ willingness to forfeit monetary reward in exchange for non-instrumental information, and examined the effect of information cost on information preference. Across cost conditions, participants chose the informative stimulus on 43.95 percent of trials (SD = 20.28), while showing good task engagement as evidenced by a low proportion of missed responses (M = 1.67 percent, SD = 1.72). On average, across cost conditions participants sacrificed 2.87 percent of available winnings in exchange for early information about the lottery outcome (SD = 3.21). A one-way repeated-measures analysis of variance (ANOVA) revealed that choice proportions were modulated by the cost of information (F(1.89, 73.60) = 65.68, p < .001; partial η² = 0.63; see Fig 2a), with information choice proportion monotonically decreasing with increases in information cost. Control analyses revealed
that behaviour was not significantly affected by the key used to select responses (left versus right arrow: $t(39) = -0.71, p = .24$) or the nominal identity of the informative stimulus (A versus B: $t(39) = 1.40, p = .08$).

We next used post-hoc $t$-tests with Bonferroni correction to assess whether participants' behavior was consistent with expected reward maximization. Expected reward maximization, which implies solely instrumental valuation of information, predicts that participants should be indifferent between the informative and non-informative stimulus when information is free, and that the non-informative stimulus should dominate the informative stimulus for any non-zero information cost [11]. In the zero-cost condition, informative stimulus choice proportion was significantly greater than the indifference point of 0.5 ($t(39) = 16.83, p < .001$). In each of the non-zero cost conditions, informative stimulus choice proportion was significantly greater than zero (1-cent condition: $t(39) = 7.17, p < .001$; 3-cent: $t(39) = 5.41, p < .001$; 5-cent: $t(39) = 4.76, p < .001$). These results indicate that participants sacrificed future reward for early information, which is inconsistent with expected reward maximization. In addition, we observed notable individual differences in patterns of information seeking behaviour (see Fig 2b), indicating heterogeneity of task strategies between participants.

**Computational model of intrinsic value of information**

To formalise the comparison between instrumental and intrinsic theories of information valuation, we implemented these theories as competing computational cognitive models, and assessed which model provided the best account of both group- and individual-level data.

The two models we assessed were termed the Expected Value of Information (EVI) model, which incorporated solely instrumental valuations of information, and the Uncertainty Penalty (UP) model, which also incorporated intrinsic valuation of information by assuming that the source of information's intrinsic value was an aversion to temporally prolonged uncertainty [13, 28–30]. Both models considered the task in a Markov Decision Process (MDP) framework, and differed only in choice of state value function [31] (see also Experimental Protocols). We found that, in addition to providing the best overall account of choices across participants
The UP model provided the best fit for a large majority of individual participants. Accordingly, a likelihood-ratio test revealed that including the participant-specific uncertainty penalty parameter \( k \) greatly improved the overall fit of the UP model relative to the EVI model (\( \chi^2(40) = 1338.34, p < .001 \)). Moreover, the UP model provided an unbiased fit to the data of all participants, including those who displayed a relatively weak overall preference for information (see Fig 4). By contrast, the EVI model systematically underestimated informative stimulus choice proportions across all participants.

Furthermore, we found that the best-fitting values of the UP’s scaling parameter \( k \) were greater than zero across participants (Wilcoxon signed-rank test: \( Z = 5.51, p < .001 \)) and, in addition, were strongly correlated with overall proportion of information-seeking choices across participants (Spearman’s rho = 0.95, \( p < .001 \)). This indicates that participants with a stronger aversion to uncertainty (higher \( k \) values) assigned a greater intrinsic value to information, and therefore made more information-seeking choices (see also SI section 1 for individual parameter estimates and parameter-behaviour correlations). Although unsurprising given the structural design of the UP model, the strength of this relationship serves to demonstrate that the UP model parameter designed to capture individual differences in information preference succeeded in doing so.

In addition, since the UP model’s implementation of intrinsic valuation of information assumes that aversion to uncertainty is a stable trait of participants, a secondary prediction of this model is that information’s intrinsic value ought to be stable across time for each participant. In order to test this prediction, we calculated information choice proportion separately in each of the seven experimental blocks in Experiment 1, and assessed the effect of block number on information preference using a 4×7 repeated-measures ANOVA with within-subjects
factors of information cost (0, 1, 3, 5 cents) and task block (1 to 7). We found no significant main effect of task block on information choice proportion ($F(3, 117.37) = 1.73, p = .16$) and no significant interaction between task block and information cost ($F(9.58, 373.59) = 1.05, p = .40$). These results indicate that informative stimulus choice proportions did not differ significantly across the task (see Fig 5), as predicted by the UP model.

Finally, we performed an additional control analysis to ensure that the relative advantage of the UP model relative to the EVI model was not simply due to its better performance in the zero-cost condition. To this end, we repeated the model-fitting procedure while excluding all trials in the zero-cost condition (that is, the models were fit solely on the basis of the 1, 3, and 5-cent conditions). The results were consistent with those obtained in the full sample, with the UP model continuing to provide a better fit to the data than the EVI model ($F(3, 117.37) = 3.12, p = .02$).

Fig 4. Individual-participant model fits. (A) Actual informative choice proportion, denoted $Pr(\text{Info})$ (horizontal axis) versus informative choice proportion as predicted by the UP model (vertical axis). Each circle indicates one participant. Euclidean distance from the diagonal (grey line) represents error in prediction. (B) Actual informative choice proportion (horizontal axis) versus informative choice proportion as predicted by the EVI model (vertical axis). Each circle indicates one participant. Euclidean distance from the diagonal (grey line) represents error in prediction. Across all participants, the EVI systematically under-predicted informative choice proportions (all participants fell below the diagonal).

doi:10.1371/journal.pcbi.1005020.g004

Fig 5. Block-wise behavioural results for Experiment 1. Mean proportion of information-seeking choices, denoted $Pr(\text{Info})$, as a function of information cost and block number across participants. Choice proportions for blocks one to seven are presented in ascending order left to right within each of the four cost conditions. Error bars represent SEM. Preference for information was static across the task.

doi:10.1371/journal.pcbi.1005020.g005
5-cent cost conditions). Once again, we found that the UP model (BIC = 718.31) provided a significantly better fit to data than the EVI model (BIC = 1031.60; likelihood ratio test: \( \chi^2(40) = 489.88, p < .001 \)). This indicates that the UP model, which incorporated intrinsic valuation of information, outperformed the EVI model even when zero-cost trials were excluded from analysis.

The effect of information rate on preference for information

A logical consequence of intrinsic valuation of information is that the temporal profile of uncertainty ought to affect participants’ preference for observing an informative stimulus. In particular, the same amount of information should have a different value to participants depending on the rate at which it resolves uncertainty. In Experiment 2, we tested this prediction among a new sample of participants. In Experiment 2, cards could be revealed at a rate of either 1, 3 or 5 seconds per card in each trial, instead of a constant rate of 3 seconds per card as in Experiment 1.

Participants in Experiment 2 made information-seeking choices on 42.89 percent of trials (SD = 24.72), indicating comparable task performance to Experiment 1. Participants showed good levels of task engagement, with low levels of missed responses (M = 1.11 percent, SD = 1.87). On average, participants sacrificed 2.75 percent of available winnings in exchange for early information about the lottery outcome (SD = 3.49). A 4×3 repeated-measures ANOVA was used to assess the effect of information cost (0, 1, 3, 5 cents) and information rate (1, 3, 5 seconds per card) on information-seeking choice proportions. The significant main effect of information cost on information-seeking choice proportions was replicated, (F(1.86, 72.70) = 84.16, p < .001; partial \( \eta^2 = 0.68 \)), and, crucially, we also found a significant main effect of information rate on information-seeking choice proportions (F(2, 78) = 3.60, p < .05; partial \( \eta^2 = 0.08 \); see Fig 6). This indicates that behaviour was modulated by the rate as well as the cost of information, though the effect size of information rate was substantially smaller than the effect size of information cost. There was no significant interaction between information cost and information rate (F(6, 234) = 1.85, p = .09) although there was a non-significant trend for the effect of information rate to be larger in the positive cost conditions (1, 3, and 5 cents) than in the zero-cost condition.

![Fig 6. Behavioural results for Experiment 2](https://example.com/fig6.png)

**Fig 6.** Behavioural results for Experiment 2. Mean proportion of informative stimulus choices (denoted Pr (Info)) as a function of information cost and information rate. Light blue bars represent the fast speed condition, medium blue bars the moderate speed condition, and dark blue bars the slow rate conditions. Error bars represent SEM. Proportions of information-seeking choices decreased as information rate slowed, particularly for positive information cost conditions (costs of 1, 3, and 5 cents). N = 40.
One potential explanation of this effect is that participants may have preferred sooner rather than later resolution of uncertainty. To explore this possibility, we formulated and compared additional computational models assuming temporal discounting of future states. For most participants, these models did not provide a better fit to data than the undiscounted UP and EVI models. However, the degree to which participants discounted future information was associated with individual differences in the effect size of information rate (see S3 Text).

**Discussion**

Participants in the present study consistently preferred an informative stimulus to a perceptually equivalent non-informative stimulus, despite the fact that information could not be used to improve future outcomes. Moreover, in many cases participants were willing to sacrifice future monetary reward in exchange for this early but non-instrumental information. Since the non-informative stimulus was always of equal or greater expected monetary reward, this pattern of results strongly suggests that participants assigned an intrinsic value to information. This stands in contrast to predictions of instrumental theories of information valuation based on expected reward maximization [10–12], but is consistent with the preference for early resolution of uncertainty posited by decision theory [16–18], and with the behavioural sensitivity to information gain proposed by active inference [19]. Although it has previously been conjectured that intrinsic valuation of information may result in willingness to pay for payoff-irrelevant information [27], this effect has not previously been demonstrated in humans using a well-controlled cognitive task.

We found that the UP model, a novel computational model of information seeking, provided a good account of intrinsic valuation of information by assuming that preference for information resulted from aversion to temporally prolonged uncertainty [13, 28–30]. Notably, we found that the model was able to capture individual differences in strength of information preference across participants, as well as providing a good account of group-level results. It is important to note that aversion to temporally prolonged uncertainty as implemented in the UP model is mathematically and conceptually distinct from the economic concept of risk aversion [32]. Risk aversion as commonly understood cannot predict the preference for information exhibited by participants in the present task, since at the point of choice informative and non-informative stimuli were associated with identical outcome probabilities, and differed only in the rate at which outcome uncertainty was resolved. Similarly, although the informative stimulus was associated with reduced payoff variance in non-zero cost conditions, a simple mean-variance tradeoff [33] does not provide a coherent account of preference for information either, since participants’ information preference was strongest in the zero-cost condition, where both mean and variance of payoffs were identical for the two stimuli. Notwithstanding this result, however, we also found that the UP model fit data well even when trials in the zero-cost condition were excluded from analysis, thus giving us confidence that participants assigned an intrinsic value to information in both zero and non-zero information cost conditions.

Furthermore, consistent with the theory that information valuation is a stable trait-level feature of individuals, we found that information preference was stable across time within the task. This would not have been expected if, for instance, participants only sought information in order to learn payoff contingencies in early blocks of the task. In addition, we found that preference for information was modulated by the rate at which uncertainty was resolved, such that participants exhibited a stronger preference for non-instrumental information when information was delivered at a faster rate. This result is analogous to the preference for faster monetary reward rate in choice behaviour [34]. Moreover, although the effect of information rate cannot be directly captured within the UP model, the direction of the information rate effect is
consistent with discounting of future information, analogous to the temporal discounting of future rewards in human judgment and decision making [35]. As such, the results of the present study are in line with the proposal that humans treat information as though it has an intrinsic reward value commensurable with (and perhaps encoded in the same neural circuits as) primary and monetary reward [2, 15, 22]. However, we also note that the results of Experiment 2 demonstrated a relatively small effect size of information rate; future research should therefore further investigate the nature and robustness of this effect.

Participants in the present study sacrificed future monetary reward in exchange for early but payoff-irrelevant information. This behaviour, which is suboptimal from the perspective of expected reward maximization, has previously been observed in pigeons and starlings [6–8]. In the present study we present for the first time a well-controlled cognitive paradigm with which to assess this behaviour in humans. We note that previous studies in human participants have reported results generally consistent with a willingness to pay for early resolution of uncertainty, such as a greater preference for a risky lottery whose uncertainty was resolved immediately relative to an equivalent lottery whose uncertainty was resolved gradually [36, 37], and a willingness to pay for immediate resolution of uncertainty rather than a 50 percent probability of delayed resolution of uncertainty [38]. Among cognitive studies explicitly assessing the value of non-instrumental information, Pierson & Goodman (2014) found that participants self-reported a willingness to pay for non-instrumental information [39]. However, this behaviour was only assessed using a hypothetical survey task, which may have confounded results given the well-documented disparity in behaviour between hypothetical and incentive-compatible choice tasks [40]. Separately, a behavioural economic study using an incentive-compatible task concluded that observing non-instrumental information was related not to intrinsic valuation of information per se, but to a desire to increase one’s post-hoc confidence regarding an earlier decision [41]. This explanation predicts that participants will only seek non-instrumental information if it provides feedback on an earlier decision. Our results are inconsistent with this explanation, since no such decision was present in the task used in the present study. The strength of our conclusions is based on a well-controlled task in which informative and non-informative stimuli were perceptually identical, and in which preferences for information were elicited in a fully incentive-compatible fashion.

The results of the present study are also conceptually consistent with the preference for early resolution of uncertainty described in economic decision theory by the Kreps-Porteus model [16], and used to account for anomalous patterns of stock pricing in finance by Epstein and Zin [18]. Our empirical and computational findings complement these theories: whereas the Kreps-Porteus model demonstrates that preference for early resolution of uncertainty is a consequence of a particular formulation of recursive utility, in the present study we present a cognitive process model which provides a psychologically plausible account of information-seeking behaviour. Specifically, our results provide evidence that information-seeking may result from an aversion to temporally prolonged uncertainty [13, 28–30]. One interesting finding in this respect was that there was a negative correlation across participants between the UP model’s information preference parameter $k$ and its response stochasticity parameter $\beta$. This correlation was such that participants who assigned a stronger intrinsic value to information also tended to exhibit greater response stochasticity. This relationship is of theoretical interest, since it has been proposed that information-seeking behaviour may result from high levels of response stochasticity in exploration-exploitation dilemmas (e.g. [25], but see also [42]), or via $\varepsilon$-greedy action selection methods in reinforcement learning [43]. Although the superior goodness-of-fit of the UP model in the present study clearly indicates that response stochasticity alone cannot account for participants’ information-seeking choices, the correlation between $k$ and $\beta$ raises the interesting possibility that intrinsic valuation of information and response
stochasticity may make separable but related contributions to exploratory behaviours. Under this hypothesis, the \( \kappa \) parameter would correspond to directed exploration, a goal-directed process aimed specifically at reducing uncertainty, whereas the \( \beta \) parameter would correspond to a more diffuse form of undirected exploration. Future research should further investigate this hypothesis.

However, it is also important to note that behaviour could also be explained by an appetitive drive for information as well as an aversion to uncertainty. Because, according to information theory, uncertainty and information are mathematical conjugates [44], aversion to uncertainty makes similar behavioural predictions to an appetitive desire for information. Accordingly, it is possible to reparametrise the UP model to explain behaviour in terms of an information value bonus, rather than an uncertainty penalty, with equivalent behavioural predictions. As such, behavioural data alone may not be sufficient to distinguish between behaviour driven by uncertainty aversion and behaviour driven by an appetitive desire for information. One possibility for future research is that, since appetitive and aversive stimuli are processed in distinct neural circuits [45], it may be possible to use neural recordings to disentangle these two potential cognitive mechanisms for information valuation.

In accounting for the results of the present study we have primarily drawn upon theories proposing that intrinsic valuation of information can result from an aversion to temporally prolonged uncertainty [18, 28–30]. However, alternative theoretical frameworks can also account for the present study’s results in terms of a positive information bonus (consistent with an alternate parametrisation of the UP model described below). For instance, it has been proposed that agents may derive utility from maintaining an internal model of the environment which is well-adapted to the statistics of natural stimuli [46, 47]. A natural consequence of this model is that agents should place a non-zero value on information about the external environment, even when no behaviour can be directly conditioned on this information (for instance, an intrinsic curiosity reward, as proposed by Schmidhuber, 2009 [48]). Similar intuitions regarding the appetitive value of information have been formalised in several general theories of cognition, including active inference theory [19, 47] and optimal Bayesian exploration [49]. Such theories can be extended to account for seemingly paradoxical attitudes towards information in other settings, such as participants’ preference for maximising entropy over choice options as well as simply maximising expected reward [50], as well as seemingly paradoxical patterns of self-deception in financial choices [51]. At a neurocomputational level, appetitive valuation of information is also consistent with the notion of dopaminergic novelty or exploration bonuses [52].

Notwithstanding the above, however, a further possibility proposed by Beierholm and Dayan (2010) is that an apparent preference for informative stimuli might, in fact, be driven by task disengagement, leading to a relatively greater decrease in the subjective value of the non-informative stimulus [53]. The paradigm tested in the present study sought to prevent such task disengagement by means of pseudo-randomly occurring ‘catch trials’, in which participants were required to make a rapid button-press response to one of the cards in either the informative or the non-informative stimulus. This manipulation helped to ensure that participants maintained task engagement even when observing the non-informative stimulus. Although this cannot conclusively rule out the possibility that participants were somewhat more engaged by the informative than the non-informative stimulus, it does ensure that participants could not fully disengage from the task during observation of non-informative stimuli. Moreover, the behavioural paradigm that we tested allows for well-controlled manipulation of task engagement: by increasing or decreasing the frequency of catch trials, it should be possible to manipulate the degree to which participants disengage during viewing of the non-informative stimulus. The Beierholm and Dayan model makes the
empirical prediction, which could be tested in future research, that information preference ought to be strongest for greater degrees of task disengagement (that is, low catch trial frequency), and that information preference ought to decrease in strength with increasing catch trial frequency.

These theoretical caveats notwithstanding, the results of the present study provide clear behavioural evidence that human participants derive utility from non-instrumental information in a manner inconsistent with traditional models of information valuation. The behavioural task assessed in the present study provides a well-controlled means for assessing intrinsic valuation of non-instrumental information, and the UP model allows for individual differences in the strength of information valuation to be quantified in a principled and mathematically tractable fashion.

Since intolerance of uncertainty has been proposed as a trans-diagnostic treatment marker for emotional disorders [34], understanding information-seeking behaviours may shed light on the symptomatology of disorders including generalised anxiety disorder and obsessive compulsive disorder [35]. For instance, the compulsive checking behaviours which are a hallmark of obsessive compulsive disorder may represent a form of pathological information-seeking behaviour. From this perspective, it might be possible to redescribe some behavioural features of obsessive compulsive disorder as an excessive intrinsic valuation of information driven by excessive levels of aversion to uncertainty. As such, we would hypothesise that individuals with obsessive compulsive disorder would exhibit a high willingness to pay for non-instrumental information in the task used in the present study.

The results of the present study also have bearing on studies of the exploration-exploitation dilemma, in which participants trade off information seeking and reward seeking [25]. A common finding in this literature is that participants seek out more information than is optimal [26]. Our results may help shed light on this finding: intrinsic valuation of information may cause participants to place a premium on information, resulting in a valuation of information in excess of its purely instrumental value. More broadly, we note that although preference for information in the present task was suboptimal from the restricted perspective of monetary reward maximisation, intrinsic valuation of information may be adaptive in more naturalistic environments. Choices in natural settings often resemble dynamic constrained optimisation problems, in that organisms are presented with epistemic uncertainty and poorly defined action-outcome contingencies. In these environments, the instrumental value of seeking information may be computationally intractable, and intrinsic valuation of information might induce a bias toward gathering information that encourages exploratory behaviour even when the usefulness of that exploratory behaviour is not immediately clear. As such, intrinsic valuation of information may induce patterns of behaviour akin to an exploration or novelty bonus [52]. Therefore, in dynamic and uncertain environments intrinsic valuation of information may be associated with profound long-run benefit, in spite of locally suboptimal outcomes in artificial task environments such as that employed by the present study. More broadly, we do not propose that there exists any single level of intrinsic information valuation that will produce optimal behaviour across all environmental conditions. For instance, a strong intrinsic valuation of information may be beneficial when exploration costs are low and overall uncertainty is high, but result in suboptimal performance in situations where exploration is relatively expensive, or where overall environmental uncertainty is low.

In summary, our results provide strong evidence for intrinsic valuation of information in humans, and we present a novel cognitive process model which suggests that aversion to prolonged uncertainty may be an important psychological determinant of this value. We show that intrinsic valuation of information can result in seemingly suboptimal behaviours, such as

Intrinsic Valuation of Information

PLOS Computational Biology | DOI:10.1371/journal.pcbi.1005020 July 14, 2016 12 / 21
a willingness to sacrifice future monetary reward in exchange for immediate but unusable information about relevant future outcomes. More broadly, our results provide a plausible psychological mechanism for human curiosity and exploration, and may explain features of decision making under uncertainty that have hitherto been considered irrational.

**Materials and Methods**

**Participants**

Participants were staff and students of the University of Melbourne. In Experiment 1, we recruited forty-one participants (15 male, 26 female; 40 right-handed, 1 left-handed), aged 18 to 31 (\(M = 22.28, \ SD = 2.63\)). In Experiment 2, we recruited 40 participants (14 male, 26 female; all right-handed) aged 18 to 32 (\(M = 22.90, \ SD = 4.04\)). Participants gave voluntary informed consent, research was conducted in accordance with the Declaration of Helsinki, and protocols were approved by the University of Melbourne Human Research Ethics Committee (ID 1341084). As compensation for participation, participants received a flat payment of AUD $10 plus all lottery winnings (lottery winnings in Experiment 1: \(M = $9.10, \ SD = $1.86\); Experiment 2: \(M = $7.15, \ SD = $0.91\)).

**Procedure**

Stimuli were presented using the Psychophysics Toolbox [56] and MATLAB R2012b (The Mathworks, Natick, MA) on a Macintosh Mini connected to an LCD monitor with resolution 1920×1080 pixels at a screen refresh rate of 60Hz.

Experiment 1 comprised seven blocks, each consisting of sixteen trials total: four trials in each of the four cost conditions (0, 1, 3, 5 cents), with win probabilities pseudo-randomised to ensure that win rates for each cost condition were identical. As is standard practice in computational modelling studies, participants were randomly assigned to one of four pre-generated trial sequences. Participants completed the task in approximately 1 hour. In Experiment 2, the rate at which cards were revealed differed between blocks. In each block, information could be revealed at a rate of either 1, 3, or 5 seconds per card. As such, the lottery delay period varied across blocks (6 seconds total in 1 sec/card blocks, 18 seconds for 3 sec/card blocks, 30 seconds for 5 sec/card blocks). Participants completed 6 blocks of 12 trials each. Each participant was assigned to one of three counterbalanced trial orders, in which no two adjacent blocks belonged to the same information rate condition. Participants completed the task in approximately forty minutes. The primary dependent variable \(Pr(Info)\) was the proportion of all choices (excluding missed responses) in which participants elected to observe the informative stimulus.

To ensure that participants maintained task engagement and attended to each stimulus type equally, approximately 10 per cent of all trials were designated as *catch trials*. In catch trials, instead of revealing a black or red card, one of the cards was revealed to be a white X, to which participants responded by pressing any key within 1.5 seconds. A successful response led to progression to the subsequent trial without penalty; failure to respond resulted in a $1 penalty. Participants who failed to respond to more than two catch trials across the experiment were excluded from all further analyses. This resulted in the exclusion of one participant in Experiment 1 (successful catch trial responses in Experiment 1: \(M = 96.88\%, \ SD = 5.56\%\); Experiment 2: \(M = 97.5\%, \ SD = 5.80\%\)). Rates of successful responses to catch trials did not differ significantly between informative and non-informative stimuli (Experiment 1: \(t(37) = 0.74, p = .46\); Experiment 2: \(t(37) = 0.61, p = .55\)). There is therefore no evidence to suggest that participants’ catch trial performance differed as a function of stimulus type.
Computational models

Models represented the task as an MDP, in which each trial was a decision problem with two actions (informative/non-informative stimuli), discrete states corresponding to different configurations of red and black cards (see Fig 7), and state transition probabilities corresponding to relative probabilities of red/black cards. Using dynamic programming, we calculated the action value of observing each of the two stimuli under varying assumptions about the nature of valuation, and used these action values to predict choice proportions. Competing models used identical MDP task representations, and differed only in the definition of the equation used to calculate action values.

The information seeking task used in the present study can be formally characterised as follows: in each trial $x$, participants chose an action $a_x$ from the set $A = \{I, N\}$, where $I$ denotes a choice to observe the informative signal and $N$ denotes a choice to observe the non-informative signal. The outcome of each trial was denoted $y_x$ and could be either a win, $y_x = 1$, with probability $P_{\text{black}}$, or a loss, $y_x = 0$, with probability $P_{\text{red}}$. By definition, $P_{\text{red}} = 1 - P_{\text{black}}$, and in the present study, it was always the case that $P_{\text{black}} = P_{\text{red}} = 0.5$. Let $c_x$ denote the cost in cents of observing the informative signal on a given trial, drawn from the set $C = \{0, 1, 3, 5\}$. Each trial’s winnings, denoted $r_x$, depended only on the action selected and the predetermined outcomes.

![Fig 7. Trial structure represented as a Markov Decision Process.](https://doi.org/10.1371/journal.pcbi.1005020.g007)
of the trial lottery, such that

\[
    r_s = \begin{cases} 
        20 - c_s, & a_s = I \text{ and } y_s = 1 \\
        20, & a_s = N \text{ and } y_s = 1 \\
        0, & \text{otherwise}
    \end{cases}
\]  

(1)

The task’s structure was implemented as an MDP by considering every different possible configuration of red and black cards as a separate state. This is a natural way of discretising trials of the task such that each state represents a perceptually distinct epoch within the trial. The general structure can be described as follows: depending on the action selected, a participant traverses one of two state trees, corresponding to the two signal types. As such, the two state trees are structurally identical and differ only in the sense that the final card state reached in the informative signal tree perfectly predicts whether the lottery outcome will be a win or a loss, whereas the final card state in the non-informative signal tree may transition to either outcome. Since transitions within the state trees depend on the relative likelihood of drawing red and black cards, state transitions are governed by the probabilities \(P_{\text{black}}\) and \(P_{\text{red}}\). This structure is illustrated schematically in Fig 7.

The structure and parameters described above give a complete description of the states, actions, rewards, and state transition probabilities of the information seeking task. As a result, standard analytic techniques of MDPs can be applied to solve this decision problem. Specifically, using dynamic programming \([57]\), it is possible to calculate the action value \(Q\) of each of the actions \(I\) and \(N\), and to use these action values to predict choice proportions for the two actions. Each of the competing behavioural models used an identical MDP framework, and models differed only in how the action values are calculated. More precisely, in each of the models assessed, action values were calculated by solving a Bellman optimality equation, where only the definition of this equation and its free parameters varied between models. Model fitting procedures were identical for each of the models.

**Expected Value of Information (EVI) model.** The EVI model assumed that agents consider solely the instrumental value of information \([11]\). As such, the state value equation for this model is simply the standard recursive Bellman optimality equation for stochastic programming in MDPs \([57]\):

\[
    V(s) = \max_a \sum_{s'} \Pr(s'|s,a) [R(s',s,a) + V(s')] 
\]

(2)

This equation can be interpreted as a probability-weighted sum over future rewards and states, where \(a\) represents an action taken in state \(s\), causing a transition to a successor state \(s'\) with probability \(\Pr(s'|s,a)\), with concurrent receipt of a transition reward \(R(s',s,a)\) (corresponding to monetary payoffs in the present task). The value of the successor state \(s'\) is denoted \(V(s')\) and is itself also calculated recursively according to Eq 2. As such, the action-value equation for this model was simply the standard recursive Bellman optimality equation for stochastic programming in MDPs:

\[
    Q(s,a) = \sum_{s'} \Pr(s'|s,a) [R(s',s,a) + V(s')] 
\]

(3)

The quantity \(Q(a)\) represents the value of taking the action \(a\) in state \(s\), causing a transition to a successor state \(s'\) with probability \(\Pr(s'|s,a)\), with concurrent receipt of a transition reward \(R(s',s,a)\) (corresponding to monetary payoffs). Note that since there was only one state in the present study from which actions could be taken (see Fig 7), we henceforth use the shorthand \(Q(a)\) to denote the value of the action \(a\).
An implicit assumption of the EVI model is that information is only valuable to the extent that it can be used by agents to increase future expected reward; that is, agents should only seek information when it increases their expected reward. This follows from the fact that state values in Eq 2 are calculated solely on the basis of the monetary reward matrix $R$. In a task such as that used in the present study, where information and expected reward are orthogonal, this model predicts that participants should display no preference for information. Specifically, the EVI model’s prediction is that when there is no cost placed on observing the informative signal participants should be indifferent between the two signals, and that when any non-zero cost is placed on the informative signal, participants should prefer observing the non-informative signal.

In describing the task as an MDP, we have adopted a somewhat different framework for modelling the structure of the decision to standard decision analysis. However, for the model presented above, the value of observing the informative signal $Q(I)$ closely resembles the decision-analytic quantity termed the Expected Value of Sample Information [EVSI; 11]. In decision analysis, the EVSI is theoretically always non-negative, since information can either increase future expected reward (in which case EVSI $> 0$) or not alter future expected reward (in which case EVSI $= 0$). In the task used in the present study, observing the informative signal had an EVSI equal to 0, since observing information could not increase future expected reward. When the cost of observing the informative signal was also considered, this meant that observing the informative signal always had a EVSI equal to zero (for $c = 0$), or less than zero (for $c > 0$).

**Uncertainty Penalty (UP) model.** The UP model was a hierarchical extension of the EVI model, and assumed that as well as seeking to maximise expected reward, agents were averse to the presence of uncertainty over time [29]. This aversion was implemented in UP model as a penalty of all states $s$ of the MDP according to the relative probabilities of winning and losing from state $s$: $P_{\text{win}}(s)$ and $P_{\text{loss}}(s)$. We defined this penalty function $H(s)$ to be the binary entropy function from information theory, since this function is zero in the case of complete certainty and maximal in the case of equal probabilities of winning and losing:

$$H(s) = (\frac{P_{\text{win}}(s)}{P_{\text{loss}}(s)} \log_2 \frac{P_{\text{win}}(s)}{P_{\text{loss}}(s)})$$

The state value equation for the UP model is created by incorporating this penalty function into normative value equation of the EVI model via an additional exponential term:

$$V(s) = \max_a \sum_{s'} \Pr(s'|s,a)[R(s', s, a) + V(s')e^{-kH(s')}]$$

Eq 5 describes the value of different states in the UP model; as such, this equation is functionally equivalent to Eq 2 from the EVI model. However, these two equations differ in that Eq 5 penalises the values of successor states according to their outcome uncertainty (which, in turn, is calculated using Eq 4). Importantly, Eq 5 allowed for the strength of the uncertainty penalty to vary between participants according to a participant-specific scaling parameter $k$. When $k$ is equal to 0, the exponential term in Eq 5 is equal to one, meaning that states’ values are unaffected by outcome uncertainty; in this case, the UP model reduces to the EVI model. For $k > 0$, an uncertainty penalty is applied to all states. Since the informative stimulus reduces uncertainty faster and in more states than the non-informative stimulus, non-informative states are penalised more heavily than informative states, inducing a preference for observing the informative stimulus. For $k < 0$, the converse is true: an uncertainty bonus applies to all states, inducing a preference for the non-informative stimulus. The UP model therefore predicts information seeking from the fact that, although monetary reward is received at the same time for each stimulus, the informative stimulus is associated with less time spent in uncertain
As an aside, we note that although the UP model used in the present study quantifies information valuation in terms of aversion to states’ uncertainty, there exist alternative specifications of the action-value equation which can produce equivalent quantitative predictions regarding behaviour. In particular, it is possible to reparametrise Eq 6 such that intrinsic valuation of information is expressed in terms of value bonus for information, rather than an aversion to states’ uncertainty (see also Sun, Gomez & Schmidhuber, 2011 [49] for a discussion of conceptual differences between information reward signals and standard reinforcement learning rewards). This is because moving from one uncertain (and therefore aversive) state to a less uncertain (and therefore less aversive) state is mathematically equivalent to receiving an information-related value bonus during the state transition. This value bonus will be equal to the amount of uncertainty which has been reduced in the transition, scaled by an individual information preference parameter:

$$Q(a) = \sum_s \Pr(s'|s,a)[R(s',s,a) + V(s')]$$

where $\hat{k}$ is a reparametrised information scaling parameter, and $I(s',s,a)$ is an information value bonus quantified as the difference in entropy of beliefs following a state transition, as follows:

$$I(s',s,a) = H(s') - H(s).$$

Notwithstanding the above, all model fits and reported parameter values in the present study are calculated using the uncertainty-aversion parametrisation of the UP model (Eq 6), not the information-bonus parametrisation (Eq 7).

Model fitting procedure. Given that the two models described above share a common MDP structure, it is possible to specify an overall choice rule and likelihood estimation procedure that is independent of how each individual model calculates action values. For every cost condition c, each of the models supplies a state-action value for observing the informative signal, $Q(I)$, and a state-action value for observing the non-informative signal, $Q(N)$. A mapping from these model-derived action values to informative signal choice probability can be accomplished separately for each cost condition using a “softmax” or Luce choice rule:

$$P_s(I) = \frac{e^{Q_s(I)}}{e^{Q_s(I)} + e^{Q_s(N)}}$$

In this equation, $\beta$ is an inverse temperature parameter, $\beta \geq 0$, governing the determinism of choices, and is constant across different cost conditions. As the value of $\beta$ increases, agents deterministically choose the action with the higher action value; for $\beta = 0$, choices are unrelated to action values, and all actions are equally likely to be selected. Applying Eq 9 with the action value of the non-informative signal in the numerator, it is also possible to show that $P_s(N) = 1 - P_s(I)$.

All models used in the present study also included an additional parameter representing the probability of making an incorrect button press. The rationale for including this additional parameter was that during post-task debriefing, some participants who otherwise behaved in a relatively deterministic fashion reported making one or more mistaken button press because of the task’s randomised response mapping. While one practice is to account for errors of this kind by the softmax equation’s $\beta$ parameter, when responses are otherwise strongly deterministic, a single mistaken button press can substantially affect model likelihood (for instance, when a participant mistakenly selects the informative signal in a condition where the model’s predicted probability of this action would otherwise have been at or near zero). To overcome
In the present study, choice probabilities were assumed to be a latent mixture of erroneous button presses (with probability \( \varepsilon \)) and accurate button presses (with probability \( 1 - \varepsilon \)). Since erroneous button presses are by definition undirected, these choices are therefore equally likely to result in the selection of the informative signal or the non-informative signal. This has the effect of placing a floor of \( \varepsilon^2 \) and a ceiling of \( 1 - \varepsilon^2 \) on choice probabilities for each option.

Therefore, \( Pr(a = I \mid c) \), the overall choice probability for observing the informative set in the cost condition \( c \), is given by simple probability calculus (note that \( Pr(a = I \mid c) = P_c(I) \) for \( \varepsilon = 0 \)):

\[
Pr(a = I \mid c) = 0.5\varepsilon + P_c(I)(1 - \varepsilon)
\]  

(10)
Since there were only two actions available to participants on each trial, the probability $P(a = I)$ can be interpreted as a binomial rate parameter. Each model’s overall likelihood can therefore be calculated as a product of binomial likelihoods across the four cost conditions in the set $C$:

$$L = \prod_{i=1}^{4} \left( \frac{n_i}{m_i} \right) \cdot \Pr(a = I|C_i)^{n_i} \cdot (1 - \Pr(a = I|C_i))^{m_i-n_i},$$  

(11)

Where $n_i$ represents the number of trials in cost condition $C_i$ and $m_i$ represents the number of times the informative signal was observed in cost condition $C_i$. Trials in which participants did not record a response were excluded.

All models were fit with maximum-likelihood estimation, using the interior point algorithm as implemented in MATLAB R2015b (The Mathworks, Natick, MA).

**Supporting Information**

S1 Text. UP model parameters. (PDF)

S2 Text. Interpretation of $\epsilon$ parameter. (PDF)

S3 Text. Model fits for Experiment 2. (PDF)

S1 Fig. $k$ parameter and behavioural data. (TIFF)

S2 Fig. $\beta$ parameter and behavioural data. (TIFF)

**Author Contributions**

Conceived and designed the experiments: DB SB CM. Performed the experiments: DB MB HW. Analyzed the data: DB MB HW. Contributed reagents/materials/analysis tools: DB. Wrote the paper: DB SB MB HW CM.

**References**


7.3 Summary

Using a novel non-instrumental information seeking task, Study 3 investigated the valuation of information in healthy human participants. Results demonstrated that participants displayed a willingness to sacrifice future reward in exchange for early but non-instrumental information about reward likelihood. This was inconsistent with the predictions of expected reward maximisation, but consistent with the proposal that humans may assign an intrinsic value to information (Grant, Kajii & Polak, 1998). Results also showed that the value of non-instrumental information was modulated by both the cost of information and the rate at which information was delivered. Specifically, participants’ valuation of non-instrumental information decreased as the cost of information increased, and increased with increases in the delivery speed of information. To account for these findings, Study 3 formulated a novel computational model of information valuation, the Uncertainty Penalty model, which provided a good account of participants’ intrinsic valuation of information by assuming that this value resulted from aversion to temporally prolonged uncertainty.

A further research question arising from these results was the identity and nature of the neural substrates of this intrinsic valuation of information. According to one prominent theory, the common currency hypothesis (Bromberg-Martin & Hikosaka, 2011), the intrinsic valuation of information observed in Study 3 is posited to result from commonalities in the neural encoding of information and primary reward within dopaminergic reward-processing regions of the brain. Using EEG data recorded from participants completing the task developed within Study 3, Study 4 went on to investigate the predictions of this common currency hypothesis in human participants.
8. Study 4: The feedback-related negativity encodes an information prediction error in decisions to seek non-instrumental information

Study 4 investigated the neural substrates of the preference for acquiring non-instrumental information that was demonstrated by participants in Study 3. Specifically, this study investigated similarities in patterns of neural processing of monetary reward and information, using EEG data collected from participants completing the same non-instrumental information seeking task as Experiment 1 of Study 3. Event-related potentials elicited by the presentation of informative stimuli in this task (red and black cards) were assessed to determine the extent to which reward and information were encoded equivalently in the amplitude of the feedback-related negativity (FRN), an ERP component commonly associated with the processing of rewarding stimuli (Holroyd & Coles, 2002; Holroyd, Krigolson, & Lee, 2011).

8.1 Introduction

Seeking information is an important drive of behaviour, and a key component of effective decision making under uncertainty (Kidd & Hayden, 2015). However, normative decision theory, which assumes that the value of information resides in its instrumental utility for acquiring future rewards (Howard, 1966; D. B. Lawrence, 1999; Raiffa & Schlaifer, 1961), provides a poor description of information seeking in humans and other animals. In particular, such theories cannot account for findings showing that animals place a positive value on information that resolves uncertainty but which cannot be used to affect future tangible outcomes (non-instrumental information).

Study 3 of the present thesis presented evidence that participants assigned an intrinsic value to non-instrumental information about appetitive future events. Moreover, the results of Study 3 demonstrate that humans, like many other species, exhibited a willingness to sacrifice part of an uncertain future reward in exchange for non-instrumental information about the reward’s likelihood (cf. Blanchard et al., 2015; Vasconcelos et al., 2015; Zentall & Stagner, 2011). These behavioural findings indicate that humans and other animals treat information as though it were of intrinsic value (Grant et al., 1998).

One recent proposal, the ‘common currency’ hypothesis, is that the intrinsic value of information might result from common neural substrates for processing of
rewarding and informative stimuli (Bromberg-Martin & Hikosaka, 2011). Single-neuron recordings from non-human primates have demonstrated that non-instrumental information is encoded within brain regions typically associated with reward processing, such as the dopaminergic midbrain (Bromberg-Martin & Hikosaka, 2009), lateral habenula (Bromberg-Martin & Hikosaka, 2011) and orbitofrontal cortex (Blanchard et al., 2015). Notably, Bromberg-Martin and Hikosaka (2011) reported that in response to informative stimuli, neurons in macaque lateral habenula encoded both reward prediction errors (RPEs; the signed difference between expected and actual reward) and information prediction errors (IPEs; the signed difference between expected and actual information). Similarly, fMRI in humans has revealed that the delivery of information is associated with increased BOLD signals within brain regions typically associated with reward processing such as the striatum (Jepma, Verdonschot, Van Steenbergen, Rombouts, & Nieuwenhuis, 2012; Kang et al., 2009). This resemblance suggests a common neural coding scheme for information and primary reward, which might result from mechanisms such as an intrinsic reward value of information (Bromberg-Martin & Hikosaka, 2011) or boosting of anticipatory utility by reward prediction errors associated with informative stimuli (Iigaya et al., 2016).

To date, many predictions of the common currency hypothesis of information valuation have not been investigated in humans. To address this issue, the present study recorded EEG data from human participants completing a non-instrumental information seeking task. Study 4 investigated whether IPEs associated with non-instrumental information were encoded in the feedback-related negativity component of the event-related potential. The FRN component has, in previous research, been proposed to encode the discrepancy between expected and actual reward (a reward prediction error; see Schultz, Dayan, and Montague, 1997). Specifically, according to one prominent theory, FRN amplitude reflects RPEs by encoding the disinhibition of neurons in anterior cingulate cortex by mesencephalic dopamine neurons (Holroyd & Coles, 2002). In support of this contention, it has been shown that FRN amplitudes are greater following negative RPEs than positive RPEs (M. X. Cohen et al., 2007; Hajcak et al., 2007). Indeed, it has been proposed that the FRN could be reconceptualised as a ‘reward positivity’ (Holroyd et al., 2011) encoding the hedonic value of stimuli relative to expectations. Premised upon this dopaminergic RPE model of the FRN, the common currency hypothesis of information valuation therefore predicts that IPEs should also be encoded in the FRN, in a comparable fashion to RPEs.
8.2 Method

8.2.1 Participants

Participants were 23 healthy right-handed participants (9 male, 14 female) aged between 18 and 32 years ($M = 23.04$, $SD = 4.15$). Participants were selected from among those who had previously completed the behavioural experiment reported as Experiment 2 in Study 3. Selection of participants was based entirely on availability to attend scheduled EEG sessions, and was unrelated to participants’ prior performance on the behavioural task. All participants provided written informed consent, and research was conducted in accordance with the Declaration of Helsinki. All study protocols were approved by The University of Melbourne Human Research Ethics Committee (ID 1341084). Participants received monetary compensation consisting of a base rate of AUD $10, plus all task winnings up to a maximum of $15 ($M = 11.48$, $SD = 1.18$).

8.2.2 Materials and Procedure

Participants completed a non-instrumental information seeking task identical to that conducted in Experiment 1 of Study 3. The task was presented on a Dell P2210 LCD monitor (1680 × 1050 screen resolution; refresh rate 60 Hz) using the Psychophysics Toolbox (Brainard, 1997) for MATLAB (The Mathworks; Natick, MA).

As in Experiment 1 of Study 3, on each trial a lottery was drawn in which participants had an equal probability of winning (receiving 20 cents) or losing (receiving 0 cents). Prior to the lottery draw, participants could choose to view either an informative stimulus, which imparted early information about the lottery outcome, or a non-informative stimulus, which was perceptually identical to the informative stimulus but imparted no information about the lottery outcome. To assess participants’ willingness to pay for non-instrumental information, a variable cost was associated with viewing the informative stimulus, to be deducted from participants’ winnings in the case of a win outcome only. Four information cost conditions (0, 1, 3, or 5 cents) were presented in a pseudo-random order, with the dependent variable the proportion of choices to observe the informative stimulus. Participants indicated choices using the index finger of their right hand to press either the left or the right button of a five-button
Cedrus response box. The left-right response mapping of informative and non-informative stimuli was pseudo-randomised across trials.

Informative and non-informative stimuli were arrays of red and/or black cards. In the informative stimulus, the relative proportion of red and black cards perfectly predicted the lottery outcome (majority black cards: win; majority red cards: loss). In the non-informative stimulus, the relative proportion of card colours was unrelated to the lottery outcome. These options were presented to participants as a choice between Set A and Set B, with the identity of the informative stimulus pseudo-randomised across trials. After making a choice, in order to ensure that participants could not forget which condition they were in, participants were informed whether the chosen stimulus was to be informative or non-informative. Cards were then drawn one-by-one at a constant rate of 3 seconds per card. Once all cards were drawn, the outcome of the monetary lottery was revealed to participants.

In order to ensure that participants maintained attention on the chosen stimulus as it was revealed, a small number of trials (~10%) were designated as ‘catch trials’ in which one card in the chosen stimulus was drawn to reveal a white X rather than a red or black card. Participants were instructed to respond to this attention check by pressing any button within 1.5 seconds. Failure to do so resulted in the deduction of $1 from overall winnings. This ensured that participants did not disengage from the task during stimulus presentation, and attended equally to both stimulus types.

Before commencing the task, participants received verbal and written instruction in the task and were asked to complete a practice task. Participants were informed that their choice of stimulus on each trial would not affect the likelihood of winning the lottery, that the probability of winning and losing was equal on each trial, and that card colours were conditionally independent of one another. Participants completed 7 blocks of 16 trials each while EEG was recorded, with a total testing duration of approximately 50 minutes.

8.2.3 Data analysis

8.2.3.1 Behavioural exclusion criteria

In line with exclusion criteria adopted in Study 3, it was determined \textit{a priori} that participants who failed to respond to more than two catch trials across the entire
experiment would be deemed to have failed an attention check. One participant failed
to respond on four catch trials, and was therefore excluded from all further analysis.
The remaining 22 participants showed good levels of task engagement, as measured by
successful responses to catch trials ($M = 98.11\%$, $SD = 3.57\%$).

8.2.3.2 EEG data acquisition and preprocessing

EEG data were acquired from 64 Ag/AgCl active scalp electrodes located according to
the International 10-20 system. Data were recorded at a sampling rate of 512 Hz using
a BioSemi ActiveTwo system using an implicit reference during recording, and were
linearly detrended and re-referenced offline to an average of left and right mastoids.
The electrooculogram was recorded from two infraorbital electrodes horizontally
adjacent to and below the left eye.

EEG data were preprocessed using EEGLAB (Delorme & Makeig, 2004),
according to the same semi-automated preprocessing pipeline employed within Study
1 and Study 2. Specifically, data were high- and low-pass filtered at 0.1 Hz and 70 Hz
respectively, and notch filtered from 45-55 Hz to remove background electrical noise.
Data were segmented into epochs from 1000ms before to 1000ms after events of
interest, and baseline corrected using a 100ms pre-stimulus baseline. An Independent
Component Analysis (ICA) as implemented in EEGLAB was used to identify and
remove components of the data related to eyeblink and saccade artefacts. Noisy data
channels were interpolated using a spline interpolation routine; no interpolated data
channels were included in final ERP analyses. Finally, an impartial artefact screening
procedure automatically excluded all epochs in which maximum/minimum amplitudes
exceeded $\pm 200\text{mV}$.

8.2.3.3 Quantification of computational variables

Study 4 sought to compare patterns of encoding of reward prediction errors and
information prediction errors in the feedback-related negativity. This section therefore
details how these computational variables were mathematically quantified. The formal
definitions provided in this section demonstrate that the key distinction between RPEs
and IPEs is that, whereas positive/negative RPEs indicate an increase in the likelihood
of winning or losing the trial lottery, positive/negative IPEs indicate that a card
conveyed more/less information about the outcome than expected, regardless of
whether that outcome was a win or a loss. Given that, according to the task design employed in Study 4, there was an equal probability of observing a red versus a black card at every point in time, positive and negative RPEs and IPEs were therefore independent of one another.

RPEs were calculated as the discrepancy between expected lottery winnings prior to observing the stimulus and actual expected lottery winnings after observing the stimulus:

\[
RPE = (\Pr(\text{win})_{post \cdot \text{winnings}}) - (\Pr(\text{win})_{prior \cdot \text{winnings}})
\]

The probability of winning at each point in time was calculated as the binomial probability that black cards would be in the majority in the informative stimulus, given conditional independence of successive card draws:

\[
\Pr(\text{win}|n,n_{req}) = 1 - \left( \sum_{k=0}^{n_{req}-1} \binom{n}{k} 0.5^n \right)
\]

Where \(n\) is the number of cards remaining to be drawn, and \(n_{req}\) is the number of additional black cards required for a majority given \(n_{black}\) already drawn:

\[
n_{req} = \begin{cases} 
3 - n_{black}, & n_{black} < 3 \\
0, & n_{black} \geq 3 
\end{cases}
\]

By definition, no information was imparted by cards in the non-informative stimulus, and in this case \(\Pr(\text{win})\) was therefore always equal to 0.5.

By analogy with RPEs, IPEs were calculated as the difference between the actual information content of a stimulus \(l\) and its expected information content \(l_{expected}\):

\[
IPE = l - l_{expected}
\]

Where, following Shannon (1948), the information content of a stimulus was defined as the entropy difference between posterior and prior beliefs:

\[
l = H(\Pr(\text{win})_{post}) - H(\Pr(\text{win})_{prior})
\]

With entropy \(H\) itself defined as the binary entropy function:

\[
H(\Pr(\text{win})) = -\Pr(\text{win}) \cdot \log_2(\Pr(\text{win})) - (1 - \Pr(\text{win})) \cdot \log_2(1 - \Pr(\text{win}))
\]
Given $n$ and $n_{req}$, the expected information content of any card can therefore be calculated as follows:

$$I_{expected} = E[I]$$

$$= \frac{H(pr(win|n - 1, n_{req} - 1)) + H(pr(win|n - 1, n_{req})) - 2H(pr(win|prior))}{2}$$

Figure 8.1 provides a schematic overview of card transitions in the style of Figure 7 from Study 3. This figure details events associated with positive versus negative RPEs and positive versus negative IPEs.
Figure 8.1. Schematic overview of task events. Segment colours within each circle denote the number of black and red cards visible at any point in time, whereas lines denote transitions between states (card draws). Grey segments denote cards whose colours are yet to be revealed (A) Overall schematic including both informative (left) and non-informative (right) stimuli. Stimuli were perceptually identical, and differed only in that the majority colour in the informative stimulus perfectly predicted lottery outcome (win, W; loss, L). (B) Informative stimulus schematic, with positive reward.
prediction errors (RPEs) denoted by solid lines, negative RPEs by dashed lines, and zero RPEs by grey lines. (C) Informative stimulus schematic, with positive information prediction errors (IPEs) denoted by solid lines, negative IPEs by dashed lines, and zero IPEs by grey lines.

8.2.3.4 ERP analysis

ERP analyses were conducted using ERPLAB (Lopez-Calderon & Luck, 2014). In line with previous research (e.g., Hajcak, Moser, Holroyd, & Simons, 2006), FRN amplitudes were calculated for each condition as the mean amplitude from 200 to 350 milliseconds post-stimulus at the fronto-central channels Fpz, AFz, Fz, FCz, and Cz. These five channels are located above the medial frontal cortex, a candidate generator for the FRN (M. X. Cohen et al., 2007; Holroyd & Coles, 2002). For the purposes of Study 4 FRN amplitude was defined in terms of its negativity, meaning that a ‘larger’ FRN was one with a greater negative amplitude (that is, a more negative mean amplitude). To avoid confusion, therefore, comparisons of FRN amplitudes between conditions describe waveforms as ‘more negative’, or ‘more positive’, rather than using potentially ambiguous terms such as ‘larger’ or ‘smaller’.

For all analyses, ANOVA degrees of freedom were adjusted using the Greenhouse-Geisser correction where the assumption of sphericity was violated. Epochs were binned for ERP analysis according to three different computational variables: positive/negative RPEs, positive/negative IPEs, and positive/negative information.

Since behavioural results showed large differences in strategies between individuals, the number of epochs available for different ERP analyses differed between participants. ERP analyses therefore only included data from participants who had at least 20 epochs of each event type under consideration (cf. Hajcak et al., 2007). This resulted in the exclusion of seven participants for card-locked analyses of the informative stimulus, three participants for card-locked analyses of the informative stimulus, three participants for outcome-locked analyses of the non-informative stimulus, and four participants for outcome-locked analyses of the non-informative stimulus. Resulting participant numbers for each analysis are detailed in Table 1.
Table 1. Number of participants included in each analysis.

<table>
<thead>
<tr>
<th>Analysis</th>
<th>Stimulus type</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Informative stimulus</td>
<td>Non-informative stimulus</td>
</tr>
<tr>
<td>RPE</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>IPE</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>Amount of information</td>
<td>15</td>
<td>19</td>
</tr>
<tr>
<td>Outcome screen</td>
<td>19</td>
<td>18</td>
</tr>
</tbody>
</table>

RPE: reward prediction error. IPE: information prediction error.

8.3 Results
8.3.1 Behavioural results

Behavioural results (Figure 8.2) replicated the overall findings of Study 3. A repeated-measures analysis of variance (ANOVA) revealed that preference for information was modulated by the cost of information ($F(1.76, 36.99) = 58.02, p < .001$), such that participants displayed a strong preference for the informative stimulus when it was available at no cost, and a non-negligible preference for this stimulus when it was available at a cost. Preference for information decreased with increasing information cost, and as in Study 3, there was considerable inter-individual variability in preference for information as measured by overall proportion of choices to observe the informative stimulus ($M = .39$, range = .15-.1). These results suggest that participants assigned an intrinsic value to the non-instrumental information imparted by the informative stimulus.
Figure 8.2. Mean proportion of informative stimulus choices (denoted $Pr(Info)$) as a function of information cost. Error bars represent the standard error of the mean.

8.3.2 ERP results

8.3.2.1 Reward prediction errors.

First, it was examined whether the amplitude of the FRN evoked by the presentation of informative cards encoded RPEs, as predicted by a prominent reinforcement learning theory (Holroyd & Coles, 2002). In line with previous studies (M. X. Cohen et al., 2007; Hajcak et al., 2007), mean FRN amplitudes from 200 to 350 milliseconds post-stimulus were analysed using a $2 \times 5$ repeated-measures ANOVA with factors of RPE (positive, negative) and electrode (Fpz, AFz, Fz, FCz, Cz), which revealed a significant main effect of RPE on FRN amplitude ($F(1, 14) = 6.09, p = .03$), with negative RPEs associated with a more negative FRN amplitude compared to positive RPEs (see Figure 8.3). This indicates that the FRN encoded RPEs in a typical fashion in the present study. Figure 8.4 presents a scalp map of amplitude differences between positive and negative reward prediction errors, demonstrating a broad fronto-central topography of this difference.
Figure 8.3. Grand average waveform for reward prediction errors at electrode Cz, time-locked to the presentation of card stimuli in the informative stimulus (stimulus presentation occurred at $t = 0$). The red line represents positive reward prediction errors, and the blue line represents negative reward prediction errors. Red and blue shaded regions denote the standard error of the mean for positive and negative reward prediction errors respectively. The FRN was quantified as the mean voltage from 200 to 350 milliseconds post-stimulus (shaded grey rectangle). Negative voltages are plotted upward.

Figure 8.4. Scalp topography of mean difference between negative and positive reward prediction errors during the FRN analysis window (200-350 milliseconds post-stimulus). Differences were calculated as negative subtract positive reward prediction errors, with greater negativities associated with ‘warmer’ colours.
We also observed a significant main effect of electrode on FRN amplitude ($F(1.62, 22.61) = 53.66, p < .001$); however, this effect did not interact significantly with the effect of RPE ($F(1.83, 25.55) = 0.98, p = .43$).

### 8.3.2.2 Information prediction errors

As in the RPE analysis detailed in Section 8.3.2.1, this study also used a $2 \times 5$ repeated-measures ANOVA to investigate the effects of IPE (positive, negative) and electrode (Fpz, AFz, Fz, FCz, Cz) on mean FRN amplitude from 200 to 350 milliseconds post-stimulus. Analogous to the effect of RPE, this analysis demonstrated a significant main effect of IPE on FRN amplitude ($F(1, 14) = 7.75, p = .01$), driven by significantly more negative FRN amplitudes in response to negative IPEs than to positive IPEs (see Figure 8.5). These results indicate that both IPEs and RPEs were encoded by the FRN: negative prediction errors—both RPEs and IPEs—both elicited more negative FRN amplitudes relative to positive prediction errors. Figure 8.6 presents a scalp map of amplitude differences between positive and negative information prediction errors. As with Figure 8.4, this scalp map demonstrates a broad fronto-central topography of this difference.
Figure 8.5. Grand average waveform for information prediction errors at electrode Cz, time-locked to the presentation of card stimuli in the informative stimulus (stimulus presentation occurred at $t = 0$). The red line represents positive information prediction errors, and the blue line represents negative information prediction errors. Red and blue shaded regions denote the standard error of the mean for positive and negative information prediction errors respectively. The FRN was quantified as the mean voltage from 200 to 350 milliseconds post-stimulus (shaded grey rectangle). Negative voltages are plotted upward.

Figure 8.6. Scalp topography of mean difference between negative and positive information prediction errors during the FRN analysis window (200-350 milliseconds post-stimulus). Differences were calculated as negative subtract positive information prediction errors, with greater negativities associated with ‘warmer’ colours.
As in the RPE analysis, the IPE analysis also found a significant main effect of electrode on FRN amplitude \(F(1.71, 23.90) = 37.83, p < .001\), but no interaction between electrode and IPE \(F(4, 56) = 0.19, p = .93\).

8.3.2.3 Amount of information

In addition, this study also examined whether the absolute amount of information delivered by stimuli was encoded in the FRN. This involved assessing the information content of stimuli independent of pre-stimulus expectations concerning information content. As such, positive information was associated with stimuli which increased participants’ certainty regarding the trial outcome (both more certain of winning and more certain of losing), whereas negative information was associated with stimuli which decreased participants’ certainty regarding the trial outcome.

This analysis can be thought of as a more general version of the IPE analysis described in Section 8.3.2.1. This is because ‘positive information’ as operationalised in this Section describes card stimuli which increased participants’ certainty, and vice versa for negative information. Stimuli associated with positive IPEs were therefore a subset of all stimuli associated with positive information, and stimuli associated with negative IPEs were a subset of all stimuli associated with negative information. As such, the positive/negative information analysis reported in this section produces results that should be considered an incremental update to the results reported in Section 8.3.2.1, rather than an entirely independent set of results.

A 2 × 5 repeated-measures ANOVA was used to assess the effect of information (positive, negative) and electrode (Fpz, AFz, Fz, FCz, Cz) on FRN amplitude. A significant main effect of information was found \(F(1, 14) = 9.59, p < .01\), with more negative FRN amplitudes for negative information (increased uncertainty) relative to positive information (increased certainty), as depicted in Figure 8.7(e). Once again, results also demonstrated a significant main effect of electrode \(F(1.69, 23.66) = 61.30, p < .001\), but no interaction effect between information and electrode \(F(2.36, 32.97) = 0.87, p = .49\).

8.3.2.4 Non-informative stimuli
As an exploratory follow-up analysis, it was also investigated whether the modulation of the FRN in response to RPEs and IPEs was unique to cards following a decision to view the informative stimulus, or whether similar patterns were observed for cards following a decision to observe the non-informative stimulus. To do this, RPEs and IPE variables were calculated for cards in each non-informative stimulus as though they had instead been an informative stimulus (because technically, according to the formulae set out in Section 8.2.3.3, IPEs and RPEs were always zero for cards in the non-informative stimulus).

This analysis revealed observed no overall modulation of the FRN by RPEs for the non-informative stimulus ($F(1, 18) = .02, p = .90$); see Figure 8.7(d). However, in an interesting parallel to the effect of IPE in the informative stimulus, this analysis also revealed a small effect of IPE ($F(1, 18) = 4.58, p = .046$); see Figure 8.7(b). There was no significant effect of absolute amount of information on FRN amplitude for non-informative stimuli ($F(1, 18) = 0.47, p = .50$); see Figure 8.7(f).

### 8.3.2.5 Outcome screens

Finally, it was also investigated whether the delivery of the lottery outcome at the end of each trial was associated with differences in FRN amplitude. Using a $2 \times 5$ repeated-measures ANOVA with factors of outcome (win, loss) and electrode (Fpz, AFz, Fz, FCz, Cz), no significant main effect of outcome on FRN amplitude across all trials was found ($F(1, 18) = 3.53, p = .08$). This was also true when analyses were conducted separately for outcome screens following an informative stimulus ($F(1, 18) = 3.74, p = .07$; see Figure 8.7(g)) and for outcome screens following a non-informative stimulus ($F(1, 17) = 0.11, p = .74$; see Figure 8.7(h)).
Figure 8.7. Grand average waveforms at electrode Cz for contrasts reported in Sections 8.3.2.1 to 8.3.2.5. Subplots in the left column present waveforms subsequent to informative stimuli, and subplots in the right column present waveforms subsequent to non-informative stimuli. Shaded regions represent the standard error of the mean for grand average waveforms. Top row (A & B): Positive and negative information

8.4 Discussion

Study 4 used the non-instrumental information seeking task developed by Study 3 to investigate the neural substrates of participants’ preference for non-instrumental information in decision making under uncertainty. Using EEG, this study assessed how both reward prediction errors and information prediction errors were reflected in the feedback-related negativity component of the event-related potential. Behavioural results replicated the overall pattern of findings reported by Study 3, consistent with an intrinsic valuation of information (Grant et al., 1998). That is, participants displayed a clear preference for acquiring non-instrumental information, despite the fact that this information was at times associated with a direct monetary cost. Analyses of the ERP evoked by informative stimuli revealed that RPEs and IPEs were both encoded in a comparable fashion in the amplitude of the FRN component.

ERP analyses showed that the modulation of the FRN during task events that elicited positive and negative IPEs was consistent with FRN modulation by positive versus negative RPEs. The FRN has traditionally been considered to encode correct and incorrect responses in tasks (Holroyd & Coles, 2002; Miltner, Braun, & Coles, 1997), as well as rewarding outcomes (M. X. Cohen et al., 2007; Hajcak et al., 2006, 2007). As such, the ERP analyses of Study 4 demonstrate a striking parallel in FRN encoding of informative and rewarding outcomes. This is conceptually consistent with the finding that firing rates of single neurons in primates respond in the same manner to positive/negative IPEs as to positive/negative RPEs (Bromberg-Martin & Hikosaka, 2011). Since FRN amplitude is thought to be related to dopaminergic projections to the anterior cingulate cortex (Walsh & Anderson, 2012), the modulation of the FRN by positive and negative RPEs has been suggested as an index of dopaminergic reward processing (Holroyd & Coles, 2002). As such, the finding that IPEs and RPEs were both reflected in a similar fashion in the FRN provides evidence in favour of the common currency hypothesis (Bromberg-Martin & Hikosaka, 2011), according to which the intrinsic value of
information might result from its representation within canonical neural reward-processing circuits.

It is important to note that, as discussed in Section 8.2.3.3, IPEs and RPEs were encoded independently of one another for the task design employed by Study 4. A positive RPE—that is, viewing a black card in the informative stimulus—could be associated with either a positive, negative, or null IPE depending on the composition of cards preceding and succeeding the event.

Interestingly, analyses of trial outcome screens revealed no significant modulation of the FRN by wins versus losses. This is inconsistent both with previous FRN research showing that rewarding outcomes modulate the FRN (M. X. Cohen et al., 2007; Hajcak et al., 2006), and also with the modulation of the FRN by reward prediction errors in response to informative stimuli in the present study. These findings could be due to underpowered statistical analyses: the primary focus of Study 4 was on event-related potentials elicited by informative stimuli, and as such the design of this study was not optimised for analysis of ERPs elicited by outcome screens. Specifically, the task used in Study 4 included only one outcome screen event per trial, whereas each trial consisted of several card-elicited reward events, meaning analyses of informative stimuli had comparatively more statistical power than analyses of outcome screens. Alternatively, it should be noted that according to one model of information-seeking behaviour, the value of reward-predictive stimuli may exceed the value of the rewarding outcome itself, as a result of the increase in anticipatory utility associated with positive predictive cues (Iigaya et al., 2016). This may have resulted in a reduced amplitude difference for wins and losses at trial outcome screens relative to information about wins and losses during informative stimuli.

Interestingly, several recent findings have challenged the RPE-FRN model of Holroyd and Coles (2002). For instance, Talmi, Atkinson, and El-Deredy (2013) found that FRN amplitude, in addition to increasing when reward was unexpectedly withheld, also increased when aversive outcomes were unexpectedly withheld. Since unexpectedly withheld aversion represents a positive RPE, the Holroyd and Coles (2002) model predicts the opposite pattern. Similarly, Hauser et al. (2014) reported that the FRN was more strongly associated with the absolute value of RPEs, rather than signed RPEs, and therefore concluded that FRN amplitudes were driven more by surprise than by outcome.
valence. The results of Study 4 may illustrate an alternative interpretation of these past findings. The findings of Study 4 suggest that the FRN encodes information as well as reward; this finding cannot be explained as a form of surprise encoding, since black and red cards were equally probable for each card draw, and therefore equally surprising according to standard operationalisations of stimulus-bound surprise (e.g., Mars et al., 2008; O'Reilly et al., 2013). Rather, it is possible that past findings demonstrating surprise encoding in the FRN may reflect a complex interaction between RPEs and IPEs. An important limitation on interpretation of these results, however, is that the ERP analyses included in Study 4 had a relatively low numbers of trials per experimental cell (participants were included in analyses if they had at least 20 events of interest remaining in each cell following pre-processing). This may have resulted in a relatively poor signal-to-noise ratio in these analyses, and therefore caution is warranted in the interpretation of patterns of significant and non-significant results.

Under the hypotheses set out above, it was not hypothesised that there would be any effects of either reward or information prediction errors on the FRN during non-informative task events. As expected, for these events there was no effect of RPEs on FRN amplitude, as well as no effect of the amount of outcome-relevant information. However, it was found that there was a small but significant difference in amplitude of the FRN elicited by non-informative stimuli during the equivalent of positive and negative IPEs, and this effect was in the same direction as that observed in informative stimuli. One possible explanation for this finding is that, although participants did not receive information about the lottery outcome in the non-informative stimulus, this stimulus may have imparted incidental distributional information. That is, the relative proportions of red and black cards in the non-informative stimulus may have allowed participants to update their beliefs regarding the generative binomial rate of card colours. Alternatively, given the simple and repetitive task design of Study 4, another possible explanation for this finding is that participants might also have been unable to suppress tracking the cards ‘as if’ they contained outcome-relevant information.

The current study is the first to investigate similarities between information and reward processing in human participants using EEG. The primary finding of Study 4, that IPEs and RPEs were both reflected in the FRN, is conceptually consistent with previous studies showing that informative stimuli are encoded in brain regions traditionally
associated with reward processing. These include the dopaminergic midbrain and lateral habenula (Bromberg-Martin & Hikosaka, 2009, 2011), the orbitofrontal cortex (Blanchard et al., 2015), regions of the striatum (Jepma et al., 2012; Kang et al., 2009) and anterior insula (Preuschoff et al., 2008). Under the common currency hypothesis, and premised upon the assumption that the FRN denotes a reward positivity (Holroyd et al., 2011), these findings suggest that acquiring information may be inherently rewarding, regardless of the instrumental use of the information provided. An expected-reward-maximising agent would not give up monetary reward for information that cannot be used to affect task outcomes. However, if information itself has an inherent motivational value, then this value can offset the monetary cost to the participant.

Two distinct neural mechanisms have been proposed which can account for this ‘common currency’ of information and reward. Bromberg-Martin and Hikosaka (2011) posited that the resolution of uncertainty may itself be inherently rewarding, meaning that information has an explicit value unrelated to its instrumental utility for future planning. This explicit value was proposed to manifest in the encoding of IPEs in dopaminergic midbrain neurons. Alternatively, Iigaya et al. (2016) noted that animals awaiting the outcome of a lottery might experience anticipatory utility (dread of expected losses and savouring of expected wins; Loewenstein, 1987). In such a scenario, Iigaya et al. (2016) proposed that reward prediction errors elicited by informative stimuli might provide an additive boost to this anticipatory utility. Such a mechanism could result in an encoding of IPEs within canonical reward-processing neurons without the necessity of assuming an explicit value of information. As such, the findings of Study 4 are consistent with the mechanisms proposed by both Bromberg-Martin and Hikosaka (2011) and Iigaya et al. (2016). ERP components recorded at the scalp do not measure the activity of dopaminergic midbrain neurons directly, and even direct measures from these brain regions have proven insufficient to distinguish between these two accounts. As such, both models propose viable candidate physiological mechanisms for the behavioural and ERP findings of Study 4. It is important to note, however, that the EEG data collected in Study 4 may not be the most appropriate data modality for resolving these theoretical questions. This is because the reward-processing neural circuits in the dopaminergic midbrain are not directly represented in EEG data recorded at the scalp, and can therefore only be measured via indirect proxy measures, such as the FRN. For this reason, future research
should assess the encoding of information prediction errors and reward prediction errors in the human brain using a neuroimaging modality such as fMRI, which has greater spatial resolution for subcortical structures implicated in dopaminergic reward processing.

In sum, Study 4 found that, as in Study 3, human participants exhibited a clear preference for acquiring non-instrumental information about future outcomes. Moreover, the neural encoding of information prediction displayed striking similarities to patterns of encoding of reward prediction errors. An updated decision theory in which information itself is a dimension of stimuli that contributes to their hedonic value may assist in explaining and predicting patterns of decision making in the presence of reducible uncertainty.
9. General Discussion

The aim of this thesis was to investigate the neural and cognitive processes underlying belief updating and information seeking in human decision making under uncertainty. These research questions fall within an overall research program which characterises as a primary computational function of the brain the construction and maintenance of probabilistic internal models of the external world. From this perspective, the study of belief updating concerns the processes by which these internal models are updated, and the study of information seeking concerns the neural and cognitive mechanisms which underlie decisions to seek out information with which to maintain well-calibrated internal models. The results of the four empirical studies reported in this thesis demonstrate that this belief-centric research perspective has substantial explanatory power for behavioural and neural data in human decision making under uncertainty.

In the present chapter, I will first review key findings from each of the four empirical studies reported in this thesis. I will then discuss the implications of these findings for the belief updating literature and the information seeking literature separately, as well as for the overall belief-centric research perspective that underpins the research program of this thesis. Finally, I will detail several limitations of the work reported in this thesis, and will consider key open questions to be addressed by future research in this field of cognitive neuroscience.

9.1 Review of key findings

Study 1 investigated the neural substrates of belief updating in perceptual learning using EEG data collected from eighteen healthy young adult participants. Specifically, this study tested the association between different aspects of belief updating as quantified by a Bayesian computational model and the event-related potential. It was found that single-trial amplitudes of the P3 component of the event-related potential were positively associated with belief update size, and that single-trial amplitudes of the SPN component were negatively associated with belief uncertainty. In the case of the P3, this component’s encoding of belief update magnitude was consistent with the proposal that P3 amplitude reflects the degree to which internal models of the environment are revised following new information (Kolossa et al., 2015; Kopp, 2008; Mars et al., 2008).
In addition, Study 1 also sought to compare the fit of two models to participants’ behaviour: a Bayesian belief updating model and a heuristic (WSLS) model. Formal comparison of these computational cognitive models indicated that, at a whole-group level as well as for a majority of individual participants, the Bayesian belief updating model provided a better account of choice behaviour. This suggested that participants’ behaviour in the perceptual learning task used in Study 1 was broadly consistent with Bayesian belief updating. This also provided empirical support for the use of a Bayesian belief updating model to quantify belief variables for use in single-trial ERP analyses.

A final noteworthy result of Study 1 was that there was considerable heterogeneity in the best-fitting computational model across different participants. In spite of the overall better fit of the Bayesian belief updating model, it was nevertheless the case that five of sixteen participants made choices better explained by the WSLS heuristic than by Bayesian belief updating. This raised the question of what factors affected the use of Bayesian belief updating versus reliance on heuristics among individual participants. Moreover, it was also unclear from the results of Study 1 alone whether the encoding of belief update magnitude in the P3 component of the event-related potential only occurred for participants who adopted a Bayesian strategy, or whether this was the case across all participants. This latter question has bearing on the interpretation of the ERP results of Study 1; specifically, this question was important in determining whether the association of P3 amplitude and belief update magnitude observed in Study 1 was evidence of a neural belief updating mechanism, or simply epiphenomenal.

Using a variant of the perceptual learning task developed by Study 1, Study 2 followed up on several unresolved questions concerning the neural substrates of belief updating. Whereas in Study 1 performance feedback in the perceptual learning task was provided solely in the form of monetary reward (0 to 25 cents per trial), Study 2 presented feedback in the form of either monetary reward or instructional directives (e.g., “Your chosen contrast was 12.5 per cent from the target). Using this performance incentive manipulation Study 2 investigated the effect of participants’ task motivation upon the use of Bayesian versus heuristic task strategies, and upon the neural substrates of feedback processing.

Analysis of behavioural data revealed that at the group level, participants performed significantly more accurately in the monetary feedback condition than in the
instructive feedback condition. Similarly, at a group level there were differences between monetary and instructive feedback conditions in the amplitudes of three ERP components: the P3, the FRN, and the late positive potential (LPP). However, a model-based cluster analysis revealed that in fact, these group-level behavioural and neural results were well explained by the presence of two distinct subgroups in the data: a ‘Consistent Bayesian’ (CB) subgroup, and an ‘Incentive-Compatible Switching’ (ICS) subgroup. The CB subgroup was composed of those participants whose behaviour was best accounted for by a model assuming the use of a Bayesian belief updating strategy in both monetary and instructive feedback conditions. By contrast, the ICS subgroup comprised participants whose behaviour was best explained by the use of a Bayesian strategy in the monetary feedback condition and a WSLS heuristic strategy in the instructive feedback condition.

Crucially, subgroup analyses of both behavioural and ERP data revealed that participants’ subgroup membership interacted with the main effect of feedback condition. For behavioural data, this interaction was such that the group-level difference between monetary and instructive feedback was present only for participants in the ICS subgroup, who switched between a Bayesian and a heuristic strategy in different feedback conditions. Participants in the CB subgroup showed no significant difference in performance accuracy between feedback conditions, consistent with the proposal that these participants adopted a Bayesian belief updating strategy in both conditions. Similarly, ERP effects were such that the amplitudes of the P3 and LPP components of the event-related potential differed between feedback conditions only for participants in the ICS subgroup, who switched strategies between conditions. This implicates these components in belief updating, and specifically suggests that the association between P3 amplitude and belief update size observed in Study 1 was more likely to be of functional significance than merely the result of an incidental confound. The amplitude of the FRN component also differed between monetary and instructive feedback, but this main effect did not interact with participant subgroup, suggesting that the FRN may have reflected feedback evaluation processes unrelated to the behavioural strategy employed by participants.

Following Study 2, the research focus of this thesis turned from belief updating to information seeking. Study 3 and Study 4 sought to investigate the nature of the value
assigned to information in decision making under uncertainty by human participants. This question was investigated over two behavioural experiments (Study 3) and one EEG experiment (Study 4) using the same basic behavioural task.

Study 3 developed a non-instrumental information seeking task to allow for incentive-compatible elicitation of participants’ preferences for acquiring non-instrumental information. In each trial of this task, participants could observe either an informative stimulus, which predicted the outcome of a lottery at the end of the trial, or a non-informative stimulus, which was perceptually equivalent to the informative stimulus but imparted no predictive information. The information contained in the informative stimulus was non-instrumental, in the sense that it resolved uncertainty about a foregone future outcome, but could not itself be used to increase any expected winnings. The use of non-instrumental information in this task was in keeping with similar studies in the animal literature (Bromberg-Martin & Hikosaka, 2009; Vasconcelos et al., 2015; Zentall & Stagner, 2011), and had the effect that preferences for observing the informative stimulus could be directly interpreted with respect to participants’ preferences concerning information itself, rather than the outcomes to which the information pertained.

The results of Study 3 demonstrated that participants placed a positive value on information imparted by the informative stimulus, and even sacrificed portions of future reward in exchange for observing this stimulus. This was inconsistent with the predictions of expected reward maximisation associated with instrumental valuation of information, but consistent with the proposal that humans assign an intrinsic value to information. The results of Study 3 also demonstrated that participants’ willingness to pay for non-instrumental information was modulated by both the cost of information and the rate at which the information was presented. Specifically, it was found that participants’ willingness to pay for information decreased as the cost of information increased, and also as the rate of information delivery slowed. To account for these findings, Study 3 formulated a computational model of information valuation, the Uncertainty Penalty model, which provided a good account of participants’ intrinsic valuation of information by assuming that this value resulted from aversion to temporally prolonged uncertainty. Importantly, this Uncertainty Penalty model provided a significantly better fit to behaviour than a competitor, the Expected Value of Information model, which instantiated theories positing solely instrumental valuation of information.
Study 4 investigated the neural correlates of this intrinsic valuation of information, using EEG data from participants completing the non-instrumental information seeking task developed in Study 3. Specifically, Study 4 investigated the predictions of one prominent theory of information seeking, the common currency hypothesis of Bromberg-Martin and Hikosaka (2011). According to this hypothesis, the intrinsic valuation of information observed in Study 3 is posited to result from commonalities in the neural encoding of information and primary reward within dopaminergic reward-processing regions of the brain. For EEG data, this hypothesis predicted that the FRN, an ERP component linked with dopaminergic reward processing (Holroyd & Coles, 2002), should encode the value of both information and monetary reward. Consistent with this hypothesis, results showed that ERPs evoked by the presentation of informative stimuli independently encoded both a reward prediction error and an information prediction error, in a comparable fashion. This result provided evidence that the intrinsic valuation of information observed in participants across three experiments in Study 3 and Study 4 may be grounded in common neural processing substrates for information and reward, as proposed by Bromberg-Martin and Hikosaka (2011).

9.2 Implications of research findings

This section considers the broader implications of the research findings of the present thesis for the study of human belief updating (Section 9.2.1) and information seeking (Section 9.2.2). These findings are interpreted in the context of an overall belief-centric research program, which unites the study of belief updating and information seeking within a broader focus on participants’ formation of internal models of the external environment.

9.2.1 Belief updating

One of the primary goals of Studies 1 and 2 of the present thesis was to investigate the neural correlates of human belief updating in the context of perceptual learning. As reviewed above, single-trial and model-based analyses of EEG data in these two studies particularly implicated that the P3 component of the ERP in Bayesian belief updating. Specifically, Study 1 found a positive association between the amplitude of the P3 and
the magnitude of belief updates, such that larger belief updates were associated with feedback stimuli which elicited larger P3 components. This relationship was consistent with previous conjecture by Kopp (2008) and Mars et al. (2008), as well as with a more recent empirical study by Kolossa et al. (2015). In addition, Study 2 found that the P3 elicited by performance feedback differed in amplitude between monetary and instructive feedback conditions only for participants who switched between a Bayesian and a heuristic behavioural strategy between these feedback conditions. Taken together, these results suggest a robust involvement of the P3 in the Bayesian updating of beliefs. From a methodological perspective, the finding that the P3 indexes belief update magnitude suggests that it may be possible for future research use the P3 as a covert marker of belief revision even in the absence of explicit report. Further research in other task domains than perceptual learning should investigate this possibility.

The observed relationship between P3 amplitude and belief update size also has theoretical implications for the interpretation of past research into the functional significance of the P3 component. Indeed, a relationship between P3 amplitude and belief update size may provide a unifying account of a number of disparate P3 results in past research. A number of distinct experimental manipulations have been associated with changes in the amplitude of the P3, and the results of the present thesis suggest that it may be possible to redescribe many of these effects in terms of the updating of an internal model of the environment. For instance, larger P3 components are typically elicited by stimuli which impart greater information (S. Sutton et al., 1967), appear infrequently (S. Sutton et al., 1965), or are contextually novel (Friedman et al., 2001). Each of these manipulations might be considered as varying the extent to which participants are required to model an internal model of the task, such as a model of stimulus probability, or the strength of the association between different stimuli. In this sense, the results of the present thesis are consistent with, and an extension of, the context updating theory of Donchin and Coles (1988). This theory proposes that the amplitude of the P3 reflects the degree to which schemata concerning stimulus context are updated. As discussed by Kopp (2008), Bayesian belief updating might be considered a more general expression of this theory.

The results of the present thesis also have bearing on the distinction between Bayesian and non-Bayesian theories of belief updating introduced in Chapter 2. To
review, although many prominent theories assume that human belief updating takes place according to Bayesian statistical principles, there also exists substantial evidence that in many domains, human participants update beliefs in a qualitatively non-Bayesian fashion (e.g., Hogarth & Einhorn, 1992; Phillips & Edwards, 1966; Sharot et al., 2011). This is particularly the case in incentivised learning tasks, as assessed in Study 1 and Study 2 of the present thesis. In such tasks, participants often exhibit conflict between Bayesian task strategies and reinforcement-based heuristics (Achtziger et al., 2015; Steyvers et al., 2009) which results in the utilisation of qualitatively non-Bayesian belief updating strategies. One of the primary goals of the present thesis was therefore to compare the fit of Bayesian and heuristic models to participants’ perceptual learning, and therefore to investigate the circumstances under which these different strategies predominated.

Taken together, the results of Study 1 and Study 2 suggest that the presence of monetary incentives for good performance motivated participants to engage in Bayesian belief updating. This is supported by the finding that a Bayesian model provided the best overall explanation of group-level behaviour in Study 1, where feedback was provided to participants in the form of monetary reinforcement. In addition, the results of Study 2 demonstrated that, in the absence of monetary reinforcement for performance, a substantial proportion of participants (those in the Incentive-Compatible Switching subgroup) switched from a Bayesian to a heuristic choice strategy. Given the greater cognitive demands of Bayesian belief updating relative to a heuristic choice strategy (Cooper, 1990), this suggests that, for many participants, the utilisation of a Bayesian task strategy was dependent upon the presence of adequate incentives to motivate performance.

Although these results may appear to emphasise the distinctness of Bayesian and non-Bayesian task strategies, it may be possible to reconcile the two perspectives within a single ecological framework which emphasises the resource limitations of human cognition. As discussed in Section 2.2.5, studies assessing the use of heuristic strategies in human judgment and decision making typically emphasise that, in many circumstances, heuristics provide a ‘fast and frugal’ alternative to optimal task strategies, which typically require greater expenditure of cognitive resources (Gigerenzer & Goldstein, 1996). As such, it is possible that human decision makers may allocate cognitive resources to different problems depending on a comparison between the
expected benefits of using a more resource-intensive cognitive strategy (such as Bayesian belief updating) and the required expenditure of cognitive resources associated with this strategy. In cases where the benefits associated with use of Bayesian belief updating are outweighed by the greater computational demands of Bayesian inference, a decision maker may reasonably elect not to engage in Bayesian belief updating. Moreover, individuals may differ with respect to the criterion employed to make this meta-cognitive strategy choice, as a result of factors such as reward sensitivity—which would affect the relative desirability of outcomes—or the availability of cognitive resources, which would affect the relative subjective cost of computationally demanding task strategies. Viewed from this perspective, the results of Study 2 suggest that strategy-switching may have resulted from a comparison between the marginal benefits of employing Bayesian belief updating and the increased computational demands associated with a Bayesian strategy relative to a heuristic strategy. Participants in the ICS subgroup, whose behaviour was consistent with Bayesian belief updating in the presence of monetary reward but a WSLS heuristic in the absence of monetary reward, would therefore be those for whom the computational costs of Bayesian inference outweighed the incentive value of good performance in the instructive feedback condition.

This perspective is in keeping with the notion of procedural rationality introduced by Simon (1976), or Type II rationality, as described by Good (1983). More broadly, such accounts of judgment and decision making, which emphasises the resource limitations of human cognition, are Bayesian in the sense of Bayesian decision theory (e.g., J. R. Anderson, 1990; Körding & Wolpert, 2006), in which decisions are made according to a weighting of the expected costs and benefits of different strategies. However, this perspective is inconsistent with stronger and more restrictive forms of the Bayesian brain hypothesis which propose that, given a particular generative model of the environment, human judgment and decision making is best explained in terms of Bayesian statistical inference (e.g., Mathys et al., 2011).

In summary, the results of Study 1 and Study 2 demonstrate that the neural correlates of perceptual learning can be explained well in terms of the updating of probabilistic beliefs. Specifically, the relationship between the P3 and belief update amplitude revealed by Study 1 suggests that this ERP component, which has been extensively studied over the past fifty years of research in cognitive neuroscience
may be understood as reflecting the output of a neural mechanism for belief updating (Kopp, 2008). In addition, the results of Study 2 suggest that in the absence of good motivation to maintain well-calibrated beliefs via Bayesian belief updating, this relationship may be abolished. Specifically, the results of Study 2 demonstrate that human belief updating may be situated within an ecologically valid framework, such that the use of a Bayesian belief updating strategy is subject to an analysis of the costs and benefits of maintaining a good model of the environment. For tasks in which maintaining a well-calibrated internal model of the environment is computationally demanding, or not associated with sufficient incentives to motivate engagement, participants may instead adopt a non-Bayesian heuristic strategy.

More broadly, the results of the present thesis also raise important general questions concerning the nature of information processing in the human brain. It has been observed that, given the resource limitations of human cognition, Bayesian approaches to decision problems in naturalistic environments require computations that exceed the information-processing capacities of even the fastest supercomputers (Sanborn & Chater, 2016). As such, the proposal of the belief-centric theory that the maintenance of internal models of the environment is a computational goal of cognition (Fiorillo, 2008) can only hold if it is also assumed that model updating is subject to constraints imposed by cognitive resource limitations. Such an assumption would be in keeping with recent studies in cognitive psychology and cognitive neuroscience that seek to explicitly quantify the computational demands of different cognitive processes (e.g., Ortega & Braun, 2013; Shenhav, Botvinick & Cohen, 2013). Consistent with the results of the present thesis, research in this field has demonstrated that incorporating additional constraints based on cognitive resource limitations can increase the explanatory power of ideal-observer models with respect to behavioural and neural data (Kool & Botvinick, 2014).

9.2.2 Information seeking

One of the primary theoretical questions addressed by the present thesis was whether human information seeking was more consistent with instrumental or intrinsic valuation of information. To review, instrumental valuation of information (e.g., Howard, 1966; D.
B. Lawrence, 1999; Raiffa & Schlaifer, 1961) proposes that information is valuable to decision makers to the extent that it can be used to obtain future tangible outcomes such as food or money. By contrast, intrinsic valuation of information (e.g., Grant et al., 1998) posits that information is assigned an inherent value by decision-makers, who should, all other things being equal, therefore prefer more information to less. In the non-instrumental information seeking task used in Studies 3 and 4, instrumental valuation of information predicted indifference between the informative stimulus and the non-informative stimulus at zero cost, and a preference for the non-informative stimulus for non-zero costs. By contrast, intrinsic valuation of information predicted a preference for observing the informative stimulus. The behavioural results of Study 3 and Study 4, which found that participants preferred to observe the informative rather than the non-informative stimulus, therefore suggest that participants assigned an intrinsic value to information to non-instrumental information.

A broader perspective on the results of the present thesis suggests that it may be worthwhile to reconsider the distinction between intrinsic and instrumental valuation of information. As discussed in Chapter 3, this procedural distinction was drawn to emphasise that theories of behaviour based solely on expected reward maximisation hold that information is only useful to the extent that it be used to maximise tangible rewards. However, if the acquisition of information is itself valuable, as suggested by the results of Study 3 and Study 4, then under these theories information should itself become a reward to be maximised like any other. From a broader perspective, therefore, the results of the present thesis suggest that information seeking may be the result of both intrinsic and instrumental valuation of information. Such a perspective would dissolve the pedagogical distinction between intrinsic and instrumental valuation of information. This expanded conceptualisation of information as reward raises interesting new theoretical questions. These particularly concern the effects on planned behaviour of computational differences between traditional rewards, which can be aggregated additively, and information-as-reward, which is additive only in expectation, and not in consumption (see Sun, Gomez, & Schmidhuber, 2011 for further discussion of this issue). These theoretical questions are beyond the scope of the present thesis, and should be addressed by future research.
Furthermore, the results of the present thesis provide insight into the cognitive factors underlying intrinsic valuation of information in humans. A prominent theory advanced in previous research is that intrinsic valuation of information might result from an aversion to temporally prolonged uncertainty (Bromberg-Martin & Hikosaka, 2011; Caplin & Leahy, 2001; Epstein, 2008). The computational modelling results of Study 3 in the present thesis were consistent with this account. Specifically, Study 3 found that the Uncertainty Penalty model, which incorporated aversion to temporally prolonged information, provided a good overall account of both participant-level and group-level behaviour. Moreover, it was found that individual differences in the UP model’s $k$ parameter, which governed the strength of aversion to uncertainty in individual participants, were positively correlated with participants’ willingness to pay for non-instrumental information. As such, results provided some support for the hypothesis that information seeking might result from aversion to uncertainty. It is important to note, however, that other cognitive mechanisms can also produce similar results to the UP model; for instance, Study 3 demonstrated that a re-parameterisation of the UP model in terms of an appetitive value of information, rather than an aversion to uncertainty, produced quantitatively equivalent behavioural predictions. More broadly, however, it is important to note that participants’ preferences can only be explained with reference to the information imparted by stimuli, since the informative and non-informative stimuli were perceptually equivalent, and differed only with respect to the information content of cards.

The results of the second behavioural experiment conducted in Study 3 suggest that the rate at which information was delivered affected participants’ valuation of information. Specifically, faster information delivery and shorter overall uncertainty duration were associated with a stronger preference for observing the informative stimulus. Interestingly, this result was contrary to the predictions of a naïve model assuming aversion to uncertainty, which would have predicted an increase in the value of information for uncertainty of a greater duration. However, a barrier to interpretation of this result is the potential confound between information rate and lottery duration in the task design employed by Study 3. In this task, trials in which information was delivered at a faster rate were also those in which information about trial outcomes was delivered sooner. As such, a stronger valuation of information for the fast information rate condition
could have resulted from preferences with respect to either the rate at which information was delivered or temporal discounting of information concerning future outcomes (cf. Green, Myerson, & McFadden, 1997). In line with this proposal, an extension of the UP model was developed with a parameter governing temporal discounting of future information, and individual differences in this parameter were found to be positively correlated with individual differences in the univariate effect of information rate on behaviour. This is also conceptually consistent with a recent study by Iigaya et al. (2016) that developed a model of information seeking behaviour incorporating temporal discounting of future information. The authors proposed a model in which preference for information result—as well as common neural encoding of information and reward—resulted from a boosting of anticipatory utility by the reward prediction errors associated with informative stimuli. Conditional on assumptions about the functional form of anticipatory utility, this model predicts that participants seek information not because of any explicit value for information, but rather because information increases participants’ ability to savour positive future outcomes. In support of this model, Iigaya and colleagues (2016) found a complex nonlinear relationship between the duration of uncertainty and preference for acquiring information. Future research in this field should therefore investigate the relationship between uncertainty duration and information valuation more extensively. One means of achieving this would be to adapt the non-instrumental information seeking task developed within the present thesis to dissociate the effects of information rate and information discounting on the valuation of information. More broadly, future research could seek to distinguish between models proposing an intrinsic value of information and those, like that of Iigaya and colleagues (2016), which propose that information seeking results from boosts to anticipatory utility. Section 9.4 below discusses an experiment which could distinguish between these two models by comparing information seeking for gains versus information seeking for losses.

Finally, the results of Study 4 of the present thesis also provided insight into the neural substrates of information valuation in humans. The primary finding of this study—that the FRN component of the ERP encodes both an information prediction error and a reward prediction error—is consistent with the proposal that valuation of information results from a common neural representation of primary reward and information (Bromberg-Martin & Hikosaka, 2011). Bromberg-Martin and Hikosaka (2011) termed
this proposal the ‘common currency’ hypothesis of information valuation, on the basis that single-neuron recordings from reward-processing brain regions in macaque monkeys encoded both reward prediction errors and information prediction errors using a common coding scheme. Study 4 of the present study tested whether a similar pattern of commonality was present in the processing of reward and information in human participants. The ERP results of Study 4 suggest that this was indeed the case, conditioned on the assumption that the FRN denotes a reward positivity elicited by dopaminergic RPE signalling, as proposed by one prominent hypothesis (Holroyd & Coles, 2002; Holroyd et al., 2011). As discussed in Section 8.4, the encoding of IPEs in the FRN was consistent with two candidate neurobiological coding schemes: an explicit value of information, as proposed by Bromberg-Martin & Hikosaka (2011), or boosting of anticipatory utility by RPEs, as proposed by Iigaya et al. (2016). These two mechanisms make equivalent predictions for the behavioural and neural data of Study 4, and future research should therefore seek to disentangle the predictions of these two different mechanistic models.

Moreover, the results of Study 3 and Study 4 also have bearing on the belief-centric research perspective. This perspective provides a principled explanation for the finding that humans assign an intrinsic value to information, even when this information is non-instrumental. If a primary computational role of the human brain is to maintain well-calibrated internal models of the environment (Fiorillo, 2008), then it is reasonable to propose that biological agents should derive utility from the degree of match between their internal model of the environment and true environmental contingencies (e.g., ‘epistemic value’; Friston et al., 2015). That is, under these assumptions it would be adaptive for biological agents to assign an explicit value to information, since this information aids in maintaining an internal model of the environment which provides a good account of observed sensory experiences (possibly via the minimisation of free energy, as proposed by Friston, 2009). In this case, seeking information would be considered both intrinsically and instrumentally valuable, since it assists in achieving what is posited to be a primary computational goal of the brain.

More broadly, therefore, intrinsic valuation of information is likely to be associated with an evolutionary advantage for biological agents in complex environments. Although quantities such as the instrumental value of information are readily calculable in simple examples in which probabilities are known and agents
possess a correct generative model of the environment, naturalistic environments rarely possess these characteristics. Rather, environmental contingencies in naturalistic environments are ambiguous and non-stationary, with the effect that the calculation of a true instrumental value of information is likely to be computationally intractable for any but the simplest decision problems (see Raiffa & Schlaifer, 1961). For more complex decisions, intrinsic valuation of information is likely to induce a bias toward exploratory and information-gathering actions (Sun, Gomez & Schmidhuber, 2011). By improving an agent’s internal model of the environment, such actions are likely to be associated with increased long-run utility as a result of outcomes which are unforeseeable at the point of the original decision. In other words, assigning an intrinsic value to information may be a heuristic which is adaptive because it reflects the fact that, in real life scenarios, even apparently non-instrumental information may be of unforeseen instrumental use in the future. As such, from the perspective of the belief-centric research program, intrinsic valuation of information might also be conceptualised as an adaptive strategy to aid decision makers in the context of cognitive resource constraints, consistent with the material discussed in Section 9.2.1.

9.3 Limitations

This section discusses limitations to interpretation of the results reported in this thesis. Specifically, this section will detail three general limitations associated with the participant samples and neuroimaging techniques employed by the present thesis, as well as separate specific limitations of the reported belief updating and information seeking studies.

A primary limitation of all studies reported in this thesis concerns participant demographics. All participants in this thesis were healthy young adults aged 18 to 35 years, and the large majority of participants were staff or students of The University of Melbourne. As such, participants in the present thesis were likely to be more highly educated and of a higher socio-economic status than the general population. This may affect the generalisability of the results reported in this thesis. For instance, the experiments reported in Study 1 and Study 2 assessed factors affecting the use of a Bayesian belief updating strategy in perceptual learning. Study 1 reported that, at a group-level, participants’ choices were better explained by a Bayesian belief updating strategy
than by a win-stay lose-shift heuristic strategy. Since Bayesian belief updating is more computationally demanding than use of a heuristic strategy, one of the factors affecting the use of Bayesian inference is likely to be the availability of participants’ cognitive resources (Steyvers et al., 2009). It is therefore probable that use of a Bayesian strategy may have been less demanding for the participant population recruited for this thesis than for the population in general. As such, it is an open question whether the broader population would also exhibit behaviour more consistent with a Bayesian model than with a heuristic model, as reported in Study 1. Notably, this potential confound is not limited to the present thesis, but may affect a great deal of recent research in cognitive neuroscience. Many recent studies assessing the use of Bayesian cognitive strategies have based their conclusions on unrepresentative participant samples, as in the present thesis (e.g., Griffiths & Tenenbaum, 2006; Vossel et al., 2015). It is therefore important to exercise caution in interpreting the patterns of results of these studies as general principles of human cognition.

A second general limitation of the present thesis is the poor spatial resolution of EEG, and the constraints that this places upon the inferences which can be drawn on the basis of EEG data. As discussed in Section 4.1, given a pattern of EEG data, it is not generally possible to identify a unique solution to the inverse problem, and therefore not possible to attribute the observed neural activity to specific cortical sources (Lopes da Silva & van Rotterdam, 2005), though the use of biophysical constraints can assist in resolving this problem in certain cases. In the present study, however, the neural regions of interest included deep subcortical structures such as brainstem nuclei and basal ganglia; for these structures, no source localisation was possible even with the use of reasonable biophysical constrains. As a result, the EEG data collected in this thesis was methodologically unsuitable for addressing several important questions concerning human belief updating and information seeking. For instance, one important question that is indeterminate on the basis of the EEG data collected in the present thesis is what regions of the brain were responsible for the differential encoding of positive and negative information prediction errors in the feedback-related negativity, as observed in Study 4. As discussed in Study 4, these ERP results are theoretically consistent with two conceptually distinct neurophysiological mechanisms for the intrinsic valuation of information. As such, the limited spatial resolution of EEG limits the degree of insight
given by the present thesis into the underlying neural mechanisms generating the observed neural data of Studies 1, 2, and 4. This limitation could be overcome by investigating these questions using neuroimaging modalities with better spatial resolution than EEG, such as magnetoencephalography, fMRI, or electrocorticography.

Finally, a third general limitation of this thesis concerns the manner in which it investigated the relationship between individual ERP components (e.g., the P3, the FRN) and cognitive variables (e.g., belief update size, information prediction errors). This thesis adopted the standard approach in this field, which was to predefined ERP components and EEG electrodes of interest. The relationship between EEG data at these locations and cognitive variables was then assessed using analyses of variance, while controlling the family-wise error rate using a Bonferroni correction. One general limitation associated with this method is that it relies on the specification of a relatively limited number of electrodes and components, and as such was restricted to a purely confirmatory mode of analysis. An alternative strategy for analysing ERP data is a mass-univariate approach, which has also been implemented with success in recent years (e.g., Groppe, Urbach & Kutas, 2011). In the mass-univariate approach, a general linear model is run at each ERP timestep across all electrodes, and correction for multiple comparisons is done using Random Field Theory (see, e.g., Kilner, Kiebel & Friston, 2005), thereby taking advantage of the strong statistical dependence between ERP data at adjacent electrodes and timesteps. This mass-univariate approach allows for a relatively more exploratory analysis of the relationship between ERP data and cognitive variables of interest, and future research into the neural mechanisms of belief updating and information seeking should therefore consider applying this analysis method as a supplementary tool to the confirmatory method employed in this thesis. Such an approach would ensure that ERP analyses are not restricted to a relatively small number of well-understood components, as in the present thesis.

One specific limitation associated with Studies 1 and 2 of the present thesis is that studies of belief updating in this thesis focused solely on the context of perceptual learning. However, as reviewed in Chapter 2, belief updating is an overarching term which also concerns such processes as probabilistic inference and sequential judgment (e.g., Griffiths & Tenenbaum, 2006; Kolossa et al., 2015). The results reported in Study 1 and Study 2 of the present thesis therefore require corroboration in these other domains.
of belief updating before it is possible to draw general conclusions concerning the neural and cognitive processes underlying the entirety of human belief updating. A recent example, which provides a good example of convergent evidence of this kind, comes from recent studies using simple probabilistic inference tasks. These studies demonstrated that, as in Study 1 of the present thesis, when new information was presented in these tasks the amplitude of the P3 component of the ERP encodes the magnitude of updates to beliefs (Kolossa, 2016; Kolossa et al., 2015).

By contrast, it is possible that other ERP results of Study 1 and Study 2 may be restricted to the specific task designs employed in the present thesis. For instance, the relationship between SPN amplitude and belief uncertainty reported by Study 1 is likely to be limited to the case in which participants anticipate the delivery of informative feedback at a known future time. That is, the nature of the SPN is such that this component is typically observed in participants passively anticipating a future stimulus (Damen & Brunia, 1994). Since this does not describe the environmental conditions in which most human belief updating takes place, further research is required to determine whether the association between SPN amplitude and belief uncertainty is robust to changes of task environment.

A specific limitation of the information seeking research presented in Studies 3 and 4 of the present thesis is that, although results demonstrated substantial inter-individual variability in valuation of information, these studies did not shed any light on what cognitive and/or personality factors might underlie these individual differences. It is not clear, for instance, whether differences in valuation of information are attributable to personality factors such as extraversion or neuroticism (cf. DeYoung, 2013). This question is naturally addressed with a correlational analysis, which would have required larger sample sizes to detect effects with conventional levels of statistical power than were available in the present thesis. Section 9.4 below discusses possibilities for investigating this question in future research.

Similarly, the design of Study 4 was unsuitable for investigating whether individual differences in the valuation of information were related to individual differences in the strength of information encoding in the FRN. There are two reasons for this unsuitability: first, as with the personality research questions discussed above, investigation of this question would have required a larger sample size than was recruited.
in the present study. More importantly, a second limitation in this respect was that the
task used in Study 4 was inappropriate for addressing this question. The nature of this
task was such that participants who preferred information more strongly were exposed to
more informative stimuli than uninformative stimuli. Since information prediction errors
were calculated solely on the basis of the informative stimulus, it was therefore the case
that more accurate estimates of the ERP waveform were obtained from participants who
sought out more information. As such, the noise in the measurement of the FRN is
correlated with the strength of preference for information, a potential covariate. This
rendered the design of Study 4 statistically unsuitable for assessing the relationship
between information encoding in the ERP and strength of participants’ preference for
acquiring information. To investigate this specific question, it would be necessary to
design a task in which exposure to informative stimuli was decoupled from the value
assigned to those stimuli by participants. This could be accomplished, for instance, by
stochastically varying the identity of the displayed stimulus on each trial.

9.4 Future research directions

Future research should seek to further explicate the neural and cognitive mechanisms
underlying human belief updating and information seeking. There are two promising
avenues by which this future research might proceed. First, future research could seek to
overcome the methodological limitations of the present thesis, following the prescriptions
set out above in Section 9.3. Secondly, future research should also seek to investigate
important new empirical questions raised by the findings of the present thesis. In this
section, I describe two belief updating research projects and two information seeking
research projects which could advance this field of research.

For belief updating, future research could involve (a) parametrically manipulating
the complexity of Bayesian versus non-Bayesian task strategies in different settings, and
(b) testing whether the relationship between belief update magnitude and P3 amplitude
observed in the present thesis also reflects the non-Bayesian belief updating effects
reviewed in Sections 2.2.1 to 2.2.4. For the first of these projects, formally quantifying
and manipulating the complexity of different belief updating tasks (cf. Murawski &
Bossaerts, 2016) would allow for a more nuanced appreciation of how decision makers
incorporate both cognitive resource demands and task motivation into their decision to
select between different task strategies. This would represent a conceptual advance over Study 2 of the present thesis, which manipulated task motivation using a factorial feedback valence manipulation. Parametrically varying the valence of feedback, for instance, or the perceptual difficulty of a learning task, would provide further insight into this issue. Secondly, future research could also assess whether the relationship between P3 amplitude and belief update magnitude demonstrated by Study 1 is specific to Bayesian belief updating, or whether non-Bayesian belief updating effects are also reflected in this relationship. For instance, using an incentivised urn-ball task (e.g., Phillips & Edwards, 1966), it would be possible to assess whether such non-Bayesian effects as conservatism and updating asymmetry are reflected in the amplitude of the P3. This would shed further light on the neural substrates of belief updating, and would provide a further point of comparison for Bayesian and non-Bayesian models of belief updating.

In the domain of information seeking, two promising potential avenues of future research would be (a) to investigate individual differences in valuation of information using larger and more diverse participant samples, and (b) to determine the effect of other task factors on participants’ preference for information in the information seeking task developed within the present thesis. The first of these proposed projects would overcome the limitation, discussed in Section 9.3, of making general conclusions concerning human information seeking on the basis of an unrepresentative participant sample. Moreover, investigating the relationship between personality factors and the valuation of information would also shed light on the potential cognitive mechanisms of information valuation. For instance, a positive association between information seeking and neuroticism would support the contention of Study 3 that aversion to uncertainty provides a good account of behaviour (Boswell, Thompson-Hollands, Farchione, & Barlow, 2013), whereas a finding that information valuation is positively associated with reward sensitivity would suggest that participants may treat information as an appetitive reward (DeYoung, 2013). In this respect, a further hypothesis is that reward sensitivity will be associated with seeking information pertaining to both potential gains (as in Study 3 and Study 4 of the present thesis) and potential losses, since the amount of information that is available is equivalent in these two cases. By contrast, an association between information seeking and neuroticism is likely to be associated more strongly with a
preference for information pertaining to potential loss outcomes, since the potential for negative outcomes is likely to result in a more aversive subjective experience of uncertainty, and therefore a greater drive to seek information.

Secondly, the non-instrumental information seeking task developed within Studies 3 and 4 of the present thesis could be modified to assess the effects of other task factors on participants’ valuation of information. There are a number of potentially interesting factors which could be manipulated in this task, including the duration of uncertainty, the manner of information delivery (piecewise versus all-at-once; cf. Hogarth & Einhorn, 1992), and a priori win probability. In addition, a theoretically important manipulation would be to compare the valuation of information concerning potential gains, as in the present thesis, with the valuation of information concerning outcomes in the loss domain. This would allow future research to compare the predictions of different models of information seeking; for example, strictly speaking aversion to uncertainty predicts information seeking for both gains and losses (Bromberg-Martin & Hikosaka, 2011; Epstein, 2008), whereas the RPE-boosted anticipation model of Iigaya et al. (2016) predicts information seeking in the gain domain but information avoidance in the loss domain. This follows from the proposal that RPEs elicited by informative stimuli might boost the signed utility of different options, rather than their absolute utility.

9.5 Conclusion

In summary, the results of the present thesis demonstrate the explanatory power of the belief-centric research program introduced in Chapter 1 of this thesis. This program posits that a primary computational goal of the human brain is the construction and maintenance of internal models of the external environment (Fiorillo, 2008; Friston et al., 2015). As a consequence, this perspective holds that many of the fundamental processes underlying human judgment and decision making are best understood as serving this goal. The present thesis adopted this rationale in investigating the neural and cognitive mechanisms of two model-related processes: belief updating and information seeking, and the results of this research has provided evidence that a belief-centric research perspective provides considerable insight into the neural and cognitive mechanisms of human decision making under uncertainty. More broadly, I also wish to suggest that this perspective licenses compelling claims regarding the nature of human cognition, and holds considerable
promise for future research into human judgment and decision making in cognitive psychology and cognitive neuroscience.

The proposal that humans and other animals maintain internal models of the environment is not itself novel; indeed, this proposal can be traced at least as far back as the work of Tolman (1948). Rather, this thesis suggests that the novel insight of a belief-centric research perspective is to consider the calibration of internal models of the environment as a computational goal of cognition. Considering human cognition from this teleological standpoint affords a new perspective on well-studied research questions in cognitive neuroscience. Across four behavioural and EEG studies, the present thesis demonstrates how this perspective reconceptualises neural activity reflected in event-related potentials in terms of the updating of probabilistic beliefs, and provides a principled framework for understanding human information-seeking behaviour.
10. References


Appendix A: The neural mechanisms of Bayesian belief updating.

The following is a commentary article published in *The Journal of Neuroscience* in December 2015. The full citation for this article is as follows:

Journal Club

Editor’s Note: These short, critical reviews of recent papers in the Journal, written exclusively by graduate students or postdoctoral fellows, are intended to summarize the important findings of the paper and provide additional insight and commentary. For more information on the format and purpose of the Journal Club, please see http://www.jneurosci.org/misc/ifac_features.shtml.

The Neural Mechanisms of Bayesian Belief Updating

Daniel Bennett
School of Psychological Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia

A central function of the nervous system is to use sensory information to infer the causal structure of the external world. According to Bayes’ rule, the optimal way of using this information is to calculate the information’s likelihood under various models of the environment, and to weight this likelihood by the strength of prior belief in each model to derive posterior beliefs. In recent years, the influential hypothesis has been advanced that Bayesian inference represents a unifying principle of neural computation (the Bayesian brain hypothesis; Knill and Pouget, 2004). This framework has been applied to many topics, including vision, motor planning, and behavioral conditioning (Courville et al., 2006; Kording and Wolpert, 2006; Yuille and Kersten, 2006), with the overarching goal of identifying how neural computations implement optimal Bayesian statistical principles.

One particular focus of this research has been Bayesian belief updating: the transformation of prior beliefs into posterior beliefs when new information is observed. Although functional magnetic resonance imaging (fMRI) and electroencephalography have been used to identify some of the regions and processes involved in belief updating (O’Reilly et al., 2013; Bennett et al., 2015; Kolossa et al., 2015), an important open question is how these regions interact to update beliefs. Addressing this question would represent a step beyond localization toward a process-based account of the neural mechanisms of Bayesian belief updating.

In a recent paper published in The Journal of Neuroscience, Vossel et al. (2015) investigated this question using fMRI data acquired from 18 healthy participants completing a Posner cueing task (Posner, 1980). In this task, participants made visual saccades toward a target—a supra-threshold Gabor patch—which appeared either left or right of a central fixation point. Before the target was displayed, the presentation of a left- or right-pointing cue predicted the target location with varying cue validity (CV), defined as the proportion of cues that correctly predicted the target. Every 32–36 trials throughout the experiment, CV was changed pseudorandomly to one of three levels (88%, 69%, 50%). Crucially, participants were not aware in advance of when or how CV would change, and so had to learn these contingencies on the basis of experience: a Bayesian inference problem.

To characterize trial-by-trial belief updating, Vossel et al. (2015) analyzed behavioral data using a Bayesian computational model, the hierarchical Gaussian filter (HGF; Mathys et al., 2011), which has previously been successfully applied to Posner cueing tasks (Vossel et al., 2014). The HGF used response speed (RS) as a dependent measure, and assumed that RSs were generated by a bipartite system consisting of a perceptual model, which tracked beliefs about changing CVs, and a response model, which mapped belief strength onto RSs. The HGF’s perceptual model took the form of a three-level Gaussian hierarchy. At the lowest level of this hierarchy were participants’ observations of validly and invalidly cued targets, assumed to be generated by a latent trial-specific CV. At the second level of the HGF hierarchy, CV evolved across trials as a Gaussian random walk to capture the fact that CVs were nonstationary. At the third level of the HGF, the volatility of the second-level random walk itself changed over time as a Gaussian random walk, capturing the fact that participants experienced periods of both stability and volatility during the task. Second- and third-level random walks were respectively parameterized by participant-specific precision parameters $\omega$ and $\theta$, which were assumed to drive individual differences in behavior. For subsequent fMRI analyses, a key feature of the HGF was that on each trial, it estimated the attention-weighted precision of beliefs, denoted $\alpha(\pi^{(t)})$, which can be interpreted as the proportion of attentional resources allocated to the cued location on a given trial.

Vossel et al. (2015) used these participant- and trial-specific values of $\alpha(\pi^{(t)})$ to identify brain regions associated with Bayesian belief updating. fMRI data were first analyzed using a general linear model (GLM) with four first-level regressors of interest: valid, invalid, left-
ward, and rightward cues. Trial-specific values of \(\alpha(\pi^{(1)})\) were then extracted from the HGF and used to parametrically modulate each of these four regressors. Next, a second-level analysis located brain regions where cue validity interacted with the parametric effect of \(\alpha(\pi^{(1)})\) [indicating different \(\alpha(\pi^{(1)})\) regression slopes for valid vs invalid cues]. This pattern was taken to be indicative of a region’s involvement in Bayesian belief updating, since invalid cues violate expectancies and signal a potential change in CV, and should therefore trigger larger belief updates than valid cues. Moreover, this relationship should be modulated by attention-weighted precision: when \(\alpha(\pi^{(1)})\) is high, as in a block with high CV, an invalid cue is a stronger belief-updating trigger than when \(\alpha(\pi^{(1)})\) is lower. Using these criteria, the second-level GLM analysis identified three regions involved in Bayesian belief updating: right anterior putamen, right frontal eye fields (FEF), and right temporoparietal junction (TPJ).

Next, to investigate how Bayesian belief updating was implemented, these three areas were designated as regions of interest for dynamic causal modeling (DCM) analysis. It was found that the best-fitting DCM model was one in which stronger beliefs [higher values of \(\alpha(\pi^{(1)})\)] were associated with decreased correlation between TPJ and FEF activity for validly cued trials, and with increased correlation between TPJ and FEF activity on invalidly cued trials (Vossel et al., 2015, their Fig. 6). This was taken to indicate a coupling between ventral (TPJ) and dorsal (FEF) streams of visual processing, suggesting that coordinated activity between TPJ and FEF may reflect transmission of a belief-updating signal. Such a signal could upregulate attention-related dorsal stream activity when expectancies were violated following invalid cues, and downregulate activity following valid cues. This conclusion is consistent with, and provides a mechanistic explanation for, previous work implicating TPJ in Bayesian updating of internal models of the environment (Geng and Vossel, 2013).

Although anterior cingulate cortex (ACC) has been implicated in Bayesian belief updating in previous research (O’Reilly et al., 2013), ACC was not one of the belief-updating regions identified by Vossel et al. (2015). This inconsistency may be driven by use of different belief-updating metrics in different studies: rather than define belief updates by an interaction between CV and \(\alpha(\pi^{(1)})\), O’Reilly et al. (2013) searched for regions encoding belief-update magnitude. These different metrics might therefore have identified distinct components of a broader-scale belief-updating network. Similarly, although ACC is thought to encode environmental volatility in learning under uncertainty (Behrens et al., 2007), Vossel et al. (2015) found no significant effect of volatility in ACC or any other brain region. One possible explanation for this discrepancy is the Posner cueing task’s very short cue–target interval (600 ms). Because of the poor temporal resolution of fMRI, this would have meant that the GLM analysis was unable to disentangle cue processing from response execution. As a result, the fMRI results of Vossel et al. (2015) may not be directly comparable to past research by Behrens et al. (2007). Future studies could disentangle these discrete task stages or by using a technique with better temporal resolution, such as electroencephalography.

A particular strength of the study by Vossel et al. (2015) was the manner in which it combined Bayesian computational modeling with fMRI data analysis. For both behavioral data and GLM analysis, the authors were able to show that the HGF fit data better than non-Bayesian competitor models: a Rescorla-Wagner learning rule and a model assuming participants knew the true CV in each block. This supports the conclusion that participants were behaving Bayes-optimally. However, an important caveat here is that the flexibility afforded to some Bayesian observer models by their parameterization and choice of prior might mean that it is not possible to empirically falsify the hypothesis that the brain acts as a Bayesian observer (Daunizeau et al., 2010; Bowers and Davis, 2012). Although it is strongly suggestive that a Bayesian model explained both behavioral and neural data better than non-Bayesian competitors, this is not logically sufficient to show that participants necessarily acted as Bayesian observers. Indeed, Vossel et al. (2015) do not make this claim.

Moreover, it remains unclear whether the assumption of Bayes-optimality in the HGF is viable in more complex environments than a Posner cueing task. Payzan-LeNestour and Bossaerts (2011) demonstrated that in complex environments, Bayesian models became increasingly computationally intractable, and no longer fit behavioral data better than non-Bayesian competitors. Furthermore, even in environments suitable for Bayesian inference, simple heuristics can provide a better account of behavior for comparable subsets of participants (Steyvers et al., 2009; Bennett et al., 2015). A potential solution to this problem is given by a recent study showing that constraints based in principles of efficient sensory coding enabled a Bayesian model to explain seemingly anti-Bayesian perceptions (Wei and Stocker, 2015). Similarly, constraining Bayesian observer models by neurophysiological principles such as capacity limits on processing may allow these models to be successfully applied to more complex environments.

In summary, the work of Vossel et al. (2015) provides a compelling synthesis of behavioral modeling and neuroimaging. By combining a sophisticated behavioral model with DCM analysis of neural data, the authors identified potential neural mechanisms of Bayesian belief updating in deployment of spatial attention. These findings represent a valuable step toward a process-based account of belief updating in the Bayesian brain.

References


Appendix B: Systematic overestimation of reflection impulsivity in the information sampling task.

The following is a commentary article in press in *Biological Psychiatry*. The full citation for this article is as follows:

Systematic Overestimation of Reflection Impulsivity in the Information Sampling Task

To the Editor:

Impulsivity, a psychological construct comprising both motor and cognitive factors, is thought to underlie important inter-individual differences in health and disease (1). In particular, reflection impulsivity, which refers to the tendency to gather and evaluate information before decision making (2), has been implicated in many psychiatric and neurological disorders (3-5). One of the standard tasks for measuring reflection impulsivity in healthy and clinical populations is the Information Sampling Task Automated Battery (CANTAB) (6). In this CANTAB version of the IST, participants sample a variable amount of information about an uncertain outcome before making a decision. The amount of information sampled before the decision gives a measure of participants’ reflection impulsivity. In this correspondence, we show that the calculation of the IST’s main outcome measure, \( P(\text{correct}) \), is based on incorrect statistical inference, resulting in systematic overestimation of participants’ reflection impulsivity and potentially inflated type II error rates. This might affect the results of numerous recent psychopharmacological, neuropsychological, and psychiatric publications that have used the IST (4,5,7).

In the IST, participants are presented with a 5 × 5 grid of closed boxes, each of which is of one of two colors. Participants indicate on each trial which of the two colors they believe to be in the majority, and can sample information before making a decision by opening boxes one at a time to reveal their colors. The task’s two outcome measures are 1) the number of boxes opened before the decision is made and 2) \( P(\text{correct}) \), or the probability that the participant’s chosen color was in the majority, based on the boxes opened at the time of response. Many studies (3-5,7) tend to focus on \( P(\text{correct}) \) as the outcome measure because it is thought to provide the more accurate estimate of reflection impulsivity, given that a given number of open boxes can be associated with very different levels of decision uncertainty, depending on the colors of the open boxes.

Let \( n_1 \) be the number of open boxes of the color the participant chose and \( n_2 \) the number of open boxes of the unchosen color. The IST calculates \( P(\text{correct}) \) as follows:

\[
P(\text{correct}) = \frac{\sum_{k=0}^{z} \binom{z}{k}}{z^2}
\]  \hspace{1cm} (1)

where \( Z = 25 - (n_1 + n_2) \), the number of boxes still to be opened, and \( A = 13 - n_1 \), the number of additional boxes of the chosen color required for a majority. Intuitively, Equation 1 enumerates combinations of boxes that would provide a majority for the chosen color (numerator) as a fraction of possible configurations of colors in the unopened boxes, assuming that each unopened box has an equal probability of being either of the two colors (denominator). Although prima facie reasonable, this assumption of equal probability mistakenly assumes that unopened box colors are conditionally independent from the boxes already opened. In fact, however, the colors of the open boxes provide important information to the participant regarding the underlying proportions of the two colors. For instance, if nine boxes have been opened and all nine are blue, it is extremely unlikely that the sixteen unopened boxes are half blue and half yellow. Because of this conditional dependence, the colors of open boxes provide information about the majority color that is not incorporated into Equation 1. As a result, Equation 1 underestimates \( P(\text{correct}) \), and therefore overestimates participants’ reflection impulsivity.

An accurate calculation of \( P(\text{correct}) \) must take into account the colors of the open boxes. Formally, this can be expressed as a Bayesian inference problem over \( \theta \), the true number of boxes of the chosen color in the entire grid. Following Bayes’ rule, the probability mass function (PMF) for \( \theta \) given the colors of open boxes is given by Equation 2:

\[
Pr(\theta = M|n_1, n_2) = \frac{Pr(\theta = M)Pr(n_1, n_2|\theta = M)}{Pr(n_1, n_2)}
\]  \hspace{1cm} (2)

where \( Pr(\theta = M) \) is the prior probability that \( \theta \) takes a given value, \( Pr(n_1, n_2|\theta = M) \) is the likelihood of the opened boxes if the true number of boxes of the chosen color were \( M \), and \( Pr(n_1, n_2) \) is a normalization constant. Under a naive prior (i.e., because participants are given no advance knowledge regarding the relative proportions of colors), this PMF is given by Equation 3 (8):

\[
Pr(\theta = M|n_1, n_2) = \frac{\binom{M}{n_1} \cdot \frac{25 - M}{n_2}}{\sum_{y=0}^{n_1} \binom{y}{n_1} \cdot \frac{25 - y}{n_2}}
\]  \hspace{1cm} (3)

A more accurate \( P(\text{correct}) \) can therefore be calculated as follows (MATLAB code available at https://github.com/danielbrianbennett/ist/):

\[
P(\text{correct}) = Pr(\theta \geq 13|n_1, n_2) = \sum_{M=13}^{25} Pr(\theta = M|n_1, n_2)
\]  \hspace{1cm} (4)

Although the error in Equation 1 is problematic in itself, a greater concern is that the size of the error is likely to vary systematically across participants. The discrepancy between the two \( P(\text{correct}) \) measures is largest for small numbers of boxes opened, and approaches zero with greater numbers of boxes opened. Consequently, the original \( P(\text{correct}) \) measure overestimates reflection impulsivity more for participants high in reflection impulsivity than for those low in reflection impulsivity.

References:

1. Clark et al. (3) and included in the widely used Cambridge Neuropsychological Test Automated Battery (CANTAB) (6).
2. Impulsivity, a psychological construct comprising both motor and cognitive factors, is thought to underlie important inter-individual differences in health and disease (1).
3. Reflection impulsivity, which refers to the tendency to gather and evaluate information before decision making (2), has been implicated in many psychiatric and neurological disorders (3-5).
4. The IST's main outcome measure, \( P(\text{correct}) \), is based on incorrect statistical inference, resulting in systematic overestimation of participants' reflection impulsivity and potentially inflated type II error rates.
5. Many studies (3-5,7) tend to focus on \( P(\text{correct}) \) as the outcome measure because it is thought to provide the more accurate estimate of reflection impulsivity, given that a given number of open boxes can be associated with very different levels of decision uncertainty, depending on the colors of the open boxes.
6. Let \( n_1 \) be the number of open boxes of the color the participant chose and \( n_2 \) the number of open boxes of the unchosen color. The IST calculates \( P(\text{correct}) \) as follows:
7. \[
P(\text{correct}) = \frac{\sum_{k=0}^{z} \binom{z}{k}}{z^2}
\]  \hspace{1cm} (1)
8. An accurate calculation of \( P(\text{correct}) \) must take into account the colors of the open boxes. Formally, this can be expressed as a Bayesian inference problem over \( \theta \), the true number of boxes of the chosen color in the entire grid. Following Bayes' rule, the probability mass function (PMF) for \( \theta \) given the colors of open boxes is given by Equation 2:
9. \[
Pr(\theta = M|n_1, n_2) = \frac{Pr(\theta = M)Pr(n_1, n_2|\theta = M)}{Pr(n_1, n_2)}
\]  \hspace{1cm} (2)
10. Where \( Pr(\theta = M) \) is the prior probability that \( \theta \) takes a given value, \( Pr(n_1, n_2|\theta = M) \) is the likelihood of the opened boxes if the true number of boxes of the chosen color were \( M \), and \( Pr(n_1, n_2) \) is a normalization constant. Under a naive prior (i.e., because participants are given no advance knowledge regarding the relative proportions of colors), this PMF is given by Equation 3 (8):
11. \[
Pr(\theta = M|n_1, n_2) = \frac{\binom{M}{n_1} \cdot \frac{25 - M}{n_2}}{\sum_{y=0}^{n_1} \binom{y}{n_1} \cdot \frac{25 - y}{n_2}}
\]  \hspace{1cm} (3)
12. A more accurate \( P(\text{correct}) \) can therefore be calculated as follows (MATLAB code available at https://github.com/danielbrianbennett/ist/):
13. \[
P(\text{correct}) = Pr(\theta \geq 13|n_1, n_2) = \sum_{M=13}^{25} Pr(\theta = M|n_1, n_2)
\]  \hspace{1cm} (4)
Empirical findings support this conclusion. Reanalysis of data from a separate sample of 109 participants (9) revealed that the size of the $p_{\text{correct}}$ error varied as a function of number of boxes opened (Figure 1). The average amount of underestimation of $p_{\text{correct}}$ using the IST formula was 7.7% (SD = 3.56%). Significant group-mean differences of approximately this size have been previously reported (3,5), and therefore this error size is clearly of practical significance.

In summary, the computation of the major outcome measure of the IST is based on a flawed statistical assumption that results in systematic overestimation of reflection impulsivity. The size of this error differs for different levels of reflection impulsivity, and therefore the overestimation may compromise statistical inferences made using this measure. In particular, past research reporting group differences in reflection impulsivity (4,5) may have significantly underestimated effect sizes, whereas studies reporting null findings (7) might have failed to detect true group differences in reflection impulsivity.

Daniel Bennett
Stuart Oldham
Andrew Dawson
Linden Parkes
Carsten Murawski
Murat Yücel

Acknowledgments and Disclosures
This work was supported by a Strategic Research Initiative Grant (to CM) from the University of Melbourne, Australia, and the National Health and Medical Research Council of Australia (Grant No. APP1021973 to MY).

We thank Marni Kras for providing a dataset for reanalysis.

The authors report no biomedical financial interests or potential conflicts of interest.

Article Information
From the School of Psychological Sciences (DB) and the Department of Finance (DB, CM), The University of Melbourne, Melbourne; and the School of Psychological Sciences and the Monash Institute of Cognitive and Clinical Neurosciences (SO, AD, LP, MY), Monash University, Victoria, Australia.

CM and MY contributed equally to this work.

Address correspondence to Carsten Murawski, Ph.D., Decision Neuroscience Laboratory, The University of Melbourne, Victoria 3010, Australia; E-mail: carstenm@unimelb.edu.au.

See also associated correspondence: http://dx.doi.org/10.1016/j.biopsych.2016.06.018.

Received May 10, 2016; accepted May 10, 2016.

References