

**Belief as a Wise Wager:
the Neural Representation of Uncertainty, Surprise
and Confidence Across Cognitive and Perceptual
Domains**

THÈSE N° 8432 (2018)

PRÉSENTÉE LE 1^{ER} MARS 2018
À LA FACULTÉ DES SCIENCES DE LA VIE
LABORATOIRE DE GÉNÉTIQUE COMPORTEMENTALE
PROGRAMME DOCTORAL EN NEUROSCIENCES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Leyla LOUED-KHENISSI

acceptée sur proposition du jury:

Prof. D. Ghezzi, président du jury
Prof. C. Sandi, Prof. K. Preuschoff, directrices de thèse
Prof. C. Mohr, rapporteuse
Prof. D. Rudrauf, rapporteur
Prof. M. Herzog, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

There are things known and there are things unknown, and in between are the doors of
perception.
— Aldous Huxley



Acknowledgements

First and foremost, I would like to thank my topic of study. Behavioral scientists face a peculiar dilemma in their empiricism; we feign objectivity while studying ourselves. Because uncertainty was my thesis topic, this PhD afforded me a learning experience that went beyond the expected, which is to say intellectual, technical and scientific development. Studying uncertainty also sent me on a night-sea journey. I believe this quality of my thesis makes me the very luckiest of doctoral students and I hope I have honored the subject adequately.

My deepest gratitude and respect go to my supervisor, Prof. Kerstin Preuschoff, who showed me a whole new way with which to examine human behavior, through computation and from whom I learned so much, more than I imagined I could. Thank you for your guidance, inspiration and trust the past 4 years; I am lucky to have had you as supervisor. My deepest thanks also goes out to my co-supervisor, Prof. Carmen Sandi. Thank you for recognizing me and taking me on board, as well as introducing me to Kerstin. I am grateful too for Prof. Bogdan Draganski, Dr. Ferath Kherif and Dr. Antoine Lutti of the LREN, for their warmth and welcome; if supervisors are parents, you were my fun uncles. Thank you for giving me a berth at the LREN. Being from a patchwork family of labs gave me the privilege of working with dozens of amazing colleagues across the Lemanic region, but I need to thank a few in particular: Estelle Dupuis, who never shied away from helping me scan; Maya Jasztrebowska, who always listened intently to my data woes; Ulrike Rimmele, Elsa Juan, Laia Morato-Fornaguera and Meltem Weger. Thank you for being there.

I must also acknowledge people who, perhaps unbeknownst to them, greatly shaped my scientific view and thinking. Prof. Zalman Amit, whose laboratory was the first I worked in, at Concordia University as well as Prof. Peter Brügger, of the Universitätsspital in Zuerich. I also want to thank my late friend, Michael Hauben, that cracked bell, whose stunning intellect was no match for his deep humanity. I wish you were here to see how prescient you were. You are greatly missed.

I also need to thank those nearest to me. First, my father, Slaheddine Khenissi, for his scathing intelligence and setting the bar high; and my mother, Lotfi Khenissi, for her love and blind faith in my abilities. My husband, Dr. Peter D. Jarowski, who made the ultimate sacrifice as an academic and left his faculty position in Surrey, England so that I may pursue my doctoral studies. And to my very, very dearest, my children, Shams, Aya and Sami: thank you for accepting "*But, I'm looking for the surprise signal in the brain!*" as an excuse these past four years. You three are the light that stays my course. *Lausanne, February 15, 2018* L.K.

Author Preface

In the mid 19th century, Hermann von Helmholtz formulated the theory of unconscious inference (Hatfield, 2002) to detail how the mind spontaneously resolves uncertainty so as to adequately reconstruct the world around it. Helmholtz's call to the mind's role in the process is almost apologetic in his treatise (Helmholtz /Southall, 1925); understandably so, as, at the time, the mind (brain) was unobservable and thus arguments on its involvement could not be proven. Helmholtz further posited that unconscious inference was not limited to visual perception but could explain all forms of sense-perception and also writes that perception is the result of a form of practice, or learning, to integrate features of a stimulus into a percept, an idea that harkens to a Bayesian learning process. Helmholtz's treatise made use of binocular rivalry examples to underline his theories.

Nearly two centuries after Helmholtz's treatise, we examine this possibility armed with two powerful tools: formal accounts of uncertainty that offer a precise means by which to capture inferential processes; and fMRI, with which we can literally look into the mind. By applying a common computational account of both perceptual and financial uncertainty during human fMRI, we aim to answer the longstanding question put forth by Helmholtz and that is whether the same neural system is responsible for inferential processes regardless of whether a decision is deliberate, as in the case of a cognitive judgment, or (seemingly) spontaneous, as in visual perception.

The following work explores several questions pertaining to decision-making under uncertainty. First, guided by previous pupillometry studies using predictive coding accounts of expected uncertainty (risk) and its error (surprise), we seek to find noradrenergic involvement in uncertainty processing, as assessed by locus coeruleus activation using fMRI. Second, we seek to investigate specific uncertainty-related variables, namely confidence, surprise and information, in a financial uncertainty task, with the aim of mapping said variables' formal definitions to specific neural representations. Third, we aim to generalize previous findings on the neural representation of uncertainty, specifically of surprise, by searching for common patterns of activation in two distinct tasks, one cognitive and objective, the other perceptual and subjective, in the same individuals. The first question presents a significant technical challenge as evidence of noradrenergic activity relies on detecting activity in a very small brainstem structure, the locus coeruleus. Nonetheless, the impact of finding evidence in favor of noradrenergic modulation of decision-making under uncertainty cannot be understated, as this link can guide our understanding of various questions in neuroeconomics, such as the impact of stress on risk assessment and taking, as well as inform our definitions of impaired

Preface

decision-making in several psychiatric and neurological diseases. The second question can better inform the use of different computational frameworks to assess various forms of uncertainty and their related variables, with an aim to both advance our understanding of human behavior but also with an eye towards applying this knowledge towards the advancement of artificial intelligence. The third question aims to provide evidence of a more philosophical debate, which harkens back to the Helmholtzian view: if a common neural system exists to process uncertainty, we can more confidently believe in the assumption that the brain is an inference machine and cast further neuroscientific questions within that simple and elegant definition of what, to date, appears an impossibly complex system to define. *Lausanne, 15*

Février 2018

L.K.

Abstract

From the moment we wake up in the morning to the day's ebb when we settle in to sleep, we are bound to the task of decision-making. Some of these decisions barely register in our consciousness, if at all, while others, less shy, take a more prominent place at our mind's table. Regardless of the importance or difficulty of the decision, few, if any, are made with perfect information: being such a small part of a large system, we can only know so much. Further, the system itself sends us noisy information for us to encode and decode as best we can. How do we do this? How do we continuously and, for the most part successfully, resolve uncertainty in order to survive and even flourish? We propose to define uncertainty in decision-making as a computational process, in line with information-processing theories of neural mechanisms. To that end, we investigate the neural correlates of uncertainty processing using functional magnetic resonance imaging (fMRI) in humans within a predictive coding framework. The field has already produced considerable evidence showing that decisions are made with the aim of maximizing utility, a process involving the dopaminergic reward system. We turn our focus to the uncertainty surrounding predictions and their concomitant errors by conducting a two-part fMRI experiment on 23 subjects. In the first session, we elicited objective, cognitive (financial) uncertainty in a gambling task. In the second session, we exposed individuals to subjective, perceptual uncertainty, in the form of visual illusions. Our fMRI results, modeled by computational definitions of surprise, confidence and information, show that 1) the brain employs computational principles to resolve uncertainty; 2) certain regions are consistently implicated in processing said uncertainty, notably insular cortex regions, across modalities (cognitive and perceptual), be it of a subjective or objective nature. These findings support the notion that the brain is an active inference machine, a paradigm within which further aspects of cognition can be investigated.

Keywords: decision-making; model-based fMRI; uncertainty; surprise; confidence; information; insula; locus coeruleus; computational psychiatry; predictive coding

Résumé

La prise de décision est une tâche que nous sommes contraints d'effectuer en continu durant nos heures d'éveil. Certaines décisions ne demandent pas trop de réflexion tandis que d'autres nous angoissent. Indépendamment de l'importance ou de la difficulté de la décision, rares sont celles faites avec une information parfaite : comme chaque individu fait partie d'un grand système, nous sommes limités dans notre capacité d'avoir une vision complète des faits. De plus, le système lui-même nous envoie des informations bruitées. Malgré cet incertain intrinsèque de notre environnement, nous réussissons à poursuivre notre chemin de façon adéquate, voire optimale. Comment parvenons-nous à continuellement résoudre l'incertain pour survivre et même prospérer ? Nous proposons de définir l'incertain dans la prise de décision comme étant un processus computationnel, en ligne avec les théories probabilistes concernant les processus cérébraux. Nous exploitons de surcroît les théories venant des sciences économiques, psychologiques, informatiques ainsi pour caractériser la surprise ; la confiance ; et l'information, en relation avec l'incertain. Afin de répondre à nos questions dans le contexte de comportement chez l'être humain, nous étudions la représentation cérébrale du traitement de l'incertain à l'aide d'imagerie en résonance magnétique fonctionnelle (IRMf), dans un cadre de codage prédictif. Des études précédentes ont déjà fourni des preuves démontrant que la prise de décisions est guidée par le but de maximiser l'utilité, un processus impliquant le système de récompense dopaminergique. Nous focalisons sur l'incertain engendré par nos prédictions, ainsi que leurs erreurs, en réalisant une expérience IRMf en deux parties sur 23 sujets. Lors de la première session, nous avons suscité l'incertain objectif (cognitif) dans une tâche de jeu. Dans la deuxième session, nous avons provoqué l'incertain perceptuel et subjectif en exposant les mêmes sujets à une illusion d'optique. Nos résultats d'IRMf montrent que 1) le cerveau utilise des principes computationnels pour résoudre l'incertain ; 2) certaines régions sont particulièrement impliquées dans le traitement de l'incertain, notamment le cortex insulaire. Ces résultats soutiennent l'idée que le cerveau est une machine à inférence statistique, un paradigme dans lequel d'autres aspects de la cognition peuvent être étudiés. Mots-clés : prise de décision ; IRMf ; incertain ; risque ; surprise ; confiance ; information ; insula ; locus coeruleus ; psychiatrie computationnelle ; théorie de l'information ; codage prédictif

Contents

| | |
|---|------------|
| Acknowledgements | v |
| Author Preface | vii |
| Abstract | ix |
| List of figures | xvi |
| List of tables | xix |
| Decision-Making Under Uncertainty – an Overview | 1 |
| 1 Decision-Making Under Uncertainty – an Overview | 1 |
| 1.1 Introduction | 1 |
| 1.2 The rational decision-maker | 1 |
| 1.2.1 Reward | 2 |
| 1.2.2 Reinforcement Learning | 3 |
| 1.2.3 Homeostasis | 3 |
| 1.3 Uncertainty | 4 |
| 1.3.1 Decision-making under uncertainty | 4 |
| 1.3.2 The human problem | 4 |
| 1.3.3 The machine problem | 5 |
| 1.4 Various Frameworks for the Investigation of Uncertainty | 5 |
| 1.4.1 Information Theory | 6 |
| 1.4.2 Expected Utility And Mean-Variance | 6 |
| 1.4.3 Free Energy | 7 |
| 1.4.4 Predictive Coding | 8 |
| 1.4.5 Bayesian Models of Uncertainty | 8 |
| 1.5 The different forms of uncertainty | 9 |
| 1.5.1 Ambiguity | 9 |
| 1.5.2 Risk | 10 |
| 1.5.3 Volatility | 11 |
| 1.5.4 Surprise | 12 |
| 1.5.5 Information (Entropy) | 16 |
| 1.5.6 Confidence | 17 |
| | xiii |

Contents

| | | |
|----------|--|-----------|
| 1.6 | The role of neuromodulators in the encoding of decision variables | 18 |
| 1.6.1 | Dopamine | 18 |
| 1.6.2 | Noradrenaline | 19 |
| 1.7 | Brain Regions Associated with the Neural Representation of Uncertainty | 20 |
| 1.7.1 | The Anterior Cingulate Cortex | 20 |
| 1.7.2 | Insula | 22 |
| 1.7.3 | Striatum | 23 |
| 1.7.4 | The Locus Coeruleus | 23 |
| 1.8 | Patient Populations | 24 |
| 1.9 | Notes on Model-Based Approaches | 26 |
| 1.10 | Conclusion | 26 |
| 2 | In Search of the LC: The needle in the haystack | 27 |
| 2.1 | Introduction | 27 |
| 2.1.1 | Does noradrenergic activity encode uncertainty processing? | 28 |
| 2.1.2 | Pupil dilation as an index of LC activity | 28 |
| 2.1.3 | Imaging the LC | 29 |
| 2.2 | Structural localization of the locus coeruleus | 30 |
| 2.3 | Pilot Study 1: Indirect LC Localization via Opto-Kinetic Nystagmus | 30 |
| 2.3.1 | OKN Experimental Procedure | 31 |
| 2.3.2 | OKN Imaging Results | 32 |
| 2.3.3 | OKN imaging Data for horizontal and vertical OKN, Group Level | 32 |
| 2.3.4 | Pilot Study 1 - Summary | 34 |
| 2.4 | Pilot Study 2: Pupil Dilation in The auditory oddball paradigm | 36 |
| 2.4.1 | Pitch versus duration deviance | 36 |
| 2.4.2 | Experimental procedure | 36 |
| 2.4.3 | Pupil Trace Preprocessing | 37 |
| 2.4.4 | Auditory Oddball Pupillometry Study: Results and Conclusion | 37 |
| 2.5 | Pilot Study 3: fMRI Sequence Comparison Study | 37 |
| 2.5.1 | Sequence Comparison Study: Experimental Procedure | 39 |
| 2.5.2 | Sequence Comparison Study: Group Level Results | 41 |
| 2.5.3 | In-scanner Pupillometry Results | 43 |
| 2.5.4 | Sequence Comparison Study: Conclusion | 44 |
| 2.6 | Summary of results | 45 |
| 3 | Of Outcomes and Expectations: the Neural Representation of Confidence, Surprise and Information | 49 |
| 3.1 | Introduction | 49 |
| 3.2 | Materials and Methods | 52 |
| 3.2.1 | Participants | 52 |
| 3.2.2 | Procedure and Task Description | 52 |
| 3.2.3 | Imaging Procedure | 54 |
| 3.2.4 | Behavioral Analysis | 54 |

| | | |
|----------|--|------------|
| 3.2.5 | Imaging Analysis | 55 |
| 3.2.6 | The Models | 57 |
| 3.3 | Neuroimaging Results | 60 |
| 3.3.1 | Confirmatory Analyses | 60 |
| 3.3.2 | Exploratory Analyses | 62 |
| 3.4 | Discussion | 71 |
| 3.4.1 | Confidence | 71 |
| 3.4.2 | Surprise | 72 |
| 3.4.3 | Surprise as learning signal | 73 |
| 3.4.4 | Confidence and Information as Precision and Accuracy | 74 |
| 4 | Are Perceptual Switches Surprising? | 75 |
| 4.1 | Introduction | 75 |
| 4.1.1 | Inferential nature of Perceptual decision-making | 75 |
| 4.1.2 | Visual Illusions: Perceptual Uncertainty Manifest | 76 |
| 4.1.3 | Neural correlates of bistable perception | 78 |
| 4.1.4 | When does the reversal occur? | 80 |
| 4.1.5 | Other Modalities | 81 |
| 4.1.6 | Conscious or Unconscious? | 81 |
| 4.2 | Perceptual Uncertainty Experiment | 82 |
| 4.2.1 | Procedure and Task Description | 83 |
| 4.3 | Necker Cube Behavioral Results | 86 |
| 4.3.1 | Behavioral Results: Necker Cube Task 1 | 86 |
| 4.3.2 | Behavioral Results, Necker Cube Task 2 | 92 |
| 4.4 | Neuroimaging Results - Perceptual Uncertainty | 101 |
| 4.4.1 | Imaging Analysis | 101 |
| 4.4.2 | General Linear Models | 101 |
| 4.4.3 | fMRI Results: Necker Cube Task 1 | 102 |
| 4.4.4 | fMRI Results: Necker Cube Task 2 | 104 |
| 4.5 | Discussion | 111 |
| 5 | A Common Neural Representation of Uncertainty in the Insula | 117 |
| 5.1 | Introduction | 117 |
| 5.2 | Categorical Univariate Analyses | 118 |
| 5.2.1 | Financial Uncertainty : Categorical Analyses | 118 |
| 5.2.2 | Perceptual Uncertainty: Categorical Analyses | 122 |
| 5.3 | Multivariate Pattern Analyses - Preliminary Analyses | 125 |
| 5.3.1 | Financial Uncertainty: Surprise (Information) Classification | 127 |
| 5.3.2 | Perceptual Uncertainty: Surprise Classification | 128 |
| 5.4 | Parameter weights correlations | 128 |
| 5.5 | Discussion | 129 |
| 5.6 | Conclusion | 135 |

Contents

| | |
|----------------------------------|------------|
| 6 Conclusion | 137 |
| 6.1 | 137 |
| 6.2 Questions Posed | 137 |
| 6.3 Impact of Findings | 138 |
| Bibliography | 141 |
| Appendix A | 169 |
| Appendix B | 181 |
| Appendix C | 193 |
| Appendix D | 201 |
| Curriculum Vitae | 209 |

List of Figures

| | | |
|------|-----------------------|----|
| 2.1 | Figure 2.1 | 31 |
| 2.2 | Figure 2.2 | 33 |
| 2.3 | Figure 2.3 | 34 |
| 2.4 | Figure 2.4 | 35 |
| 2.5 | Figure 2.5 | 38 |
| 2.6 | Figure 2.6 | 39 |
| 2.7 | Figure 2.7 | 41 |
| 2.8 | Figure 2.8 | 42 |
| 2.9 | Figure 2.9 | 43 |
| 2.10 | Figure 2.10 | 44 |
| 2.11 | Figure 2.11 | 45 |
| 2.12 | Figure 2.12 | 46 |
| 3.1 | Figure 3.1 | 53 |
| 3.2 | Figure 3.2 | 53 |
| 3.3 | Figure 3.3 | 57 |
| 3.4 | Figure 3.4 | 59 |
| 3.5 | Figure 3.5 | 60 |
| 3.6 | Figure 3.6 | 61 |
| 3.7 | Figure 3.7 | 62 |
| 3.8 | Figure 3.8 | 63 |
| 3.9 | Figure 3.9 | 64 |
| 3.10 | Figure 3.10 | 64 |
| 3.11 | Figure 3.11 | 65 |
| 3.12 | Figure 3.12 | 66 |
| 3.13 | Figure 3.13 | 67 |
| 3.14 | Figure 3.14 | 67 |
| 3.15 | Figure 3.15 | 68 |
| 3.16 | Figure 3.16 | 69 |
| 3.17 | Figure 3.17 | 70 |
| 3.18 | Figure 3.18 | 70 |
| 3.19 | Figure 3.19 | 71 |

List of Figures

| | | |
|------|-----------------------|-----|
| 4.1 | Figure 4.1 | 85 |
| 4.2 | Figure 4.2 | 86 |
| 4.3 | Figure 4.3 | 87 |
| 4.4 | Figure 4.4 | 88 |
| 4.5 | Figure 4.5 | 89 |
| 4.6 | Figure 4.6 | 90 |
| 4.7 | Figure 4.7 | 91 |
| 4.8 | Figure 4.8 | 92 |
| 4.9 | Figure 4.9 | 93 |
| 4.10 | Figure 4.10 | 94 |
| 4.11 | Figure 4.11 | 95 |
| 4.12 | Figure 4.12 | 96 |
| 4.13 | Figure 4.13 | 97 |
| 4.14 | Figure 4.14 | 99 |
| 4.15 | Figure 4.15 | 100 |
| 4.16 | Figure 4.16 | 101 |
| 4.17 | Figure 4.17 | 102 |
| 4.18 | Figure 4.18 | 103 |
| 4.19 | Figure 4.19 | 103 |
| 4.20 | Figure 4.20 | 105 |
| 4.21 | Figure 4.21 | 106 |
| 4.22 | Figure 4.22 | 106 |
| 4.23 | Figure 4.23 | 107 |
| 4.24 | Figure 4.24 | 108 |
| 4.25 | Figure 4.25 | 109 |
| 4.26 | Figure 4.26 | 110 |
| 4.27 | Figure 4.27 | 110 |
| 4.28 | Figure 4.28 | 111 |
| 5.1 | Figure 5.1 | 119 |
| 5.2 | Figure 5.2 | 119 |
| 5.3 | Figure 5.3 | 120 |
| 5.4 | Figure 5.4 | 121 |
| 5.5 | Figure 5.5 | 122 |
| 5.6 | Figure 5.6 | 123 |
| 5.7 | Figure 5.7 | 124 |
| 5.8 | Figure 5.8 | 125 |
| 5.9 | Figure 5.9 | 126 |
| 5.10 | Figure 5.10 | 128 |
| 5.11 | Figure 5.11 | 129 |
| 5.12 | Figure 5.12 | 130 |
| 5.13 | Figure 5.13 | 135 |

List of Tables

| | | |
|-----|---|-----|
| 2.1 | Oddball Task Parameters | 37 |
| 3.1 | Decision Variables Across Sessions | 55 |
| 3.2 | ConfidenceBrainstemCluster | 65 |
| 3.3 | Surprise Brainstem Cluster | 68 |
| 4.1 | Correlation values for individual dwell time to stimulus bias ratios | 98 |
| 4.2 | MLE for Gamma Fit Shape and Scale Parameters, all stimuli classes | 98 |
| 5.1 | Main Effect of Risk in 7 ROIs | 120 |
| 5.2 | Main Effect of Surprise in 7 ROIs | 121 |
| 5.3 | Main Effect of Perceptual Ambiguity in 7 ROIs | 124 |
| 5.4 | Main Effect of Perceptual Bias in 7 ROIs | 126 |
| 5.5 | Table of p-values for Correlation of Surprise Related Beta Weights Across Tasks | 130 |

1 Decision-Making Under Uncertainty – an Overview

1.1 Introduction

Humans and animals engage in decision-making continuously during their waking hours. Decisions can come in myriad forms across sensory and cognitive domains and can be viewed in terms such as important or trivial, snap or deliberative. Regardless of these various features, the decision process has a common framework, wherein an agent is confronted with more than one possibility at one time and, following an event, usually an action, obtains an outcome. In the following chapter, we detail the process and lay the groundwork of our experimental questions concerning decision-making under uncertainty. We first describe the commonly held view of reward-guided decision-making before introducing the problem of uncertainty. We then detail the different conceptual prisms through which uncertainty is examined, before defining several different forms of uncertainty that are pertinent to our own research framework. We end with a short review of brain regions we believe are especially relevant to uncertainty processing in the context of our research study.

1.2 The rational decision-maker

The rational decision-maker, or *homo economicus*, is an agent that always acts to maximize his gain and minimize his costs (von Neumann & Morgenstern, 1945). Deviations from this goal are dubbed irrational yet individuals show a wide variety of irrational biases in their decision-making (Kahneman, 2014). Rather than framing these common behaviors as irrational ones, it behooves us to further clarify how rewards and costs are viewed. After all, humans and other agents are generally more complex than mere reward-maximizing machines. For instance, human decision-making is inevitably subject to affective modulation (Smith, 1759). In more recent times, neuroscientific investigations into decision-making using neuroimaging generated the somatic-marker hypothesis, whereby emotionally salient signals mediated decision processes (Damasio et al., 1994).

As mentioned, rational choice theory is challenged by observed human decision-making,

meaning that individuals do not always act to maximize their interest. For instance, Kahneman and Tversky (1979) famously set out to elucidate on decision-making under uncertainty. Uncertain options appear to incur a penalty relative to sure options, even if the former have a higher average return, dubbed expected utility. Kahneman and Tversky formulated Prospect Theory to explain this ?irrationality? and in sum suggest that humans are loss and risk-averse (Kahneman & Tversky, 1979). Another known bias is found in the framing effect, whereby the same option offered in different contexts will prompt differing choices, again often violating axioms of expected utility. A simple gambling experiment framing sure versus uncertain options in terms of either wins (a positive frame) or a loss (negative frame) found subjects predictably shied away from options framed as a loss.. It has been suggested that individuals may not rely on computations when making decisions under uncertainty or in complex environments, but may resolve such quandaries by relying on affective processes or heuristics (Gilovich et al., 2002). Imaging data confirmed an amygdalar role in the framing effect bias, suggesting limbic system involvement in decision-making; further, subject-specific rationality correlated with increased activity in the orbitofrontal cortex (de Martino et al., 2006). an area associated with the valuation process (Padoa-Schioppa & Assad, 2006). While affect-driven deviations from rational decision-making have been observed, it is hypothesized that irrationality can be explained in other ways too. Current research applies computational accounts of various decision sub-processes (O'Dogherty et al., 2007), with an eye towards capturing hidden variables beyond mere reward maximization to explain observed biases in choice behavior. The implication of this line of research is that humans are in fact rational ? only our limited knowledge of latent variables and hidden processes makes behavior appear irrational.

1.2.1 Reward

The most parsimonious manner by which to describe decision-making is to frame it as a means to maximize fitness, which in turn leaves us with the conclusion that a drive to maximize rewards and concomitantly minimize costs guides decision-making. This goal is taken as a given in economics, where research is performed under the assumption of homo economicus, or rational man, who acts only to increase his benefit (von Neumann & Morgenstern, 1945). In psychology, this drive has been found experimentally, first in Pavlovian conditioning, then operant condition and formalized by the Rescorla-Wagner model (Lieberman, 2000) <insert equation here>. Conditioning is a form of learning and one can see that decision-making can also be viewed as a learning process. If a given decision leads to reward, it is more likely to be repeated, but if it leads to a cost, it is more likely to be eschewed in the future. While economists understood reward-guided decision-making to be obvious, psychological research on the phenomenon provides empirical evidence supporting the assumption.

1.2.2 Reinforcement Learning

Given the extensive research on conditioning in humans and animals, it was not long before the information science community turned to the Rescorla-Wagner model with the aim of simulating the process so as to teach a machine to learn (Sutton & Barto, 1988). Sutton and Barto borrowed from the Rescorla-Wagner model amongst other sources to contribute to the array of algorithms dubbed Reinforcement Learning algorithms. The reinforcement-learning problem was outlined in the context of artificial intelligence research. Sutton and Barto's aim was to characterize a learning process that better approximates a 'real' learner than do other machine learning algorithms that generally fall into supervised and unsupervised learning categories. Sutton and Barto point out that the former is limited (the agent needs to be told what is correct) and the latter doesn't necessarily seek a goal, or reward, but a pattern. Reinforcement learning starts with an agent that is forced to interact with her environment and in so doing learns to reach her goals. Crucially, this agent can integrate uncertainty into the process and can therefore act in new contexts without additional, explicit instruction. Amongst the different reinforcement learning algorithms developed, temporal difference learning is one whose computational account has been most readily applied to animal behavior (Houk et al., 1995; Schultz, 1998). In temporal difference learning, an agent learns the value of an action depending on a subsequent event's temporal distance relative to the action. Importantly, the agent makes a prediction on a future event that is iteratively adjusted by the error between the actual event and its prediction. While psychological research in animals inspired reinforcement learning, it has developed independently; fittingly, we can now borrow from its developments to better refine models of human learning and decision-making processes, creating a highly beneficial loop between the two fields.

1.2.3 Homeostasis

Maximizing reward or utility is a clear means by which an agent can ensure survival. The goal of decision-making then, regardless of the type of decision, can be seen as a way to maintain homeostasis (Korn & Bach, 2015), a specific physiological state required for an agent to survive. Given the ubiquity of 'irrational' decision-making (Taleb, 2001; Kahneman, 2014), we must question whether it is reward alone that acts to maintain homeostasis. Indeed, it has been suggested that minimizing uncertainty may be a critical factor for an agent (Friston, 2010). This introduces the exploration-exploitation dilemma, where an individual must consider the implications of pursuing a known rewarding policy, and in so doing disregard risky options, when other unknown actions may prove to be even more rewarding. This dilemma is therefore encapsulated by two competing drives, 'staying safe' or 'maximizing reward', though these strategies nonetheless share the ultimate goal of maintaining homeostasis.

1.3 Uncertainty

Uncertainty is an inevitable feature of our environment that we are confronted with continuously because an individual only has access to partial information and available information is noisy. This unavoidable limitation is essential to our understanding of how we perceive the world: in interacting with the environment, we can only approximate its reality (Gregory, 1997). Research on decision-making has emphasized reward-guided behavior, however, the implications of the predictive nature of such behavior is that it is inherently probabilistic, and thus by extension, includes uncertainty. Research on decision-making under uncertainty has been implicated in the field of economics (Kahneman & Tversky, 1979) as well as in other fields. The advent of neuroeconomics now allows us to scrutinize its neural dimension. Below we review the problem of uncertainty in more detail.

1.3.1 Decision-making under uncertainty

We have seen that the decision-making process requires quantities related to reward maximization. Specifically, an agent needs to: predict a reward when selecting an action; assign a value to the outcome; and must distinguish between the two. Given that predictions are by their nature probabilistic, we can also deduce that other variables are at play in the process, namely the uncertainty surrounding an agent's estimation process. Different forms of quantitative uncertainty exist, each of which may 1) have a specific impact on behavior; 2) possess a distinct neural correlate; 3) may in principle be impaired.

1.3.2 The human problem

So far we have hypothesized that human behavior is driven by reward maximization specifically but more generally as a means to maintain homeostasis. However, an agent does not have a complete map of the environment and therefore must estimate the value of actions leading to desired outcomes. This estimate, being probabilistic, must entail an integration of uncertainty. How does an individual compute and integrate uncertainty? At what level of the process does uncertainty manifest? Where in the brain are these quantities computed? What type of scenario demands uncertainty considerations? Are these computations labile in time or in the face of disease? Do these processes differ within and across individuals? Finally, are they encoded by a specific neurotransmitter? Statistical principles and quantities give us a means by which we can test the process of uncertainty integration. We must assume that, given our species' success, humans do not simply rely on guesswork (Parr & Friston, 2017).

How then does a human make approximations on which course is best? It is evident that a learning process must be at play. Yet, in introducing the concept of uncertainty integration into the decision-making processes, we are confronted with a slew of potential considerations. Does an agent merely engage in trial-and-error decisions until an optimum is reached? Such a model-free approach may be straightforward in its implementation, but is also costly, time-

consuming, and ultimately rigid, an unsuitable feature in a dynamic environment. Does an individual then generate a map of his environment, which allows for simple changes in one element of the model over another? Such model-based learning may be more advantageous for a human, but is also subject to a significant initial cost (Glaescher et al., 2010). What can be seen from such approaches is that they are not free of uncertainty. Errors can be minimized but can never converge to 0. Ergo, uncertainty remains to be accounted for.

How does the brain implement these processes? What regions are involved in encoding different facets of uncertainty when faced with a decision? Crucially, how is decision-making affected when one of these brain regions is impacted by disease? Latent or hidden variables are factors that we cannot observe and uncertainty fits such a specification well. However, we can infer their existence, dynamics and effects via observed variables (Vilares & Kording 2011). In detailing the role of different uncertainty variables in the brain, we can better inform other fields in which they may be of use, such as in neurofinance, neuroeconomics and machine learning. At a more immediate and compelling level, this work will help us identify specific failures in the cognitive domain, and in so doing offer precise therapeutic approaches for decision-making impairments.

1.3.3 The machine problem

Uncertainty is not only a feature of the human experience. Current research and development on artificial intelligence and pattern recognition also encounters the quandary, namely how to model an intelligent agent or machine that can confront and address uncertainty. The question of uncertainty is glaringly obvious in the machine learning field because the field itself is founded on statistics and probability, whereas the idea that humans act according to such principles is not necessarily a given. An example of an artificial model developed to emulate human intelligence is the belief, desire and intention (BDI) model that is endowed with intent and belief; where belief resides, so must uncertainty. While formal accounts of reinforcement learning can be applied to human research, they are formulated in the context of artificial intelligence. Questions on how such systems confront uncertainty form an active line of research (Blum & Langley, 1997; Lorini, 2006; Proedrou, 2002). How best to capture and quantify different forms of uncertainty can thus lead to more efficient machines. In the case of surprise, for instance, if the phenomenon is indeed a dynamic experience that may lead to learning, its reflection in an intelligent agent would be of great value (Demey, 2015).

1.4 Various Frameworks for the Investigation of Uncertainty

Because uncertainty presents a problem in a wide variety of contexts, several frameworks from different fields have been put forth to capture and define associated quantities. All the following frameworks assume probabilistic models and thus include uncertainty and its related variables (Feldman & Friston, 2010). Key in this thesis is the view that the brain is an inference machine, an old idea whose conception is found in the 11th century scientist Alhazen's treatise on optics.

tise on visual perception, and further developed in Hermann von Helmholtz' work (Gregory, 1973) and is now, with the technological advances afforded us by neuroscience, actively being investigated by several researchers. The brain as inference machine hypothesis highlights the following: as the organ receives various stimuli, it must estimate what the latter predicts. In essence, the brain's job is to formulate hypotheses about its environment and to test these hypotheses against the evidence (Gregory, 1997; Friston, 2012). Researchers in the neuroscience of decision-making specifically have employed several different frameworks within which to study this notion. While these frameworks are in some studies compared to one another, they are neither mutually exclusive nor canonical. Below we focus our non-exhaustive review on frameworks that are especially relevant to our research, namely, information theory; the mean-variance theorem; Bayesian frameworks; signal detection theory; predictive coding; and the free energy principle.

1.4.1 Information Theory

In the communication sciences, uncertainty is described as the problem inherent to a system that seeks to reconstruct a signal after its reception. That is, how is uncertainty integrated when a system encodes and then decodes a signal? In 1948 Claude Shannon (Shannon, 1948) described the problem in a signal-processing context and in so doing birthed the field of information theory. When a system receives a signal, or information, it must then interpret the signal as accurately as possible given its inherent noise, or uncertainty. This uncertainty is dubbed the entropy of the system. Importantly, information comes with a probability distribution. The more improbable the signal, the more surprising it is; the negative log of its probability captures this quality of the signal. The information theoretic notion of entropy is a more slippery concept to map to human experience. Entropy represents the amount of information needed to resolve an uncertainty and is, in Shannon's original paper, represented by bits. While information theoretic formulations of uncertainty were not, in their initial conception, meant to characterize the human experience, they have been recently applied to studies on uncertainty in humans (Mars et al., 2005; Strange et al., 2008; O'Reilly et al., 2013). Indeed, information theory is framed in elegant formulations, compelling a researcher to consider its application in the study of human behavior.

1.4.2 Expected Utility And Mean-Variance

In economics, two contrasting theories on utility maximization have been examined in the context of pure expected utility (the probability of an outcome p times its value at outcome, U) relative to a mean-variance framework, whereby an agent is guided towards a utility maximization simply by knowing the mean and variance of a distribution (Kroll et al., 1984). Thus in using a mean-variance analysis, an agent accounts for risk (as captured by the variance) in addition to her expected utility, and seeks to either minimize this risk for a desired return or maximize her return for a given level of risk. Mean-variance theory was cast in the framework of financial decision-making, with the general conclusion that investors should diversify their

portfolios with the aim of minimizing risk (Markowitz, 1954). While this course may seem obvious in modern-times, the concept is not as intuitive as straight utility maximization (von Neumann & Morgenstern, 1945) and this may indeed reflect risk's second-order mathematical definition. How does the question of mean-variance impact neural processes of decision-making? Learning through expected utility can be a lengthy process, because an option's expected value doesn't necessarily reflect the range of options. However, mean-variance theory would facilitate learning, as it does not entail repeated trial-and-error (D'Acremont & Bossaerts, 2008). Importantly, behaving according to the mean-variance theorem would allow us to walk around making decisions with only two values, namely the mean and the variance of options, leaving us neurally unencumbered. If humans deploy a mean-variance analysis in computing option values, then we would expect distinct correlates for expected utility and variance, which have indeed been found in the brain (Preuschoff et al., 2008; Tobler et al., 2007). These studies thus applied a mean-variance framework in the context of neuroimaging studies on human decision-making.

1.4.3 Free Energy

The free-energy principle posits that a system seeking equilibrium must minimize its free energy, a quantity characterized by both a surprise measure and entropy. Because a biological system needs to maintain homeostasis, it follows that the probability of being in a given state must have low entropy, meaning a small number of possible states. As seen above, entropy is an information-theoretic quantity formalized by the negative log of a probability of the event. According to this principle, a biological system would seek to minimize this average surprise in order to maintain homeostasis. In its formal account, free energy imposes a bound on surprise; thus there is an upper limit to the number of states entropy provides. In a predictive coding context, minimizing free energy consists of minimizing the prediction error and in a Bayesian brain consists of updating prior beliefs. Crucially, in both cases, the system seeks to converge on an optimal prediction for a given cue. The Free Energy Principle is a valuable framework in that it integrates several models that share theoretical elements, and this at times from different fields, to explain the brain. Especially pertinent models include predictive coding, information theoretic and Bayesian models (Friston, 2010).

A cornerstone of the Free Energy Principle is that surprise minimization drives behavior, because the brain seeks to reduce its states to a set number and thus conserve energy. The Free Energy Principle has been proposed as a unifying brain theory. While this thesis does not claim to do as much, Free Energy is a compelling corollary to our investigative questions, because it takes into account questions on surprise, belief and information. It must be noted however that the premise Free Energy rests on, that a system seeks to minimize surprise so as to maintain homeostasis, may be challenged when contrasted with the exploration-exploitation dilemma.

An example used to illustrate the drive to minimize surprise is in considering a fish in water (low

surprise, homeostasis) versus a fish out of water (high surprise, deviating from homeostasis) (Friston, 2010). This example illustrates the necessity of minimizing surprise and free energy. At first glance, one summarily concludes that the fish out of water lives an undesirable outcome. However, when one considers that a fish out of water that survives the outcome may well eventually evolve towards a human state, one can reconsider the premise of staying in the water as a uniquely desirable state. In more general terms, free energy minimization represents the drive to minimize error so as to maintain homeostasis. However, there may be a competing drive to allow error above and beyond its minimum, so as to explore, or invite surprise. Evolution itself is driven by error, albeit stochastically. Some individuals, on the other hand, actively and consciously seek error, and aim for surprise, presumably with the intent to expand their repertoire of possible states, or increase their entropy. While the exploration-exploitation dilemma is a topic of investigation in the decision sciences, to our knowledge it has not been addressed in the framework of the free energy principle.

1.4.4 Predictive Coding

The predictive coding framework is a model formulated by Rao & Ballard in the context of visual information processing. In brief, higher level cortical areas make predictions on stimuli, sending these messages to lower-level visual areas, which in turn send messages back up, in the form of prediction errors to indicate a mismatch between a prediction and the actual percept. The premise lies in the correlation of a quantized stimulus, such as a pixel, with those nearby. In such a case, adjacent pixels share features and thus an approximation of many such stimuli can be made, rather than have costly and redundant mapping of one neuron say, per pixel. At several layers of a hierarchical model network, predictions are made and then tested by lower level layers, down to the lowest (sensory) levels, which capture the initial image (Rao & Ballard, 1999). Predictive coding makes use then of the framework where expectations are predictions, and outcomes are prediction errors, the latter which update and tune subsequent predictions. The concept is simple yet powerful enough to account for neural processes. Notably, predictive coding allows for the examination of uncertainty in a single iteration of a decision (prediction \rightarrow outcome \rightarrow error) and thus provides a parsimonious means by which to examine uncertainty in an experimental setting.

1.4.5 Bayesian Models of Uncertainty

Normative models of the brain and behavior first 1) hypothesize on the brain's role; 2) build models that would achieve that role and 3) test these models against observed behavior. In a sense, this manner of examining brain and behavior takes a reverse engineering approach. Thus in this view, the brain's role is that of an inference machine, which approximates the truth of the world by estimating its descriptive features, or parameters. A compelling framework within which to cast this model of the brain is in Bayesian models (Knill & Pouget, 2004; Beck et al., 2008; Ma & Jazayeri, 2014). The brain makes a prediction on the environment given available evidence and updates this prediction as new evidence is obtained. In its canonical

form, the Bayesian model encapsulates the problem of decision-making under uncertainty though not all problems fit the model. These theories also imply that human behavior is Bayes optimal. People adjust their inference based on evidence before them. Not only are we acutely aware that this optimality doesn't hold for individuals, but there is significant evidence that our priors are not so mutable and our choice preferences sticky (Dave & Wolfe, 2003; Taleb, 2001; Kahneman, 2014).

1.5 The different forms of uncertainty

Broadly speaking, uncertainty represents the unknown. By this definition, a study of the phenomenon would be limited. Fortunately, various forms of uncertainty can be identified and quantified (Bland & Schaefer, 2012). Below, we briefly review the following forms of uncertainty and their related quantities, namely: ambiguity; risk; volatility; surprise; confidence and information.

1.5.1 Ambiguity

Ambiguity is a form of uncertainty whereby an agent does not know the underlying distributions of the problem she faces (Knight, 1921). In this case, the outcome probabilities associated with specific actions are unknown. An example of ambiguity is the response 'maybe' to a yes/no question. While we understand the answer to be either yes or no, we cannot predict how likely the answer will be yes, respectively, no. At best, we can attribute chance probabilities to each possible outcome. Ambiguity is thus an unknown unknown. A key feature of ambiguity is that it can be a form of reducible uncertainty, meaning that over several tries, one can uncover the specific outcome probabilities associated with an event. Thus, in some contexts one can learn the probabilistic structure of a system by repeated sampling. In the case of the 'maybe' response, such an answer may carry information if it comes from a known acquaintance (eg: maybe probably means yes/ maybe probably means no). Ambiguity is generally thought of as an unpleasant state of being (Lauriola & Levin, 2001). If an agent has no guiding information on how best to decide, that agent will try to leave the ambiguous state, either by avoiding irreducible uncertainty or by learning. The Ellsberg paradox illustrates how a preference for risky choices over ambiguous ones violates subjective expected utility theory. Individuals consistently prefer choosing an option with known probabilities (risk) over an option with unknown contingencies, even if odds are winning are low in the first option, while the odds of winning may be high in the second (Segal, 1987).

Neural Correlates of Ambiguity

As ambiguity is a distinct form of uncertainty, it may also inhabit a specific region of the brain. A study investigated the possible neural differentiation of known uncertainty (risk) from ambiguous uncertainty examined a control group and a group of patients with lesions in the

orbitofrontal cortex (Bechara et al., 2000). This patient group was selected because previous studies on their performance in a decision-making paradigm, the Iowa Gambling Task, showed deficits in the latter. Ambiguous conditions elicited OFC and amygdalar activation relative to the risky condition while risky decisions activated the dorsal striatum (caudate). Importantly, dorsal striatal activation scaled with the expected value of reward in this study while evidence of such value tracking did not appear in the OFC or amygdala. The authors suggest this finding provides evidence that ambiguity impacts the expected value of a reward. Time-courses of activation in the three regions shows a rapid response in the OFC and amygdala and a comparatively delayed response in the dorsal striatum to uncertainty. In the patient group, subjects could not differentiate between certain; ambiguous and risky decisions, which is noteworthy because such an agent will act in accordance with expected utility theory (Hsu et al., 2005).

1.5.2 Risk

When sampling reduces ambiguity in a problem, the resulting uncertainty becomes a known unknown, or risk. Specifically, outcome probabilities associated to a set of options are known. Thus an agent possesses some information that guides her choice behavior. The distinction between ambiguity and risk is an important one because a large body of research on human decision-making tests the process with the use of the Iowa Gambling Task (IGT) (Bechara et al., 2004). In the IGT, subjects select cards from different decks to obtain a desired outcome; with each deck representing an (initially unknown) probability of win/loss. Several studies have found impaired performance in the IGT associated with specific populations, such as in Parkinsonian patients; adolescents; schizophrenia patients; and patients with OFC lesions (Ryterska et al., 2013; Hooper et al., 2004; Shurman et al., 2005; Fellows & Farah, 2004). While this task is thought to investigate risky decision-making, it includes ambiguity initially, which should be reduced to risk at some point in the task (if risk processing specifically is unimpaired). The point at which subjects learn the probabilistic structure of the task may differ across individuals, as it will depend on a subjective learning rate. Indeed, some healthy subjects never learn the underlying structure of the IGT. Therefore, studies using the IGT as a means to assess decision-making should be viewed with caution, as the source of any observed impairment is confounded by several quantities (risk, ambiguity and learning rate, in addition to reward valuation). Thus some patients thought to be impaired in risk-processing may in fact have specific difficulties with ambiguity, and not risk per se (Buelow & Suhr, 2009). A more precise way to characterize risk is by taking it to be the variance of a probability distribution (Weber et al., 2004; Preuschoff et al., 2006). In this way, risk is unsigned and not burdened by psychological constructs of what bad may come in predicting an event.

The Neural Correlates of Risk

Where is risk encoded? Several studies have found risk signals in the orbitofrontal cortex (Bechara et al., 2000; Krain et al., 2006). The orbitofrontal cortex also has a well established role

in valuation proper, which may integrate risk values overall. A study looking at risk and its error specifically examined neurons in the orbitofrontal cortex of the monkey (O'Neill & Schulz, 2013). Pains were taken to control for value related variables but it is worthy of note that the risk prediction error calculated in the above study investigated the interval between the timepoint prior to cue and following cue, and not at outcome. A search for neural correlates of risky financial decision-making finds the ventral striatum (NAcc) implicated in risky choices as well as in errors resulting from the latter, while risk-free choices and their errors were reflected in anterior insula activation. Here, a risky choice is defined as one which has an associated probability of reward but whose expected value is higher than a certain option. Importantly in this study, NAcc activity predicted a risky choice and anterior insula activity predicted a safe, or certain choice, irrespective of the outcome. Interestingly, in this study, ACC activation correlated with a measure of uncertainty, or conflict and specifically when probabilities spanned midpoints between maximal uncertainty ($p=0.5$) and low uncertainty (0.9, 0.1) (Kuhnen & Knutson, 2005).

While many studies characterize risky decision-making with probabilistic tasks, we focus on studies that formalize risk as the variance of probability distribution. A study investigating risk as variance and its corollary errors in an fMRI experiment employing the IGT found risk prediction errors correlating with activity in the ACC, inferior frontal gyrus, insula, striatum and amygdala (Li et al., 2010). The authors note that IFG involvement may be due to action selection specifically. Among other studies examining risk as variance, risk signals have been found in posterior parietal cortex, cingulate cortex, striatum, amygdala, and insula (Huettel et al., 2006; Preuschoff et al., 2008; Tobler et al., 2007; O'Neill and Schultz, 2010; d'Acremont et al., 2009).

1.5.3 Volatility

Volatility is yet another form of uncertainty that has implications in economic forecasting (Cont, 2006) but is also applied more generally to human decision-making behavior (Payzan-LeNestour et al., 2013). Volatility can be thought of as unexpected uncertainty and relates to the unknown engendered by a change in a known probabilistic environment. A volatile environment thus includes a measure of sudden changes in the generative model of the observer; it is of note that the canonical method of quantifying volatility in finance is by taking the standard deviation of trade returns over the course of a year. Thus volatility must be integrated over several trials. In a reinforcement-learning context, an environment that is recognized as a volatile one should prompt a higher learning rate (Kennerley et al., 2007; Niv et al., 2012). Further, a volatile environment prompts agents to use the most recent events to update their models, while in stable environments, models should reflect the average of individual events. While volatility is not a focus of our study, we briefly describe it above because it can be confused with surprise, in that it is unexpected uncertainty. We propose that volatility can only be derived from a model while surprise can arise from a single event.

1.5.4 Surprise

Surprise is a response to an expectation violation. This response may prompt amazement, bewilderment, dismay, humor, shock and a slew of other subsequent responses, but its most parsimonious definition is that surprise arises from an unexpected event. Thus below, we detail surprise at its minimum, and how best to model the latter in a mathematical sense. We do not address the emotional quality of surprise or its consequence, or even its uniqueness to sentient beings, rather what are the necessary and sufficient conditions for the phenomenon.

Surprise occurs at the outcome of a process. Because surprise is an unexpected event, it also encapsulates uncertainty and is thus a measure of the latter. Questions on the formulation and neural representation of surprise take pride of place in this thesis, as it is one of our primary lines of investigation. Below, we treat the problem of surprise to date in the context of information theoretic, Bayesian and predictive coding points of view. One feature of surprise is that, while it can be described with a mathematical account, it is also, in the conscious human experience, more readily identified as an affective state. Thus, we feel surprised. Further, in the affective context of surprise, we experience physiological changes and even a physical response, such as a gasp or startle. With regards to the quantification of surprise, the phenomenon has been defined according to two dimensions. In the first, surprise is a stimulus-bound response to an unexpected event. In the second, this response can act as a learning signal that shifts one's expectations. Surprise quantification is linked to information but the exact general nature of the relationship across different fields such as information theoretic domains and the human experience, remains elusive.

Given the uncertain environment within which we operate, we hypothesize that a response to a surprising event is crucial to a system, as it may signal an important change in the environment. A ready example of this is found in climate change denial. While the weather forecast is consistently probabilistic, surprising meteorological events have been on the rise (even in California). Categorizing these events as freak occurrences rather than a signal of system-wide change may have potentially devastating consequences, including loss of life. Thus our brains must be uniquely tuned to register surprising events so we can better predict what our environment will offer in the future.

Surprise as Prediction Error

Several authors cite an unsigned prediction error as surprise (Glaescher et al., 2010; Preuschoff et al., 2011; Hayden et al., 2011). Surprise can be captured by the prediction error because the latter signals a difference between an outcome and an expectation. Studies have found that the prediction error is modulated by surprise ? that is prediction errors depend on the context in which they occur. Improbable prediction errors are then those that occur in a stable environment. Such prediction errors can be said to be more surprising than prediction errors that occur in a volatile environment. One study investigating prediction error responses under different cognitive loads in a dual-task paradigm found that prediction error responses in

one task were not altered by the cognitive load demand of another concurrent task. Here surprise was taken to be the unsigned prediction error in an EEG study employing an oddball paradigm (Garrido et al., 2016). Interestingly, in this study surprise is the mismatch negativity response, or the difference in EEG signal between a standard and odd tone within a given window following the sound (150 ms). This response did not alter during an N-Back task (a task with a high cognitive load), suggesting that our sensitivity to surprise is a low-level function that operates automatically.

Shannon Surprise

The field of information theory was born in 1948, with the publication of Claude Shannon's seminal paper "A Mathematical Theory of Communication" (Shannon, 1948). Shannon studied the problem of encoding and decoding transmitted signals over channels to determine how much information (entropy) is necessary to sufficiently reconstruct a signal. Importantly, events with a low frequency of occurrence encapsulate the most information and the most surprise. This surprise is quantified as the negative log of its probability. Thus surprise can be viewed as an item of information. What is the relationship between Shannon surprise and information? The former indicates a quantity associated with an outcome that conveys some amount of information, where the most probable outcome is the least informative. Shannon entropy on the other hand presides over the model space, encapsulating an average of surprise over possible outcomes and their associated probabilities. Shannon entropy can thus be computed prior to an outcome. It is important to stress that Shannon borrowed the term surprise for his theory but did not apply the concept to human surprise. More recently however, neuroscientists have seized the opportunity to investigate surprise via its information theoretical account (Strange et al., 2005; Mars et al., 2008), as the latter plausibly explains the human experience as well. This characterization of surprise is not without its detractors however.

Bayesian Surprise

If a surprising event defies our current expectations, it may alter our future expectations. That change can cast surprise as a learning signal. More specifically, if surprise changes our expectations of a given cue, one can argue that a Bayesian process is at play, whereby our prior distributions regarding a given cue change into posterior distributions. If one considers the zen koan of a tree falling in a forest, we can pose the question: if an event is improbable but you don't learn from it, is it surprising? Some have proposed that surprise can only be quantified by evidence of learning. Thus, a Bayesian model of surprise has been proposed to challenge information theoretic accounts of surprise (Itti & Baldi, 2005; Itti & Baldi, 2009; Itti & Baldi, 2010; Bencomo & Bellagoun, 2014). Specifically, Itti & Baldi posit that not only must a surprising event result in a shift in belief distributions but that the shift is surprise itself. To quantify this update, they employ a distance measure, specifically a Kullback-Leibler (KL) divergence or relative entropy. In this context, surprise will decrease in a stable environment

as the distance measure converges to 0 with increasing learning trials. A distinction between Shannon Surprise and a KL divergence model of Bayesian Surprise is that the former is taken over the data space while the latter measures model differences. Itti & Baldi take the example of a uniform prior distribution where each model is as likely as the next to explain the data. The probability of the data given the model is then equal to the probability of the data.

For a highly improbable event, the Shannon Surprise tends to infinity bits. However, using Bayesian surprise, we see that the posterior distribution is equal to the marginal distribution, and thus we are left with 0 units of surprise ('wows'), as the distribution does not shift. Intriguingly, Itti and Baldi argue that improbable events may be surprising but not useful if all other events are also improbable and thus low probability alone should not define surprise. They underline this quandary by taking the example of snow (white noise) on a television screen, relative to an unexpected television program. The first has the highest amount of Shannon Surprise, or information, but is the least surprising because it is the most boring, while the second will capture our attention and is thus more surprising, namely it will shift our Bayesian prior. In their discussion, they touch upon a crucial facet of information, decision-making and uncertainty, which has yet to be appropriately treated in the behavioral sciences, and that is relevance. Itti & Baldi therefore propose that information includes three components: Surprise (as defined in Bayesian terms); Shannon entropy; and relevance (Baldi & Itti, 2010).

From the information above, the quantification of surprise remains a matter of debate. KL divergence can be seen as the data packet that updates a model while Shannon information can be viewed as a signal to reorient the system. From a human perspective, surprise is associated with an improbable event, be it in the form of Shannon surprise or a prediction error. Learning from such an event is possible but not necessary. Similarly, surprise may not be necessary for learning. The two are often correlated however in experimental paradigms. Below, we review key studies examining this dual aspect of surprise.

The P300 event related potential in EEG studies is often viewed as a 'surprise' signal. While its occurrence correlates with the presence of an unexpected stimulus, its computational account is not well elucidated. A study investigating trial-by-trial variations in signal amplitude using information theoretic models found stimulus-bound surprise best explained fluctuations in the P300 signal. Surprise in this study was quantified by Shannon surprise but data were also fit to a Bayesian model of surprise, where the latter is captured by the KL divergence. The two models differ in that KL divergence measures the difference between a posterior and a prior distribution following an event and thus captures an average quantity of surprise summed over all probabilities, whereas the information theoretic model captures surprise as a change probability related to one outcome. This distinction is not merely theoretical: the KL divergence implies that the model changes due to one event, an inefficient and thus unlikely mechanism by which humans integrate surprising information. The information theoretic model of surprise was found to explain P300 variations better than the KL divergence or a categorical model (where surprise was formulated as the probability of the event, with less probable events being more surprising). Further, in this experiment, the task was presented in

blocks. Information theoretic surprise was quantified within blocks (a forgetting condition) and across blocks. They found that the forgetting model of surprise best explained the data (Mars et al., 2008).

To tease out their respective contributions (if indeed they are dissociable) an experiment was performed investigating surprise as Shannon information versus surprise-induced updating as a KL divergence. Here the key variable dissociating the two facets of surprise was *relevance*, thus it was assumed that only relevant surprising events would lead to an update. The paradigm applied was a saccade response task where surprising events included stimulus presentation in an improbable location; update-worthy events were presented in a different color relative to 'one-off', irrelevant events. Interestingly, saccade reaction times for surprising events increases but for update trials, it is the dwell time that increases. No significant effect of prior entropy was observed. Shannon surprise correlated with posterior parietal cortex while KL divergence correlated with ACC activity (O'Reilly et al., 2013).

To distinguish the neural correlates of a surprising event's potential dual roles, Strange and colleagues (Strange et al., 2005) applied a stimulus bound information theoretic measure to an event as well as an average entropy measure to the same event. To investigate the contributions of both measures in the neural response to an event, they performed an fMRI experiment where subjects engaged in a probabilistic learning task involving the selection of an item from a pool of four stimuli. Their aim was to identify hippocampal responses to entropy alone, after controlling for stimulus-bound surprise. The surprise bound to each choice reflects the frequency with which a particular stimulus had previously been encountered. Entropy activated bilateral regions in the cortex and thalamus, but interestingly did not provoke activity in the striatum and insula (Strange et al., 2005). Their results also show a hippocampal-specific response to entropy that is not found for surprise. The modulation of the hippocampus by entropy is viewed as learning that such an event can occur (Strange et al., 2005, Schiffer et al., 2012), which supports the notion that surprising events act as learning signals.

A feature of surprise is that it is costly to a system. Experimental evidence shows an increase in reaction time with unexpected events for instance (O'Reilly et al., 2013). Surprise-induced updating would further tax the system, and can be viewed as a switching cost (Monsell, 2003). While the first cost cannot be avoided, the second one can. Thus perseverance in maintaining what is perhaps an outdated belief is a common occurrence (Nickerson, 1998). Indeed, the notion that humans are Bayes optimal is severely challenged by evidence pertaining to choice preferences (Dave & Wolfe, 2003; Sharot et al., 2010; Sharot et al., 2012). The results above suggest that in humans, surprising events will not necessarily entail learning. This conclusion is supported by research in the artificial intelligence field, specifically in the modeling of belief-desire-intention (BDI) agents. Because these models attempt to simulate human affective and motivational dynamics, surprise is considered. In one framework, surprise is defined as having two components, one that signals a mismatch and another, later 'second-order' component involved in revising a belief. One can see how these two functions of surprise can map to information theoretic or prediction error surprise and KL divergence, respectively.

Lorini & Castelfranchi indeed argue that not all surprising events will lead to a revision of belief, or update in the Bayesian sense (Lorini & Castelfranchi, 2006).

The Neural Correlates of Surprise

As we have seen above, formal accounts of surprise vary. However, fMRI studies on surprise generally identify similar neural correlates. Surprise, as absolute risk prediction error, was found encoded in the anterior cingulate cortex (Hayden et al., 2011). Surprise as absolute reward prediction error may correlate with the reward prediction error (RPE) however and thus a question emerges on the differentiation of the separate impact of both errors on the system, either temporally or spatially (Fouragnan et al., 2017). Learning from an RPE will inform which actions to repeat in the future while learning from surprise should guide an agent towards allocating attentional resources. While studies have found distinct neural representations of RPE and surprise, their respective dynamics, due to fMRI's poor temporal resolution, remain elusive. In a paradigm employing fMRI and simultaneous EEG during a probabilistic reversal-learning task, RPE valence and unsigned RPE (Surprise) overlapped temporally but with a different spatial distribution. Surprise engaged dorsolateral prefrontal cortex, bilateral insula, medial prefrontal cortex, inferior frontal, supramarginal, angular and precentral gyrii. A recent study examined surprise-related neural activity in contrast to a prediction confirming related inactivity using a visual task in monkeys. The researchers found that outcomes confirming a prediction elicited a significant neural suppression while surprising events yielded a slight, non-significant neural enhancement (Ramachandran et al., 2017).

1.5.5 Information (Entropy)

As seen above, investigating surprise necessitates a consideration of information theoretic quantities, namely surprise but also information, or entropy. Shannon entropy is the amount of information needed to obtain a correct response in units of bits. Thus, in an uncertain context, it would be the amount of information necessary to answer a question; formally, it is the lowest number of yes/no questions required to resolve uncertainty. When applying these measures to human behavior, we can enter a slippery slope. Surprise can be informative, but what of the entropy of a decision? An EEG study investigated the possible dissociation of Shannon entropy and Shannon surprise in relation to task switching in a modified version of the Wisconsin Card Sorting Task (Kopp & Lange, 2013) with the aim of differentiating the P300 index into its P3B and P3A components. They find the P3b component present for all task switches, which correspond to surprise, while the P3a was modulated by entropy, increasing with the most uncertain outcomes. An fMRI study employing Shannon entropy as an uncertainty measure in an economic task found neural correlates of the quantity in pre-supplementary motor Cortex, mid-cingulate cortex and thalamus (Goni et al., 2011).

1.5.6 Confidence

When one considers uncertainty, one has to consider confidence. Confidence can be thought of as a measure of certainty, with an emphasis on the term measure. Confidence, like uncertainty, ambiguity, and surprise, can be defined in quantitative terms or as an emotional response. While we may denote confidence with a point estimate, the latter usually implies an interval; that is, we explicitly experience confidence as a graded value, not a binary one (Kiani & Shadlen, 2009; Meyniel et al., 2015). Confidence is resolved when outcomes are certain. Thus confidence can only be sampled in the Bardo state between prediction and outcome. The importance of confidence in decision-making should be outlined here, as this computation is what permits a decision-maker to act; it is the belief that one is correct. Therefore, without confidence, action becomes difficult, if not impossible.

For instance, one can view the phenomenon of learned helplessness as an outcome of having no confidence. When animals cannot learn contingencies between cues and aversive outcomes, that is, when conditions are ambiguous; risk is maximal; and confidence is lowest, they eventually cease to respond to a cue at all and resign themselves to a state of learned helplessness (Seligman, 1972). The importance of confidence in decision-making has even been highlighted as the most important variable with respect to decision-making (Pouget et al., 2016).

Given the importance of confidence in decision-making (Goette et al., 2015; Bendahan et al., 2017), its neural corollary is the topic of current study. In a study using signal-detection theory, yet another framework within which to investigate uncertainty, confidence signals have been found encoded in the rat orbitofrontal cortex using neural recordings (Kepecs et al., 2008). This study introduced uncertainty as task difficulty, which in this specific case represented ambiguous odors; thus a 50/50 odor mixture was the most difficult to categorize. Subjective confidence was estimated as the distance between a trial sample's category and a memory of previous samples categorizations. Further, confidence correlated with a willingness to wait for a reward. Intriguingly, a similar study in humans did not find linear increases in OFC with decision confidence computed as a measure of discriminability but in medial prefrontal and anterior cingulate cortices (Rolls et al., 2010). Perceptual confidence has been linked to signals in the human and macaque striatum (Ding & Gold, 2012; Hebart et al., 2016). In an fMRI paradigm investigating confidence and a decision variable using signal detection theory in a random dot motion task, Hebart and colleagues find striatal signals relating to increases in decision confidence and insula signals with decreases in confidence. Here, the decision variable is taken to be the amount of sensory evidence in favor of one option. The confidence is the distance between the decision variable on the i th trial and a subject's criterion; confidence is quantified by self-report. A key element of the study was to examine the relationship of self-reported confidence to accuracy, for which the authors find a linear correspondence. Interestingly, without explicitly alluding to confidence as inverse variance, Lebreton and colleagues formalized confidence as a second order function of value assignment. They found that confidence values were encoded in the ventromedial prefrontal cortex, an area adjacent

to the orbitofrontal cortex (Lebreton et al., 2015).

While several authors emphasize the subjective quality of confidence, confidence can be computed in an objective manner. The relationship between the two quantities, subjective and objective confidence, has recently been investigated and objective confidence has indeed been found to correlate with subjective accounts (Sanders et al., 2016).

Thus uncertainty arises in different forms and engenders other related variables, such as confidence, surprise and information. It is of note that ambiguity can sometimes be reduced to risk; that confidence and risk are two sides of the same coin, depending on the operationalization method used; that surprise and volatility, two forms of unexpected uncertainty, are not equivalent though they can be related.

1.6 The role of neuromodulators in the encoding of decision variables

Decision-making is a dynamic process. Within the frameworks discussed above, encoding different values requires a means by which to do so. In the human brain, this process can plausibly be attributed to different neuromodulators. Two key neuromodulators of decision-making are reviewed here: dopamine and noradrenaline. Dopamine's importance in encoding decision-making variables is of such weight that we cannot by-pass a cursory review of its role. We review noradrenaline in more detail, however as we hypothesize that it plays a specific role in uncertainty processing.

1.6.1 Dopamine

While extensive evidence supports the notion that dopamine codes reward and its prediction in a monotonically increasing manner (Schulz et al., 1997; Schulz, 1998), its role in uncertainty encoding is less clear. In one study, dopaminergic neurons were found to encode uncertainty in primates (Fiorillo et al., 2003). Fiorillo and colleagues recorded 188 midbrain dopamine neurons during a Pavlovian task. No activity was recorded when a cue fully predicted a reward but probabilistic reward did elicit expected activity. Interestingly, this activity persisted even after training, when the animal had presumably learned the contingencies of the task. Cues embodied a range of reward probabilities, including the most uncertain at 0.5, and Fiorillo and colleagues note an increase in sustained activity between cue and outcome that scaled with increasing uncertainty (or converged to 0.5). Intriguingly, reward and uncertainty-related activations appeared in the same population of dopaminergic neurons but the dynamics of phasic activity related to reward magnitude differed from the uncertainty related activity in that the latter was slower and sustained. The studies this thesis is based on also point to a dopaminergic involvement in uncertainty processing: specifically risk (as variance) and risk prediction error signals were found encoded in the striatum, implicating dopamine in a financial uncertainty task (Preuschoff et al., 2006). Thus dopamine likely plays a role in

uncertainty encoding, however, it may be specific to certain forms of uncertainty and not others. Entanglements at the neural and computational levels offer a distinct opportunity for further experimental investigation.

1.6.2 Noradrenaline

Noradrenaline is a hormone and neurotransmitter (Doya, 2008). Here, we will focus on its role in the central nervous system as a neurotransmitter that engages the sympathetic nervous system. In the brain, noradrenaline is produced by two small brainstem nuclei that together form the locus coeruleus (LC). Noradrenaline is synthesized from tyrosine, as is dopamine; and its direct precursor is the latter, only one enzymatic step away. Thus the chemical structure of dopamine and noradrenaline are similar (Briand et al., 2007). Noradrenaline's primary general role is in arousal and alertness; its activity during sleep is muted. Its role in the fight or flight response and more specifically in response to stress, has been widely noted (Glavin, 1985). Globally, noradrenaline functions to increase metabolic expenditure. Crucially, noradrenergic dynamics vary, making it an ideal candidate as a neuromodulator (Doya, 2008). While noradrenergic neurons are concentrated in a very small nucleus of the brainstem, they project widely to both the spinal cord and throughout the brain. More recently, noradrenaline's role in cognitive and behavioral processes has been investigated. For instance, Cahill & McGaugh have found a noradrenergic modulation of memory formation (Cahill et al., 1996; Rimmele et al., 2016). Studies in rats have found noradrenergic responses to reward but more importantly for our research, to contingency changes in probabilistic learning tasks (Bouret & Sara, 2004; Aston-Jones et al., 1997). A shift in a task's structure implies uncertainty; therefore a noradrenergic response to the latter may signal uncertainty encoding (Yu & Dayan, 2005; Aston-Jones & Cohen, 2005; Preuschoff et al., 2011).

Neurophysiological recordings of noradrenergic activity show acute and replicable spikes in response to stimuli (Lestienne, 2001), an ideal quality in a neuromodulatory candidate. Further, noradrenaline is widely viewed as being necessary for attention (Robbins, 1984). It has been hypothesized that through arousal, noradrenaline enhances synaptic plasticity, which would facilitate learning but also support its role in signaling uncertainty. Evidence of noradrenergic involvement of salience comes from rat studies, where unexpected stimuli evoked LC activity that resolves when the stimulus was not paired with a relevant consequence (such as reward) (Herve-Minielle & Sara, 1995). Animal studies on noradrenaline show LC neurons firing in two ways: a tonic mode and a phasic mode, mirroring mid-brain dopamine in this capacity. The phasic mode is seen to fire for novelty, which, again, relates to uncertainty. In reviewing studies on noradrenaline, Aston-Jones formulated a theory on its role and suggested that noradrenaline encodes an adaptive gain for relevant neural circuits in the face of stimuli that require increased attention; and this adaptive gain facilitates improved performance (Aston-Jones, 2005). A signal detection task study in monkeys shows a phasic noradrenergic response to rewarding cues and not distractors, suggesting that the NE response is specific to relevant stimuli, and not merely salient ones (Aston-Jones et al., 1994). Further, this response

is found highest when an animal performs well (Aston-Jones et al., 1997). Pharmacological challenge studies show that the α_2 receptor agonist clonidine (administered directly via microinfusion) lowered tonic NE activity, which in turn enhanced phasic activity and improved task performance in a monkey (Rajkowski et al., 1994). The most notable contribution of Aston-Jones' adaptive gain model is that it offers a mechanism by which agents engage in exploration and exploitation. A 'greedy' agent guided solely by utility will repeatedly exploit a given policy that certainly yields reward; however, in exploring, an agent may lose utility in the interim while pursuing an even more rewarding policy. The balance between these two courses may be reflected in noradrenergic phasic and tonic modes. Elevated tonic NE activity promotes 'random' behavior and exploration while phasic firing, which correlates with focus on a given task and concomitant high performance, may model exploitation.

What happens when noradrenaline is affected? In Parkinson's disease, noradrenergic activity is affected before dopaminergic deterioration (Rommelfanger & Weinshenker, 2008). Parkinson's disease profiles include behavioral changes, including cognitive impairment, decision-making deficits, and apathy. While therapeutic approaches have focused on dopaminergic replacement therapy to resolve motor symptoms, the implications of noradrenergic failures have been by the by neglected (Loued-Khenissi & Preuschoff, 2015).

1.7 Brain Regions Associated with the Neural Representation of Uncertainty

Neuroimaging research on decision-making has yielded a widespread area of brain regions involved in the process, including limbic regions, frontal and prefrontal cortices, parietal cortices and the striatal system (Ernst & Paulus, 2005). Guided by recent studies, we describe areas thought to be especially involved in decision-making under uncertainty. Here, we briefly detail regions examined in the experimental parts of this thesis, to explain our focus on them.

1.7.1 The Anterior Cingulate Cortex

The anterior cingulate cortex (ACC) is a medial brain region that spans the cortex from the posterior part of the corpus callosum through to the anterior portion. The ACC had previously been enlisted as a limbic region, not unlike the insula (Bush et al., 2000) however the anterior cingulate cortex in particular has been implicated in a variety of high-level cognitive functions, such as planning and control (Ridderinkhof, 2004). The region was hypothesized to play a prominent role in conflict monitoring (Botvinick et al., 2001) as it correlates with error commission, uncertainty and impulse inhibition. Considerable empirical evidence has since widely supported this function of the ACC (Pochon et al., 2008; Botvinick et al., 2004). It is thus a region that often appears in studies on decision-making, as the latter process implies the inclusion of conflict. Shenhav et al. (2013) attempt to clarify the ACC's function without reducing it and suggest the region is specifically responsible for cognitive control, that is,

1.7. Brain Regions Associated with the Neural Representation of Uncertainty

staying a course when competing influences vie for the system's attention or resources. They frame the ACC's role as that of an evaluator of cognitive control demands, or the region that computes the expected value of control. Studies show ACC activity scaling with task difficulty (Venkatraman & Huettel, 2012).

The ACC may also be linked to negative outcomes, such as pain (Shackman et al., 2011) however, as Shenhav and colleagues note, conflict can be viewed as a negative, unpleasant state. Intriguingly, ACC activity correlates with both magnitude and probability of reward in both neural recordings and neuroimaging studies (Wallis & Kennerley, 2011; Knutson et al., 2005; Bartra et al., 2013). The ACC has specifically been implicated in the reinforcement-learning framework. Several studies have found ACC activity in relation to prediction errors, from feedback related negativity signals in EEG (Miltner et al., 1997; Nieuwenhuis et al., 2005). ACC has also been found to encode reward prediction in addition to its errors (Amiez et al., 2006; Rushworth & Behrens, 2008). Interestingly, these signals are tied to the prediction error and not necessarily to a monetary loss in economic paradigms. Further, ACC activity is tied to prediction errors regardless of sign (Bryden et al., 2011; Hayden et al., 2011), suggesting it specifically plays a role in surprise (Cavanagh et al., 2012; Behrens et al., 2007).

Interestingly, ACC activity is found in imaging studies investigating exploration over exploitation in humans (Daw et al., 2006; Amiez et al., 2012; Cavanagh et al., 2012). If ACC signals a surprising event, it may also prompt an exploratory response to reassess a structure's contingencies. In addition to co-activating with the insula, the ACC projects to the striatum (Choi et al., 2012; Haber & Knutson, 2010), which is a central actor in the predictive coding model of learning and decision-making. Finally, the ACC has strong connections with the LC, main site of noradrenergic production in the brain, which itself may signal the need for exploration in the face of an error, or surprise, or a conflicting outcome of some sort (Aston-Jones & Cohen, 2005; Gilzenrat et al., 2010; Jepma & Nieuwenhuis, 2011; Murphy et al., 2011). A study probing decision-making performance in ACC lesioned monkeys examined whether activity is linked to error alone or to encoding a history of action-outcome contingencies to formulate predictions. Lesioned monkeys did not show an immediate impairment in performance on a reward-guided learning task, but could not integrate trial outcomes into a policy. Further, in a dynamic foraging task where rewarding outcomes were probabilistic, lesioned monkeys also displayed impaired performance on average and not for specific trials (Kennerley et al., 2006). The ACC's role in integrating prior outcomes and generating predictions was further found in a paradigm that did not necessarily elicit an error (Brown & Braver, 2005). In a task where subjects could either recognize an error or received feedback on an error from an external source, the ACC was implicated in both (Holroyd et al., 2004). As uncertainty begets conflict, we expect the ACC to be implicated in our studies in particular, as the region is clearly implicated in processing conflict at a higher level.

1.7.2 Insula

The insula is a bilateral, cortical region that has been implicated in myriad functions (Nieuwenhuys, 2012) but is generally involved in interoception, the process by which internal signals, or stimuli, are interpreted into declarative states. Interoception is thought necessary for emotional awareness (Wiebking et al., 2015; Zaki et al., 2012). The insula has previously been found to play a role in uncertainty signals, specifically prediction errors (Preuschoff et al., 2008). If one considers that uncertainty engenders a feeling, then one can see how the insula is implicated in guiding decisions (Singer et al., 2009). A study investigating the impact of interoceptive awareness on insular activity in the Iowa Gambling task found a correlation between individuals high in interoceptive awareness and decision-related insular activity (Werner et al., 2013). A defining feature of the insula is its membership to the brain's salience network, a collection of regions, primarily the anterior cingulate cortex and insula, which correlate with engagement in resting-state fMRI. This fact attributes the insula with more than affective processes and suggests that its role is central to human experience (Menon and Uddin, 2010).

A notable feature of the insula is that it likely does not have an animal homologue (Craig, 2009) though this hypothesis has been vigorously challenged and others believe monkeys do have an insular-like structure (Nieuwenhuys, 2012). Menon and Uddin note that the common factor across studies in which the insula appears is the presence of a deviant stimulus amongst other, presumably expected, stimuli. In terms of connectivity, the insula along with the ACC receive afferent projections from the spinal cord, which brings bodily signals to the cortex via the thalamus (Craig, 2002, 2003). Projections from the insula descend into the brainstem. This characterization of insular connectivity is limited however in that it was identified in animals. Another intriguing feature of the insula and the anterior cingulate cortex is that both include von Economo neurons (Allman et al., 2010), the latter which have large axons and can thus relay information rapidly. While several studies have implicated the ACC in decision-making tasks, the insula has appeared less often. In addition to its involvement in risk and risk prediction error signals specifically (Preuschoff et al., 2008), it has been implicated in uncertainty in several other studies (Critchley et al., 2004; Xue et al., 2010; Jones et al., 2011; Weller et al., 2009).

Menon & Uddin suggest that the insula does indeed have a cognitive control function beyond its apparent involvement in emotional processing, but that the region is responsible for transient signals that are then relayed to the ACC and PFC regions and thus acts as a switching region, a hypothesis that is supported by its inclusion in the salience network, which deactivates the 'default-mode' network in resting-state fMRI studies. This hypothesis is further supported by an EEG-fMRI study investigating the temporal dynamics of insular activation in an oddball paradigm (Crottaz-Herbette and Menon, 2006). While some propose a functional division in the insula along the anterior and posterior axis, with the former playing a role in awareness and the latter more relevant to visceral states, such spatial differentiation remains elusive. The insula has further been implicated in exploratory states, which may

1.7. Brain Regions Associated with the Neural Representation of Uncertainty

be confounded with uncertainty-inducing, or surprising outcomes (Laureiro-Martinez et al, 2014). In the context of our work, we zero in on the insula especially as it relates to the ACC; and because it's role in awareness and interoception may play a particular role in translating computations, which we cannot assume occur explicitly, into states of awareness (Singer et al., 2009; Medford & Critchley, 2010).

1.7.3 Striatum

The striatum's role in learning and specifically in reinforcement learning has been studied extensively (Hollerman & Schulz, 1998; O'Dogherty et al., 2003; McClure et al., 2003; Schulz, 1998; Abler et al., 2006; Pessiglione et al., 2006; Balleine et al., 2007; Daw et al. 2011). While several regions have been implicated in prediction error encoding, a meta-analysis of studies on the topic reveals that it is the striatum that is singularly implicated in the process (Garrison et al., 2013). Expected utility for instance, which is captured in reward prediction, elicits increased striatal activity in a monotonic fashion (Tobler et al., 2007). Both reward prediction and its error are encoded in the striatum and evidence suggests uncertainty is integrated in the region as well (Fiorillo et al., 2003) as does previous work highlighting the region's involvement in uncertainty-related prediction and prediction errors (Preuschoff et al., 2006). The striatum's function is not limited to reward based learning however. Motivation and incentive salience, or desire, play a role as well although the long-held view of the striatum as a pleasure center is no longer widely accepted (Berridge, 2013).

Beyond decision-making and learning, the dorsal striatum mediates motor function. As such, in Parkinson's disease, whose neural hallmark is dorsal striatal degeneration, motor deficits are the most dramatic and impactful symptoms. While a dorsal-ventral pattern of degeneration in relation to motor symptom etiology is clear in Parkinson's patients, such a mapping for functionally distinct cognitive symptoms is less obvious. Some have suggested that dorsolateral striatum mediates 'goal-directed' learning while ventromedial captures trial and error learning but such distinctions, borrowed from animal studies, are difficult to verify in humans (Doherty et al., 2014). A review of studies investigating functionally distinct roles for dorsal and ventral striatal regions in reinforcement learning finds conflicting evidence for such a spatial mapping (Garrison et al., 2013). Thus the whole striatum emerges as a prime candidate within which to look for neural signals of uncertainty.

1.7.4 The Locus Coeruleus

The locus coeruleus (LC) is a nucleus of the brainstem that is the primary site of noradrenaline secretion in the brain. The nuclei are present bilaterally and are shaped into two long and narrow structures beginning just under the IVth ventricle on the superior dorsal lateral pons. Its mean dimensions are 14.5x2.5x2 mm (Fernandes et al, 2012). The region is easily seen on histological slices because it is pigmented by neuro-melanin, a by-product of its synthesis, which gives the region its name. Activity in the nucleus is generally related to alertness:

LC activity is low during sleep cycles. The LC projects widely to cortical regions, as well as sub-cortical regions that are key to motivational states, such as the thalamus, nucleus accumbens and amygdala. Importantly, LC noradrenergic neurons interact with dopaminergic neurons in the substantia nigra. In Parkinson's disease, LC degeneration is widespread and observed substantia nigra degeneration may result from the latter (Srinivasan & Schmidt, 2003; Rommelfanger & Weinshenker, 2007; Loued-Khenissi & Preuschoff, 2015). The LC is not visible with standard in-vivo imaging techniques in humans. However, a specific T1-TSE sequence that exploits the gradient offered by the nuclei's pigmentation has succeeded in localizing the region with MRI (Sasaki et al., 2006; Keren et al., 2009). The LC has been implicated in the exploration-exploitation trade-off (Laureiro-Martinez et al., 2014). Pupil dilation is, to date, taken to be a proxy for LC activation (Murphy et al., 2014). Given previous work finding increased pupil dilation in response to unexpected uncertainty or surprise (Preuschoff et al., 2011), we hypothesize that the LC may be specifically involved in uncertainty signaling.

1.8 Patient Populations

A quality of neurological and psychiatric conditions is that they often entail cognitive deficits, subtle impairments in the cognitive domain that are generally placed under the rubric of executive dysfunction (Petersen et al., 2014). Executive dysfunction is usually identified with the use of neuropsychological tasks that are meant to capture specific functions such as planning and working memory. Model-based accounts of learning and decision-making offer a means by which we can capture such subtle deficits (Maia & Frank, 2011). Indeed, as we know that reinforcement-learning mechanisms rely on striatal dopamine, disorders in which dysfunction of the system are known hallmarks become prime candidates within which to examine specific sub-processes that may be impacted. Examples of such conditions include Parkinson's disease, schizophrenia and drug addiction. For instance, schizophrenic patients have been found to show a reduction in striatal response to prediction errors (Waltz et al., 2009). In Parkinson's disease, deficits in tasks such as the IGT may reflect a decreased learning ability with respect to probabilities specifically (Sinz, et al., 2008; Euteneuer et al., 2009; Delazer et al., 2009; Gleichgerrcht et al., 2010) and have been linked putatively to reward processing deficits (Aarts et al., 2012; Ryterska et al., 2013; Muhammed et al., 2016). Thus, several studies have attempted to characterize dysfunctional risk processing in decision-making tasks. Pertl and colleagues (Pertl et al., 2015) tested patients with mild cognitive impairment (MCI) on a range of decision-making tasks, including the Game of Dice Task, which requires risk processing. They found patients performed worse than controls in the Game of Dice specifically but performed comparably to controls in other tasks. Parkinson's disease patients are thought to have impaired reward sensitivity (Muhammed et al., 2016) however, this may be modulated by apathy. Thus it is not clear that Parkinsonians are impaired in reward processing per se; they may in fact have trouble assigning meaning to cues because of co-morbid apathy, a state of affairs that may present outwardly as dysfunctional reward sensitivity. The role of apathy in decision-making impairments in Parkinsonians was also found in a study using the Balloon

Analog Risk Task (Buelow et al., 2013).

Anxious individuals are thought to be impaired in their decision-making processes, perhaps by being especially risk-averse (Charpentier et al., 2017). Using a reinforcement-learning framework, Browning and colleagues (Browning et al., 2015) find that anxious individuals have a specific deficit in updating their predictions in volatile versus stable environments. In their study, anxious individuals were impaired in their ability to adjust to volatility in an environment with potentially aversive outcomes, a feature reflected in a decreased pupillary response to volatility and in learning rates between stable and volatile tasks. The anxious individuals maintained similar learning rates in both environments while the control group increased their learning rate for the volatile tasks (Browning et al., 2015).

A recent study investigating the dissociation between confidence and action in a predictive inference task found a distinct impairment in patients suffering from obsessive-compulsive disorder (OCD). Vaghi and colleagues sought to study the evolution of confidence and action in a learning task. OCD patients often engage in repetitive behavior that is exaggerated relative to need (such as cleaning or, in the opposite case, hoarding), yet often recognize that their behavior is dysfunctional (van den Hout & Kindt, 2003). The authors exploited this dissociation between action and belief, as OCD patients' belief are seemingly accurate, while their behavior contradicts the latter. If confidence is a belief about the state of the world, then an action should relate to that belief. Here, the possible dissociation between action and confidence was investigated. Following an error in a spatial task, patients' learning rates were consistently higher than controls', though they scaled correctly with error magnitude. Further, patients' confidence matched controls, indicating that this part of the decision process was not altered by their pathology. Thus if patient behavior is driven by prediction errors rather than confidence estimates, it suggests that patients may use a model-free strategy to navigate their environment rather than a model-based one. Importantly, in this study confidence measures are obtained via self-report and are thus subjective (Vaghi et al., 2017). Autistic patients are known to have difficulty seeing illusions. Thus, within the frameworks reviewed above, they may possess a distinct deficit in their inference processes (Lawson et al., 2015). This possibility was investigated in the context of a Bayesian framework, and the results suggest that autistic patients may not form priors, and thus cannot have posteriors in the form of say, a perceptual switch (Pellicano & Burr).

Finally, while it is not a disease, stress, both of the acute noradrenergic and delayed, cortisol-mediated kind, may have specific effects on uncertainty processing (Goette et al., 2014; Bendahan et al., 2017). A very recent study found adults subjected to early-life stress showed a deficit in learning with regards to probabilistic losses in a gambling task. Specifically, they showed decreased neural activity in loss anticipation and increased response at loss outcome. These results are remarkable in that they come from a longitudinal study and are conducted on people who are not necessarily stressed in adulthood (Birn et al., 2017). At a less dire level, stress effects on financial decision-making can have a significant impact, for instance in traders, and is thus currently an important topic of research in neurofinance (Kandasamy et

al., 2013).

1.9 Notes on Model-Based Approaches

The capture of cognitive variables in human behavioral research is generally enhanced with the use of a formal account. Many variables of interest can be thought of as latent variables, not readily observable or operationalized even with well-thought out tasks. Such model-based approaches thus offer a more precise tool by which to discern hidden variables contributing to an overt signal on a trial-by-trial basis. A caveat to this approach is that several models may be viable candidate algorithms for the brain. One manner in which to test these models, usually against behavioral data, is to perform a model comparison test, usually either an Aikake or Bayesian Information Criterion test. However, the latter tests do not disprove a hypothesis but rather provide evidence in favor of one model over another. This point is a crucial one to consider; in choosing specific models to explain a given phenomenon, we may be leaving a viable candidate out of our set list. Indeed, we might even be leaving out the best candidate for a process (Mars et al., 2012). It is of this author's view that model-based methods be guided by the same principles as hypothesis testing in general, and not merely, say, the number of free parameters in the model. A choice of model therefore needs to be anchored, formally or empirically, in previous studies, even if those come from different fields. Further, an understanding of human and neural behavior ought to provide a frame of reference to a researcher. For instance, while Bayesian methods are a popular means by which to model human behavior, we must remember that decades of psychological research, if not millennia of human history, would suggest that the human, with his many biases, is decidedly not Bayes optimal. Our view is that the models reviewed above need not compete against one another. They provide useful tools by which to address the problem of inference, by nature of being probabilistic. One can note that they are not necessarily mutually exclusive: for instance, Free Energy incorporates Shannon Surprise and Bayesian Surprise. Thus we review them here to provide context to the story of uncertainty, not to pit one model against another.

1.10 Conclusion

The review above details the context of our research questions, by first exposing the sometimes hidden importance of uncertainty in decision-making. We then outline the different model frameworks that uncertainty is cast in, as it is a problem many fields are confronted with and not a unique feature of human behavior. We briefly delineate the studies on the neural representation of uncertainty, focusing on key regions and neuromodulators. And finally, we provide a short summary of how these frameworks can perhaps better explain ill-defined deficits in several patient populations.

2 In Search of the LC: The needle in the haystack

2.1 Introduction

If uncertainty is an inevitable and ubiquitous factor in the process of decision-making, there must be a neural system responsible for tracking it. As dopamine encodes reward signals in decision-making and learning, it is possible that another neuromodulator tracks uncertainty-related variables. One emerging candidate for the latter is noradrenaline (Yu & Dayan, 2005; Doya, 2008), a monoamine whose chemical structure closely resembles dopamine. Indeed, noradrenaline synthesis is one enzymatic step removed from that of dopamine. Thus in this chapter, we seek to investigate this potential link by training our sights on the locus-coeruleus noradrenergic system (LC-NE). One overarching aim in this work is to find evidence of noradrenergic involvement in uncertainty processing, as indexed by a significant BOLD response in the locus coeruleus (LC), the main site of noradrenergic neurons in the brain. This goal is ambitious and fraught, given the LC's size and location in the brainstem. In this chapter, we detail a series of pilot studies performed to determine if LC localization is feasible with the use of MRI.

The ground-breaking findings on dopamine's role in reward-guided learning and decision-making (Bayer & Glimcher, 2005; Glimcher, 2011; Schultz et al., 1997; Schultz, 2000) have prompted a search for similar functions among other neuromodulators. Noradrenaline (NE) in particular has emerged as potential candidate in uncertainty signaling (Yu & Dayan, 2005; Doya, 2008; Preuschoff et al., 2011). Noradrenaline has a dual role in the central and peripheral nervous system. In the former, it engages the sympathetic nervous system, mobilizing the body for action, by prompting increased heart rate, blood pressure and glucose utilization. In the latter, it increases vigilance and alertness. Thus noradrenaline orients the system towards a salient signal. Noradrenergic neurons are few, in spite of the neurotransmitter's importance to the system. As mentioned, most of these neurons are located in a small brainstem area called the locus coeruleus (LC), which takes its name from the pigmentation of its cells due to the presence of neuromelanin in noradrenaline, a by-product of its synthesis. While the locus coeruleus is small, its projections are widespread across the brain and into the cortex.

Previous work investigating risk and risk prediction errors in a pupillometry task finds increases in pupil dilation with larger risk prediction errors (surprise) (Preuschoff et al., 2011; Nassar et al., 2012; Kloosterman et al., 2015). As pupil dilation is an index of LC-NE activity (Murphy et al., 2014), the finding suggests that the LC encodes surprise specifically but may more generally encode uncertainty variables.

2.1.1 Does noradrenergic activity encode uncertainty processing?

In examining neuromodulatory roles for different kinds of uncertainty, Yu & Dayan hypothesized that acetylcholine (ACh) and NE signal expected uncertainty (risk) from unexpected uncertainty (surprise). Noradrenergic responses to novel stimuli or a sudden change in the environment are acute and resolve rapidly following exposure (Aston-Jones & Cohen, 2005) while ACh dynamics are slower. Yu & Dayan propose a computational framework whereby acetylcholine interacts with noradrenaline in a top-down manner, while large, bottom up signals from NE increase ACh levels. We hypothesize that uncertainty, and surprise specifically, is encoded by noradrenaline. As we cannot determine the direct relationship between the two in healthy humans without, at the least, introducing a pharmacological challenge, we propose to attempt to localize LC activation in fMRI as a proxy for noradrenergic activity. LC neuronal firing is closely coupled to noradrenergic release (Berridge & Abercrombie, 1999). While many regions are innervated by projections from the LC, such as the insula and the ACC, those large cortical regions have heterogeneous makeup and thus their activation cannot be said to reflect noradrenergic activity alone.

2.1.2 Pupil dilation as an index of LC activity

In non-human primates tonic pupil diameter is correlated with baseline LC activity (Rajkowski et al., 1993; Rajkowski et al., 1994). Single unit cell recordings in monkeys, show LC activity predicts a pupil response (Joshi et al., 2016). In humans, Murphy et al (2014) demonstrated a correlation between LC BOLD response and the pupil reflex (see below). Studies investigating the role of LC-NE in behavioral processes have used pupil dilation as an index of noradrenergic activity (see Nieuwenhuis et al., 2005 for an overview of earlier studies). In more recent studies, pupil dilation as a proxy of LC noradrenergic activity correlates with uncertainty and surprise specifically, as captured in a financial uncertainty task (Preuschoff et al., 2011).

In the perceptual domain, pupil dilation increases just prior to a reported perceptual switch for an ambiguous stimulus (Einhäuser et al., 2008), which suggests that perceptual uncertainty may also be modulated by the LC-NE system. The relationship between LC-NE activity and pupil response remains a putative one, due to the difficulty in imaging the LC in vivo. Thus, great care must be taken when attributing a cognitive process to one mediated by the LC-NE system via the pupillary response.

2.1.3 Imaging the LC

Establishing the relationship between pupil dilation and LC-NE activity in humans using functional imaging is problematic at best. The locus coeruleus is a very small bilateral region at the posterior border of the brainstem, which introduces cerebrospinal noise in its signal. The LC's small size and location has sparked controversy surrounding its identification (Minzenberg et al., 2008; Astafiev et al., 2010) in fMRI studies. The unique challenges to its structural and functional identification with MRI (Keren et al., 2015) has prompted the neuroimaging community to proceed with great care and caution when studying the LC. One recent study reports a correlation between the LC BOLD response, localized with the use of a probabilistic map (Keren et al., 2009) with pupil dilation in an oddball task (Murphy et al., 2014). The experiment examined two key questions relating to fMRI of the LC, and that is whether or not to apply a smoothing kernel to such a small set of voxels and whether or not to apply physiological correction to fMRI time-series. It has been hypothesized that the LC's location in the brainstem makes it particularly susceptible to cardiac and respiratory artifacts on the one hand. However, physiological correction may be challenging in relation to functional imaging of the region for two reasons: 1) noradrenergic activity in the central nervous system may be correlated with peripheral activity, meaning LC activity may be positively correlated with an increase in heart rate and respiration; 2) adding physiological regressors to a design matrix unnecessarily could wash out what is expected to be a small signal change in the LC, even though evidence suggests there are no significant physiological artifacts in brainstem imaging. The Murphy study (Murphy et al., 2014) examines LC activity with and without smoothing and physiological correction. Further, they employ a neuromelanin sensitive sequence (Sasaki et al., 2006) to structurally localize the LC within subjects. While this study was successful, the task employed was a very simple one while our global aim is to link latent variables linked to a high-level process. Further, the study trained its functional imaging sequence on the brainstem, whereas we aim to identify cortical structures in addition to the LC. Therefore, our first step in addressing our research questions was to determine how best to reliably capture LC-NE activity with the use of MRI.

In the following sections, we detail pilot studies related to the functional and structural localization of the LC in healthy humans using pupillometry and fMRI. In study 1, we attempt to indirectly localize the LC by administering a visual stimulus known to reliably elicit BOLD activity in neighboring brainstem nuclei. In study 2, we perform a pupillometry study to confirm the suitability of a localizer experiment, the auditory oddball task, to ultimately directly localize the LC using fMRI. In study 3, we administer this task during fMRI acquisition using three different sequences in a within-subject design to determine which one best captures a BOLD response in the LC.

2.2 Structural localization of the locus coeruleus

The first quandary we encounter in our objective to link noradrenaline to uncertainty signaling via locus coeruleus BOLD response is in structurally identifying any significant cluster that may arise in the region. Common neuroimaging atlases, such as AAL (Tzourio-Mazoyer et al., 2002) and the Neuromorphometrics atlas (Neuromorphometrics, Inc.) do not include a map of the LC or subregions of the brainstem. Indeed, the LC, while clearly visible on histological slices due to its pigmentation is not visible in T1, T2 or PD weighted images (Sasaki et al., 2006). A recent attempt has been made however to exploit the paramagnetic properties of iron-rich neuromelanin and thus provide a contrast for both the LC and the substantia nigra (Sasaki et al., 2006). The fast spin echo sequence has been dubbed neuromelanin-sensitive and its clear advantage is that it is a non-invasive means by which to examine LC and substantia nigra morphometric changes. Keren and colleagues (Keren et al., 2009) used this sequence to produce a probabilistic map of the LC (Figure 2.1). The group used a high resolution (0.4 mm by 0.4 mm in plane resolution) neuromelanin sensitive sequence on 44 subjects, ranging in ages from 19 to 79 years old and subsequently validated spatial coordinates with LC identification in post-mortem human brains, obtaining a correlation of $r = 0.90$. Based on this procedure, two probabilistic maps normalized to the MNI152 template image, were made available for general use, the first encompassing 1 standard deviation from the mean map and the second, two standard deviations from the mean map. As mentioned, the atlas is a probabilistic map; in the figure below, one can see that the right LC 'skips' a slice for instance, which is not reflective of an anatomical feature. Further, the process by which it was generated is conservative. This may cause difficulty especially in the transverse plane, where the LC spans, with the Keren atlas, 3 mm. Nonetheless, given that no neuromelanin sensitive sequence was available for our use and given the great care with which the Keren atlas map, we decided, as have others (Murphy et al. 2014; Naegeli et al., 2018) to employ the 2 standard deviation map to confirm LC activity in our subsequent studies. Specifically, we will train our search on peak voxels identified in the Keren study as being shared by 95% of subjects at the following coordinates: -4.7, -37.3, -27 (left LC); and 5.8, -37.2, 27 (right LC). The range of LC voxels found in the study are as follows (in MNI coordinates) : for left LC, $x = -2.5$: -7; $y = -36$: -39; $z = -18$: -33; for right LC, $x = 4$: 8; $y = -36$: -39; $z = -18$: -33. The range in the y direction is especially narrow; thus we will scrutinize voxels falling within an additional error margin of ± 4 mm.

2.3 Pilot Study 1: Indirect LC Localization via Opto-Kinetic Nystagmus

In a first instance, we sought to verify that small brain stem nuclei could be localized and activated at all using fMRI. As the link between uncertainty and LC is a putative one, imaging the LC itself would be a poor proof of concept that would eventually turn into a circular argument. We thus turned to the much better established link between eye movements,

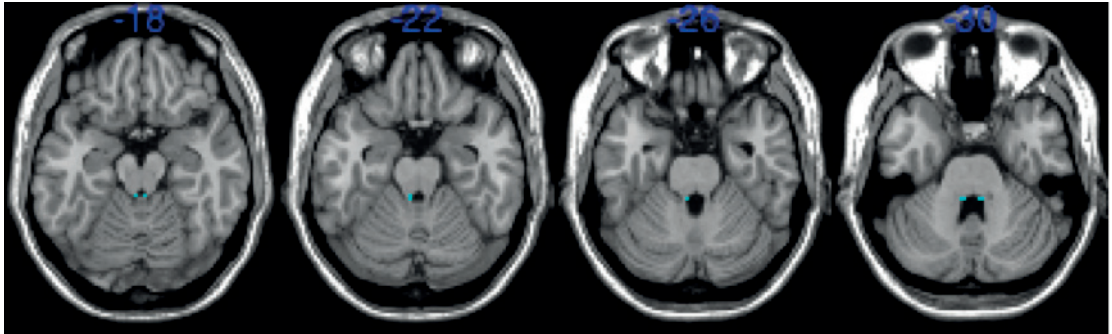


Figure 2.1 – Probabilistic LC atlas map (Keren et al., 2009). The map in blue is overlaid on transverse slices of the canonical T1 image. At -26, the right LC "disappears", underscoring the probabilistic nature of the map. Overall, this map illustrated the very limited spatial extent of the region, highlighting the difficulty of assigning significant BOLD responses to the LC.

specifically the optokinetic nystagmus (OKN) reflex and brainstem nuclei. OKN refers to the involuntary periodic saccades healthy individuals make when viewing what appears to be moving lines. For instance, when looking at a moving train, our eyes follow the train in its forward translation before returning to a point of origin, where the saccade is repeated again (horizontal OKN). A vertical OKN (vOKN) can also be induced when viewing a stimulus moving in an upward or downward motion. Animal studies have identified an optokinetic complex dubbed the cortico-pretectopontine-olivo-cerebellar system (Simpson et al., 1998), and linked brainstem vestibular nuclei to the OKN (Mazzena et al., 1974; Buettner & Buttner, 1979). OKN is thought to implicate the oculomotor nucleus (superior); dorsolateral pontine nucleus (ventral); and tegmental pontine reticular nucleus (ventral) (Mustari et al., 1994). We tried to replicate findings by Bense et al. (2006) that identified active brainstem nuclei adjacent to the LC, in response to an OKN induction. This fMRI study investigated different BOLD responses in hOKN and vOKN to determine if a pattern similar to that found in non-human animals (Hoffmann et al., 1988) could be identified in humans. Bense and colleagues found clusters in the dorsal medullary and pontine areas of the brainstem while vOKN correlated with clusters in the dorsal ponto-mesencephalic region of the brainstem. While the OKN is not of direct interest to our research question, the paradigm provided a means by which we could assess the feasibility of 1) imaging small brainstem nuclei and 2) potentially verify the accuracy of the functional mapping of LC later on..

2.3.1 OKN Experimental Procedure

We recruited 11 subjects for the experiment (4F; average age 24.33 years). In a first instance, following intake, subjects were fitted with cardiac and respiratory sensors in the scanner bore (BioPac, BioPac Systems Inc, Goleta, CA). We then proceeded to acquire a localizer scan (voxel size 0.5 x 0.5 x 7 mm; 1 slice; FOV 250 mm; TR = 8.6 ms; TE = 4 ms, FA = 20 °), followed by a gre-field mapping scan (voxel size 3 x 3x 3 mm; 64 slices; FOV 192 mm; TR = 1020 ms; TE = 10 ms; 2mm slice thickness; FA 90°), to correct for magnetic field inhomogeneity. We then

acquired whole brain functional data during the induction of an optokinetic nystagmus (OKN) for 11 subjects. We used a standard fMRI sequence with the following parameters: 3.5 mm resolution; TR= 2.46 s; 2D EPI; matrix size 64 x 64; 39 slices; 117 volumes (FOV = 224 mm; slice TE = 63.1 ms; FA = 90 °). Participants were instructed to passively view the stimuli presented. The stimuli were comprised of sinusoidal gratings, moving grids of black and white stripes, in either a vertical (VOKN) or horizontal (HOKN) direction. Stimuli were presented for a block of 32 seconds in duration, followed by a rest condition of 32 seconds. OKN stimuli were projected onto a mirror in the scanner bore. Each OKN inducing stimulus was presented 3 times, totaling 9 trial blocks per experiment. Following EPI acquisition, we acquired individual anatomical T1 volumes (MPRAGE, voxel size= 1 x 1 x 1 mm; FOV = 256 mm; (slice) TR = 2000 ms; TE = 2.39 ms; FA = 9 ° ; matrix size 256 x 256).

2.3.2 OKN Imaging Results

To establish whether there was any brainstem activation related to OKN similar to that reported by (Bense et al., 2006), we addressed two particular points of discussion with respect to fMRI of brainstem regions concerning physiological correction and smoothing. Our preprocessing pipeline included: generation of voxel displacement maps (VDM); applying VDM to functional scans before warping and realigning to subject's mean functional image; bias field correction; co-registration to individual subject anatomical scan; and segmentation (6 tissue probability maps) and normalization to the MNI152 template. We then analyzed group data unsmoothed; smoothed to 3 mm (just under the width of the Keren LC map in the transverse plane); and smoothed to 8 mm (a more standard Gaussian filter). Generally speaking, it is suggested that when looking for a response in a small region, one uses a small smoothing filter. Of course, a small smoothing filter in turn skirts the edge of what is required for random field theory (Friston et al., 1995). In practice, the choice of smoothing kernels is not well defined for smaller regions. Murphy and colleagues did however address the smoothing versus non-smoothing question with regards to the LC and we opt to take this course as well. We further address another point of contention with regards to brainstem fMRI and that is whether or not to correct for physiological noise. Our intuition is again, in line with Murphy et al.'s concerns, in that LC response may be correlated to cardiac and respiratory rates. Nonetheless, we apply retrospective physiological correction (RETROICOR, Glover et al., 2000) to our data, to compare corrected versus uncorrected response signals in the brainstem. The RETROICOR suite of functions adds 14 nuisance regressors to the design matrix of an analysis, in addition to the 6 standard nuisance regressors related to head movement. All in all, we obtain 12 group level data sets: hOKN and vOKN unsmoothed; smoothed at 3 mm FWHM; smoothed at 8 mm FWHM; with and without physiological correction.

2.3.3 OKN imaging Data for horizontal and vertical OKN, Group Level

A second level analysis was performed on 11 subjects to investigate brainstem activation in response to horizontal and vertical OKN stimulation. Due to the study's small sample size

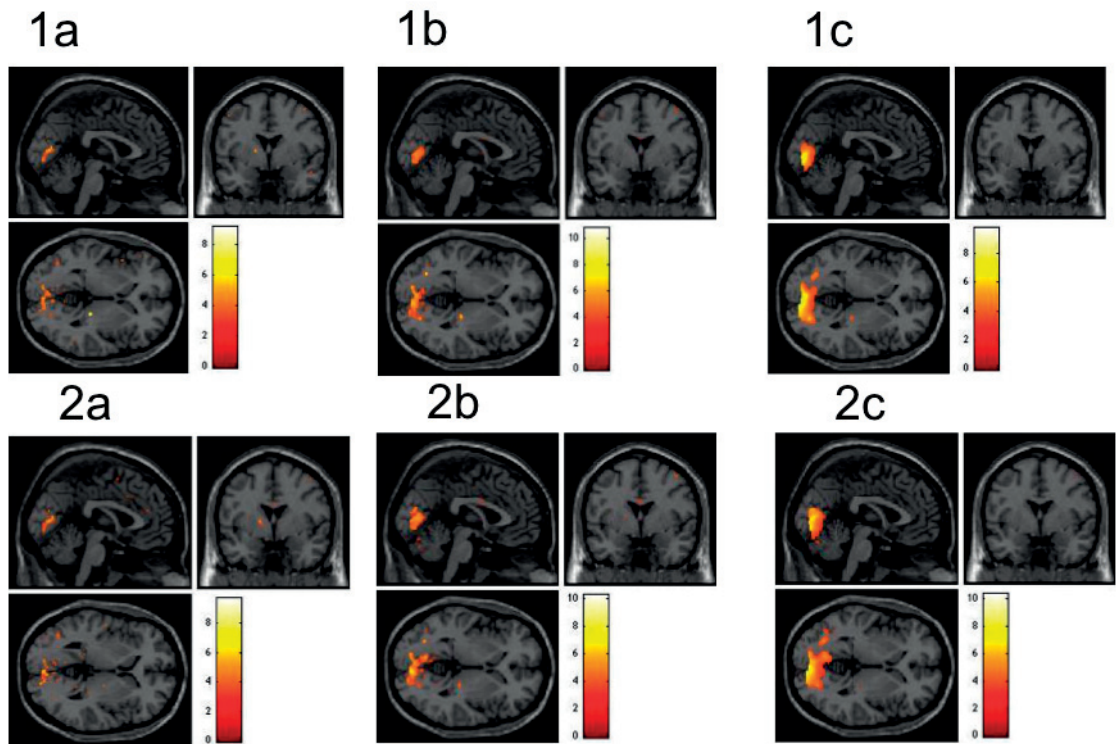


Figure 2.2 – Statistical map of horizontal OKN in 11 subjects ($p=0.001$, unc.). The top row (1) shows statistical maps without physiological correction. The bottom row (2) shows physiologically corrected statistical maps. Column one shows unsmoothed maps; column 2 shows data smoothed to 3 mm FWHM; column 3 shows columns smoothed to 8 mm FWHM. Significant clusters in the occipital lobe are seen in all six statistical maps; none in the brainstem.

and our expectation of a weak response in the brainstem, we apply a low threshold of $p = 0.001$, uncorrected to our data. All contrasts for all smoothing kernels as well as physiological corrected and uncorrected data, yielded reliable activation in the visual cortex (bilateral occipital fusiform gyri). Group level analysis did not yield activation in the brainstem. Below, we describe results for hOKN and vOKN output from different pre-processing conditions.

Results for hOKN show increased activation with higher smoothing filters as expected. Further, higher activation is seen in physiologically corrected data (see Figure 2.2). Yet no brainstem activity appears in our analyses above.

Like hOKN, statistical maps for vOKN in all pre-processing conditions show significant clusters in the occipital lobe. Signal strength is increased with smoothed data (columns 2 and 3 in Figure 2.3). Interestingly, a cingulate response appears prominently in the 8 mm FWHM condition. On a more relevant note with regards to our question of interest, we see a small brainstem cluster in physiologically uncorrected data for images smoothed to 3 and 8 mm.

A close up view of these significant voxels in the brainstem/cerebellum appear admittedly

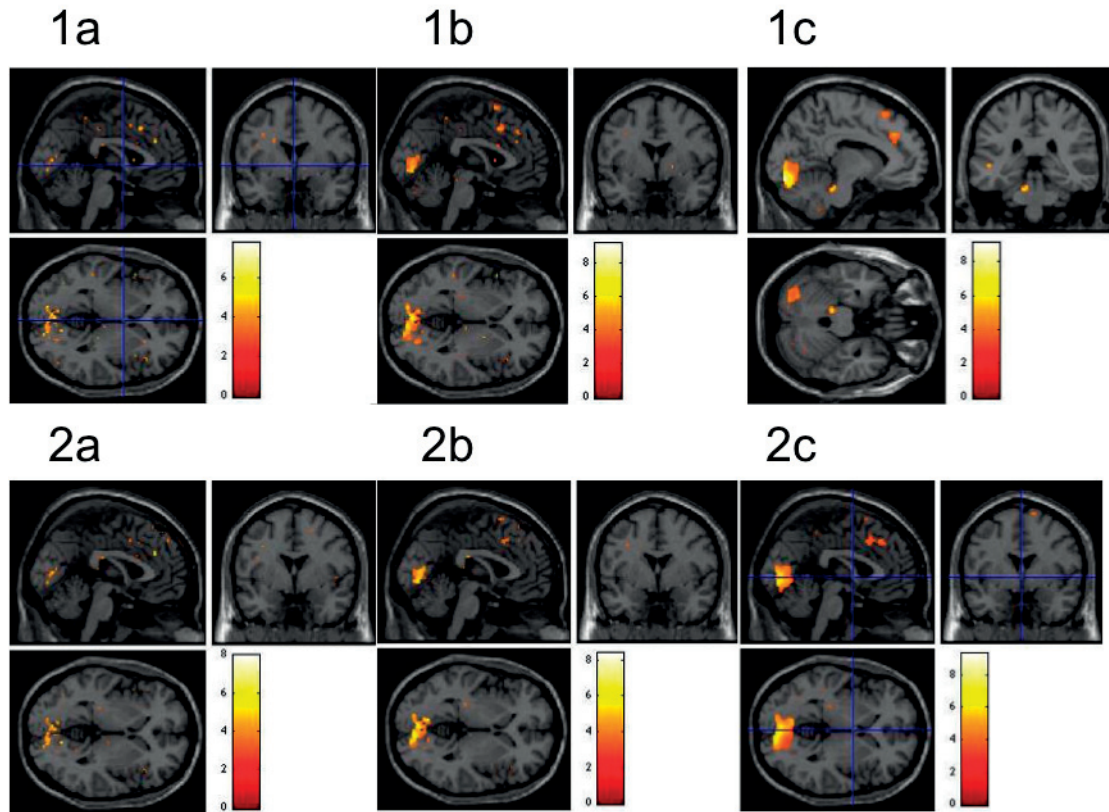


Figure 2.3 – Statistical map of vertical OKN in 11 subjects ($p=0.001$, unc). The top row (1) shows statistical maps without physiological correction. The bottom row (2) shows physiologically corrected statistical maps. Column one shows unsmoothed maps; column 2 shows data smoothed to 3 mm FWHM; column 3 shows columns smoothed to 8 mm FWHM. Significant clusters in the occipital lobe are seen in all six statistical maps as well as in the cingulate, especially in column 3; a small cluster in the cerebellum/brainstem border is noted for data smoothed to 3 and 8 mm and uncorrected for physiological regressors.

noisy for maps smoothed to 3 mm (Figure 2.4, panel A). However, panel B shows what appears to be a reliable cluster, in a clean brainstem. No stray voxels appear but for this cluster on the edge of the left cerebellar peduncle, lateralized relative to where we would expect the left LC. Of course, it may be indeed an artifact of pulsation, as this cluster does not survive physiological correction.

2.3.4 Pilot Study 1 - Summary

The study above suggests that indirect localization of the LC by inducing an OKN is not feasible. In a first instance, our group results yielded only one significant cluster in the brainstem, for data smoothed to 8 mm. This cluster does not survive physiological correction; given the cluster's location at the boundary of the 4th ventricle, this disappearance may indeed signal

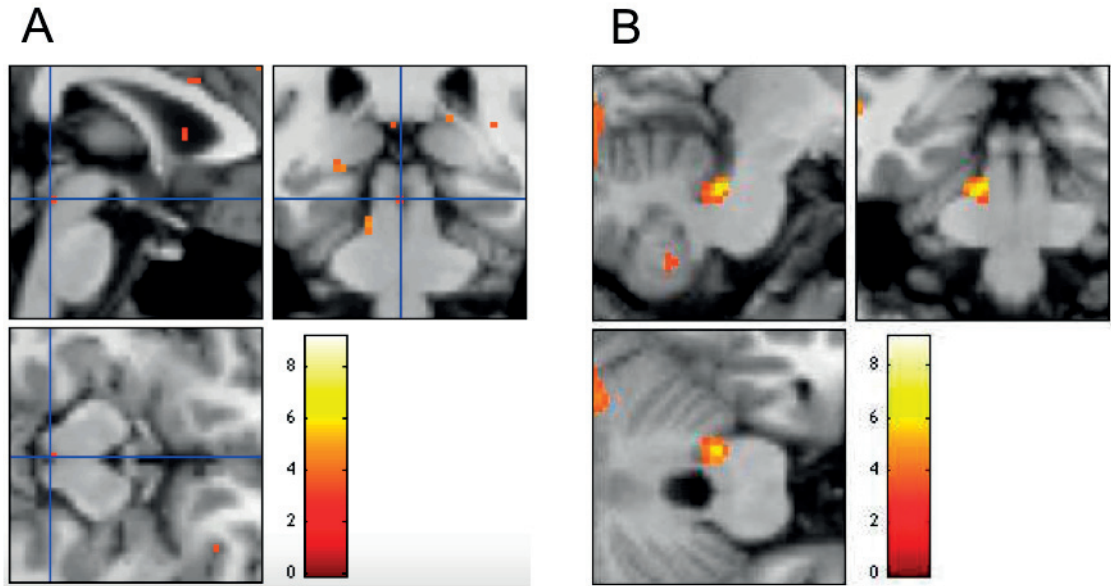


Figure 2.4 – A close-up view of the brainstem in statistical maps of vOKN smoothed to 3 mm and 8 mm FWHM ($p = 0.005$, unc.) The data above were not subject to physiological correction. Panel A crosshairs at (1; -31; -13); Panel B crosshairs at (-9; -32; -26).

its identity as an artifact. What is clear from our data is that we do not replicate the findings in the Bense paper (Bense et al., 2006). Thus brainstem imaging is fraught even when using basic sensory stimuli such as moving gratings. In a second instance, the original paper investigating OKN-induced brainstem activity concedes that clusters found encompass several nuclei, which, in the case of vOKN includes the LC, though it is not limited to it. Given the difficulty in attributing a cluster to a specific brainstem nucleus, we must conclude that an indirect localization of the LC is not a practical endeavor. Third, we probed two questions that emerge in pre-processing fMRI for small, brainstem regions, concerning 1) smoothing; 2) physiological correction. With regards to smoothing, statistical maps in unsmoothed data appear too noisy to be reliable. Further, BOLD response is, in hOKN, as expected, higher in smoothed relative to unsmoothed data. The question then becomes which smoothing kernel to use? The 3 mm kernel also yields questionably noisy data. Indeed, in vOKN, stray voxels are seen in the brainstem, which may simply be noise. As for physiological correction, our only significant cluster in the brainstem is washed away with the inclusion of the 14 cardiac and respiration regressors. On the one hand this suggests the correction did what it was meant to do; on the other hand, as LC activity is correlated with autonomic function, we may not want such a "correction". Therefore, we will continue to explore these two pre-processing questions in our next pilot study. Finally, we employed a standard EPI sequence in this study. However, in consideration of how difficult capturing a cognitively-driven BOLD response in the LC is likely to be, we may benefit from more tailored sequences. Our next study will thus probe the sensitivity of alternative sequences on brainstem BOLD response. In sum, given the difficulty in adequately localizing any nucleus in the brainstem, as well as the lack of consistency across

subjects, we opted to abandon indirect functional localization in favor of direct functional localization.

2.4 Pilot Study 2: Pupil Dilation in The auditory oddball paradigm

Given our conclusions on Pilot study 1, we set about finding a task that would directly stimulate LC activity. The auditory oddball task emerged as a reliable candidate for our search (Nieuwenhuis et al., 2005). The oddball paradigm is the simplest of detection tasks. A deviant stimulus (the oddball) is introduced in a stream of standard stimuli (standards). Previous research has shown that it elicits the P300 event related potential in EEG studies, as well as pupil dilation. It has also been found to elicit LC activity (Murphy et al., 2014). Most importantly, under our (and other) models of uncertainty and surprise that we employ in financial gambling tasks, the deviant stimuli generate large signals, i.e., the oddball task likely poses an upper limit to the statistical power of our gambling task. Indeed the oddball has been used to elicit a very basic form of surprise (Timmerman et al., 2017). If we cannot find strong BOLD responses with this task then the more subtle effects that we expect in the card game (Chapter 3) are unlikely to emerge. Thus we first administered an auditory version of the task on a small group of subjects outside of the scanner to determine if indeed pupil dilation in relation to the oddball would be found significant relative to standard tones before administering the task inside the scanner.

2.4.1 Pitch versus duration deviance

While previous studies have found both pupil dilation and LC activity correlating with oddball stimuli, different deviant features can confer oddball status to a stimulus. Studies examining mismatch negativity (MMN) event-related potentials in comatose patients employ the auditory oddball paradigm (Juan et al., 2016). The types of deviants used in the study include: 1) pitch; 2) duration; 3) location. According to these studies, duration deviants induce the strongest MMN signal. While EEG studies cannot map signals to brainstem regions and fMRI cannot identify the mismatch negativity signal (150 ms post-stimulus), due to poor temporal resolution, we hypothesized that a stronger mismatch signal would elicit stronger pupil dilation and thus more significant LC activity.

2.4.2 Experimental procedure

The task was coded in Matlab using the PsychophysicsToolBox (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007). Pupil data were recorded with a monocular configuration (left eye) on an Eyelink 1000 eyetracker (SR Research, Ottawa, Ontario) sampled at 1000 Hz. Subjects were seated facing a screen and a keyboard, with the screen placed 30 cm away from the eyetracker head mount. Subjects placed their chins on the chin rest and rested their foreheads against the headmount. Following a 9 point calibration, the task began. Subjects were asked to press a

2.5. Pilot Study 3: fMRI Sequence Comparison Study

key on a keyboard for standard tones, and another key for oddball tones while viewing a black fixation cross on a grey screen in a darkened room. Sounds were played over speakers in a quiet room. All experiments were performed on campus at the Ecole Polytechnique Fédérale de Lausanne.

Table 2.1 – Oddball Task Parameters

| Task | Trials | ISI | Odd Frequency | Deviance | Standard Pitch | Odd Pitch | Standard Duration | Odd Duration |
|--------|--------|------|------------------|----------|-------------------|--------------|----------------------|-----------------|
| Task 1 | 500 | 1800 | 20% | Duration | 1000 Hz | 1000 Hz | 100 ms | 175 ms |
| Task 2 | 500 | 1800 | 20% | Pitch | 1000 Hz | 1500 Hz | 100 ms | 100 ms |

2.4.3 Pupil Trace Preprocessing

Pupil data were analyzed as follows. Blinks were first removed from pupil data via cubic spline interpolation. The latter were flagged as beginning with a faux-saccade start message preceding the blink start message and ending with a faux-saccade end message following the blink end message. Time-series were partitioned into trials then separated into odd and standard trials. Time-series for each condition were then averaged across trials. Baseline measures were taken as being the last 100 ms of the preceding trial. Individual time-series averages were then meaned across subjects to yield a grand mean.

The traces above show a reliable difference in pupil dilation for oddball versus standard tones (Figure 2.5). Indeed the traces are similar between task conditions. We examine significant differences in pupil dilation for both task conditions to determine if one yields more differences than the other (Figure 2.6).

2.4.4 Auditory Oddball Pupillometry Study: Results and Conclusion

Time-courses of the pupil trace were similar for both deviants (Figure 2.5). However, the rates of statistically significant tests are higher for the duration oddball (341 significantly different time-points in the long deviant condition versus 219 in the high pitch condition). In spite of these statistical tests, participants reported greater subjective ease in identifying the pitch deviant in the experiment. Given the additional noise participants are subjected to in an MRI scanner, we opted to err on the side of caution with respect to these subjective reports and employ a pitch deviant version of the oddball in our fMRI LC localization task.

2.5 Pilot Study 3: fMRI Sequence Comparison Study

We set out to test the oddball paradigm, which has previously yielded promising results in terms of LC-NE driven pupil dilation and LC localization in fMRI studies. The aim of this pilot

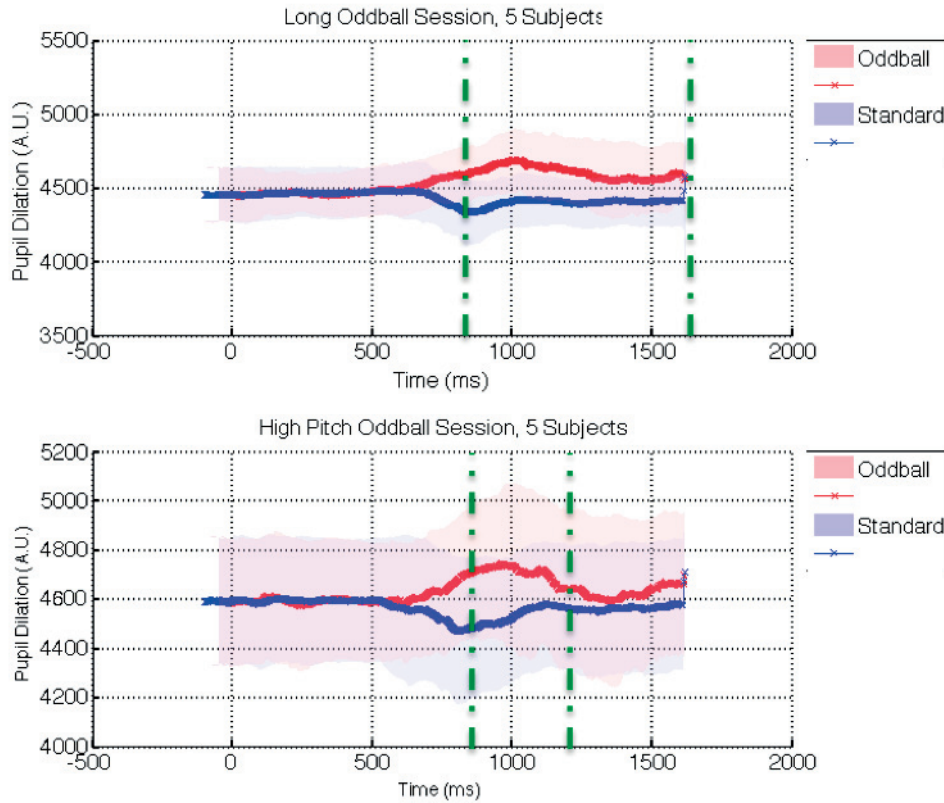


Figure 2.5 – Grand mean average time-courses for the two oddball tasks. Red curves indicate oddball trials while blue curved denote standard trials. The top panel shows time-courses for the duration deviant while the bottom panel represents pitch deviants. Significant differences in time-courses are found in the time window bound by the dashed green lines (starting at 800 ms following the start of the sound).

was to localize LC functional activity with fMRI and, as we expect this endeavor to be difficult, we sought to test different fMRI sequences in the same subjects performing the same task, to determine if one sequence would offer a better signal to noise ratio (SNR) overall in identifying LC activity. The known trade-offs in fMRI are temporal resolution, spatial resolution, spatial extent (which is linked to temporal resolution) and SNR. Thus, while it may appear that simply increasing the spatial resolution of our sequence would aid us in identifying the LC, we would incur a significant cost in SNR, as well as time. To determine what the best sequence would be to perform our experiments on uncertainty, we ran a pilot imaging study to test three different sequences. . Specifically, we used 1) the same standard sequences employed in the OKN experiment; 2) a high spatial resolution sequence; 3) and a multi-echo sequence. The first sequence, as we saw above, did not yield promising data for BOLD response in the brainstem; nonetheless, it is a standard sequence and thus provides a 'control' condition for our sequence comparison study. An obvious attempt to mitigate the difficulty in imaging the LC because of its size is to increase spatial resolution. Of course, the problem with such a

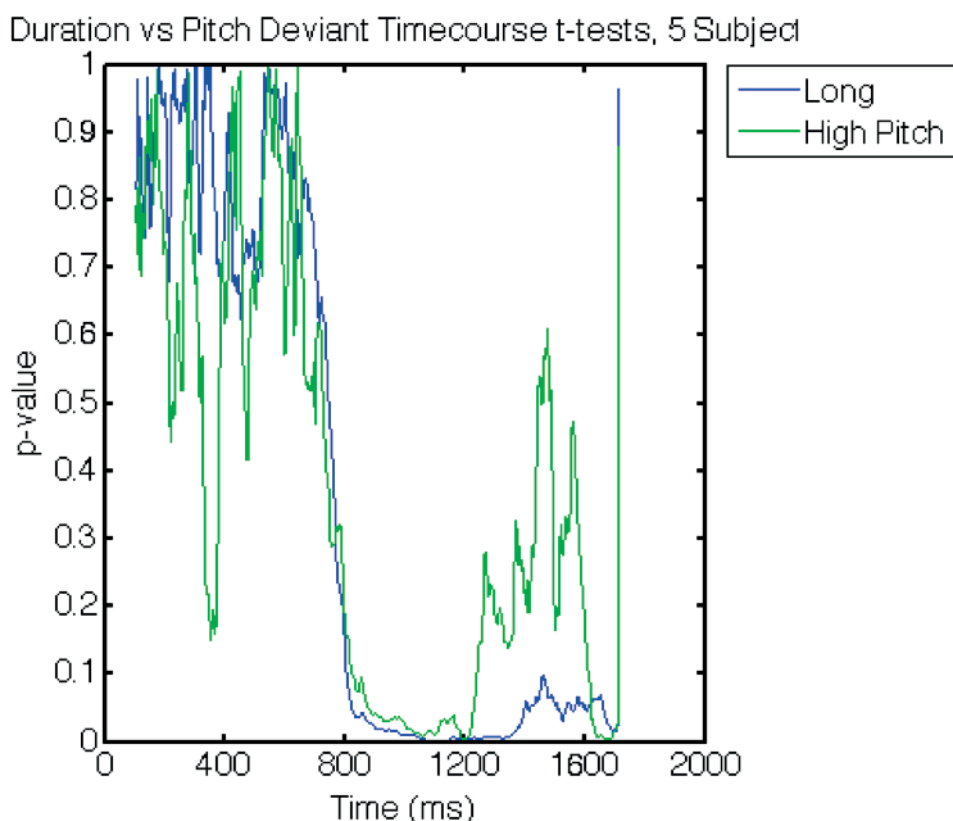


Figure 2.6 – Representation of t-tests between standard and oddball tones for each task (duration deviant in blue, pitch deviant in green). P-values are on the y axis while trial time is on the x-axis. Significant tests appear at roughly 800 ms following the start of the sound. Oddball effects appear to resolve faster for pitch deviants (1200 ms) while duration deviants continue to the end of the trial (1600 ms).

straightforward course is that high spatial resolution comes with an increase in repetition time, which complicates eventual task design and more importantly lowers the SNR. Indeed, such a high resolution sequence could not easily be accomplished for a whole brain sequence without a concomitantly impractical length of time on task. The multi-echo sequence presents an interesting option however. Multi-echo sequences, as the name implies, include more than one echo time (TE) per repetition time (TR); thus several volumes are acquired per effective volume, resulting in an increase in data, and thus SNR. The net effect of a multi-echo sequence is a brain-wide increase in SNR with repetition times and spatial resolution similar to the standard sequence (Poser et al., 2006).

2.5.1 Sequence Comparison Study: Experimental Procedure

Fourteen subjects were recruited from the Universit'e de Lausanne and EPFL campus (Mean age: 20.73; 6 F). Participants provided informed consent and were screened for MR compatibil-

Chapter 2. In Search of the LC: The needle in the haystack

ity before entering the MR room. In the bore, subjects were fitted with cardiac and respiratory sensors (BioPac, Systems Inc., Goleta, CA). Once settled in the bore, we performed a 3-point eyetracking calibration (Eyelink Eyetracker 1000 S). We chose a 3 point calibration because the bottom of the "screen" could not be detected by the camera (monocular setup, right eye). Instructions were provided on a screen, projected into the bore through a mirror. Sounds were embedded within the instructions. Sound was delivered via MR compatible headphones. Participants were instructed to provide a button press for one kind of sound (the 1000 Hz) and another button press for the other sound (1500 Hz tone).

All scans were acquired on a Siemens 3T Prisma, at the Centre Hospitalier Universitaire Vaudois in Lausanne, Switzerland. Subjects performed 3 blocks of the task, with one sequence assigned to each, in a counterbalanced fashion across subjects, such that each subject was scanned with all three sequences. Sequences used included a standard sequence; a high resolution sequence; and a multi-echo sequence.

- Sequence 1: 'Multi-Echo'; voxel size = 3 x 3 x 2.5 mm; 2DEPI; Multi-Echo sequence; TR = 1.92 s; 34 slices; FOV 192 mm; (slice) TR = 80 ms; (slice) TE (1) = 17.4 ms; FA = 90 ° ; matrix 64 x 64; 417 volumes (800 seconds), no dummies.
- Sequence 2: 'High Resolution'; voxel size = 1.5 x 1.5 x 1.5 mm, TR = 3.528 s, 48 slices; FOV = 192 mm; (slice) TR = 63 ms; slice (TE) = 30.8 ms; FA = 15 ° ; matrix = 128 x 128; 227 volumes (800 seconds), no dummies.
- Sequence 3: 'Standard'; 3.5 mm, TR = 1.262 s, 20 slices; FOV = 224 mm; (slice) TR = 63.1 ms; (slice) TE = 30 ms; FA = 90 ° ; matrix 64 x 64; 634 volumes (800 seconds), 5 dummies.

Because the high-resolution sequence did not cover the whole brain, all sequences above were of limited spatial extent, so as to more adequately compare LC activity across sequences (See Figure 2.7). The bounding box for the sequences was restricted to the lower portion of the brain encompassing the brainstem (Figure 2.7). Task parameters were as follows: 450 trials, 15% oddball, 1.6 s ISI, 1500 Hz Odd Tone, 1000 Hz Standard tone, 100 ms sound duration. Scans were preprocessed and analyzed using SPM 12. Our preprocessing pipeline included: generation of voxel displacement maps (VDM); applying VDM to functional scans before warping and realigning to subject's mean functional image; bias field correction; co-registration to individual subject anatomical scan; and segmentation (6 tissue probability maps) and normalization to the MNI152 template. We performed analyses on both smoothed (8 mm FWHM, 3 mm FWHM) and unsmoothed images. On the one hand, smoothing increases signal-to-noise ratio (Friston, 1995); on the other hand, we considered that activity in the LC might be lost when a smoothing kernel is applied. Thus we examined both.

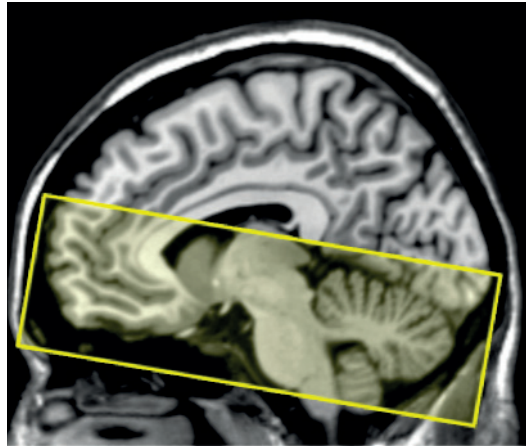


Figure 2.7 – Representation of bounding box used for the sequence comparison study. All three sequences were limited in their spatial extent, so as to better compare responses in the brainstem region across sessions. Brain coverage was thus not guided by centering on the AC-PC line; rather, we took care to include as much as the brainstem and thalamus as possible.

2.5.2 Sequence Comparison Study: Group Level Results

In the following analyses, we used a statistical threshold of $p=0.001$ uncorrected. We first determined whether or not any voxels in the area of the brainstem known to include the LC survived thresholding before applying small volume correction analyses using the Keren LC atlas (Keren et al., 2009).

Sequence 1

For the standard sequence at a 3.5mm resolution ('Standard'), no significant voxels survive in the brainstem in images smoothed to 3 mm and 8 mm FWHM and, both uncorrected and corrected for physiological artifacts (Figure 2.8).

We conclude therefore that this standard sequence should not be used in our future studies.

Sequence 2

In the high-resolution sequence, we find voxels adjacent to the edge of the LC atlas. in images smoothed to 3 and 8 mm FWHM, corrected and uncorrected for physiological artifacts ($p=0.005$ uncorrected) (Figure 2.9).

The high-resolution sequence yields promising results with significant voxels found just outside the boundary of the LC map (Figure 2.10). As in the OKN study, smoothing appears to yield more reliable statistical maps, with data visibly less subject to noise (Figure 2.9). Physiological correction does not remove cerebellar/brainstem activation, nor does it impact signal intensity. While results are significantly improved on the standard sequence, we must

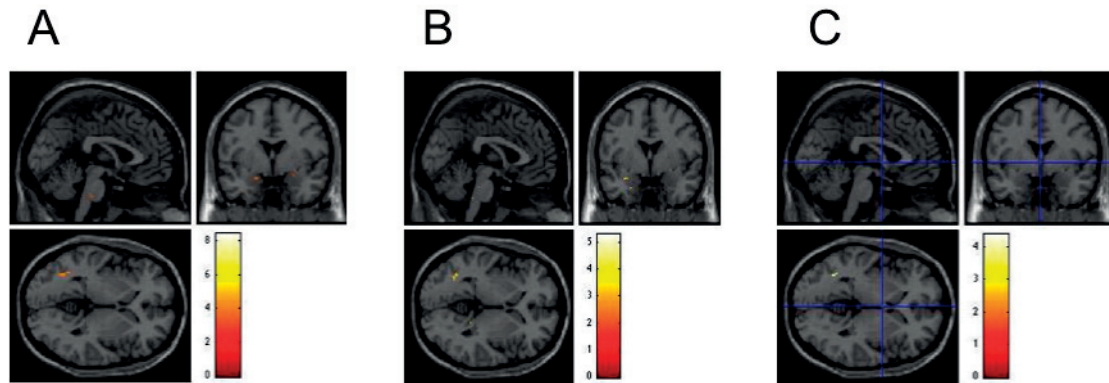


Figure 2.8 – Statistical maps for Oddball > Standard tones. All maps overlaid on the canonical T1 image were thresholded at $p = 0.005$, uncorrected. Panel A shows results smoothed to 3 mm FWHM, uncorrected for heartbeat and respiration; Panel B shows results smoothed to 3 mm FWHM, physiologically corrected; Panel C shows results smoothed to 8 mm FWHM, physiologically corrected.

concede that they remain imperfect. Of greater concern is the eventual difficulty of developing a whole brain sequence using high-resolution parameters.

Sequence 3

In the multi-echo sequence ('Multi-echo'), we find a small cluster of voxels at $(-2, -44, -22)$, again in the vicinity of the LC atlas but that are not included in the map. These results were found in physiologically corrected images only (both at 3 and 8 mm FWHM).

The multi-echo sequence yields results in the brainstem for physiologically corrected data in data smoothed to 3 and 8 mm FWHM (Figure 2.11). As in the high-resolution sequence, the significant voxels found lie at the border of the LC atlas, which is a conservative map (Figure 2.12). As expected, only data smoothed with a large filter yields visually compelling results. The 3 mm images simply appear noisy. Tellingly too, is that the voxels adjacent to the LC are not the same ones found in the high-resolution sequence. Nonetheless, the important takeaway from these results is that the multi-echo sequence can detect a BOLD response in the brainstem, while the standard sequence could not.

While the method Keren and colleagues used to characterize the LC is reliable, namely through neuro-melanin sensitive MR on a large cohort and validated against post-mortem brains, it is worthwhile to note that the map as applied to a canonical T1 image, over which statistical maps are overlaid, sees a portion of the atlas covering the ventricle. Given the narrowness of the probabilistic LC map in the transverse direction, moving forward, we decide to focus on clusters namely surrounding the coordinates $(-5, -37, -27)$ and $(6, -37, -27)$ in the left and right hemispheres respectively, the coordinates for which 95% of participants in the Keren et al study show peak voxels. Nonetheless, even if significant voxels in the LC vicinity are found,

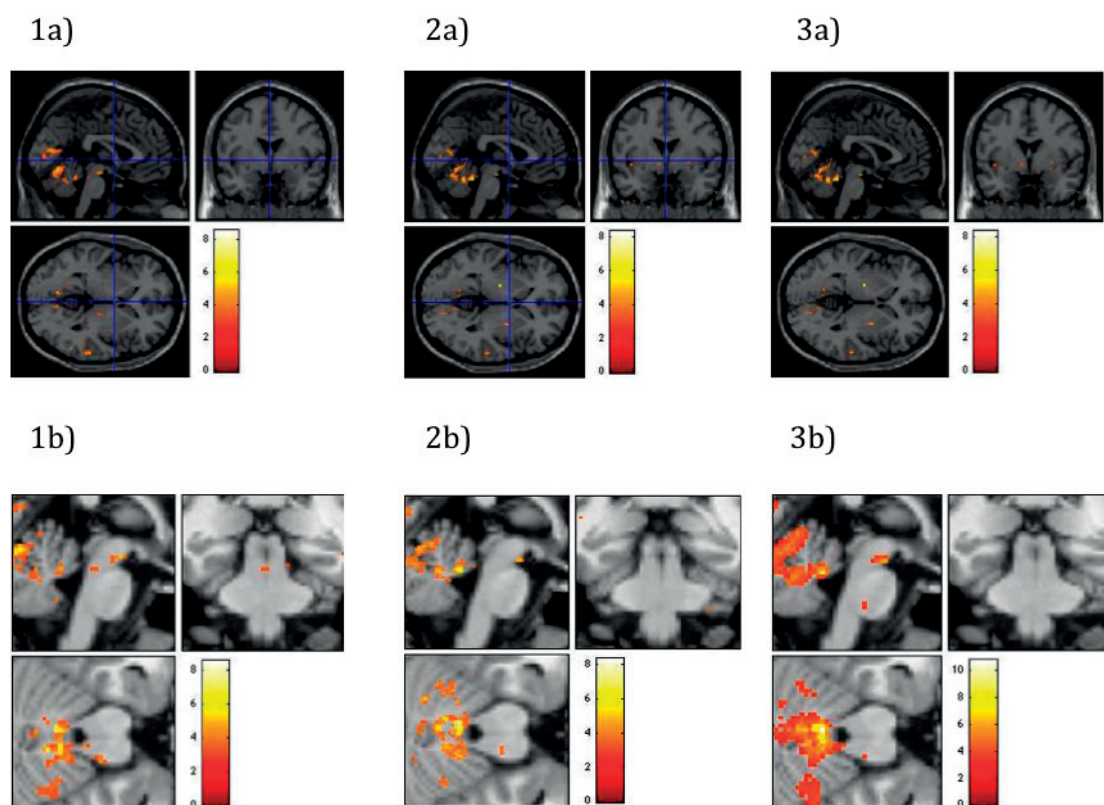


Figure 2.9 – Contrast for Oddball > Standard tones ($p=0.005$, uncorrected). The top row shows whole brain statistical maps (3mm FWHM, no physiological correction; 3 mm FWHM, with physiological correction; 6 mm FWHM, with physiological correction). The bottom row shows the same contrast for the same preprocessing steps as those in row 1 but zoomed in on the following voxels (0 -31 -22). Significant voxels appear in the vicinity of the LC, notably in the cerebral peduncle in all 3 images.

we will also examine the rest of the brainstem to determine if other nuclei appear active, as in the Bense et al. study (Bense et al., 2006).

2.5.3 In-scanner Pupillometry Results

We performed the same analyses on pupil data acquired in the scanner as we did on pupil data acquired outside the scanner. Out of 13 subjects, 12 had usable pupil recordings in sessions 1 and 2; and 10 were included in analyses for session 3. While the data acquired in the scanner is noisier than that collected outside of the scanner, we still see a significant difference in oddball versus standard tone pupil trace in the same time window as in the behavioral study, namely between 600 and 800 ms (Figure 2.13). This difference is dampened however in session 3, which is perhaps indicative of a vigilance decrement.

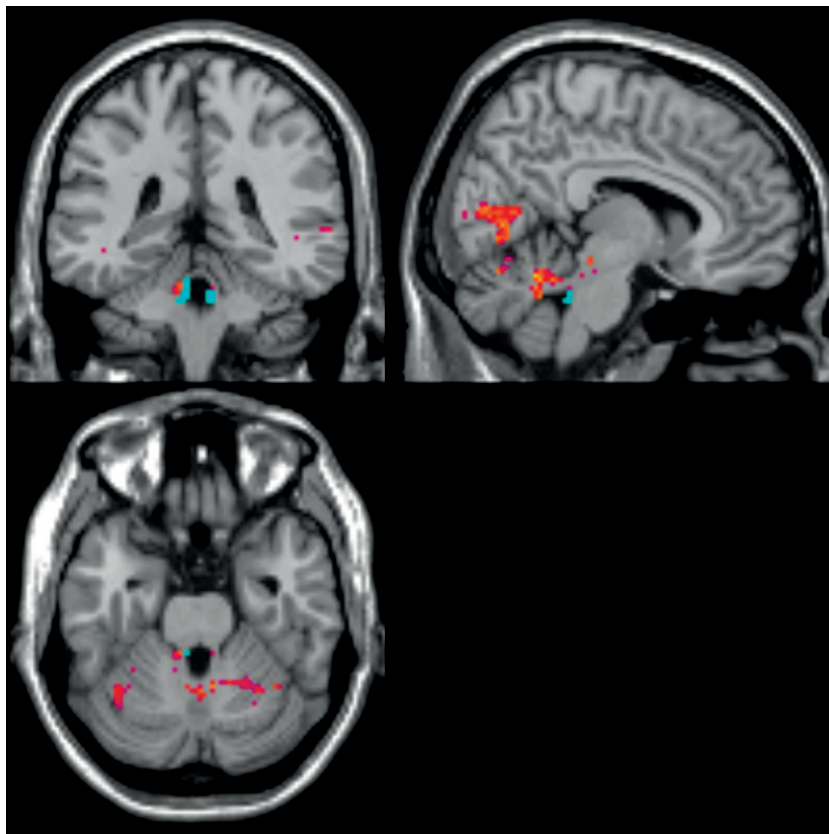


Figure 2.10 – Statistical maps of Oddball > Standard tones (3 mm with physiological correction) in the high-resolution condition ($p=0.005$, unc.) in red and fuchsia overlaid on the T1 canonical image. The LC atlas is also overlaid (in light turquoise). Significant voxels are seen in all three orientations, bilaterally, just adjacent to the LC map.

2.5.4 Sequence Comparison Study: Conclusion

The fMRI results above informed our choice of EPI sequence for experimental studies investigating uncertainty. Of the 3 sequences, we selected the multi-echo sequence because it captured BOLD responses in the brainstem and also because it could be reasonably tailored for a close-to-whole brain analysis, which allows us to investigate other regions of interest in our study. The high-resolution sequence on the other hand, given its long repetition time (TR), would prove problematic in the whole-brain domain, even though it yielded the most promising results with regards to LC localization. We opt to retain the smoothing process in our subsequent analyses as well, given our aim to examine cortical regions but also in light of the brainstem results found above. Data smoothed to 3mm data yields scattered voxels across the region, which may indeed reflect noise. Physiological data correction may improve results as well though we opt not to employ it in light of our questions of interest regarding surprise and sympathetic system engagement.

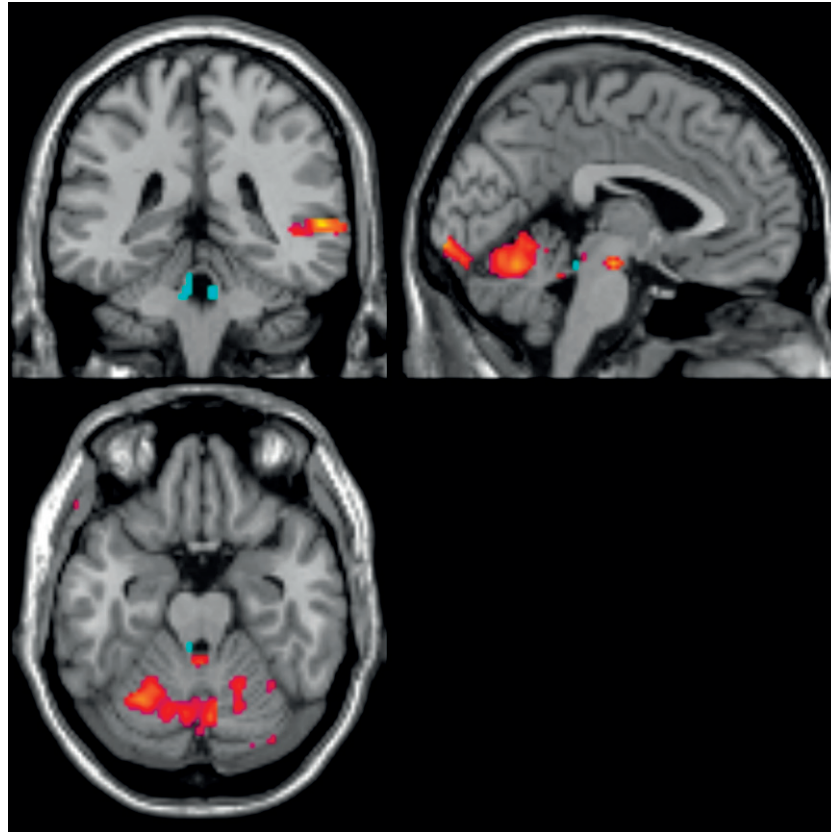


Figure 2.11 – Statistical maps of Oddball > Standard tones (8 mm with physiological correction) in the high-resolution condition ($p=0.005$, unc.) in red and fuschia overlaid on the T1 canonical image. The LC atlas is also overlaid (in light turquoise). Significant voxels are seen just adjacent to the LC map notably in the sagittal orientation.

2.6 Summary of results

Taken together, results of our three pilot experiments provide guidance in how to proceed experimentally in our study. First, with regards to structural localization in future studies, we opt to link any brainstem activation to a probabilistic LC atlas (Keren et al., 2009). Second, based on results above, we determine that indirect localization of the LC via OKN stimulation is not feasible. Third, in light of the LC's role in autonomic function, we must consider that applying physiological correction to data may remove relevant variance and thus compromise results. Thus, we will not apply physiological correction later on in our pre-processing pipeline. Fourth, we replicate findings that the auditory oddball paradigm reliably elicits pupil dilation, within and without the scanner. Fifth we determine that pupil data acquired in the bore is far noisier and more difficult to process than pupil data acquired outside the scanner. Sixth, in considering whether or not to apply a smoothing kernel to our image data, we decide to retain the smoothing process, given that unsmoothed data yielded 1) noisy statistical maps; 2) we aim to investigate larger structures that benefit from smoothing. Seventh, we determine that using a multi-echo sequence, developed to increase signal to noise ratio and specifically to rescue

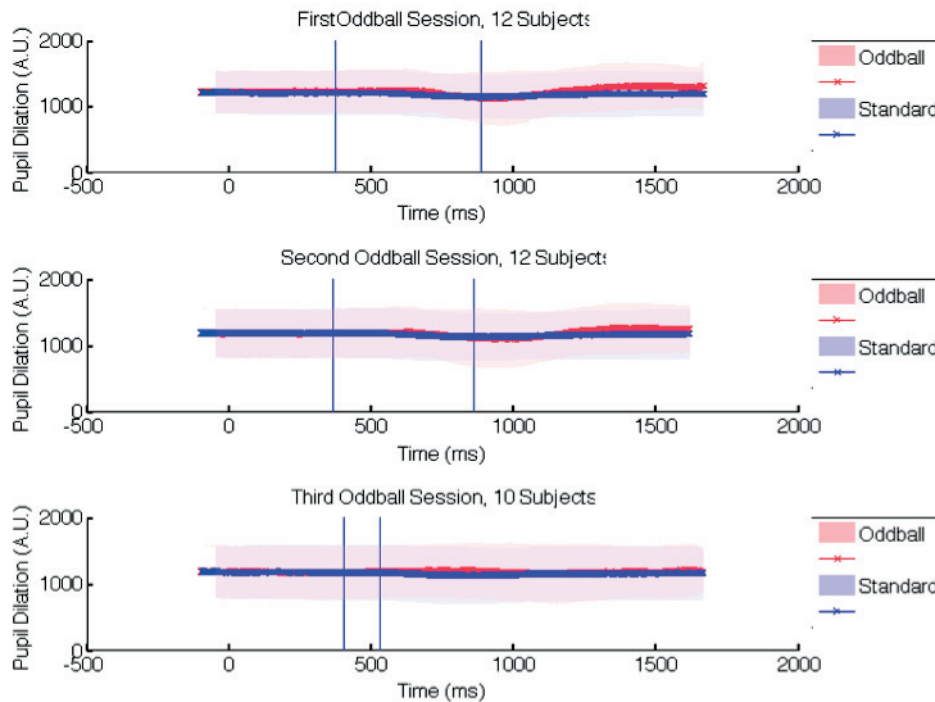


Figure 2.12 – Grand mean time courses for odd (red) versus standard (blue) tones, acquired in the scanner, for three separate imaging sessions.

signal drop-out in orbitofrontal regions, is preferable to other available sequence options when taking into consideration the tradeoffs inherent to fMRI research namely, SNR, time, spatial resolution and spatial extent as well as the susceptibility that is potentially introduced by LC's proximity to CSF.

The LC lies in the dorsal pons. It is part of the reticular formation in the brainstem. Subregions ventral to the LC include the medial lemniscus, the central tegmental tract, the medial longitudinal fasciculus and reticular formation nuclei (nucleus reticularis centralis superior and nucleus reticularis tegmenti pontis); in the dorsal direction, the superior cerebellar peduncle; and in the superior direction, the inferior and superior colliculi followed by the periaqueductal grey (PAG). With respect to fMRI, we do not expect significant BOLD responses arising from tracts, (although see Gawryluk et al., 2014 for a review on white matter fMRI activations in the corpus callosum) two neighboring pontine nuclei must be considered: the trochlear and motor nuclei of the trigeminal nerve. The superior cerebellar peduncle is an obstacle to localization efforts, as it not only lies just behind the LC but also flanks the region. Finally, with respect to inferior and superior colliculi, these regions are easily identifiable on T1 images and thus significant clusters found within their bounds can be assigned correctly. The PAG is superior to the aforementioned colliculi, and thus, again, can be easily classified as being 'out of LC'. Interestingly with respect to the PAG, it encompasses a cholinergic nucleus, the laterodorsal

tegmental nucleus. As seen above, Ach represents another neuromodulator of interest with respect to decision-making under uncertainty. BOLD response in the PAG has been linked to aversive prediction errors (Roy et al., 2014) as well as uncertainty-modulated pain response (Yoshida et al., 2013). In summary, the dorsal wall of the pons provides a natural boundary for the LC, as does the inferior colliculus. Bounds in the ventral and inferior direction are less clear and thus clusters crossing the atlas bounds should be subjected to significant scrutiny.

The question of whether or not to apply physiological correction to fMRI data from the brainstem is a pertinent one (Harvey et al., 2008; Brooks et al., 2013). It is of note that in the Brooks study, signal variation due to physiological effects was low (2%) in the region of the LC specifically. Applying physiological correction to our data may prove problematic. If LC activity correlates with increases in cardiac and respiratory rates, for instance in the experience of surprise, then correcting for such artifacts may in fact divert pertinent variance away from what is expected to be a difficult to detect, small signal change. As our subsequent fMRI paradigms will be event-related, we expect low frequency noise to be removed by high-pass filter. Finally, the most pertinent recent study investigating LC bold responses found no differences in physiologically versus non-physiologically corrected LC BOLD responses (Murphy et al., 2014). Thus, we will not be including physiological correction in our future pre-processing pipelines.

The process of smoothing fMRI data is a means of increasing SNR as well as a method to model data for parametric tests. With a region as small as the LC, applying a smoothing kernel may well have the effect of obliterating any significant voxels in the region. If only 2 or 3 voxels display a significant BOLD response, then averaging their intensity across neighboring voxels may have the net effect of erasing a significant response. On the other hand, not smoothing data ignores the problem of noise in the signal and presents a problematic state of affairs with regards to parametric tests down the line, as the application of random field theory mediates family-wise error correction. In our auditory oddball pilot experiment above, we find unsmoothed results indeed yield questionable statistical maps, increasing our odds of assigning significant LC activity to noise. Further, Murphy and colleagues also compared smoothed (6 mm FWHM) and unsmoothed data and found that unsmoothed data reduced the significant cluster in the region of the LC (Murphy, 2014). As we aim to investigate cortical regions as well as the LC in future studies, we opt, based on the reasoning and evidence above, to include smoothing in our pre-processing pipeline.

The controversy surrounding LC localization with the use of MRI is acute (Minzenberg et al., 2008; Astafiev et al., 2010) and elicits strong, almost partisan views on what does and what does not constitute the LC. On the other hand, the drive to assign the LC-NE complex to a specific cognitive function is significant. As these high-level relationships are difficult to measure in non-human animals, that is, non-invasively through fMRI, our methodology with respect to measuring the LC remains limited. What is remarkable in the LC debate is that one side will assign LC BOLD response to brainstem areas that are suspiciously far from the LC (e.g. the inferior/superior colliculus) (Minzenberg et al., 2008) while the other side categorically

negates the possibility of an LC-specific BOLD response in fMRI results. The middle road in the debate thus becomes razor sharp for what is a researcher to do with a significant cluster in what appears to be the vicinity of the LC? Such clusters should not be ignored. In an attempt to avoid controversy, researchers have taken to equivocating and identifying such clusters as being 'in the vicinity of the LC', a path we take in the rest of the thesis. It must be stressed that, in our case, 'in the vicinity of the LC' encompasses a specific definition based on the LC Keren atlas. The LC Keren atlas is a probabilistic map of the region and thus we will anchor our future experiments on the bilateral voxels that are shared by 95% of the subjects tested in the study. We feel this conservative approach to LC functional identification is sufficient to make the claim that a significant cluster in the brainstem may fall in the vicinity of the LC.

As mentioned, the search for a neuromodulator responsible for uncertainty signaling is a compelling quest. Our and other's views on noradrenergic involvement in uncertainty have been guided by a series of relationships: pupil dilation to risk processing; LC to pupil dilation to risk processing; LC to NE, to LC-NE to pupil dilation and risk processing; and therefore NE to risk processing. A chink in any of the above links obviously compromises the final association of NE to risk processing. What has not been addressed in our view is the potential pitfall in assigning noradrenergic activity to LC. While all LC neurons produce noradrenaline, and it is known that the LC projects to nearly all regions of the brain save for the basal ganglia (Nagai et al., 1981), its effects downstream appear heterogeneous (Schwartz & Luo, 2015). Subsets of neurons within the LC have been found to differ along a ventral-dorsal gradient, in both morphology as well as co-expression of neuropeptides (galanin and neuropeptide Y). The LC is also heterogeneous in terms of neurotransmitter receptors, with $\alpha 1$ receptors being more abundant in the anterior portion of the region, relative to $\alpha 2$ receptors in the posterior region (Chamba et al., 1991). Acetylcholine also elicits LC neuron firing, further complicating a narrow LC-NE relationship (Egan & North, 1986). As mentioned, LC efferent projections are widespread however evidence suggests that neurons projecting to the forebrain are concentrated in dorsal LC while those projecting to the cerebellum and spinal cord are found in ventral LC (Loughlin et al., 1986). In another study, anterior cells were found projecting to the hypothalamus while posterior cells projected to the thalamus. Cells projecting to the cortex were found across the whole region (Mason & Fibiger, 1979). It is in my view the relationship between LC response and specific noradrenergic activity that merits more scrutiny. While LC neurons do not project to the substantia nigra; globus pallidus or striatum, they have been found to stimulate dopaminergic neurons in the hippocampus (Yamasaki & Takeuchi, 2017; Wagatsuma et al., 2018). Further, LC stimulation modulates dopamine downstream via the ventral tegmental area (VTA) (Park et al., 2017). Thus even if an uncertainty-related BOLD response is found in the LC, it may not be linked to noradrenaline *per se*, but in fact dopamine. The LC itself, again in spite of its small size, represents a heterogeneous set of neurons (Schwartz & Luo, 2015). It is indeed perhaps too simplistic to assign noradrenergic signaling to the LC. Therefore, in our future studies, while we will seek to determine if BOLD responses correlate with the LC, as a means of informing the question, we will make no strong claims on noradrenergic involvement *textit per se* in uncertainty signaling.

3 Of Outcomes and Expectations: the Neural Representation of Confidence, Surprise and Information

Abstract

How does confidence relate to surprise? And surprise to information? All are at play in the decision-making process albeit at different time-points. Surprise manifests at the outcome of a decision, when all variables are revealed, while confidence arises between a decision and its outcome, in the face of partial information. Both phenomena are prime candidates for neuroscientific study, because they can be formalized by a computational account and also because they represent affective states. By exploiting this dual representation of confidence and surprise as computation and feeling, we can investigate their common neural representation in the insula. Further, information theoretic accounts of surprise cast it as information, but formal accounts also show that more information correlates with less surprise (Schwartzbeck et al., 2015; Itti & Baldi, 2009). To address this quandary, we explore the neural representation of information as a confidence error, by using an information theoretic account to capture its signal at the outcome of a trial, while controlling for surprise as prediction error. We investigate these questions in the context of a gambling task in a model-based functional magnetic resonance imaging experiment. Specifically, we assign confidence as inverse prediction risk; surprise as absolute prediction error; and information as Shannon entropy, in parallel to reward and reward prediction errors at distinct time-points a task. We find a common insular involvement for all three variables of interest; LC, striatal and anterior cingulate activation for surprise; and cuneus and frontal lobe correlates for information, specifically. These results support the notion that the insula plays a central role in prediction errors; objective confidence, as precision, is reflected in regions previously linked to uncertainty; the brain encodes surprise and information as distinct quantities.

3.1 Introduction

What purpose does surprise serve? From a phenomenological perspective, surprise serves to alert an organism to a change in her environment. A surprising event is thus one that violates

Chapter 3. Of Outcomes and Expectations: the Neural Representation of Confidence, Surprise and Information

expectation, or, in other words, a signal of an improbable event. In addition to its role of alarm bell, surprise could be used as a learning signal, in which some information is gained from the surprise itself. In Claude Shannon's seminal paper on Information Theory (Shannon, 1948), he elegantly demonstrates the notion of surprise as information. While Shannon's work alluded to signal processing and the problem of encoding and decoding information, we can, some 70 years later, borrow from his work and apply the concept of surprise as information in the context of human behavior.

As natural environments are probabilistic, our expectations and predictions reflect estimates of what we believe will happen. These estimates may contain information about the (expected) uncertainty. As probabilities, expectations are endowed with a variance, or the second moment of its probability. Previous work has characterized this variance as expected uncertainty, or risk, inherent to a reward (or utility) prediction (Preuschoff et al., 2006). The outcome of a reward prediction elicits a learning signal in the form of a reward prediction error in the brain, a process that relies on the striatal dopaminergic system and that has been well-characterized in previous literature (Schulz et al., 1997; Hollerman & Schulz, 1998; Abler et al., 2006; Glaescher et al., 2010; Pessiglione et al., 2006; Garrison et al., 2013), both computationally and empirically. Previous work has further identified a risk prediction error related to risk, which we term surprise, in the anterior insula as well as in the striatum (Preuschoff et al., 2006; Preuschoff et al., 2008). Just as an agent learns through the reward prediction error, which serves to adjust subsequent reward predictions, surprise as risk prediction error should serve as a learning signal that tunes subsequent risk prediction. Unlike the reward prediction error, surprise does not necessarily have a hedonic valence. We expect surprise to occur independently of reward, so long as it violates the confidence an agent has in her predictions. Thus a loss can be, affectively and computationally, just as surprising as a reward. Information theory posits that surprise is information, a plausible consideration in light of the fact that a surprising event can add to our a priori knowledge of the world. However, it has been noted that surprise and information, as quantified by an information theoretic framework, may be differentiated into a stimulus-bound surprise and a Bayesian surprise, or information gain (Itti & Baldi, 2009). Namely, maximal information (in the form of entropy) can carry the least surprise, as quantified by an improbable event. Entropy in information theory is the quantity by which a system can reconstruct a message; that is, given a problem, how many yes/no questions are needed to resolve it? When uncertainty is minimal, that is when an agent has a degree of confidence in the matter, we can see that entropy, or the information gain, will be low and surprise, if the outcome is unexpected, will be high. However, when uncertainty is maximal, and confidence low, the amount of information provided by an outcome is also maximal, while the surprise carried by the outcome is low, as expectations are not well defined. In viewing this question through the Free Energy Principle framework, we see that these two quantities together define free energy. A question worthy of study then is to determine if surprise and information are differentiated in their neural representation.

And what of confidence? Confidence, like surprise and uncertainty, is often thought of as a feeling but its mathematical definition has been extensively used in the fields of statistics and

economics (Manski, 2004; Cesarini et al., 2006) and has more recently attracted interest in the field of neuroscientific examination of decision-making (de Martino et al., 2013; Rolls et al., 2010; Kepecs et al., 2008; Kiani & Shadlen, 2009). Pouget and colleagues (Pouget et al., 2016) write that confidence is the probability of being correct given the evidence, which distinguishes it temporally from certainty, and further posit that, in the case of a continuous variable, can be represented by the inverse variance of said variable. Confidence as inverse variance fits with our definition of risk as variance; the higher the risk, the lower the corollary confidence. Such a measure is also called precision and is an objective measure of confidence. Most studies on confidence in decision-making employ a subjective measure of confidence, obtained via self-report or inferred from such measures as reaction time. While studies above implicate the orbitofrontal and parietal cortices in confidence encoding, Lak and colleagues (Lak et al., 2017) recently found that midbrain dopamine encodes confidence in addition to reward prediction in monkeys; similarly Hebart and colleagues (Hebart et al., 2015) also find striatal involvement in confidence, supporting a dopaminergic role in uncertainty. Specifically, they employed a choice-dependent model of confidence in a task exploiting perceptual ambiguity.

If one is to consider confidence as the degree to which we believe we are correct, we must consider the confirmation of our belief. This is distinct from surprise or reward prediction error but relates to the information contained in a given trial, or what is dubbed entropy or average surprise in information theory. If confidence is precision, then information is accuracy; at the outcome of a gamble, we can view this information gain as a confidence error. We do not expect this quantity to engage regions related to a sympathetic response as we would for surprise, nor do we expect this quantity to induce conflict in an agent, even though it is a prediction error. Crucially, we examine this variable at the outcome of a trial to solve the following quandary: are surprise and information one and the same or are they neutrally distinct? If information gain is a confidence prediction error, is it represented in the brain? Does such a quantity even register at the outcome of a trial? While surprise can be informative, it has been argued that the greater the information content, the lower the surprise; thus we can differentiate them in a formal account. We examine whether disassociating the two variables yields distinct patterns of neural activity. If they do, it better supports the notion that information, or entropy, is indeed different from surprise.

In the following study, we examine three main questions in the context of cognitive (financial) uncertainty. We seek to examine the neural representation of distinct uncertainty-related variables, notably confidence, surprise and information, as formalized by precision (inverse variance), absolute prediction error and Shannon entropy. Specifically we hypothesize that 1) surprise as error-detection will elicit a BOLD response in the insula, striatum, anterior cingulate and the LC; 2) that objective confidence signals, will be reflected in the insula; 3) that information-related neural signals will be distinct from surprise. We first seek to confirm the existence of these signals, distinct from reward-related variables, in regions related to risk and reward processing. We test our hypotheses using fMRI within the context of a gambling paradigm that elicits both reward and risk predictions as well as their concomitant errors while controlling for motivational, learning and motor effects, to determine whether these

Chapter 3. Of Outcomes and Expectations: the Neural Representation of Confidence, Surprise and Information

changes in neural activity are confirmed within a set of brain regions (notably the striatum and insula) known to be implicated in reward learning and risk signals (Preuschoff et al., 2006; Preuschoff et al., 2008; Kuhn & Knutson, 2005; Hsu et al., 2005; Lak et al., 2017).

3.2 Materials and Methods

To address our research questions, we employed a gambling task performed during functional magnetic resonance imaging (fMRI) acquisition. We used an auditory version of a card game (Preuschoff et al., 2006), where subjects had an equal probability of winning or losing 1 CHF (1 USD) at each round, starting with an initial endowment of 25 CHF. At each round, subjects were instructed to place a bet via manual button press on whether a second card drawn from a deck of ten cards would be higher or lower than a first card drawn from the same deck. The bet is made prior to any card being drawn, at which point outcome probability is at .5, or maximal uncertainty and the expected value of reward is 0. Following the bet, subjects heard the value of the first card. After a 5.5 second interval, subjects heard the value of the second card. After another 5.5 s interval, subjects were instructed to report whether they had won or lost the round. An incorrect response incurred a penalty of 25 ¢ off a round's total payoff. Each trial lasted approximately 25 s. Inter-trial intervals durations were randomly jittered (2-5s). All possible card pairs, excluding pairs of identically valued cards, were presented to each subject in a random order, totaling 90 trials per experiment. These 90 trials were divided into three sessions of 30 trials each, to give subjects a chance to rest in between sessions. Each session began with a new 25 CHF endowment. The order in which specific card pairs were presented was randomized; thus no two subjects played the same sequence of gambles. Throughout the experiment, subjects viewed a black fixation cross on a grayscale screen projected into the MRI bore. Auditory stimuli were coded using the voice of 'Alex' from Mac OSX text-to-speech function. At the end of the experiment, subjects were paid the resulting payoff of one of the experimental blocks. The entire experiment was conducted in English. The Ethics Committee of the Canton of Vaud, Switzerland, approved the experiment.

3.2.1 Participants

Twenty-five healthy participants (10 F, average age 25.13 years) were enrolled in the experiment. Participants were recruited via paper and online advertisements targeting the student populations of Ecole Polytechnique Fédérale de Lausanne and Université de Lausanne. Exclusion criteria included metal implants, previous psychiatric illness, and psychotropic drug use within the past year. Inclusion criteria included proficiency in English.

3.2.2 Procedure and Task Description

Subjects were sent electronic versions of information relating to the risks and benefits of the experiment the night before their scheduled session. On the day of the experiment, this infor-

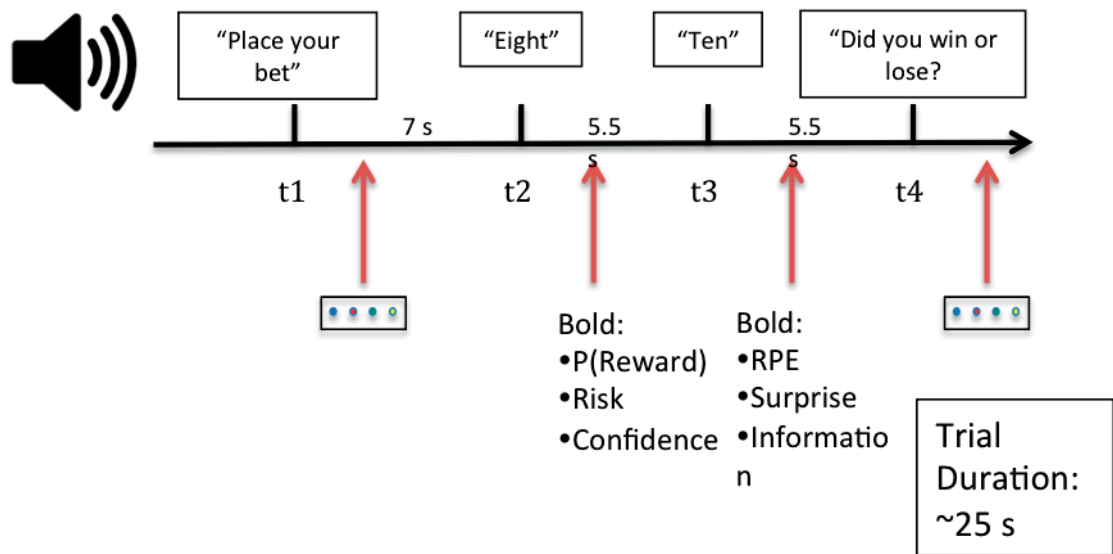


Figure 3.1 – Sample Trial Timeline. Subjects are first instructed to choose whether a second card drawn from a deck of 10 cards will be higher or lower than a first card (t1). At t2, subjects hear the value of the first card, whereupon they can make a prediction on their expected reward, risk, respectively confidence. At t3, subjects hear the value of the second card, at which point they know whether they have won or lost the round, and incur a reward prediction error, surprise and a confidence error (information). At t4, subjects are asked to report whether they have won or lost as a means of assessing their attention. Trials lasted roughly 25 s each. Inter-trial jitter was randomly set to 2-5s. Subjects played 90 round of the trial divided into 3 blocks of 30 trials.

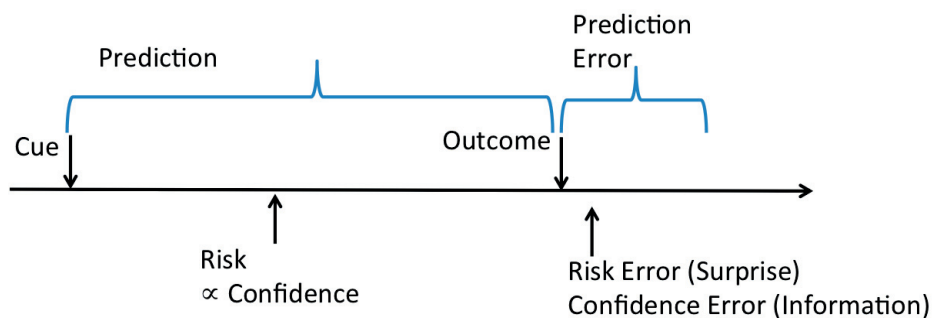


Figure 3.2 – Uncertainty Related Variables in a one-shot decision process. If the brain as an inference machine, we expect it to make predictions related to uncertainty when faced with a cue: first, risk at cue; followed by risk and confidence, post-decision; followed by errors relating to both predicted risk (Surprise) and confidence (Information gain)

mation was reviewed with the subject before obtaining their written and informed consent. Subjects were then asked to undergo screening prior to the scanning session to ensure their compatibility with the procedure, as well as to collect relevant information on their medical history. We then provided an oral explanation of the gambling task before proceeding to the

Chapter 3. Of Outcomes and Expectations: the Neural Representation of Confidence, Surprise and Information

scanner room. The task was coded in Matlab (Matlab and Statistics Toolbox Release 2013a, The MathWorks, Inc., Natick, Massachusetts, United States) and the PsychophysicstoolBox (Brainard, 1997; Pelli, 1997; Kleiner et al, 2007) and incorporated the Eyelink Toolbox functions (Cornelissen et al., 2002). Following the end of the experimental session, subjects were paid for their time and debriefed. Payment of their task-related payout was reserved for a subsequent second experimental session, to lower rates of attrition.

3.2.3 Imaging Procedure

All scans were acquired on a Siemens 3T Prisma, at the Centre Hospitalier Universitaire Vaudois in Lausanne, Switzerland. Once in the scanner room, subjects were fitted with a heart-rate monitor on their finger and respiration rate sensor around their mid-section, for physiological monitoring via BioPac (BioPac Systems Inc., Goleta, CA). Once settled in the bore, we acquired first a localizer scan, followed by a gre-field mapping scan to correct for magnetic field inhomogeneities. We then alerted subjects to the beginning of the task and functional sequence. In a first instance, prior to MRI acquisition, we performed a 3-point calibration on the in-scanner eye-tracker (EyeLink SR 1000, SR Research, Ottawa, Ontario), as we aimed to collect pupillometry measures related to task activity. Following this, we displayed task instructions on the screen projected onto a mirror within the bore. Subjects could then signal their readiness for the scan, upon which the first of three EPI sequences was launched. As the task was auditory in nature, once the EPI sequence was launched, we presented subjects with a black fixation cross centered on a grayscale screen throughout the experiment. The task itself was sounded with the use of Mac OSX text-to-speech function, with the voice of 'Alex', transmitted to the subject via MR compatible headphones. Parameters for our sequence were: 2D EPI, Multi-Echo sequence, 3 x 3 x 2.5 mm resolution, with a TR=2.72 s (34 slices; slice TR = 80 ms; slice TE = 17.4 ms; FOV = 192 mm; FA = 90; matrix size 64 x 64). Following the three task sessions, subjects were told they could close their eyes and rest while we acquired their T1 structural MRI (MPRAGE; 1 x 1 x 1 mm; slice TR = 2000 ms; slice TE = 2.39 ms; FOV = 256 mm; FA = 9; matrix size = 256 x 256). The task related session lasted approximately 36 minutes while the imaging session lasted approximately 50 minutes; the whole experimental session, including intake and debrief lasted approximately 90 minutes.

3.2.4 Behavioral Analysis

All 25 subjects completed the task. Of these, 4 were found to have made more than 3 errors in at least one session (an error being either a false report of the outcome at the end of a trial, or a missed bet). These subjects were excluded from subsequent analyses, as errors of commission and omission may reflect a lack of attention to the task and thus compromise result interpretability, given that no action is required from the time-points of interest. Average task-related payout per participant was 29.57 CHF; across all sessions and subjects, payoffs were in the range of 13-39 CHF.

Table 3.1 – Decision Variables Across Sessions

| Variable | F | p | Mean Session 1 | Mean Session 2 | Mean Session 3 |
|-----------------------|---------|----------|-------------------|-------------------|-------------------|
| Predicted Reward | 0.0389 | 0.9619 | 0.0100 | 0.0104 | 0.0013 |
| Risk | 27.3517 | <0.0001* | 0.5921 | 0.6666 | 0.5167 |
| Confidence | 27.35 | <0.0001* | 0.0004 | -0.0740 | 0.0759 |
| Risk Prediction Error | 13.4 | <0.0001* | -0.0720 | -0.0457 | 0.1092 |
| RPE | 3.4 | 0.0337 | -0.0247* | -0.0546 | 0.0558 |
| Absolute RPE | 5.603 | 0.0037* | 0.5561 | 0.6438 | 0.5713 |
| Shannon Entropy | 24.6024 | <0.0001* | 0.6364 | 0.7047 | 0.5623 |

As our task was designed to be completely random, we performed post-hoc analyses on potential differences for several variables of interest across sessions. We performed an ANOVA to determine if any one session contained more of one type of card value for card 1 and found no significant differences across sessions ($F = 0$, $p = .996$). We then performed ANOVAs on the mean differences of higher bets and lower bets across sessions and found no significant differences ($F = .19$, $p = .8324$ and $F = 0.2$, $p = .8204$, respectively), indicating that subjects understood the random nature of the task and did not 'switch' strategies across sessions. We then sought to determine if there was a difference in bet choices and found a significant bias for selecting a higher bet in all sessions ($F = 34.69$, $p < 0.001$). We cannot conclude that there is a bias towards selecting higher bets in spite of this result, because buttons indicating a higher bet were consistently on the right side for all subjects. Thus our significant results may reflect handedness rather than bet preference.

Finally, we performed F tests on all decision variables across sessions to determine if differences existed in levels of exposure to our respective variables of interest.

We took the decision to fully randomize card pair presentations within and across experimental sessions to ensure neither order effects, nor inadvertent learning were introduced into the study. All subjects experienced a different sequence of card pair presentations. We expected no differences in decision-making variables means across sessions. Post-hoc tests contradict that expectation and thus preclude us from undertaking tests on across-session effects.

3.2.5 Imaging Analysis

Scans were preprocessed and analyzed using SPM 12. We first generated voxel displacement maps (VDM) and applied these to functional data. We then warped and realigned data to the mean image before apply a bias-field correction. Then data were co-registered to individual anatomical volumes before being segmented (6 class tissue probability maps) and normalized to the MNI152 template and smoothed to 8 mm FWHM. The resulting images were then used for analysis. The course we took to perform analyses on fMRI data was as follows. We designed three model-based GLMs. In the first, we modulated card 1, offset by 1 second, with

Chapter 3. Of Outcomes and Expectations: the Neural Representation of Confidence, Surprise and Information

reward prediction, as well as confidence; and card 2, with regressors separated by win or loss outcomes, offset by 1 second, with surprise as risk prediction error as well as reward prediction error. In the second GLM, we modulated card 1 with the same parameters as those above but at card 2, we applied absolute reward prediction error as surprise and added a third parametric modulator, that of Shannon Entropy, as a measure of information; further, as winning RPEs correlate with absolute RPEs, we did not separate regressors according to outcome. We offset the sampling of the BOLD response to take into account the time duration of the card value's sounding, which lasts 1 second. The difference between risk prediction error and absolute reward prediction error are that the first is a quadratic function, while the second is linear. They are qualitatively the same otherwise. In a third GLM, we maintained the same design as GLM 2, but we swapped the third parametric modulator at card 2, that is information, with a Kullback-Leibler (KL) divergence measure, which would reflect a potential 'learning' across the session. We applied these designs at the single subject level before performing a random-effects analysis at the group level. We then tested for whole brain activation, as a means to inform our analyses, first at $p=0.05$, FWE corrected, then at $p=0.001$, cluster-level corrected. Should cluster-level correction yield significant voxels in specific contrasts of interest, namely confidence, surprise and information, we proceed to using non-parametric statistical tests, as a means to overcome both false positives (with cluster level correction) and false negatives (with whole-brain family wise error correction; see Eklund et al., 2016). Because we are using model-based fMRI to investigate cognitive processes, we do not expect large changes in the BOLD response. Therefore, we identified 7 a priori regions of interest (ROIs) (13 bilateral, but for the locus coeruleus) in advance of any analyses to apply small volume corrections. We constrained our choice of regions to the striatum (caudate, putamen and nucleus accumbens); anterior cingulate cortex; insula (posterior and anterior); and the locus coeruleus. All ROIs employed in our analyses came from the Neuromorphometric Atlas except for the locus coeruleus map. The latter ROI employed came from identification of the region with a neuromelanin-sensitive MRI sequence from Keren and colleagues (Keren et al., 2009). Given results from our pilot data suggesting this map may not be suitable for analysis, we created two LC ROIs centered at the peak voxel identified by the Keren paper (within which 95% of individual regions were contained). In addition, our specific investigation of the locus coeruleus eschewed cluster-level correction, as we do not expect to find a cluster for such a small region.

It is worthy of note that many brain regions have been implicated in decision-making processes, such as the amygdala, dorsolateral and ventro-medial prefrontal cortices and orbitofrontal cortices however, we constrained our choice of a priori regions to better focus on those specifically relevant to our paradigm, as found in previous studies employing the same task (insula and striatum), as well as regions we hypothesize to be implicated in uncertainty processing (anterior cingulate and the locus coeruleus).

The designs above were repeated for each of the three sessions before being concatenated for estimation. The two primary time-points of interest were at 1 second after the start of card 1's sounding and at 1 second after the start of card 2's sounding. After hearing the value of card 1,

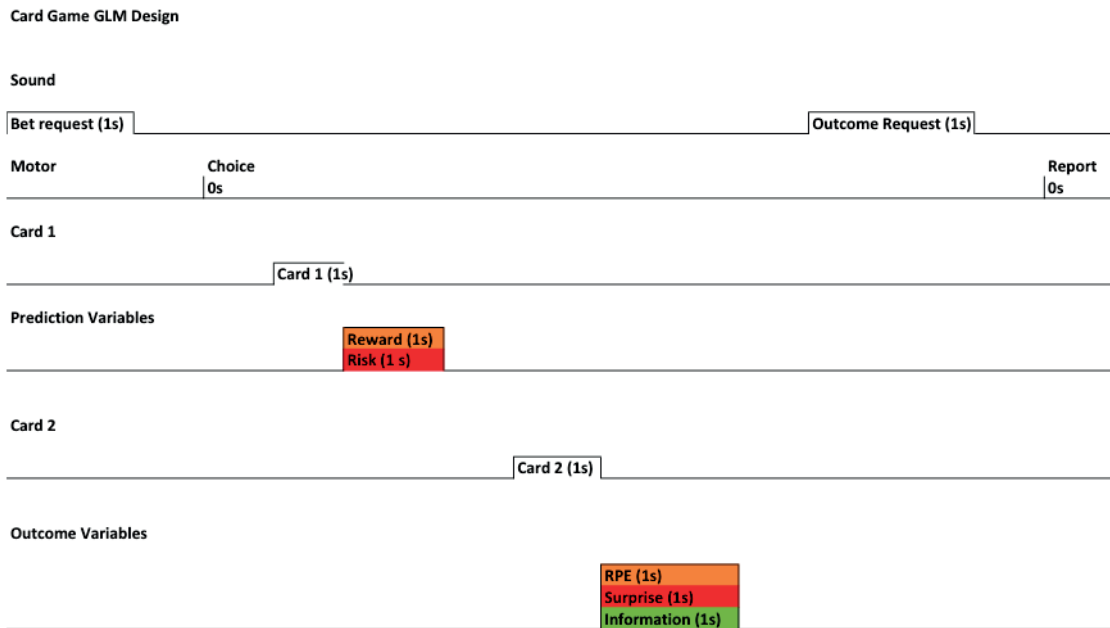


Figure 3.3 – The general linear model. Model-based fMRI analyses of the Card Game. Regressors included in the model are presented above, including a sound regressor (bet and report instructions); a motor regressor (Bet Choice and report responses); a regressor for Card 1’s sounding; a regressor for Card 1’s offset, modulated by prediction values; a regressor for Card 2’s sounding; and a regressor for Card 2’s offset, modulated by error values. Motion correction regressors of no interest are also included in our design matrices but are not shown here. Parametric modulators are represented in colored blocks. Outcome variables were separated into winning and losing trials in a first instance, to check win related BOLD responses.

the subject can make a reward prediction, a risk prediction and a concomitant confidence estimate. After hearing card 2, the subject can experience a reward prediction error as well as surprise and information. Because confidence and risk prediction after card 1 are negatively correlated, only confidence was included as a parametric modulator to card 1.

3.2.6 The Models

Our computational model for reward prediction after card 1 reflects the probability of winning given the bet placed (higher or lower), and card 1’s value. The reward prediction error at card 2 reflects the trial outcome (win or loss) minus the reward prediction. Risk prediction (uncertainty) at card 1 is taken as the variance of the reward prediction; and confidence is taken to be the inverse of this risk prediction (a precision term). At card 2, two numerically equivalent models quantified surprise as error: the risk prediction error, computed as the actual minus expected reward prediction error and the absolute reward prediction error. At card 2, information was also included as parametric modulator in the form of Shannon entropy, an information theoretic measure. While we have reviewed several possible models

Chapter 3. Of Outcomes and Expectations: the Neural Representation of Confidence, Surprise and Information

for surprise, including Shannon Surprise and Bayesian KL divergences, we do not employ the latter two, as they are better suited to learning tasks. Because our task begins with an equal probability of outcome and ends with a terminal state that is independent of prior and future outcomes, we do not expect any learning to occur. In other words, the model space of the task is confined to a trial; the trial begins with a flat prior and ends with a deterministic outcome. Trials are assumed to be independent. Where there may be a learning effect is in the unlikely event that a subject counts card pairs as they are presented, because each possible card pair is only presented once. Should a subject deduce that each card pair is only presented once and retain card pair values as the trial proceeds, then we may expect the model space to expand to the experimental session. We nonetheless controlled for the possibility that a subject counted cards during the experimental sessions. We computed a Bayesian update measure by employing a Dirichlet counting process, as per Strange et al. (Strange et al., 2005), where wins were counted across a session, and included this measure of learning or divergence in a general linear model as a parametric regressor at Card 2. No significant voxels emerged, even when lowering the threshold to $p=0.05$, uncorrected.

The following quantities are computed in the prediction phase of the experiment (t2 in Figure 3.1):

$$E(Reward) = \sum p(outcome | card1, Bet) * x_{outcome} \quad (3.1)$$

where x denotes the trial reward.

$$Risk = E[(E(Reward) - Reward)^2] \quad (3.2)$$

$$Confidence = E(Reward)^2 - Risk \quad (3.3)$$

We opt to use this value of confidence as opposed to the true inverse of the risk because risk values do include zero values (when the outcome is sure).

The following quantities are computed in the outcome phase of the experiment (t3 in Figure 3.1):

$$\delta = Reward - E(Reward) \quad (3.4)$$

$$Surprise = |\delta| \quad (3.5)$$

$$RiskPredictionError = \delta^2 - Risk \quad (3.6)$$

$$Hs = -p_{win} * \log_2(p_{win}) - ((1 - p_{win}) * \log_2(1 - p_{win})) \quad (3.7)$$

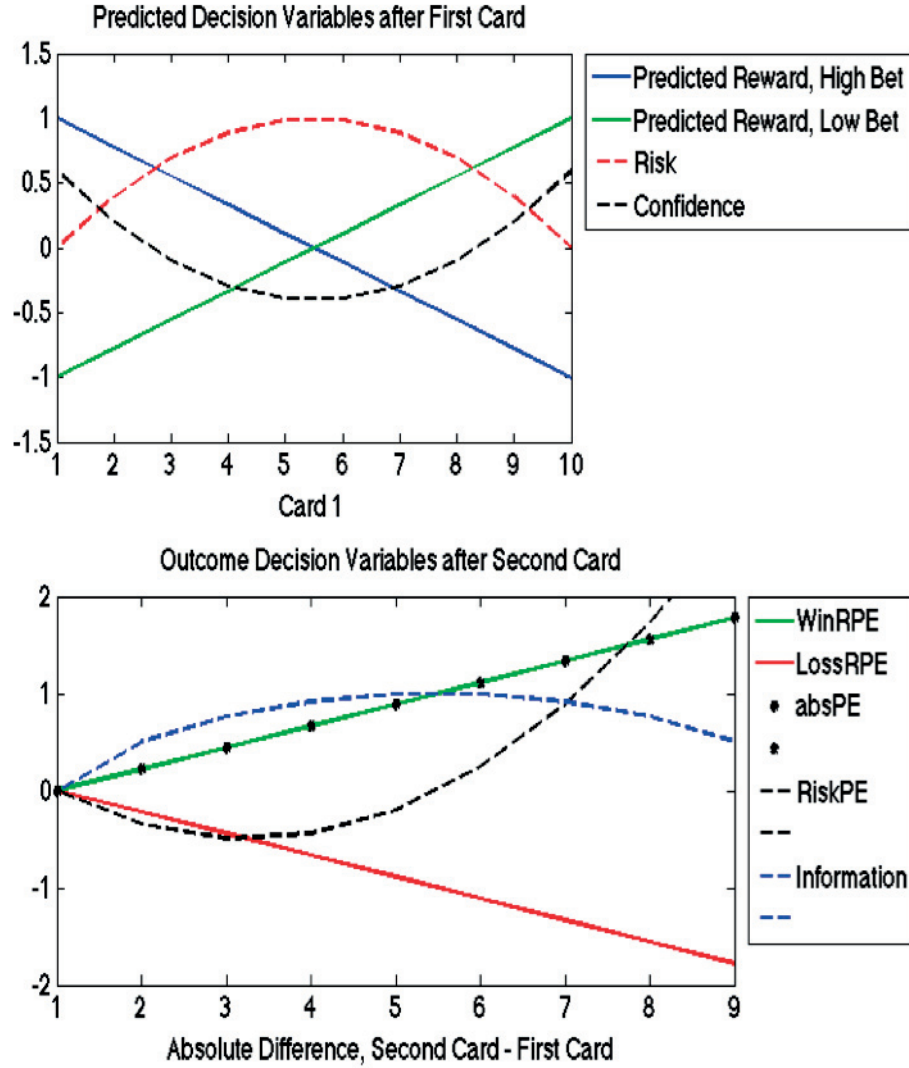


Figure 3.4 – Graphical representation of the computational accounts employed in our model-based fMRI study. The top plot shows prediction values computed at t2: reward prediction for higher bets (blue) and lower bets (green), as well as confidence (black dashed lines) and risk (red dashed lines). The bottom plot shows decision variables computed at t3, at the trial outcome for reward prediction error in winning trials (green), loss trials (red), as well as unsigned errors, notably surprise (absolute prediction error in black dots; risk prediction error in black dashes) and information (blue dashes).

$$p(win_i) = \frac{\sum_1^i Wins + 1}{\sum_1^i Outcomes + 1} \quad (3.8)$$

Above, we define information as being a confidence error. One manner in which to further define such an error is to cast the confidence error explicitly as 1 (true outcome) minus the

Chapter 3. Of Outcomes and Expectations: the Neural Representation of Confidence, Surprise and Information

confidence of the true outcome at t2.

$$\text{ConfidenceError} = 1 - \text{Confidence} \quad (3.9)$$

When we compute such a value, we find that an explicit confidence error is simply a scaled form of our information value (Shannon entropy in equation 3.7, above). We see this relationship clearly in Figure 3.4.

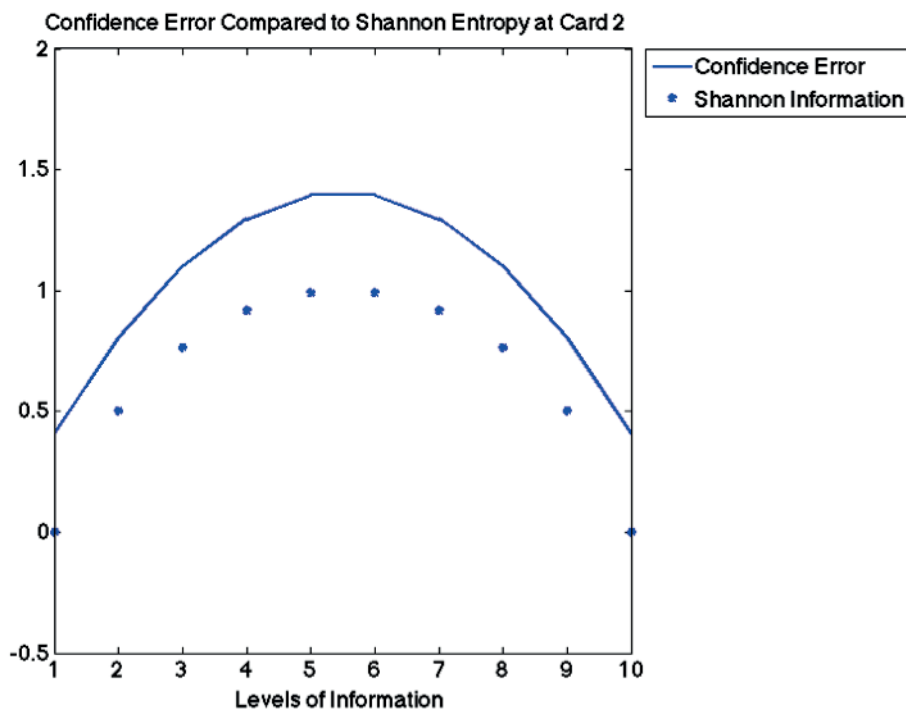


Figure 3.5 – Explicit confidence error and Shannon information plotted for the 10 possible outcomes at Card 2. We see that confidence error and Shannon information present the same quantities, scaled relative to one another.

3.3 Neuroimaging Results

3.3.1 Confirmatory Analyses

To validate our data and replicate previous findings in reward-based decision making we first inspect the (known) BOLD response to auditory stimuli; reward prediction; reward prediction error and reward.

Contrast Sound

As expected, our auditory stimuli yield highly significant clusters in bilateral auditory cortex at $p = 0.05$, FWE.

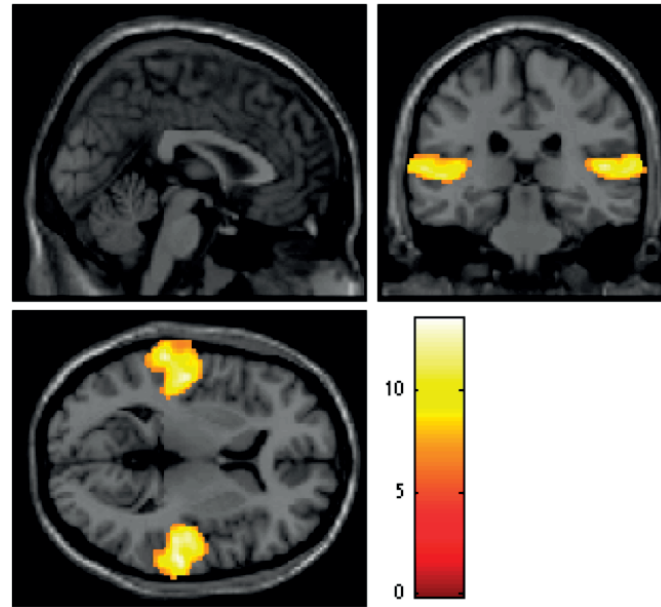


Figure 3.6 – Statistical maps for the auditory contrast. Here, we plot statistical maps thresholded at $p = 0.05$, FWE for auditory contrast overlaid on the canonical T1 image (0, -28, 6). Event onsets included in the regressor include the sounding of bet instructions, of card 1, of card 2 and the instruction to report trial outcome. Significant clusters are found bilaterally in the auditory cortex.

Contrast Win > Loss

To test for reward related activity, regressors for winning outcomes were separated from regressors for loss outcomes and contrasted against one another. These regressors were taken one second after the start of the sounding of card 2's value with the aim of confirming reward related neural activation. At $p = 0.001$, cluster level corrected ($k = 35$), we obtain a significant cluster of 1344 voxels in the region of the anterior cingulate cortex, a region implicated in reward processing, though we did not detect striatal activation as expected.

We also test a Loss > Win contrast and found no significant clusters, in line with previous findings (Preuschoff et al., 2006).

Expected Reward

We then tested for reward prediction activation following card 1 ($p = 0.001$, $k = 27$). No significant clusters emerged at the whole brain level. Our SVC analyses show significant activations in

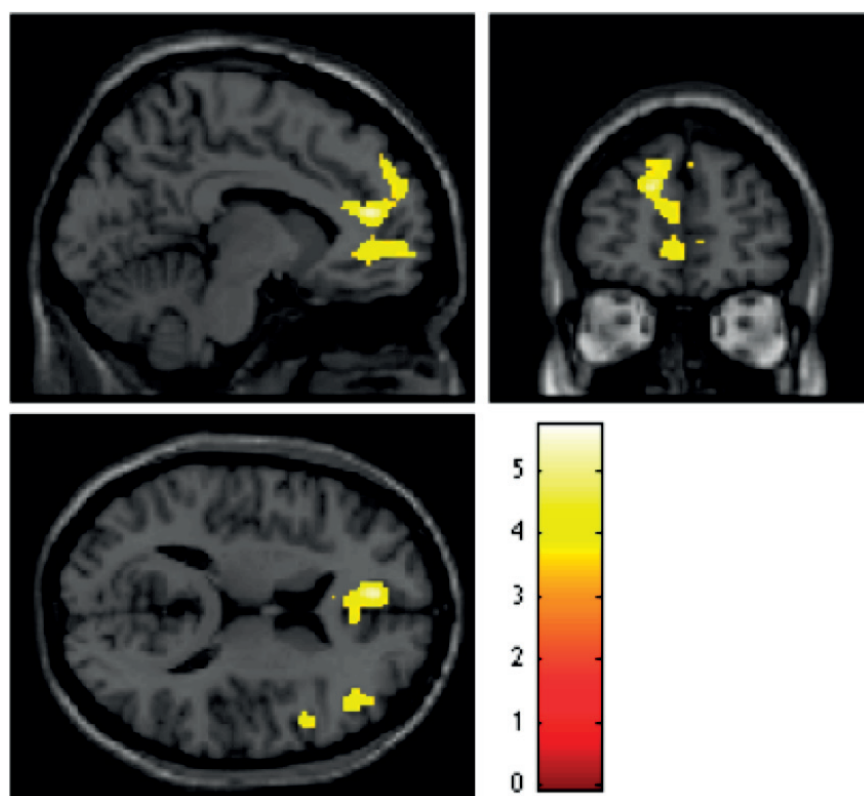


Figure 3.7 – Statistical map thresholded at $p = 0.001$, cluster level correction $k = 35$, for win > loss contrast (-8, .49, 14). Maps are overlaid over the canonical brain. A large significant cluster in the anterior cingulate is visible on the map. No significant clusters in the striatum were found however.

right dorsal (caudate, putamen) and ventral striatum (See Appendix A, Table A.1).

Reward Prediction Error

We then tested for activation following card 2 for the reward prediction error, to determine if a neural pattern similar to ones found previously would emerge, namely in striatal regions. At the whole brain level ($p = .001$, $k=27$), significant clusters are found in the left putamen and right caudate extending into the left and right ventral striatum (see Appendix A, Table A.2).

3.3.2 Exploratory Analyses

The following analyses concern our questions of interest, centered on investigating the neural representation of uncertainty related variables, namely confidence, surprise and information.

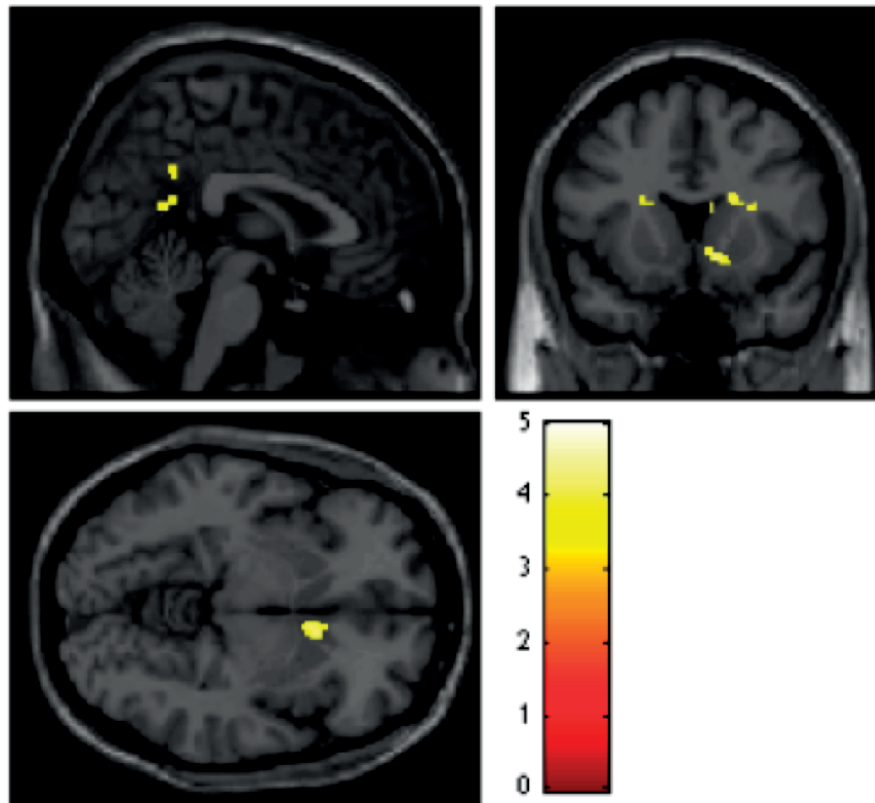


Figure 3.8 – Statistical map for expected reward, thresholded at $p=0.001$, cluster level corrected ($k = 35$) (2, 13, -6) and overlaid on the T1 canonical image. A cluster in the right striatum is visible in the image but is not significant at the whole brain level. SVC analysis using the same threshold reveals significant clusters in the right dorsal striatum extending into right nucleus accumbens.

Confidence at Card 1

We performed a t-test on the offset of card 1's sounding parametrically modulated by confidence, computes at t_2 (See equation 3). Recall that at the beginning of each trial, the risk prediction is the same; the latter changes following card 1, as the subject updates her risk. Confidence here is orthogonal to reward prediction (experienced during the same time interval) but is anti-correlated with risk. Using this definition of confidence in the human realm is ideal, because in our task, confidence is objective. Thus we need not concern ourselves with potential subjective effects. Its relation to surprise is also clear: the higher this confidence at card 1, the higher the surprise will be at card 2. No significant voxels were found for risk prediction at the group level (a negative t-test contrast on the confidence at card 1). We then performed a test on confidence values at card 1. Whole brain analyses reveal bilateral activation in the angular gyrus ($p = 0.001$, $k = 31$), right middle frontal gyrus and anterior insula, as well as left middle frontal gyrus and superior frontal gyrus. SVC analyses confirm a significant cluster in the right anterior insula (See Appendix A, Table A.3).

Chapter 3. Of Outcomes and Expectations: the Neural Representation of Confidence, Surprise and Information

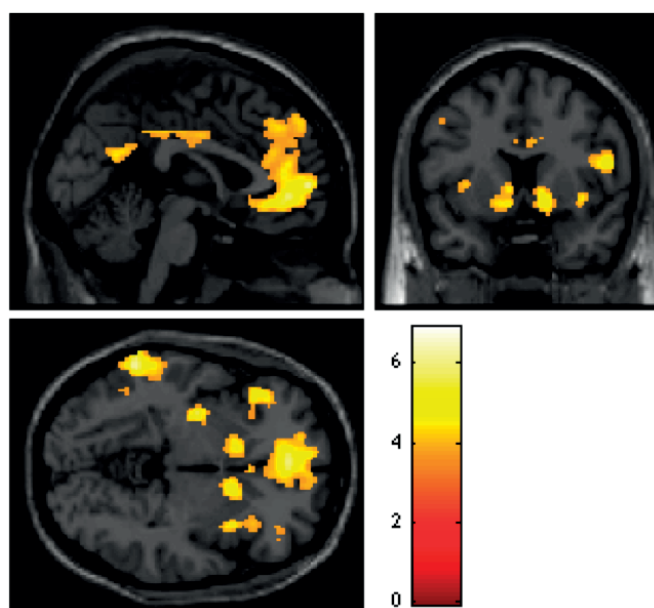


Figure 3.9 – Statistical maps for the RPE contrast, thresholded at $p=0.001$, cluster level corrected with $k = 27$ (0, 0, -4) overlaid on the T1 canonical brain. We find the expected pattern of BOLD responses here, including in the ACC as well as bilaterally in the striatum.

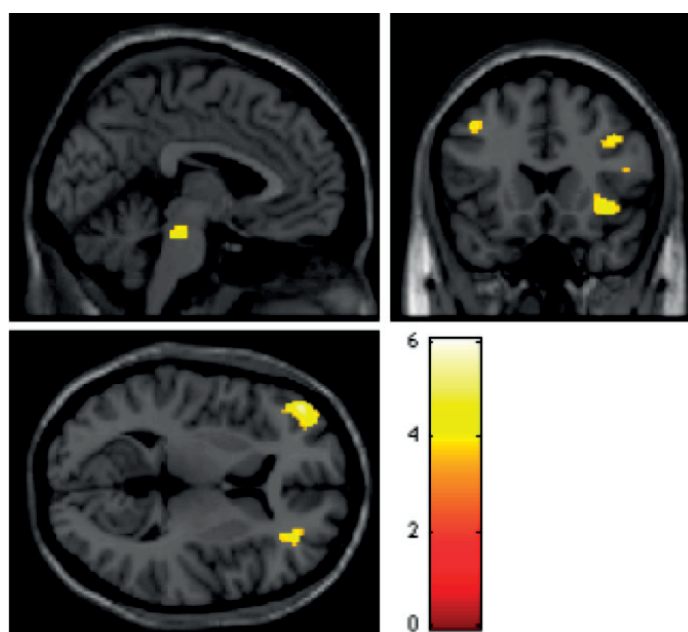


Figure 3.10 – Cluster level corrected statistical maps for confidence, (thresholded at $p = 0.001$; $k = 31$) (-2, 18, 7). The map is overlaid on the T1 canonical image. A significant cluster in the right anterior insula is seen in the coronal orientation

The statistical map of the confidence contrast shows a cluster in the brainstem that may be in the region of the LC. We performed an SVC analysis on the region using an LC centered

volume at (-5 -37 -26) and (6 -37 -27) with a 10 mm radius.

Table 3.2 – Confidence Brainstem Cluster

| cluster p(FWE-corr) | cluster p(FDR-corr) | cluster equivk | cluster p(unc) | peak p(FWE-corr) | peak T | peak equivZ | peak p(unc) | x | y | z |
|------------------------|------------------------|-------------------|-------------------|---------------------|-----------|----------------|----------------|----|-----|-----|
| 0.015 | 0.348 | 25 | 0.348 | 0.009 | 4.41 | 3.64 | 0 | 0 | -30 | -24 |
| | | | | 0.023 | 3.89 | 3.32 | 0 | -6 | -30 | -18 |
| 0.011 | 0.26 | 36 | 0.26 | 0.007 | 4.57 | 3.74 | 0 | 4 | -34 | -34 |
| | | | | 0.009 | 4.41 | 3.64 | 0 | 0 | -30 | -24 |

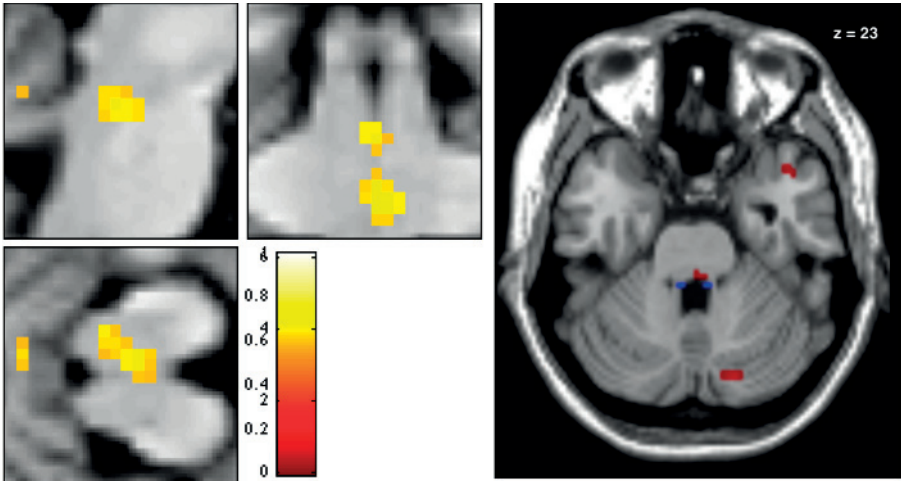


Figure 3.11 – Close up image of t cluster in the region of the LC for the confidence contrast at the whole brain level ($p = 0.001$, $k = 31$). The right panel shows the cluster in question, while the right panel shows the cluster's position relative to the LC Keren atlas (Keren et al., 2009). As in Chapter 2, the cluster sits just at the border of the LC map. However, we note that the cluster does extend into the anterior portion of the brainstem into other nuclei.

Though we find significant clusters with that definition, we cannot conclude that this cluster is in fact in the LC, because it extends anteriorly in the brainstem (see Table 3.6 for peak coordinates). Recall however that the Keren map is probabilistic but subject LC's spanned the following coordinates: for left LC, -2.5:-7; -36:-39; -18:-33; and right LC, 4:8; -36:-39; -18:-33. With that in mind, the second peak voxel found in table 3.6 with our left LC centered ROI would fall into the LC proper. We cannot say as much about the right LC, notably in that it's y coordinate, at -34, falls out of the map's narrow range (-36:-39).

We then performed statistical non-parametric tests on the contrast, as a more stringent method of investigating confidence related neural signals. We find significant clusters in bilateral angular and middle temporal gyrii, as we as left middle frontal gyrus and right anterior insula ($p = 0.05$, FWE) (See Appendix A, Table A.4).

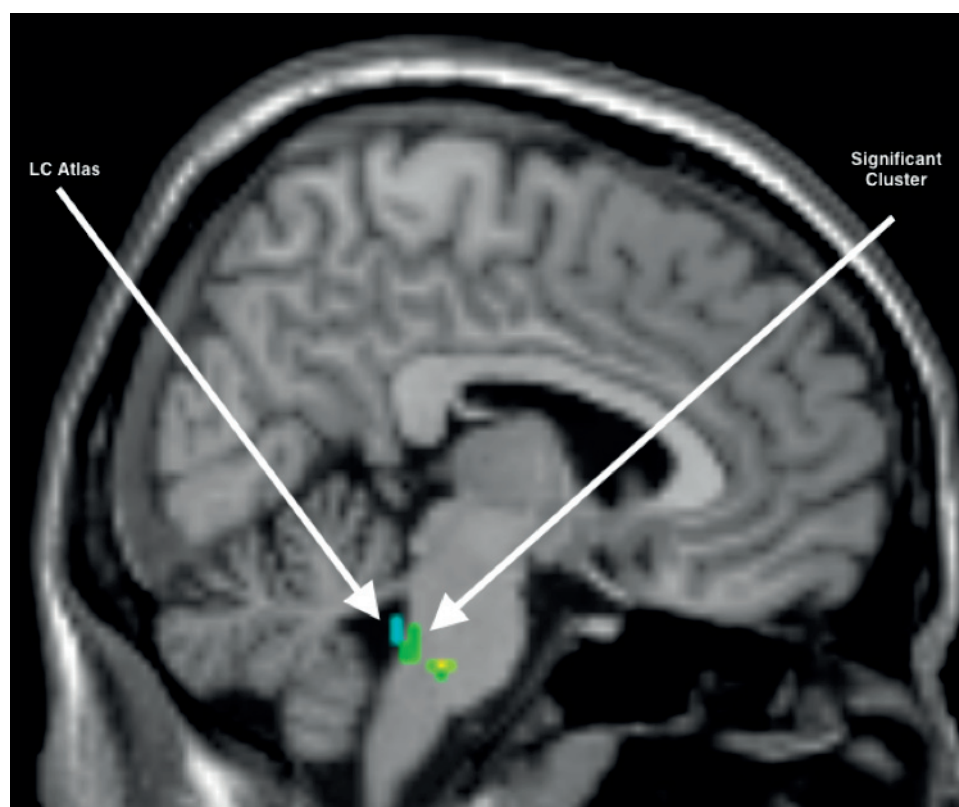


Figure 3.12 – Sagittal view of brainstem cluster for the confidence contrast in the region of the LC overlaid on the canonical T1 image (Green and yellow map). The blue cluster shows the LC atlas. Again, the cluster is clearly just adjacent to the LC map, yet we cannot conclude that it is in fact and LC BOLD response captured for the confidence contrast

Surprise following Card 2

Here, we examine neural activity that correlates with Surprise following card 2. In the first design, we apply the risk prediction error as a measure of surprise to replicate results from previous experiments (Preuschoff et al., 2006). No significant voxels were found for this model. We then use a different model of surprise to investigate related activity beyond the reward prediction error: absolute (unsigned) reward prediction error. Surprise as absolute RPE yields significant activation bilaterally in the insula, right caudate and right ACC, as well as several other cortical regions. SVC analyses confirm bilateral activity in anterior cingulate, insula as well as dorsal striatum (see Appendix A, Table A.5).

As in confidence, we find a significant cluster in the region of the LC. We perform an SVC analysis without cluster level correction at $p=0.001$, centered at $(-5 -37 -26)$ and $(6 -37 -27)$, respectively, with a 10 mm radius for each. We find significant activity for the LC in left and right ROIs using this ROI.

Just as in confidence, peak voxels found lie just outside the narrow definition of LC y coordi-



Figure 3.13 – Thresholded non-parametric statistical map ($p = 0.05$, FWE) for confidence overlaid on an axial slice of the T1 canonical image ($z = 34$). We see a cluster in the right anterior insula and left middle temporal gyrus.

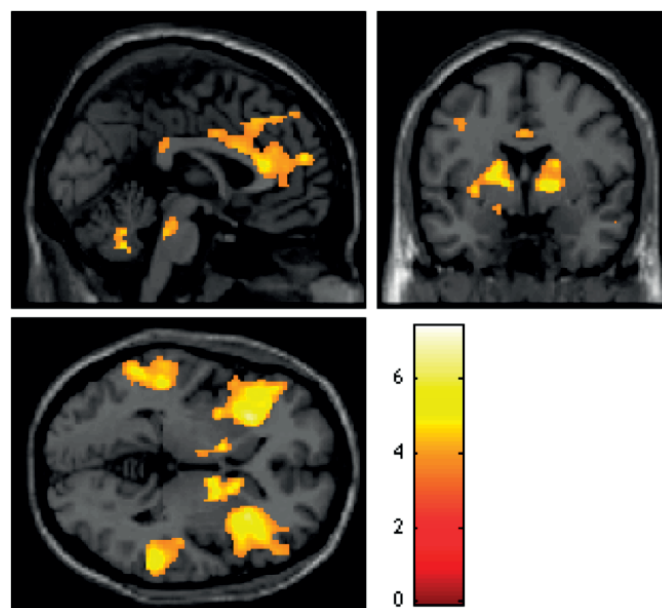


Figure 3.14 – Statistical maps of the surprise contrast overlaid on the T1 canonical image, thresholded at $p = 0.001$, cluster level corrected ($k=34$) (0, 0, 0). The image shows significant clusters in the ACC (sagittal view), bilateral striatal clusters (coronal view); and bilateral anterior insula clusters (axial view). A brainstem cluster also appears in the sagittal view.

nates from the Keren study. This limitation poses a challenge in assigning the observed cluster

Chapter 3. Of Outcomes and Expectations: the Neural Representation of Confidence, Surprise and Information

Table 3.3 – Surprise Brainstem Cluster

| cluster p(FWE-corr) | cluster p(FDR-corr) | cluster equivk | cluster p(unc) | peak p(FWE-corr) | peak T | peak equivZ | peak p(unc) | x | y | z |
|------------------------|------------------------|-------------------|-------------------|---------------------|-----------|----------------|----------------|----|-----|-----|
| 0.003 | 0.089 | 98 | 0.089 | 0 | 6.34 | 4.64 | 0 | -6 | -34 | -28 |
| 0.010 | 0.494 | 43 | 0.247 | 0.002 | 5.19 | 4.08 | 0 | -2 | -32 | -28 |
| | | | | 0.007 | 4.55 | 3.72 | 0 | 6 | -30 | -18 |
| | | | | 0.016 | 4.05 | 3.42 | 0 | 2 | -28 | -22 |
| 0.035 | 0.892 | 1 | 0.892 | 0.038 | 3.57 | 3.1 | 0.001 | 8 | -34 | -34 |

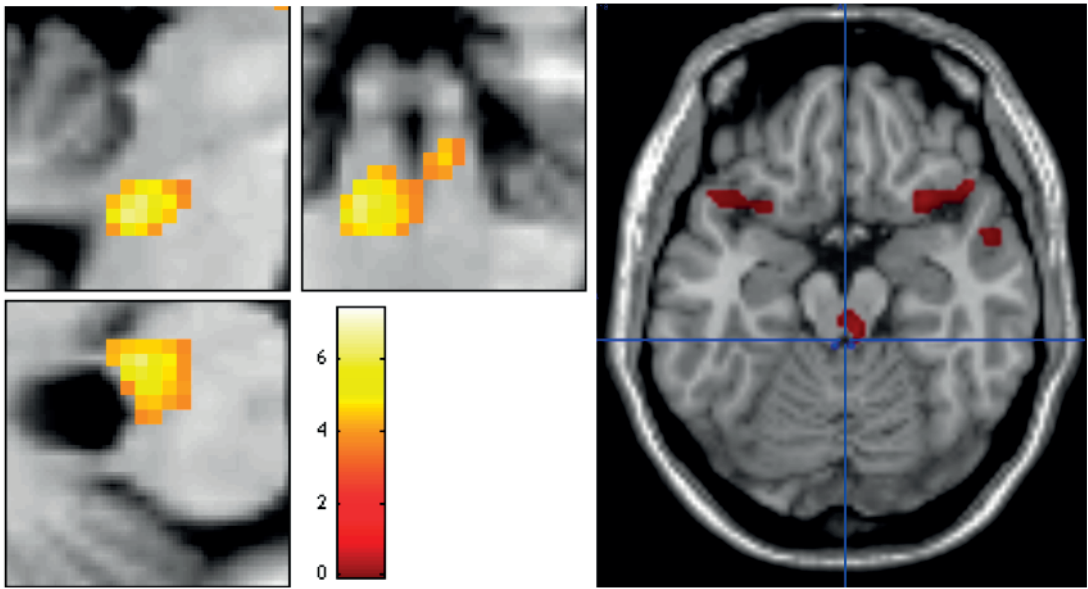


Figure 3.15 – Close up view of significant clusters in the region of the LC. The right panel shows the thresholded statistical map for surprise (red) overlaid on the canonical T1 image, as well as the Keren LC atlas (blue). The left panel shows the cluster extending posteriorly towards the left cerebellar peduncle in the axial view, as we would expect the LC to do. The coronal slice shows clusters bilaterally, sitting below the inferior colliculi.

to the LC.

We then performed non-parametric permutation tests for surprise. This test yielded significant clusters in the following regions ($p = 0.05$, FWE): bilateral superior temporal gyri, anterior cingulate cortex; right posterior orbital gyrus (extending in to the right anterior insula); and left anterior insula, inferior/middle frontal gyrus, middle temporal gyrus and caudate (see Appendix A, Table A.6).

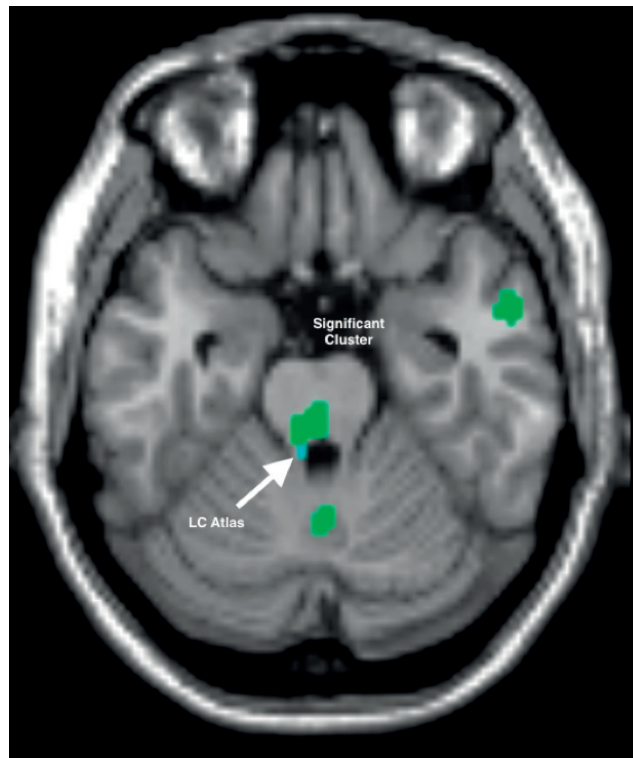


Figure 3.16 – Surprise contrast thresholded statistical map ($p = 0.001$) overlaid on the T1 canonical image (green). Also visible on the image is the Keren LC atlas (blue). As in the confidence contrast above and the oddball contrasts in Chapter 2, a cluster is found just at the anterior border of the map, precluding firm conclusions on LC localization.

Information After Card 2

We then examined BOLD responses for the information at Card 2. As in the other contrasts, we first perform a whole-brain a cluster level corrected analysis ($p = 0.001$, $k = 31$) which yielded significant clusters in bilateral occipital gyri; left supramarginal and superior frontal gyri; and right putamen and superior frontal gyri and then an SVC analysis with the previous analyses' threshold which yielded significant clusters in bilateral posterior and anterior insulae and dorsal striatum (caudate and putamen) (see Appendix A, Table A.7).

A nonparametric statistical test also revealed a distinct pattern of activation for information as compared to surprise, however, information gain also shows common activations with surprise, namely in the right anterior insula. This test yielded bilateral clusters in the cuneus; left supramarginal and lingual gyri, central operculum, calcarine cortex; and right anterior insula, superior occipital gyrus, superior frontal gyrus, postcentral gyrus, putamen and cerebellum (see Appendix A, Table A.8).

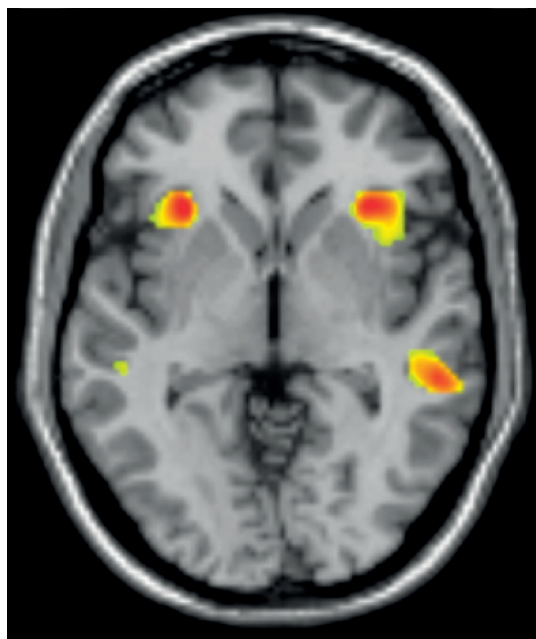


Figure 3.17 – Statistical maps of the non-parametric statistical maps for the surprise contrast overlaid on canonical image. The image shows clear bilateral insular engagement as well as a cluster in the right middle temporal gyrus.

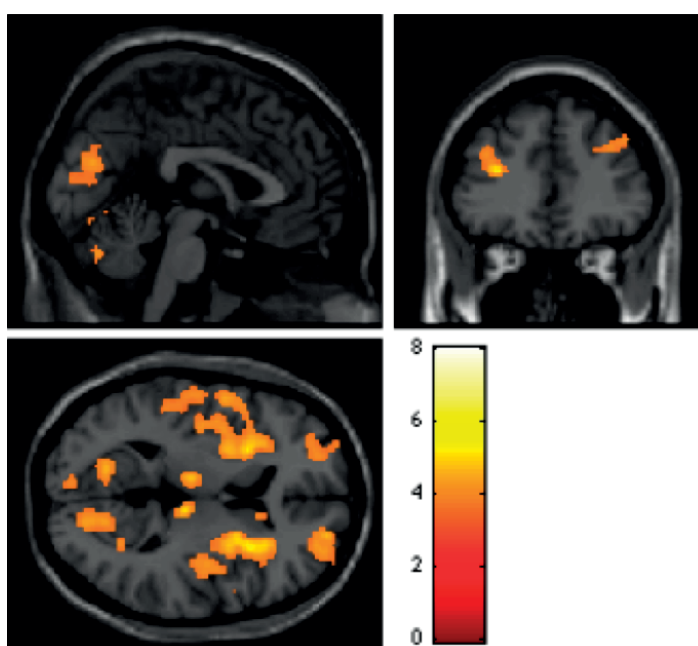


Figure 3.18 – Statistical map of the information contrast, thresholded at $p = 0.001$, cluster level corrected ($k = 31$) overlaid on T1 canonical image (0, 37, 9). The image above shows bilateral clusters in the insula and right putamen as well bilateral frontal gyri (axial view).

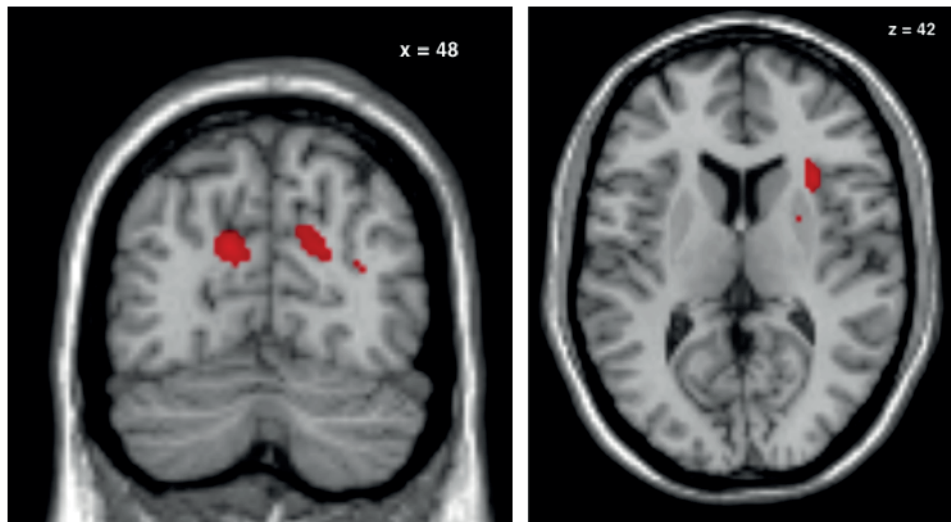


Figure 3.19 – Statistical maps of the non-parametric tests for the information contrast ($p = 0.05$, FWE) overlaid on the T1 canonical image. The coronal slice ($x = 48$), shows bilateral activity in the cuneus while the axial view ($z = 42$) shows a cluster in the right anterior insula.

3.4 Discussion

The results above show that 1) objective confidence correlates with the insula, a region previously linked to uncertainty processing; 2) surprise elicits activity in the insula, striatum, anterior cingulate cortex; 3) Information, as distinct from surprise, exhibits common activations in the insula and distinct responses in frontal gyri and cuneus for information. By using a formal account of all three measures while controlling for reward-related decision variables as well as task-related phenomena, such as overt learning and motor action, we link confidence, surprise and information to distinct neural correlates and further show an insular involvement in all three.

3.4.1 Confidence

Our results support the notion that confidence, or the degree to which one believes she is correct, occupies a particular role in decision-making variables. Confidence measures in human studies often suffer from being a self-reported, subjective measure. Here, objective confidence is easily captured by the inverse variance computed after card 1. We do not replicate the findings for neural signals of risk at the same time-point (Preuschoff et al., 2006). This negative finding may be due to several factors: first, we employ an auditory version of the same task; second, we use a different EPI sequence; third we test a different cohort. It is not unreasonable to suppose that sensitivity to risk versus confidence will differ across individuals, as one is the degree to which one believes she is incorrect and the other is the degree to which one believes she is correct. Thus a risk-averse individual may be more preoccupied with the former and in principle, this attribution of importance towards one form of uncertainty over

Chapter 3. Of Outcomes and Expectations: the Neural Representation of Confidence, Surprise and Information

another may be reflected in the BOLD response. Importantly, we note that a negative group result for prediction risk does not imply an either/or weighting of risk versus confidence. We assume an individual considers both variables in the decision-making process, only we did not find a BOLD response reflecting each in our study. Our sample size precludes us from investigating the risk versus confidence distinction in individuals but we can support the notion of confidence over risk, should the two compete for space in neural representation. Confidence guides our belief towards what we think will work, while risk guides our belief on what will go wrong. In our day to day, it may be beneficial to weight predictions more on what will work rather than on what will not, as the latter cannot guide our actions. In addition, as we used an auditory version of the task, the prediction period of 5.5 seconds instead of the original 7 seconds in the visual paradigm may have hampered the development of a (prediction) risk BOLD response.

The key confidence-related neural correlates found in our study are the insula and the middle temporal gyrus (MTG). We hypothesized that the insula would play a role in confidence assessment, as it is a computation that translates to a feeling, in line with the region's role in interoception. Confidence-related BOLD responses in the MTG offer an intriguing result as well. The region has previously been implicated in a prospective memory task (Okuda et al., 2003) as well as subjective confidence (Chua et al., 2006). Indeed, a meta-analysis of studies on confidence finds a consistent role left medial temporal gyrus for (subjective) confidence (White et al., 2014). While our initial research question was guided by the search for LC-NE involvement in uncertainty encoding, we did not expect confidence per se to elicit brainstem activity in the vicinity of the LC. Our SVC analyses using LC centered spherical ROIs yield significant clusters (at $p = 0.001$, uncorrected) however, the cluster extends into the anterior portion of the brainstem, which we have seen in chapter 2 provides an especially problematic area for which to assign specificity. The question remains as to what this brainstem cluster reflects, as our design matrices account for auditory, motor and first order effects such as reward prediction. With regards to confidence, it is of note that we do not find orbitofrontal cortex or striatal involvement for confidence as in other studies. This may be due to the fact that we control for valuation, which implicates the OFC, as well as reward prediction, which calls on striatal activation.

3.4.2 Surprise

The study of surprise has implications for both pattern recognition algorithms as well as human behavioral study. In machine learning, surprise details an unexpected outcome from a generative model whose computational definition can come in many forms. In humans, surprise can be characterized as a physiological and affective response to the unexpected. Formal accounts of the phenomenon include Shannon surprise (Shannon, 1948); Bayesian surprise (Itti & Baldi, 2009); and a predictive coding account of surprise (as absolute prediction error or risk prediction error). Shannon surprise, or the negative log of an event's probability, represents average surprise over time, which is not easily mapped to our paradigm.

Bayesian surprise captures the difference between prior (expected outcomes) and posterior distributions (actual outcomes) but, as an update, does not necessarily reflect an alerting signal. Importantly, Bayesian surprise must assume a shift in probability distributions, which subjects should not expect in our task, as none occur but for the difference between before and after card 1. Incidentally, this value of Bayesian surprise is equivalent to our prediction risk calculation at Card 1, which we input into our design and which yielded no significant results at acceptable thresholds. Studies investigating computational models of surprise fit to human behavior remain sparse and generally elicit surprise by introducing saliently different stimuli at low frequency (Michelon et al., 2003; Parris et al., 2009), however such stimuli incorporate phenomena that are related but distinct from surprise, such as novelty and incongruence. Here, by controlling for both the reward prediction error and the information gain in a gambling task, we can confidently assert that our measure of surprise captures surprise as error-detection specifically. The fact that activation patterns include the anterior insula meshes well with the notion of surprise as a feeling (Singer et al., 2009); surprise-related activity in the ACC corresponds to an event's conflict induction; the caudate's role in surprise fits well with its involvement in learning. Our initial aim was to identify an LC BOLD response for surprise that would corroborate previous studies correlating pupil response to surprise (Preuschoff et al., 2011). While we identify a cluster in the vicinity of the LC, it extends anteriorly into the brainstem, precluding a firm identification of a surprise-related LC BOLD response.

3.4.3 Surprise as learning signal

By design, our task does not include explicit learning or decision-making (except for time-points of no interest, including bet placement and trial outcome report) so as to allow for a clear picture of latent decision variables to appear in the BOLD response; specifically, all trials start with an equal probability of winning so no learning or strategizing can occur. As each trial starts with a flat prior, we did not employ a Shannon surprise metric to capture surprise, opting instead for absolute values of the prediction error. Nonetheless, in our model, we expect learning signals in the forms of prediction errors to arise merely in response to expectations made after cue presentation. Thus, surprise (as well as the reward prediction error) may encode a learning signal that adjusts predictions for a given cue. Should we not see a surprise signal, we can hypothesize that learning will be diminished, an interesting, if distressing, state of affairs that may suggest dysfunction and even pathology, such as in problem-gamblers or other forms of addiction. Indeed, Parkinson's disease, whose neural signature begins in dorsal striatal degeneration, is known to increase the risk of addictive and compulsive behaviors (Voon et al., 2006). While some evidence points to the role of dopamine replacement therapy in the etiology of PD-related addictive behaviors (Vaillancourt et al., 2013, Cools et al. 2001), studies on impaired reward processing have yielded mixed results (Schott et al., 2007; van Eimeren, 2009; Euteneuer et al., 2009). Our findings thus offer an opportunity to investigate whether it is specifically an inability to experience surprise that is compromised in patients.

Chapter 3. Of Outcomes and Expectations: the Neural Representation of Confidence, Surprise and Information

3.4.4 Confidence and Information as Precision and Accuracy

The brain is often cast as a type of machine and thus applying concepts from the engineering sciences may aid us in elucidating its function. Confidence and information in our study were viewed through the prism of the brain as an inference machine, within which an agent computes several measures to best resolve the problem of uncertainty. The application of Shannon surprise to human behavior is appealing because the formulation is simple and intuitive. However, while surprising events can be informative, information may not be surprising. It has been shown that when information (in the form of Shannon entropy) is high, surprise is low. In the case of an ambiguous prediction, say hovering around 0.5, the information provided by an outcome will be high, as the outcome can take several possible states and will resolve the maximum uncertainty. By contrast, in such a case, surprise will be low, as we have no dominant expectations. Thus equivocating surprise with information may be counterintuitive, a quandary highlighted by Itti & Baldi (2010). In addition, this distinction is formalized in the Free Energy Principle as both quantities (surprise and entropy r) together define Free Energy. While Free energy is usually formalized specifically as Shannon surprise and a KL divergence, it does not preclude its formulation as absolute prediction error and Shannon entropy. Indeed, we cast the latter as a type of learning, as information, or confidence error, in our paradigm represents an information gain at the trial outcome, relative to the confidence (in Shannon terms, the redundancy (Shannon, 1948)) at card 1. We differentiated the two phenomena and control for their respective contribution to the neural signal at the same time-point. Our study clearly disambiguates the two values by using a predictive coding account of surprise instead of Shannon surprise. As mentioned, the information gain at outcome can be viewed and computed as a confidence error. It is the distance between truth and confidence. Indeed, one minus confidence yields a quantity that is equivalent to Shannon entropy, multiplied by some constant. Our imaging results further offer an intriguing story regarding information gain on the brain; the insula remains correlated with information, again suggesting that this region is crucial for translating lower-level prediction error signals into states of awareness (Singer et al., 2009). Other information-related regions include frontal regions and the cuneus. The cuneus in particular has previously been implicated in learning rates (Payzan-LeNestour et al., 2013). Notably, the anterior cingulate cortex in information-related statistical maps is absent. This underlines the quality of information gain, in that it does not represent a conflict, even while being a prediction error.

We conclude here that objective confidence is to information, what precision is to accuracy. The information gained from an outcome in a given probabilistic environment encapsulates the difference between a confidence prediction (the precision) and the certainty of the terminal state. Like in other cases, it may be the information processing that is specifically impaired in various decision-making dysfunctions and thus it is worthwhile to track its emergence in probabilistic paradigms.

4 Are Perceptual Switches Surprising?

4.1 Introduction

Sensory perception is an automatic process that, in healthy humans, is not subject to deliberation and does not necessarily require conscious awareness. Thus the feature of uncertainty in sensory perception may be puzzling at first glance, since we perceive our environment with such ease and efficiency. Nonetheless, this deception is likely a credit to unseen work performed in the brain. If we consider first the noise surrounding the stimuli we perceive, then we must account for perceptual uncertainty processing. Second and more obvious, when we concede that individuals perceive the same stimuli differently, we must again deduce that the brain integrates uncertainty about a stimulus, with different outcomes for different people. Third, an individual does not necessarily map the external world to an internal representation in a consistent manner over time, which hints at an inferential process with regards to perception. Finally, when we consider the phenomena of visual illusion, the evidence of perceptual uncertainty is laid bare: if we can give two names to a same stimulus, which one is the true name? In questioning the mechanisms of visual perception specifically, the 11th century scientist Ibn Al Haytham (Alhazen) was the first to propose that visual perception requires an inferential process (Howard, 1996; Pelillo, 2016) that leads to a judgment, albeit a hidden one. Alhazen's treatise was prescient in that he challenged the notion that 1) visual perception was a holistic, instantaneous process; 2) suggested the mind (soul) resolved the process, not merely the eye. Later on, Hermann von Helmholtz (von Helmholtz/Southall, 1925) coined the term unconscious inference with regards to visual perception, and posited that such a process supported perception. With the use of fMRI and a predictive coding account of bistable perceptual switches, we peek under the hood of observed phenomena to capture the neural processes implicated in the perception of ambiguous visual stimuli.

4.1.1 Inferential nature of Perceptual decision-making

What is meant by perceptual decision-making? Can we choose to perceive the environment as we want? It is important to distinguish here and elsewhere that perceptual decision-making

Chapter 4. Are Perceptual Switches Surprising?

generally removes the 'I' from the agent. That is the decision is not declarative or intentional and its conscious component appears spontaneous. Nonetheless, unbeknownst to us, a decision-process is taking place. The surrounding environment is noisy and thus perception is probabilistic - a given stimulus can take on more than one state to the observer. In order to function however, the observer must infer what the state is. Thus we can apply the framework of the brain as an inference machine to perceptual problems as well. While it seems that visual perception is passive and automatic, the process involves a construction of sorts, because the input to the visual systems is riddled with noise and there is a variability inherent to the external to internal mapping of stimuli. Thus an inferential process is likely at work resolving uncertainty (Gregory, 1997; Alais & Blake, 2005). With some stimuli, such as visual illusions, this ambiguity is made flesh in that the inferential process breaks through the consciousness barrier. With such stimuli, we can actually report on perceptual changes. These illusions generally have more than one conformation and the most common type are bistable illusions, such as the Necker Cube (Necker, 1832). How do bistable illusions arise? Stimuli that can't be resolved into one representation offer the viewer more than one perceptual interpretation. That is the stimulus 'flips', or 'reverses' from one percept to another. This feature is a property of the unchanging stimulus. Thus if the stimulus is stable, why is our perception not? What happens when we perceive a switch in an illusion?

4.1.2 Visual Illusions: Perceptual Uncertainty Manifest

Visual illusions offer an ideal way by which to examine perceptual inference questions, as the change in perception in the face of a stable stimulus clearly delineates an object's ambiguity and thus the probabilistic nature by which we recognize it as one thing or another. Illusions are not limited to the visual domain and can also come in many forms, including motion from stability; multi-stable illusions and binocular rivalry (Gregory, 1997). Below, we will briefly go over studies on the latter two.

Binocular Rivalry

When two objects are simultaneously presented to each eye, a perceptual conflict occurs, where stimuli compete for conscious representation. This phenomenon is termed binocular rivalry and is a form of perceptual uncertainty. Einhäuser and colleagues (Einhäuser et al., 2014) investigated brain regions involved in either passively viewing or actively reporting perceptual changes. Perceptual changes were predicted by pupillometry. By using fMRI, they determined how the perceptual switch alone (in the passive condition) engaged specific brain regions and how these differed from brain regions responding to the active condition, where subjects had to report the change when it occurred. They find frontal areas implicated in the active condition only, suggesting that monitoring for the reporting of the switch as opposed to the switch per se drives neural activity. The switch itself implicated lower-level brain areas. The stimuli used to induce binocular rivalry were dynamic and static gratings. By measuring the optokinetic nystagmus induced by the moving grating, a subject's dominant percept could

be inferred. For the static grating, pupil dilation was measured as an index of perceptual change.

One advantage of binocular rivalry is that many different stimuli can elicit perceptual bistability, including simple stimuli that engage low-level visual areas and more complex images that can easily and predictably be dissociated in imaging studies (Sterzer et al, 2007). Thus these stimuli can induce a conflict at both the visual and perceptual levels. Intriguingly, the temporal dynamics of binocular rivalry are similar to those of bistable illusions, as are the effects of interrupting stimulus presentation, which suggests a common mechanism for perceptually ambiguous stimuli. These stimuli differ from bi-stable stimuli in other ways however; for instance, an intermediate or mixed versions of the stimuli can be perceived. Binocular rivalry further differs from ambiguous stimuli in that it is in fact two different stimuli that are presented to a subject. Imaging studies using binocular rivalry show evidence of activation for the suppressed image in regions implicated in specific stimulus categories (e.g. faces). This feature is found in visual illusions as well (Moradi & Heeger, 2009). However, in a study using the Rubin face-vase figure, it was found that fusiform face area activity prior to stimulus presentation was higher in trials where the face percept was ultimately reported, which provides evidence for a model in which random fluctuations in neural systems bias one perceptual state over another (Hesselman et al., 2008). Rees (Rees, 2007) points to the extrastriate visual cortex as a candidate for conscious visual perception, however, activity in certain regions is not always associated with conscious experience.

Dynamics of Perceptual Decision Making

Above, we hint at the importance of perceptual switch dynamics. Predictable, periodic perceptual switches hinder the notion that an inferential process is at play to some degree, because such a time constant would imply a true, binary switch, from one receptive field to another. But evidence of such a regular dynamic is not found. Two theories are offered in the literature regarding the dynamics of the neural impetus to switch. As mentioned, one view is that an adaptation occurs to one percept and those neurons, once fatigued, hand over the reins to another set of neurons that are specific to an alternative percept. This model assumes a periodicity to the perceptual alternation, as there is no reason to believe that these processes vary in time. The model is termed an oscillator model (Moreno-Bote et al., 2007). A challenge to this model is the attractor network, where attractor states represent alternative percepts as troughs in the system; noise drives perception towards one or the other minimum ('attractor') state in a stochastic manner. Both models have found support in experimental studies (Moreno-Bote et al., 2007). A significant challenge to the oscillator model is that dominance times, or time spent perceiving a given percept, are not fixed but vary stochastically. This feature of perceptual multistability supports an attractor network model of perceptual switches, where random noise accumulates over time until a threshold is reached, at which point conscious perception lands in a different trough. In the case of bistable illusions, a first percept is 'chosen' at random; fluctuating activity will then, over time, eventually overcome the decision barrier and settle

Chapter 4. Are Perceptual Switches Surprising?

in an alternative local minimum, which is the other perceptual state. Thus, when viewing a percept in a given state, perception destabilizes over time (on average, 5-6 s) (Kornmeier & Bach, 2012). Morena-Bote et al. (2007) propose a hybrid model of attractor and oscillator models that describes observed data.

Modeling Perceptual Decision Making

Can Bayesian frameworks explain perceptual uncertainty? As seen above, the theory that multistability arises from neural adaptation alone is no longer a widely held view. Evidence points to an interplay between top-down and bottom-up processing (Gregory, 1997). Thus it is possible that, in making a perceptual decision on a particularly ambiguous stimulus, many factors are integrated to generate a probabilistic model of a percept. In such a case, ambiguity may be resolved through Bayesian inference. Sundareswara and Shrater (2008) investigated this possibility by using a Necker Cube paradigm. The authors biased the dominant percept of the Necker Cube by surrounding the stimulus with flankers of disambiguated Necker Cubes. Such contextual stimuli have previously been found to influence the amount of time spent in a given percept (the one in concordance with the flankers' orientation) and this bias was found in their behavioral experiment. They further replicated the known bias that individuals tend to see the cube in its from-above orientation more often. The increase in time spent in a biased percept and a decrease in time spent in the opposing percept indicated to the authors that ambiguity was decreased. The authors exploited this variation to propose a Bayesian inference model using Markov Renewal Processes that explained switching behavior by attributing differences in percept probability based on time spent in a given perceptual state. Weinhauer et al. (2017) have revisited Bayesian frameworks to provide a neurocomputational account of bistable perception, this time casting the switch as a prediction error. In using a predictive coding account to model perceptual switches in Lissajous figures during an fMRI experiment, they find neural correlates in the anterior insulae as well as inferior frontal gyrii, which supports our hypothesis of uncertainty processing occurring in this region in particular. While a predictive coding account has been set forth to explain bistability in binocular rivalry (Hohwy et al., 2008; Dayan, 1998), the two studies above are, to our knowledge, the only empirical studies ones investigating visual perceptual switches as prediction errors.

4.1.3 Neural correlates of bistable perception

Competing theories on the neural correlates of visual bistable perception include the hypothesis that the switch in perception arises from low-level visual areas versus the notion that it is non-sensory, cortical processes that resolve ambiguity. In support of high-level modulation of perceptual switches, an fMRI study investigating auditory bistability, Kondo & Kashino (2007) found prefrontal cortex activation specific to endogenous perceptual switches, as opposed to a change in perceptual stimulus.

In binocular rivalry, the BOLD response in the V1 region of visual cortex can predict changes

in perception (Haynes & Rees, 2006). Indeed, TMS over V1 has been found to cause perceptual alternations (Pearson et al., 2007). Evidence for Lateral Geniculate Nucleus (LGN), another early visual processing region, involvement in perceptual switches has also been found (Wunderlich et al., 2005). Taken together, these findings suggest low-level visual processing in perceptual switches, however, electrophysiological studies found no evidence of early visual processing regions' involvement on perceptual fluctuations (Sterzer et al., 2009). When faced with perceptual decisions, macaques show an increase in activity in lateral intraparietal (LIP) neurons, which becomes steeper with more difficult decisions. Further this study showed that response (in this case an eye movement) was consistently initiated when activity summed to a specific threshold, which suggests that the decision process happens over time with the accumulation of information (Huk & Shadlen, 2005). There is conflicting evidence regarding implicated brain regions depending on the methods used to assess neural activity and this conflict may result from the fact that electrophysiological methods measure spike trains while perceptual processes reflect changes in local fields potentials; because fMRI and EEG are more likely measuring local field potentials rather than neural spikes (Logothetis et al., 2001), evidence culled from studies using the latter may better capture activity related to perceptual processes.

Event related potentials in the parietal region roughly 250 ms before a button response were found with motion reversals in an EEG study using stroboscopic stimuli as well as Necker Cube Switch reversals (Struber and Herrmann, 2002) and with orientation reversals of the Necker cube (Struber et al., 2001). Evidence of decreased neural activity is found in ambiguous versus disambiguated stimulus variants (Leopold and Logothetis, 1999; Kornmeier and Bach, 2006; Moreno-Bote, 2007; Pitts et al., 2010). Kornmeier and Bach (2009) postulate that this dip in neural activity may reflect the so-called flatness of attractor states more ambiguous states should have a lower barrier of transition than disambiguated variants. In studies investigating stimuli that spontaneously switch, extrastriate visual activity is correlated with conscious perception (Tong et al., 1998). Most functional neuroimaging studies consistently show activity in parietal and frontal regions in response to perceptual switches (Sterzer & Kleinschmidt, 2007; Sterzer et al., 2009; Kanai et al., 2010; Britz et al., 2011; Megumi et al., 2015; Zaretskaya et al., 2010; Kornmeier & Bach, 2012) though not all (Fraessle et al., 2014). These findings lend support to a top-down modulation of visual processing from higher-order cortical areas though it has been argued that fronto-parietal involvement may reflect downstream activation from visual areas following a perceptual change. Temporal resolution in fMRI does not readily allow for a temporal disambiguation in brain activity although Sterzer & Kleinschmidt (2007) have found a right PFC activation prior to a V5/MT activation during a perceptual switch. An EEG study also found that right parietal activity came before a perceptual switch in a Necker Cube paradigm (Britz et al., 2009). Taken together, a review of the neural correlates of bi-stable perception presents an inconclusive picture. Evidence for low- versus high-level neural region involvement is mixed. As we have seen, Sterzer, Kleinschmidt and Rees (2009) suggest that cortical activity drives the perceptual switch while other groups have found low level regions driving the phenomenon (Kornmeier & Bach, 2004; Kornmeier & Bach, 2005). One must

Chapter 4. Are Perceptual Switches Surprising?

consider the fact that different studies employ varied bi-stable stimuli, each of which may be endowed with non-generalizable features. From a predictive coding perspective, such an ambiguous state of affairs indeed offers an opportunity to investigate perceptual switches as prediction errors, as predictive coding includes bi-directional processes in its formulation (Rao & Ballard, 1998).

4.1.4 When does the reversal occur?

The most straightforward way to report a perceptual switch is via button press, which introduces sources of error with respect to the motor response and reaction time. That is, how good of a temporal precision does the manual response provide in relation to the conscious experience of the switch? The perceptual switch necessarily occurs before the response. This question was addressed in an EEG study of the Necker Cube. By using depth cues, the Necker cube was disambiguated into its respective percepts and reaction times measured for the latter. Median response time was 616 ms (Kornmeier et al., 2011).

Another manner in which to assess true reversal times is by interrupting stimulus presentation (Orbach et al. 1966). Necker cubes were presented for varying durations with varying ISIs. It was found that switch frequency decreased with ISIs longer than 400 ms, however, for ISIs shorter than 400 ms, reversal rates increased as compared to continuous presentation. Intriguingly, switches occurred close to onset for ISIs shorter than 400 ms. O'Donnell et al (1988) exploited this feature of Necker Cube reversals in an EEG paradigm. By presenting ambiguous and unambiguous cubes, they found a fronto-parietal positivity in both exogenous and endogenous stimulus change that occurred roughly 550 ms after onset of a disambiguated cube and 650 ms after onset of an ambiguous cube. It is hypothesized (Struber and Hermann, 2002) that this positivity indicated a consciousness of the reversal, leaving the switch to occur sometime between 0 and 650 ms after stimulus onset. Kornmeier and Bach (2012) point out that all EEG experiments find a slow positivity from parietal to frontal dominance which occurs between 400 and 500 ms after stimulus onset and 100-150 ms prior to response in the onset' paradigm. In the manual response paradigm, this positivity peaks at 250 ms prior to motor response. The parietal positivity according to several authors correlates to a P3b component, which may reflect an awareness, or a consciousness of an event (Dehaene and Changeux, 2011).

Kornmeier and Bach used a Necker Cube Paradigm with a presentation time of 800 ms, to allow for detection of a P300 potential and to minimize the chance of a switch occurring in that time, and an ISI of 400 ms. Subjects were asked to report their percept following presentation (during the ISI to control for ERP signals related to motor-planning) in two conditions. In one, they were asked to report a current percept that differed from a previous one (indicating a reversal) and in the other, a current percept that was the same as the last one (indicating a stability). They further presented subjects with disambiguated versions of the cube in a control condition. The endogenous reversal condition elicited a positivity (RP) at

130 ms post-stimulus onset that was most pronounced at the occipital lobe. This signal has been found across different bi-stable stimuli. The authors suggest this signal arises from a decision conflict, which would contribute to an increase in the response time for perceptual switches in ambiguous stimuli (Kornmeier & Bach, 2005). In studies investigating perceptual switches against so-called 'replay' conditions, where disambiguated versions of the stimuli are presented as a control condition to the bi-stable percept, numerous studies report a right lateralization in fronto-parietal areas (Brascamp et al., 2015; Britz et al., 2008). Finally, support for frontal and parietal cortices' involvement in perceptual reversal comes from lesion patients, who experience attenuated perceptual switches (Rees, 2001; Kanai et al., 2010; Scoccia, 2014).

4.1.5 Other Modalities

While most research has focused on visual perception, multistable perception exists for auditory as well as tactile stimuli (Gregory, 1997; Kondo et al., 2017). Evidence suggests that, while intra-subject variability exists across modalities, these phenomena emerge from common mechanisms. For instance, Kondo & Kashino (2007) found that dwell times for auditory percepts follow a gamma distribution, just as they generally do in visual paradigms. Auditory multi-stability presents in two general classes: auditory streaming and verbal transformations (Warren & Gregory, 1958). While common points may exist between different sense modalities, the auditory multistability field presents a research area that can stand on its own; wherefore, a detailed review extends beyond the limits of this thesis.

4.1.6 Conscious or Unconscious?

When the perception of a stimulus changes while the stimulus stays the same, we can put forth questions on consciousness. When are we conscious of the switch? That is when are we able to recognize and declare that a perceptual switch has occurred? Is there a process immediately preceding the switch that can be attributed to unconscious processes, notably relating to the resolution of ambiguity? Kornmeier & Bach (2012) found that a decision on a percept occurs at least 340 ms before a subject can respond. In reviewing several EEG studies on bistable perception, they suggest that somewhere between 250-500 ms before a key press, conscious recognition of a switch occurs, though the switch may happen before then. Further, 1000 ms prior to a button response is when gamma power increases and alpha-power decreases, activity that suggests higher-level cognitive processes (Kornmeier & Bach, 2005). The authors posit that a percept switch differs from a perceptual choice. This distinction may specifically relate to each event's quality of consciousness that is the unconscious decision bound and the conscious one.

Finally, what does a perceptual switch represent? Our nominal answer to this question is that a perceptual switch represents uncertainty in a stimulus. One possibility is that the switch encapsulates a perceptual prediction error. As mentioned, a very recent fMRI study examined this possibility. Weilhhammer and colleagues employed a model-based fMRI experiment em-

playing Lissajous figures, a bistable illusion, by fitting a predictive coding model to behavioral data. Their prediction error model elicited a significant activation in the right anterior insula (Weilnhammer et al., 2017), a region implicated in other forms of prediction error.

4.2 Perceptual Uncertainty Experiment

We endeavored to investigate the neural correlates of perceptual uncertainty by employing a stimulus that induces a bi-stable illusion. We specifically sought to capture a similar pattern of neural activity for perceptual switches by casting these as prediction errors, just as we do in the financial uncertainty task and performing the experiment on the same subjects that performed the financial uncertainty task. We examine the question of perceptual inference in an fMRI paradigm by exploiting the ambiguity of the Necker Cube, a structure drawn by Swiss crystallographer Louis Albert Necker in 1832 that induces an optical illusion, an overt form of perceptual uncertainty (Necker, 1832).

The Necker Cube is a transparent cube whose vertices are visible to a subject. Thus an individual can perceive the cube in one of two conformations, from below or from above, with the lower face or upper face appearing in the foreground. The (conscious) perceptual switch between the two percepts happens spontaneously. The ambiguity inherent in the illusion thus offers a means by which we can induce perceptual uncertainty in a laboratory setting. The Necker Cube in its standard conformation can be said to be ambiguous: that is, there's an about equal probability of perceiving the cube in one way, or in another. In order to better capture different levels of uncertainty and thus mimic our cognitive uncertainty task, we modified the Necker Cube so as to bias a subject's perception towards one conformation or another. By warping the angles of the vertices in an incremental fashion, we introduced different levels of bias to the cube (Einhäuser et al., 2004). It is of note that 1) the ambiguous Necker Cube is already biased towards a from above percept; 2) the Necker Cube's ambiguity is vulnerable to individual subjectivity (Einhäuser et al., 2004). Thus, while we objectively bias the cube towards one percept or another in an incremental fashion, it is not guaranteed that an individual's subjective perception will match this objective distortion.

In order to investigate a perceptual form of uncertainty that most closely matches our financial uncertainty task, we administered two separate tasks to subjects. The objective of the first task was to obtain a participant's individual psychometric curve fit to our objectively biased stimuli. In the second task, we present biased stimuli from the participant's own subjective curve to determine what happens in the brain when a given subject experiences a switch for a stimulus biased to a given degree or, what the neural correlates of increasing (decreasing) uncertainty are for a perceptual stimulus. As in the Card Game, because the switches happen spontaneously, we control for neural activity related to decision-making variables of no interest (such as deliberation, anticipation and learning), thus capturing a succinct account of neural activity related to uncertainty processing. Further, our task offers no monetary incentive, thus eschewing reward-related variables.

In the following experiment, we investigate the following questions: 1) In applying a formal account of risk, as variance, to viewing the Necker Cube, will we observe a BOLD response? Will this BOLD response occur in the insula and striatum? 2) Can we model the perceptual switch as a prediction error? Will such a modulation yield a BOLD response? Specifically, can we apply a measure of surprise to the perceptual switch and in so doing observe similar neural patterns to that found for surprise in the financial uncertainty task?

4.2.1 Procedure and Task Description

Sessions for the perceptual uncertainty task were scheduled to occur between 5 and 10 days following the financial uncertainty experiment. Subjects were sent electronic versions of information relating to the risks and benefits of the experiment the night before their scheduled session. On the day of the experiment, this information was reviewed with the subject before obtaining their written and informed consent. Subjects were then asked to undergo screening prior to the scanning session to ensure their compatibility with the procedure, as well as to collect relevant information on their medical history, including, specifically, any history of neurological or psychiatric disorders or psychotropic drug use in the past year. We then provided an oral explanation of the Necker Cube tasks before proceeding to the scanner room. The experimental paradigm was coded in Matlab using the Psychophysics ToolBox (Brainard, 1997; Pelli, 1997; Kleiner et al., 2007) and incorporated the Eyelink Toolbox (Cornelissen et al., 2002) functions. Following the end of the experimental session, subjects filled out two questionnaires: the trait component of the State and Trait Anxiety Inventory (STAI-T, Spielberger & Sydeman, 1994) and the DOSPERT risk preferences scale (Weber et al., 2002). They were then paid for their time as well as their payout from the Card Game, and debriefed.

Participants

The study was conducted on 23 subjects (9 f) with a mean age of 25.13 years. Twenty of the subjects also participated in the Financial Uncertainty Study. For seventeen of the above subjects, MPM maps (1.5 mm) as well as diffusion (NODDI) data were also acquired. Pupil data was obtained for 20 of the 23 subjects. Participants were recruited via paper and on-line advertisements targeting the student populations of Ecole Polytechnique Fédérale de Lausanne and Université de Lausanne. Exclusion criteria included metal implants, previous psychiatric illness, and psychotropic drug use within the past year. Inclusion criteria included proficiency in English.

Imaging Procedure

All scans were acquired on a Siemens 3T Prisma, at the Centre Hospitalier Universitaire Vaudois in Lausanne, Switzerland. Once in the scanner room, subjects were fitted with a heart-rate monitor on their finger and respiration rate sensor around their mid-section, for

physiological monitoring via BioPac (BioPac Systems Inc., Essen, Germany). Once settled in the bore, we acquired first a localizer scan, followed by a gre-field mapping scan to correct for magnetic field inhomogeneities. We then alerted subjects to the beginning of the task and EPI sequence. In a first instance, prior to EPI acquisition, we performed a 3-point calibration on the in-scanner eye-tracker (EyeLink 1000, SR Research, Ottawa, Ontario), as we aimed to collect pupillometry measures related to task activity. We then displayed task instructions on the screen projected within the bore. Subjects could then signal their readiness for the scan, upon which the EPI sequence was launched and task 1 commenced. Task 1 lasted 20 minutes. At the end of task 1, a subject's individual psychometric curve was derived and input to task 2, a process which lasted roughly 1-2 minutes. Following this, we alerted subjects that task 2 would start shortly, whereupon we performed the eyetracking calibration again before launching the second EPI sequence, along with task 2. Parameters for our EPI sequence were the same as those used in the Card Game (see Chapter 3 for detailed sequence parameters): 2D EPI, Multi-Echo sequence, 3 mm resolution, with a TR=2.72 s. Following the task sessions, subjects were told they could close their eyes and rest while we acquired MPM maps (1.5 mm) as well as diffusion data (NODDI sequence). As all subjects performing the Necker Cube task had previously performed the Card Game task, we did not reacquire a subject's individual anatomical scan (T1 MPRAGE, 1mm resolution; see Chapter 3 for detailed sequence parameters). The task related sessions lasted approximately 40 minutes while the imaging session lasted approximately 50 minutes; the whole experimental session, including intake and debrief lasted approximately 90 minutes.

Necker Cube - Task 1

In task 1, subjects were presented with 21 different versions of the Necker Cube. The stimuli differed in the angles of their vertices that biased the stimulus towards being perceived from above (+1) or from below (-1), and this at regular intervals of 5% from the fully ambiguous Necker Cube (at 0). Each of the 21 stimuli was presented 10 times, for 3 seconds each. Following the 3-second presentation time, subjects had a two second time window during which they could report their last percept (in case a switch had occurred during the viewing period). Thus each trial lasted a total of 5 seconds and subjects responded to 210 trials in total (Figure 4.2).

Upon completion of the task, subjective percentages of percept reports were tallied for each stimulus (a frequency of a given percept out of 10 same stimuli presented) and fit to a sigmoidal curve. This fitting procedure allowed for the creation of subjectively biased stimuli for the subsequent task. Importantly, a subject's true ambiguous cube value was determined as the inflection point of the curve; this point is dubbed the point of subjective equality (PSE) and represents the point at which a subject perceived the cube from above or from below in an equiprobable manner(see Figure 4.3).

$$\frac{1}{1 + e^{-k*(x-PSE)}} \quad (4.1)$$

4.2. Perceptual Uncertainty Experiment

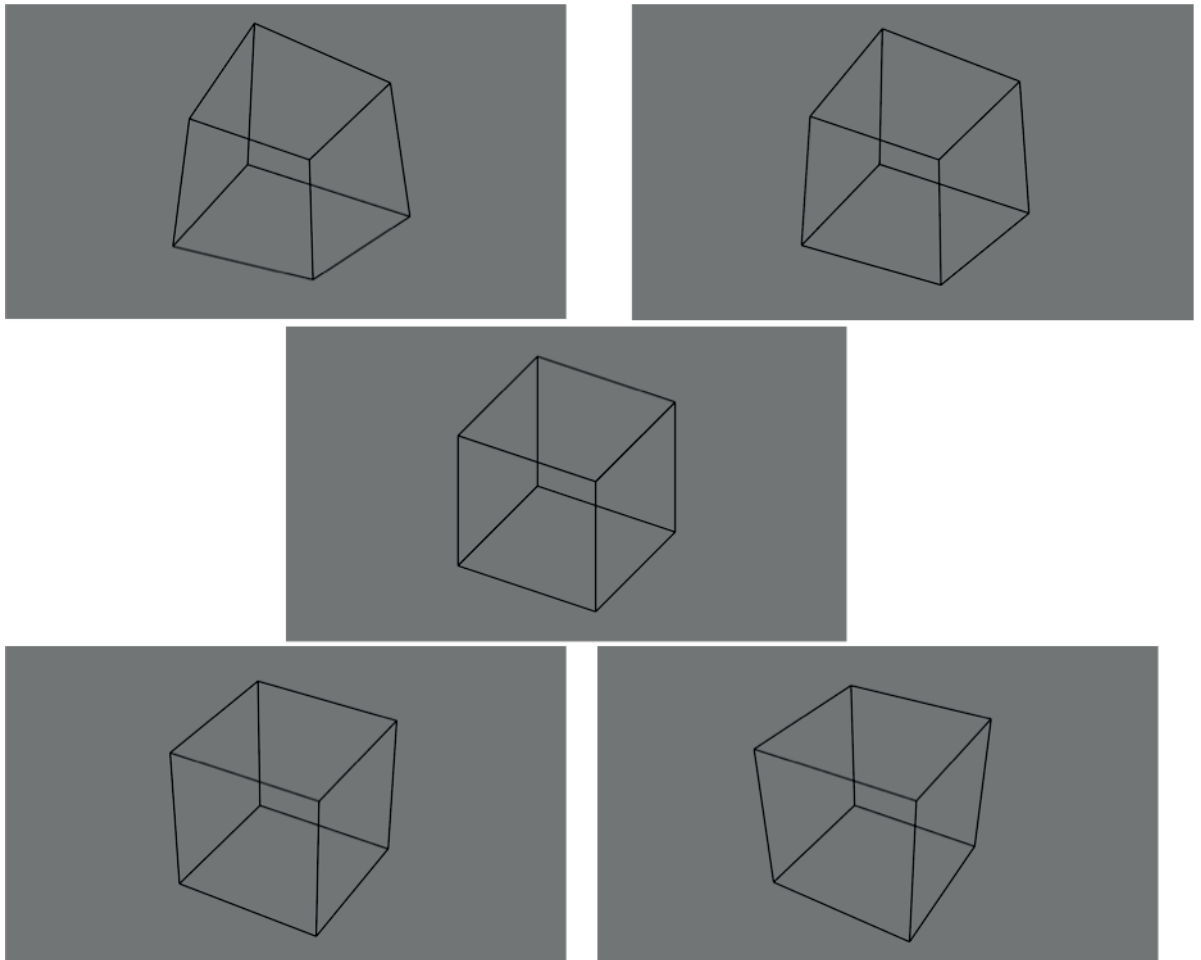


Figure 4.1 – The Necker Cube Stimuli. The two top cubes show stimuli biased towards a from below view (100% bias towards from above in the left panel, 50% bias towards from below in the right panel). The fully ambiguous cube is seen in the middle. The bottom row shows stimuli biased towards a from above view (50% bias towards from above in the left panel, 100% bias towards from below in the right panel).

In task 2, subjects were presented with 5 different kinds of stimuli, subjectively biased at 20%, 35%, 50%, 65% and 80% towards a from above percept, based on their individual psychometric curve. Stimuli were presented on a grayscale screen projected into the bore. These stimuli were presented in 11 trials, for 80 seconds at a time, in a random order. Each stimulus type was presented twice, but for the fully ambiguous cube, which was presented 3 times. During the trial, subjects were asked to report their first percept with a manual button press and to indicate perceptual switches via button press as they arose, if they arose. Button presses were made on an MRI compatible response box. Individual subjects were asked to report "from above" responses on one side of the button box, and "from below" responses on the other side. To avoid lateralization confounds, these sides were switched for each subject. Four of the 11 trials also included a "replay" condition, where heavily disambiguated Necker Cube stimuli

Chapter 4. Are Perceptual Switches Surprising?

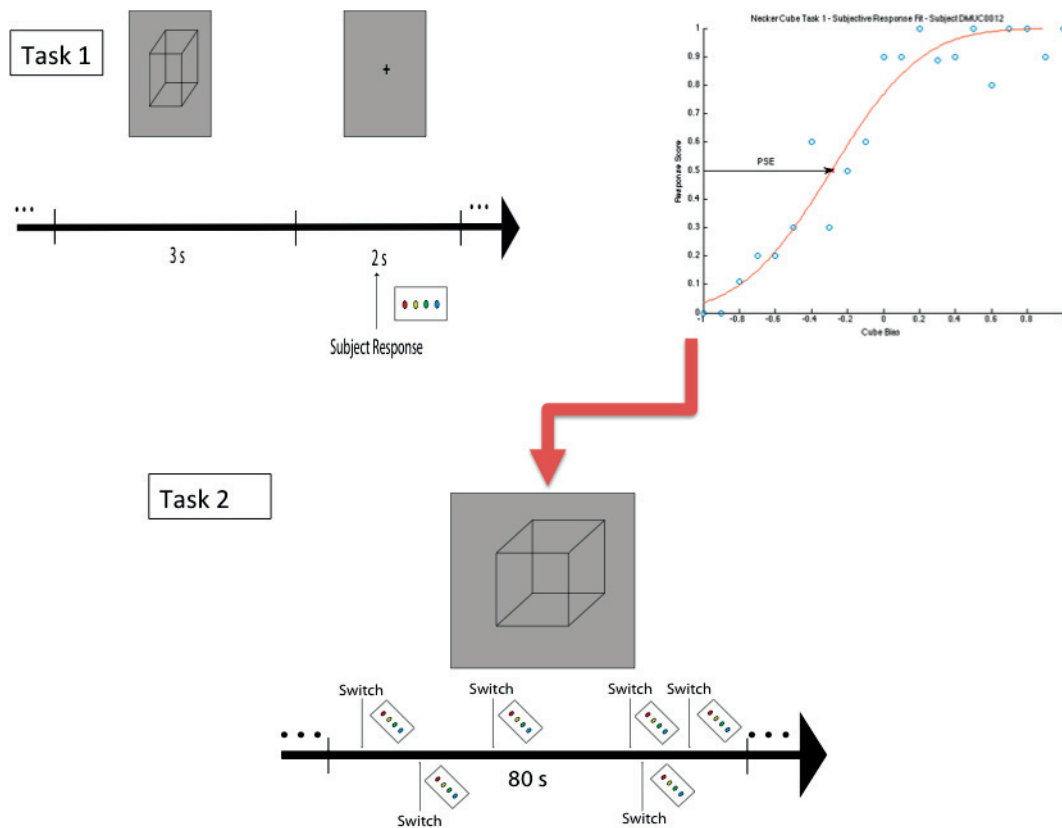


Figure 4.2 – An overview of the perceptual uncertainty experiment. In task 1, subjects are presented with 21 versions of the cube, 10 times each, for a total of 210 trials. Subjects report their percept following a 3 second viewing period. Results of this task generate an individual psychometric curve, which is then used to derive 5 subjective stimuli, biased towards a 20%, 35%, 50%, 65% and 80% likelihood of being viewed from above. In task 2, subjects are presented with each of these 5 stimuli for 80 seconds at a time, and asked to report their initial percept, and then subsequent perceptual switches, if and when they arise.

were presented at the same time-points that prior switch reports were made in the previous experimental trial.

4.3 Necker Cube Behavioral Results

4.3.1 Behavioral Results: Necker Cube Task 1

In a first instance, we plotted the distribution of several measures relating to the individual results of the fitting task. As our manipulation of the Necker Cube constitutes a novel task, we deemed the following tests necessary to ensure that the stimulus manipulation did indeed yield the desired and expected outcomes. Subject responses from task 1 were fit to a sigmoid curve to determine subjective perceptions of objective stimuli. From this fit, several parameters can

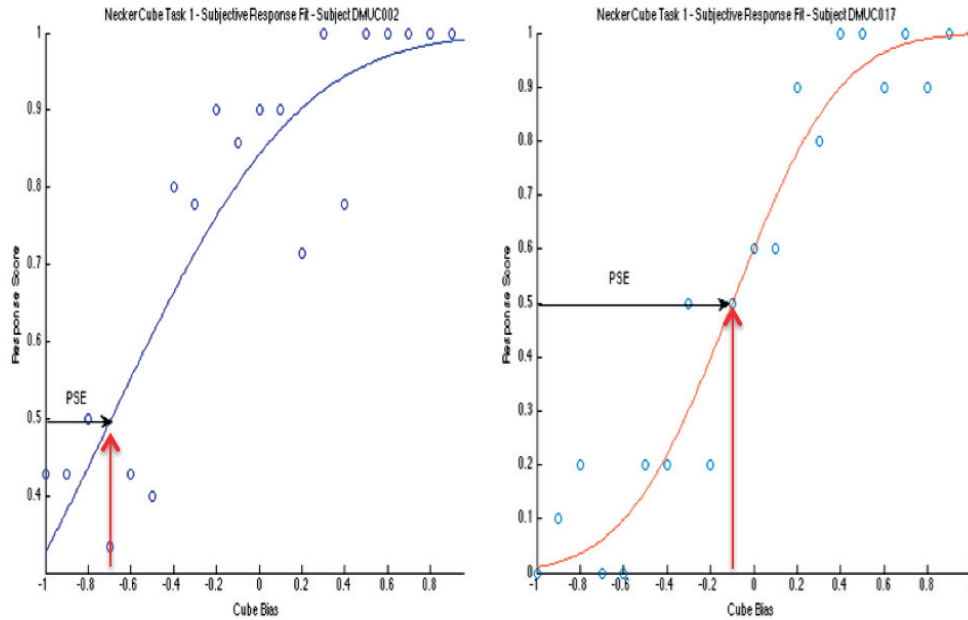


Figure 4.3 – Representative subject psychometric curves. The left panel shows an example of a poor fit. The right panel represents an example of a good fit, with a point of subjective equality approaching a 0 bias, or an objectively ambiguous Necker Cube.

be obtained, including: the point of subjective equality (PSE), which is the particular stimulus bias value at which a given subject perceives the Necker Cube from above or from below equiprobably; and the k parameter for a given subject, which determines the steepness of an individual's response curve (Figure 4.4). We further plot the distribution of two goodness-of-fit measures: the sum of squared errors (SSE); and R^2 (Figure 4.5). Finally, we plot the distribution of STAI-T anxiety measures (Figure 4.6), as one of our peripheral interests lay in assessing the impact of anxiety on uncertainty-related behavior.

The histogram of PSE values across 22 subjects shows a spread spanning from -1 (stimulus most biased towards being viewed from below) to 0.5 (stimulus objectively biased towards being viewed from above 75% of the time), confirming our expectation that such a point is indeed subjective. At the same time, we see that most subjects' PSEs (17) cluster just below and at 0 (the true ambiguous Necker Cube). This result suggests that a fully subjective ambiguous cube is biased slightly towards a from below perspective (Figure 4.4).

In plotting correlation coefficients of subjective responses to the fit curve, we find again a range of values, underscoring intersubject variability in perception. At the same time, most values (15/22) exceed 0.8, indicating a good fit between objective stimulus values and subject

Chapter 4. Are Perceptual Switches Surprising?

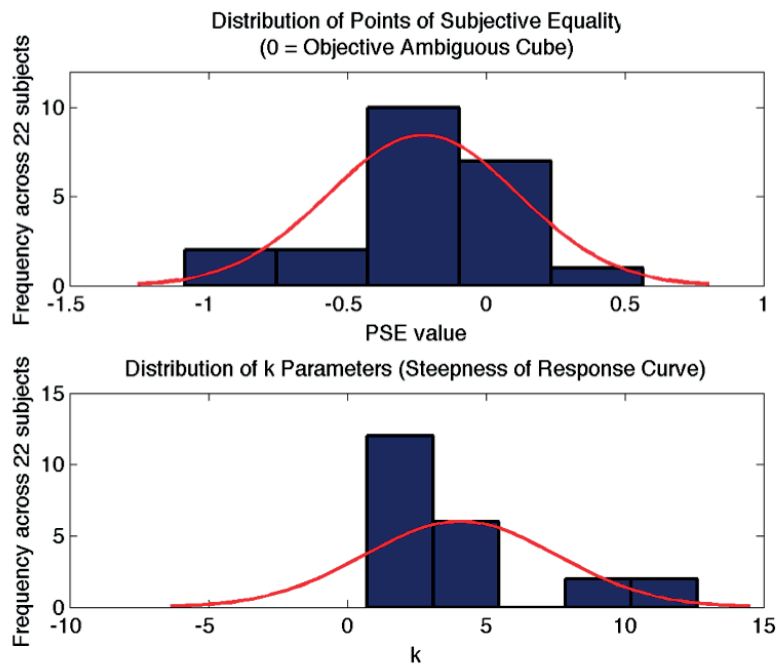


Figure 4.4 – Top Plot: Distribution fit of point of subjective equality (PSE) values across subjects ($n = 22$). PSE values constitute the inflection point of the sigmoid curve, or the point at which subjects are equally likely to perceive the cube from above or from below. This point is thought to vary across individuals and differ from the "true" ambiguous Necker Cube. We plot a histogram of PSE values above to gain a picture of the measures spread across subjects. We see above that most PSEs cluster at or just below zero (17/22 subjects). One subject in particular has a PSE hovering at -1 (a stimulus most biased towards being viewed from below). The variability in PSE values confirms inter-individual differences in Necker Cube perception. At the same time, a high frequency of PSE values clustered around the true PSE suggests that subjective Necker Cubes do not stray far from objective versions. Bottom plot: k parameter spread across subjects. This value indicated the steepness of the curve. The histogram shows both variability across subjects as well as a positive value for all subjects.

perception (Figure 4.5).

In plotting psychometric measures of trait anxiety (STAIT), we find a variance again across subjects (Figure 4.6). For STAIT scores, it must be noted that scores ranged from 40-54, a span that may not detect high or low anxious individuals. The STAI-T scale was not formulated to assess clinical anxiety but scores that may signal anxiety symptoms have been suggested to be anywhere from 39 to 55, which would place all subjects above in a high anxious group. It is important to stress that normative STAI-T scores vary with age, gender and geography. We do not, then, assume that our cohort is particularly anxious. Further, no subjects were native English speakers and thus their responses to the inventory may have been more conservative for reasons of language comprehension. All in all, the distribution of the scores collected would allow us to probe individual anxiety profiles in relation to other variables, however,

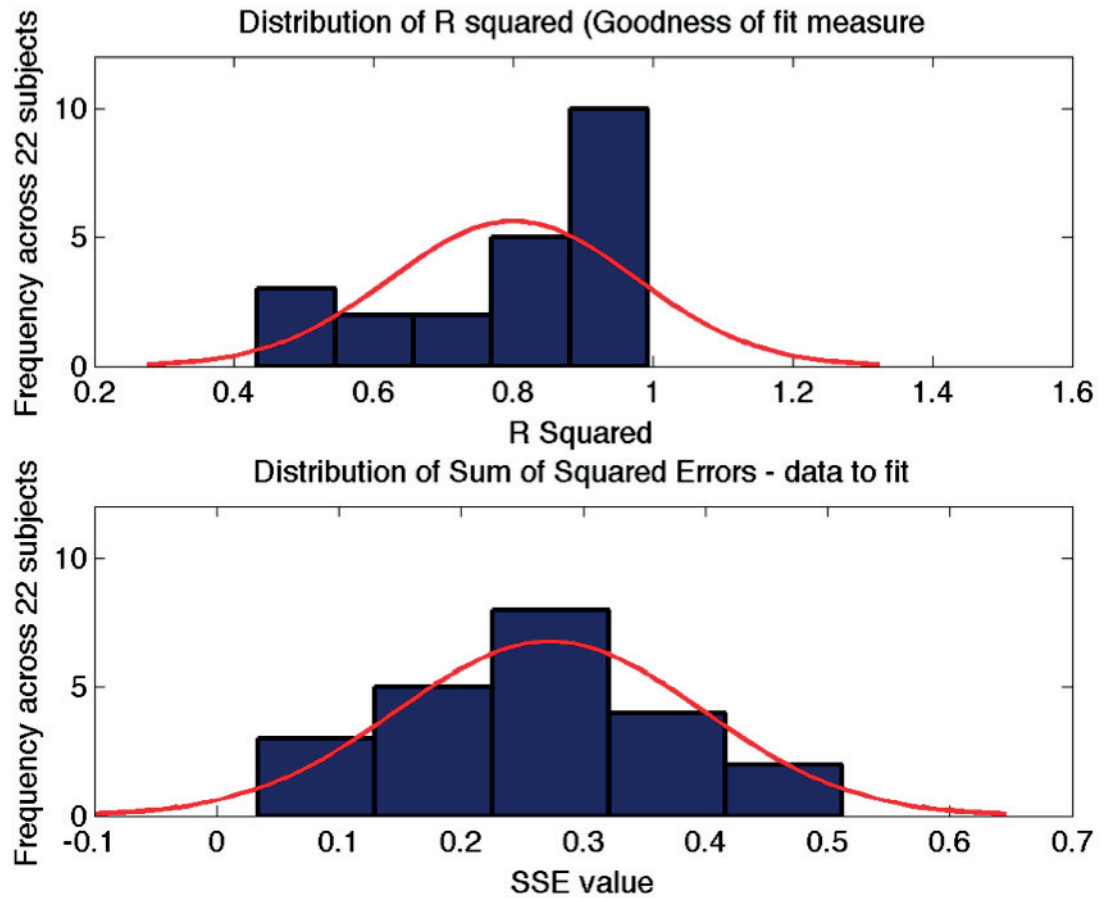


Figure 4.5 – Distribution of goodness of fit measure of individual curves against a logistic fit. We fit individual values of correlations between subjective values and the logistic function as a means of assessing subjective measures' fits. Most regression coefficients ((15/22) exceed 0.8 and all exceed 0.4. The histogram plot confirms that the manipulation introduced in task 1 conforms to expectations (that is behavioral data roughly follows the logistic model). On the other hand, we again see a variability in individual response curves, which underscores the subjective nature of the task. Distribution of sum of squared errors (SSE) measures of individual subjective fit curves to the logistic fit. Here, we plot the SSE values in a histogram to gain a picture of subjective responses' fit. SSE values appear to follow a normal distribution across subjects. Of note is the low value of SSE's across the experimental pool.

significant investigations into the STAI-T would require an expanded sample size.

Reaction Times

Next, we checked to see if any differences were found in reaction times between percepts and across stimulus bias values. A t-test was conducted on the average reaction time to button press for from above versus from below percepts and yielded no differences ($p = 0.1790$, $t =$

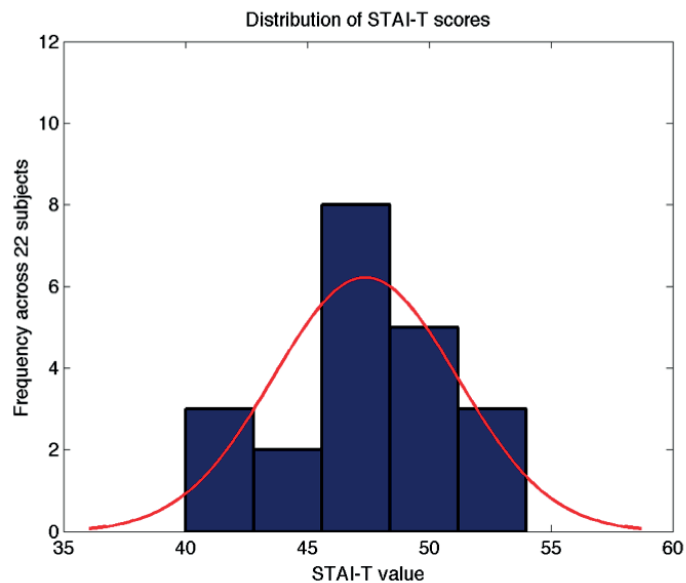


Figure 4.6 – Distribution of STAIT (Trait Anxiety) scores across subjects. Responses to the anxiety scale were acquired following the experimental session in 21 subjects. Subject scores ranged from 40 to 54. We plot these values to a histogram above to determine if scores provide a range, and what the description of the histogram suggests. The histogram above does show variability but does not appear to follow a normal distribution. Indeed, while the STAI-T inventory was not developed to assess clinical anxiety, a higher score indicates a higher likelihood of experiencing anxiety symptoms. Here, our data show that all subjects fall in the higher end of the STAIT spectrum (scores greater than 40).

1.39, $df = 21$). A second t-test was performed on the average reaction time to button press for most ambiguous versus least ambiguous stimuli and yielded no differences ($p=0.9917$, $t = -0.1$, $df = 21$).

A 2-way (Subject x Bias Level) ANOVA was performed on individual subjects' average reaction time per stimulus. There was no effect of stimulus on reaction time ($F = 1.32$, $p = 0.1636$, $df = 20$) but there were significant differences across subjects ($F = 47.87$, $p < 0.0001$, $df = 21$) (Figure 4.7).

The result above again underscores the subjectivity of Necker Cube perception, as reaction times in Task 1 differ across individuals but not for other dimensions such as stimulus value or percept.

Percept Bias

Perceptual reversals for the Necker Cube are known to show a bias for the from above percept (Troje & McAdam, 2010; Kornmeier et al., 2009; Sundareswara & Schrater, 2008). To determine whether our data corroborates this feature of the stimulus, we tested for the frequency of "from

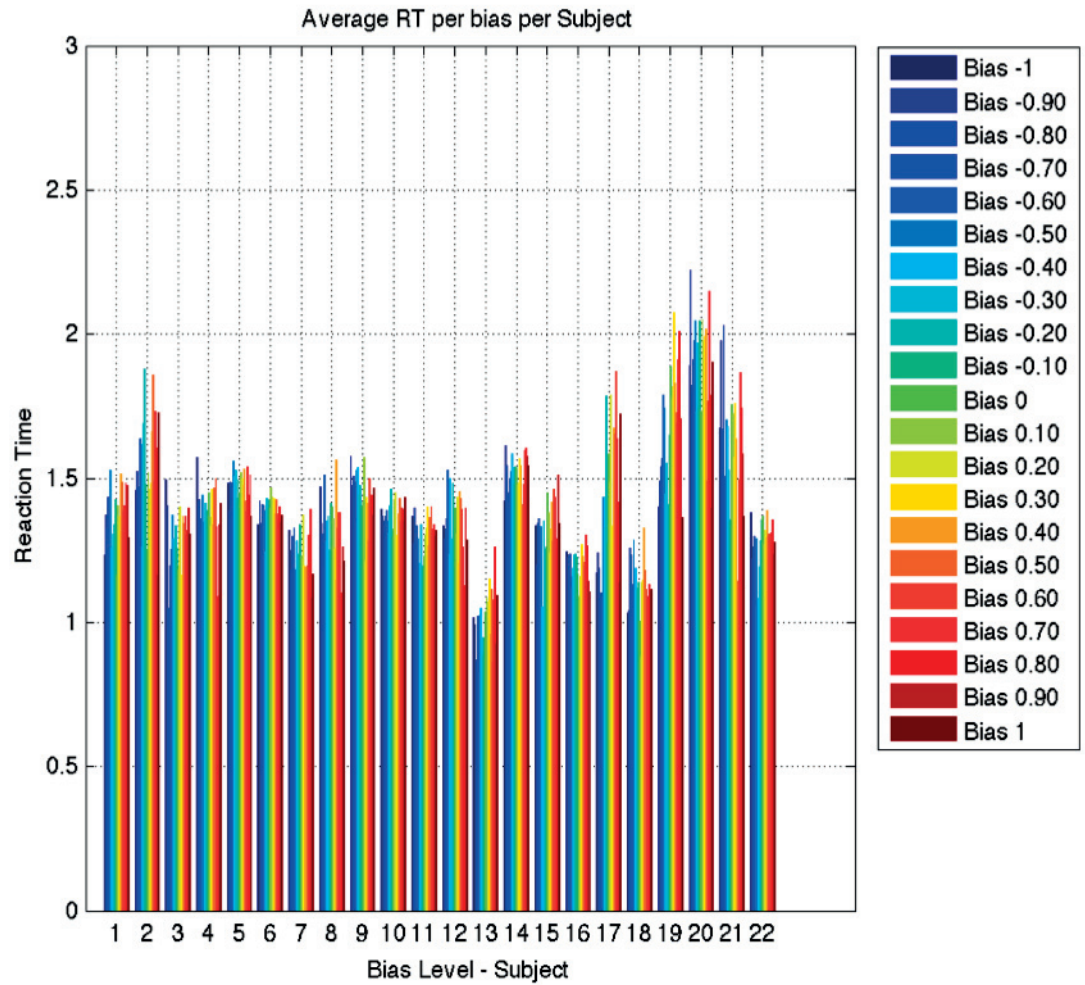


Figure 4.7 – Average Subject Reaction Time Per Stimulus Type in task 1. Task 1 includes 21 different biases of the Necker Cube Stimulus. Above, we computed each subject's average reaction per stimulus type, obtaining 21 different values for 22 subjects. The y-axis shows average reaction time; the x-axis shows stimulus type grouped for each subject. Each color denotes a stimulus type. A 2-way ANOVA yields no significant differences in reaction time across stimuli but does yield significant differences in reaction time across subjects.

above" to "from below" percept selection in Task 1. Subjects selected from above percepts significantly more frequently than they did from below percepts ($p = 0.0023$, $t = 3.47$, $df = 21$).

The results of this test above thus show a significant bias towards viewing the cube from above versus from below (Figure 4.8), replicating previously reported findings. This bias portends a potential confound in other results as the assumed symmetry in perception may not be present. In attractor state parlance, the from-above state may present a deeper trough than the from-below state. A formal investigation of oscillator, attractor and hybrid models of bi-stable

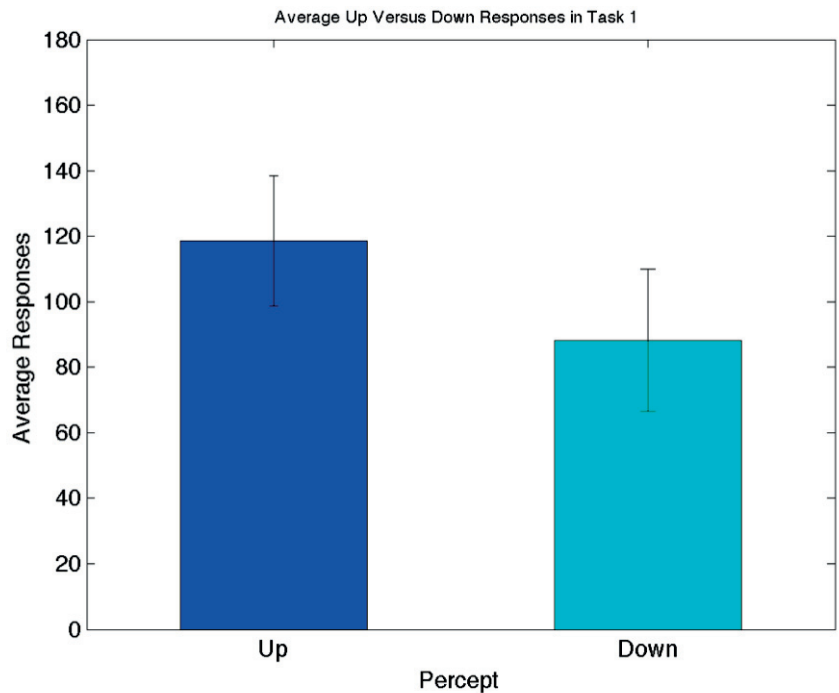


Figure 4.8 – Average Number of "from above" versus "from below" reports in Necker Cube Task 1 in 22 subjects. From-above response frequency averages are shown in dark blue while average frequency of from-below responses are in light turquoise. Average frequencies are denoted on the y-axis. Results show a significant preference for from-above percept response across subjects ($p = 0.0023$).

dynamics is beyond the scope of this thesis. Nonetheless, we describe a possible model of bi-stability that may better describe Necker Cube dynamics (Figure 4.9).

4.3.2 Behavioral Results, Necker Cube Task 2

In task 2 we first tested behavioral data to determine if the more ambiguous stimuli elicited more perceptual switches, as we would expect it to. We performed an ANOVA on 3 levels of ambiguity: fully ambiguated; partially ambiguated (cube bias at 35%/65%) and disambiguated (cube bias at 20%/80%). While no significant differences were found ($F=1.93$, $p = 0.1545$, $df = 54$), a post-hoc t-test yielded significant differences for average switch frequency between the most and least ambiguous conditions ($p < 0.001$, $t = 5.27$, $df = 18$) (Figure 4.10).

We then tested to see if a preference for from above percepts was replicated in Task 2. To do so, we conducted a t-test on the frequency of first percepts in each trial being from above versus those reported as from below. Subjects again showed a bias towards a "from above" percept ($p=0.0027$, $t = 3.47$, $df = 18$) (Figure 4.11).

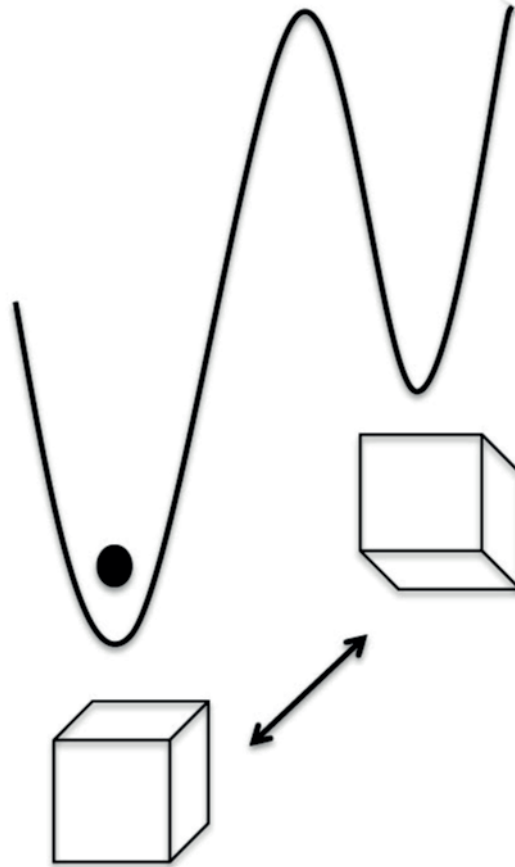


Figure 4.9 – A schematic representation of Necker Cube bi-stability, incorporating findings above on an overall preference for the from-above perceptual state. The black ball represents an agent’s perceptual state. Two minima are described here, one for the from-above conformation and the other for the from-below view. We propose that the perceptual states are not equivalent; namely, the attractor state for the from-above percept is deeper. Thus the perceptual state in the from-above conformation needs more noise accumulation to escape to the alternative state; the corollary would imply that it is easier to escape the from-below perceptual state, as its barrier to transition is lower.

Perceptual Switch Dynamics

The dynamics of perceptual reversals remain a topic of considerable interest and study (Moreno-Bote et al., 2007; Zhou et al., 2004; Shpiro et al., 2009). Below, we describe analyses relating to the amount of time spent in one state or another (the dwell time, or dominance time) prior to a perceptual switch. We do so to ensure that 1) we replicate a gamma distribution of dwell times between switches; 2) to determine if biased cubes in particular also display such a distribution. Biased cubes incurring the same dynamic as fully ambiguous cubes lend support to the notion that perceptual switches in biased cubes are qualitatively similar to those in the ambiguous version, thus making comparisons across ambiguity levels more credible. We first tested to see if the amount of time spent in a from above perceptual state differed

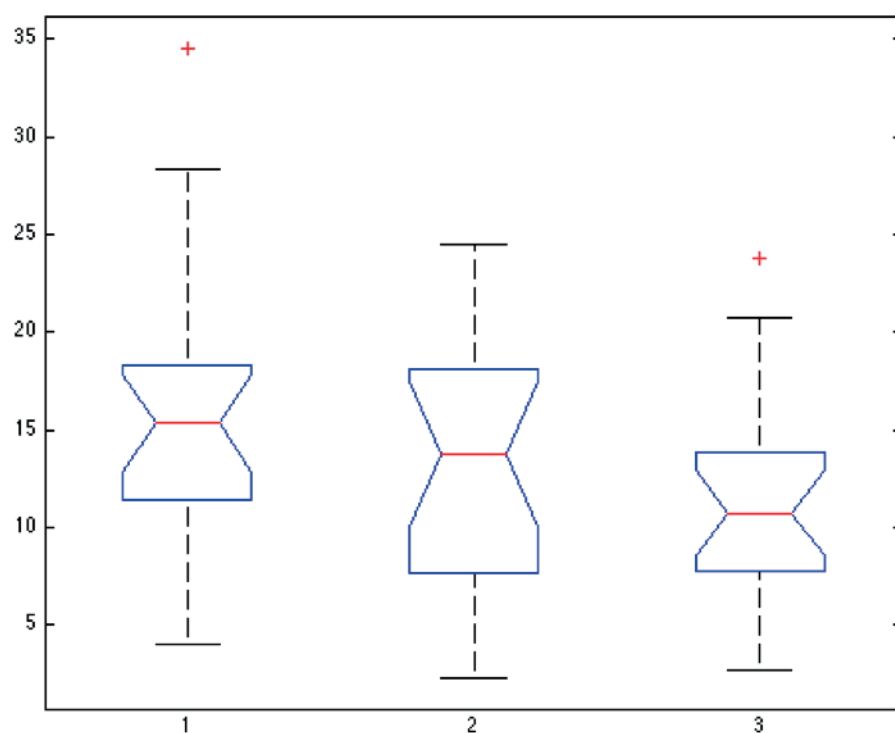


Figure 4.10 – Mean switch frequency for each of three levels of ambiguity in Necker Cube Task 2 in 19 subjects. An ANOVA test shows no significant difference in means while a post-hoc t-test between least and most ambiguous stimulus types shows significantly more switches in the most ambiguous condition relative to the least ambiguous condition ($p < 0.001$).

from time spent in a from below perceptual state, across all conditions. A t-test revealed a significant difference in average dwell times spent in from above versus from below states in 19 subjects ($p=0.0025$, $t=3.5$, $df=18$). Again, a preference was found for the from above state (Figure 4.12).

In summarizing our findings across task 1 and task 2, both in perceptual reports and in dwell times between from-above and from-below states, we confirm that a significant bias towards from

We then sought to find a perceptual dominance effect for different bias levels by calculating the ratio of the amount of time spent in the from above percept over the amount of time spent in the from below percept for each stimulus condition (per Einhäuser et al., 2008). A cube biased towards a from-above percept should elicit longer dwell times in the from above state. The ratio of experimental dwell times should follow a linear pattern with the bias level. Results show that dwell times for the from-above percept increase with the subjective bias value of

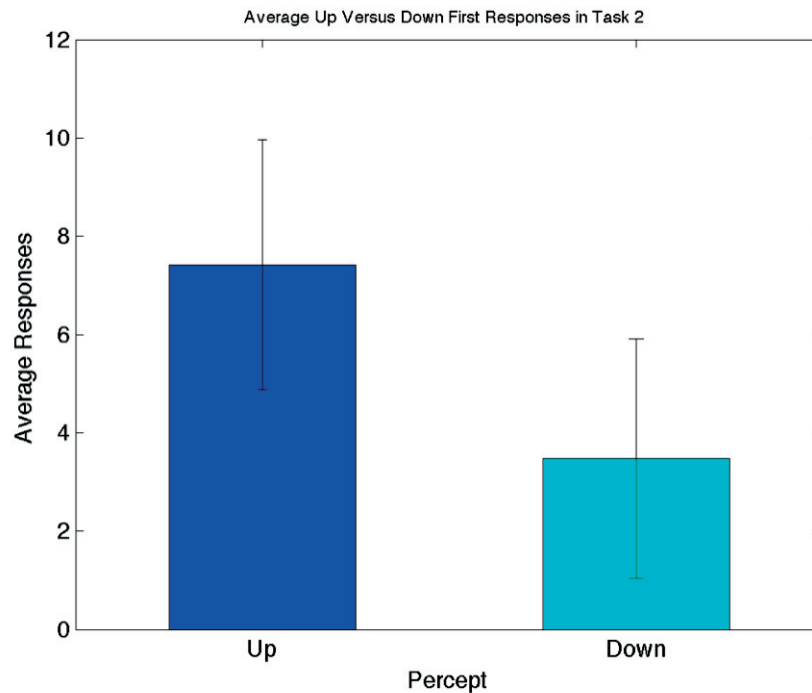


Figure 4.11 – Average frequency of "from above" and "from below" perceptual assessments for first responses within a trial in Task 2. From-above responses are represented in the dark blue bar graph while from-below responses are contained in the light turquoise bar graph. The y-axis represents task averages across 19 subjects. Recall that in Task 2, subjects are asked to provide their first percept in a trial; therefore the responses represented above are not perceptual switches. Subjects reported stimuli to be from-above more frequently than they cast them as from-below percepts ($p = 0.0027$).

the stimulus across subjects, though the values found also support an overall preference for the from-above state irrespective of stimulus bias (Figure 4.13). When looking at individual subjects, we find a linear relationship ($r > 0.80$) between subjective fit values and dwell time ratios for 13 out of 18 subjects (Table 4.1).

An ANOVA investigating the effect of stimulus bias and perceptual choice on dwell times shows a significant main effect of stimulus bias ($F = 5.3$; $p = 0.0003$, $df = 4$) and percept ($F = 97.84$; $p < 0.0001$, $df = 1$) as well as a significant interaction of these effects ($F = 13.59$; $p < 0.0001$, $df = 4$).

To ensure that our manipulation held, and that the subjective uncertainty captured by fit values obtained from task 1 were replicated in task 2, we performed a correlation between subjective fit values and relative dominance ratios. Specifically, we correlated averaged subjective fit values from task 1 to average relative dominance ratios computed from task 2 across subjects that were ultimately included in our analyses in task 2, and subjects that experience more than 1 switch in each ambiguity condition in task 2, totalling 18 subjects. These measures are highly correlated ($R = .99$, $p = 0.002$, Figure 4.14).

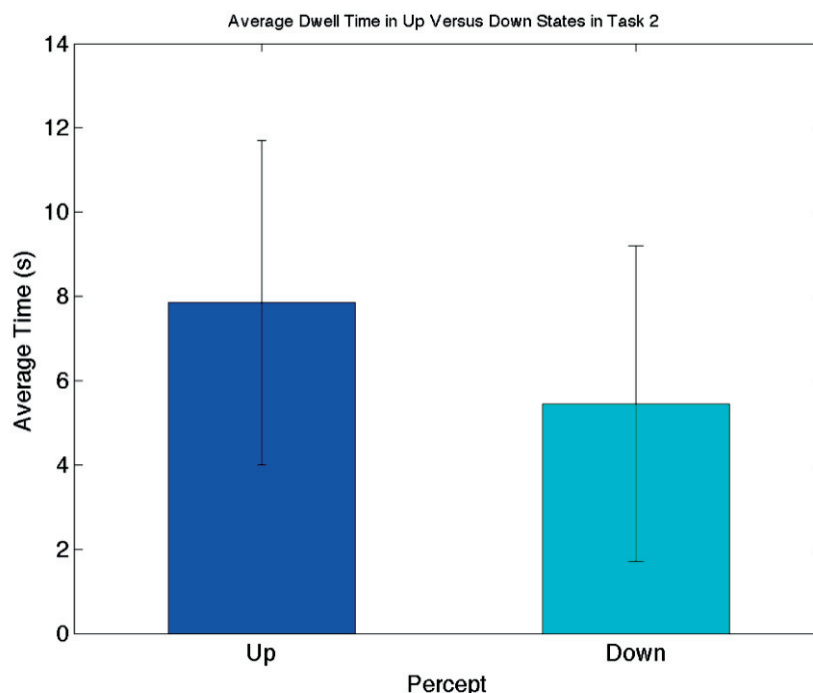


Figure 4.12 – Bar graph representing the average amount of time spent in a from-above state (dark blue graph) versus a from-below state (light turquoise graph) in 19 subjects. The average dwell time is plotted on the y-axis. Subjects spent significantly more time in the from-above state on average across the experimental session ($p=0.0025$).

Previous studies on the Necker Cube have yielded several ongoing debates, one of which is the emergence of a drift in dominance times with time on task (Zhou et al., 2004; van Ee, 2005). That is, some studies have found evidence of an increase in dominance duration with time on task which would support some form of learning in relation to the perceptual switch. To determine if there may be a change in dominance times with time-on-task, we split each 80 second trial into 4 time bins of 20 seconds each and obtained the mean dominance time for each time bin. We then performed an ANOVA to search for differences in mean dominance duration depending on time interval and bias (Figure 4.15). No main effect of interval was found, ($F=0.62$; $p=0.6046$, $df=3$); a main effect of bias was found, in line with our results above ($F=5.7$; $p=0.0002$, $df=4$) and no significant interaction between interval and stimulus bias was found ($F=0.42$; $p=0.9557$, $df=12$).

We finally examined the distribution of dwell times, both across all conditions as well as within the 3 different ambiguity levels (at 50%, 35%/65% and 20%/80%) to 1) determine if our results replicated previous findings that fit dwell time distributions best to a gamma distribution; 2) to see if the expected gamma distribution was found for biased Necker Cubes. All latencies irrespective of stimulus bias followed a gamma distribution (Figure 4.16) with scale parameters falling within the expected range of 3-5 (Gigante et al., 2009). Interestingly, the latter increases

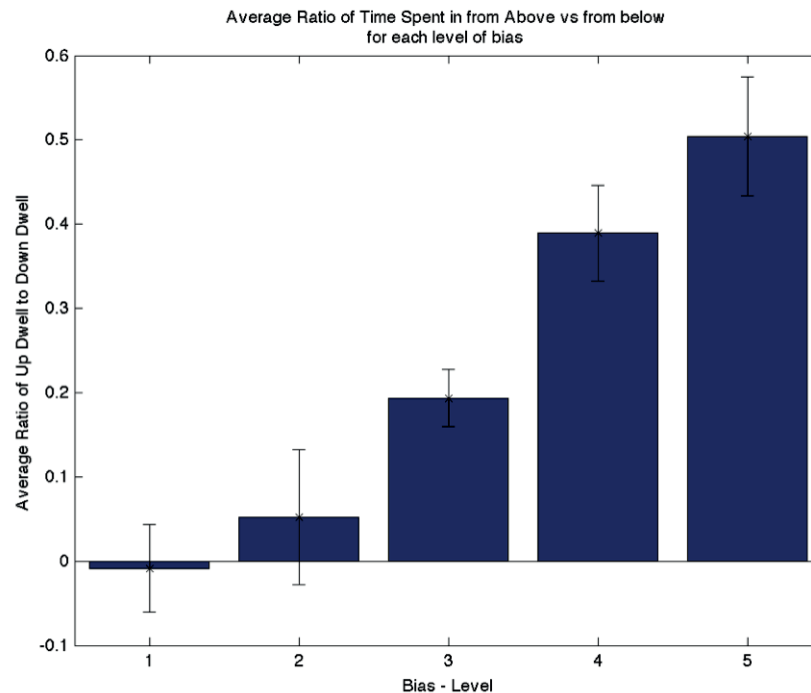


Figure 4.13 – Ratio of dwell time in "from above" to dwell time in "from below" states, according to stimulus value (bias) across subjects ($n = 19$). Each bar in the plot represents the average ratio across subjects for a given bias level, with 1 denoting least biased and 5 denoting most biased towards a from-above percept. Ratios were normalized and bounded to values between -1 to 1, with -1 indicating all viewing time spent in a from-below state and 1 representing all viewing time in a from-above state. Our data show that stimuli most biased towards a from-below state (bar graph 1) yields a dwell time ratio of 0 - indicating that subjects spent an equal amount of time in the two perceptual states. We assume this is due to the overall preference subjects show for the from-above state, irrespective of stimulus bias. Nonetheless, we see in the plot above that time spent in the "-above" perceptual state increases with increasing stimulus bias for that percept, and this in an linearly increasing manner.

with cube disambiguation, a feature that may not be spurious but that has yet to be analysed in this thesis.

A last sanity check with respect to the data concerned correlations between: different goodness of fit measures and parameters obtained in Task 1; possible correlations with the latter and two personality measures, risk taking and trait anxiety; and finally possible correlations between both fit measures, parameters, personality traits and Task 2 results, the latter which include average dwell times and switch frequencies.

Of 22 subjects who completed Task 1, 3 were removed from analysis in Task 2. Of these, 2 did not complete the Risk questionnaire scale and 1 did not complete the anxiety scale. We then plotted a correlation matrix with measures for 17 subjects to determine significant

Chapter 4. Are Perceptual Switches Surprising?

Table 4.1 – We list correlation coefficients and associated p values for individual subjects when plotting individual dominance ratios per stimulus bias to the stimulus bias. Here we sought to ensure that a majority of subjects conformed to the linear relationship in figure 4.13, to ensure the relationship holds beyond the possibility of outlier values. Of 18 subjects (3 were excluded from Task 2 for poor performance; an additional subject was excluded due to not experiencing switches while viewing the most highly biased stimulus), 12 show r^2 values >0.8 .

| r | p |
|---------|--------|
| 0.9520 | 0.0125 |
| 0.9054 | 0.0344 |
| 0.9851 | 0.0022 |
| 0.3755 | 0.5334 |
| 0.9438 | 0.0158 |
| 0.8958 | 0.0397 |
| -0.1730 | 0.7809 |
| 0.4503 | 0.4467 |
| 0.9695 | 0.0064 |
| 0.8659 | 0.0578 |
| 0.9115 | 0.0312 |
| 0.6769 | 0.2094 |
| 0.9239 | 0.0249 |
| 0.6131 | 0.2715 |
| 0.8803 | 0.0488 |
| 0.8769 | 0.0509 |
| 0.7340 | 0.1579 |
| 0.8880 | 0.0442 |

Table 4.2 – Maximum likelihood estimates for gamma distribution parameters of dwell times across subjects in all stimulus conditions as well as ambiguous, partially ambiguous and disambiguated stimulus conditions.

| | MLE (Gamma Distribution) | |
|--|--------------------------|----------|
| | k | θ |
| All Dwell Times | 1.5457 | 3.2922 |
| Ambiguous Stimulus Dwell Times (50%) | 1.5975 | 2.8688 |
| Partially Ambiguous Stimulus Dwell Times (35%/65%) | 1.5943 | 3.1817 |
| Disambiguated Stimulus Dwell Times (20%/80%) | 1.4834 | 3.7866 |

relationships between dwell times, fit parameters and questionnaire measures (Figure 4.17). A significant portion of early studies on the Necker Cube and bistable illusions sought to identify personality measures that correlate with subjective perceptual experience of ambiguous stimuli (Lidberg et al., 1978; Rogers, 1988), and thus we sought to determine in a first instance if our data suggest such a relationship with anxiety and/or risk profiles.

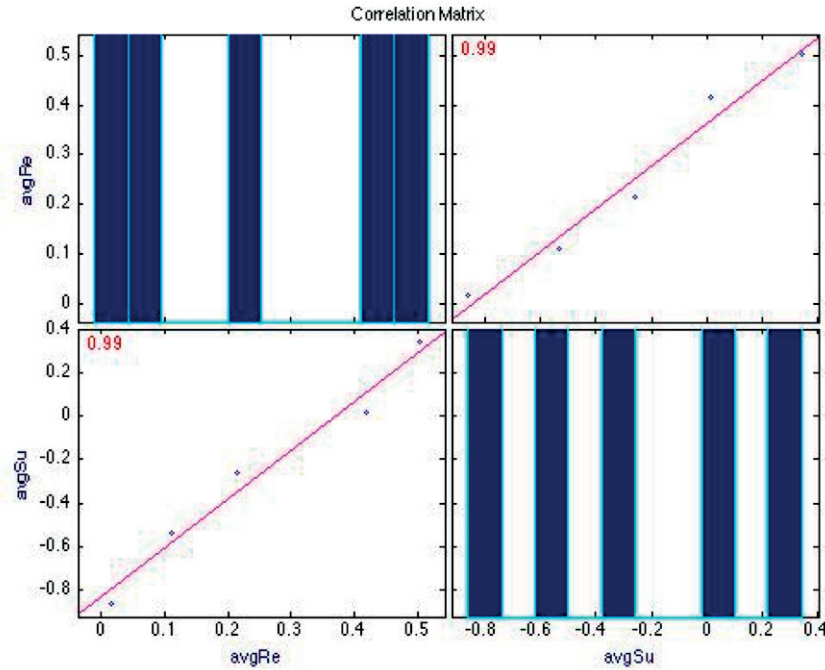


Figure 4.14 – Correlation matrix of subjective fit values to relative dominance ratios. Results show that stimulus values pulled from task 1 are significantly correlated with the dwell time of a given stimulus bias in task 2 ($r=0.99$; $p=0.002$).

Both goodness of fit measures (SSE and R^2) are negatively correlated ($r=-0.64$, $p=0.0057$). Fit parameters (k and the PSE) are positively correlated ($r=0.51$, $p=0.0380$); k and SSE are negatively correlated ($r=-0.80$, $p<0.001$). Thus the lower the SSE, the steeper the curve and the higher the R^2 .

Similarly, behavioral data from Task 2 were significantly correlated with one another. Average dwell time correlated significantly with its mode, variance, and switch frequency, as expected. Anxiety measures did not correlate significantly with behavioral data though anxiety scores do show a relationship with switch frequency ($R=0.43$, $p=0.0810$) and concomitantly, a decrease in the dwell time average ($R=-0.38$, $p=0.1585$). While the correlation was not significant, it must be stressed that the low sample size ($n=17$) coupled with a comparatively narrow range of scores (40-54) may dampen the relationship. Given the interest in anxiety's effects on uncertainty in decision-making (Hartley & Phelps, 2012) and on perceptual rivalry (Meredith, 1967), we deem this finding to be worthy of further investigation. Risk scores on the other hand show no relationship with any of the above measures.

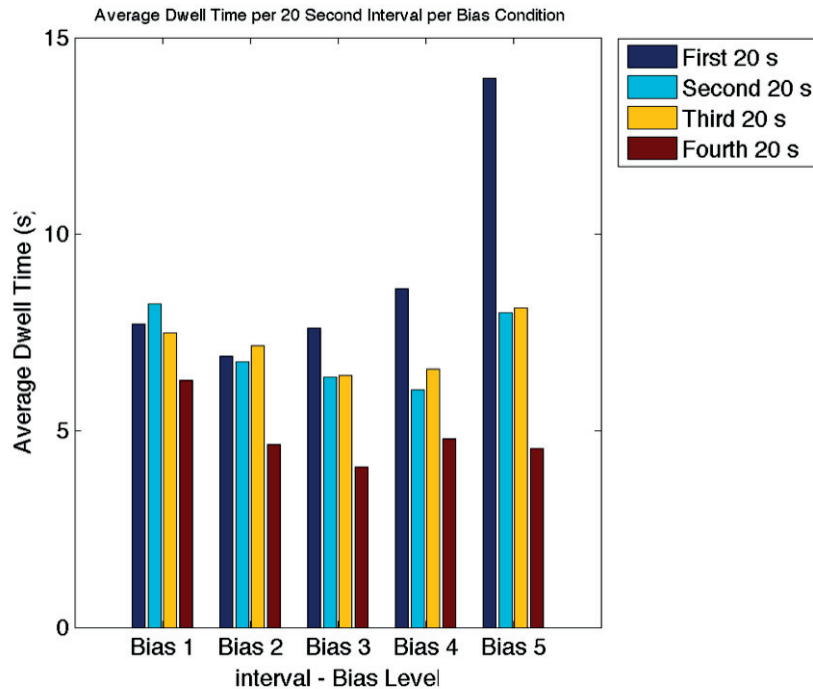


Figure 4.15 – Average Dwell Time in 1st through 4th time intervals (20 seconds each) across trials in Task 2, differentiated according to stimulus bias value (1-5 corresponding to 20%-80%). An ANOVA shows no significant drift in dwell time, that is dwell times do not vary across 20 second trial time bins, indicating no habituation or learning takes place during a trial. The y-axis represents average dwell time while the x-axis describes stimulus bias levels (1:5) split into separately colored time bins.

Data Quality Control

Although we control for subjectivity in Task 1, the nature of the experiment is such that falsifying responses is challenging. We excluded one subject from task 1 based on their behavioral data, which saw all Necker Cubes classified as being from above. No other subjects were removed, irrespective of their psychometric curve parameters or goodness of fit, as we have no other objective criterion by which to distinguish bad results (the subject removed presented an extreme case which we can confidently classify as not following task instructions). In task 2, we exclude 3 additional subjects on the basis of their behavioral data. Specifically, we excluded subjects whose switch responses were the same in more than 20% of their switch responses (i.e. 20% of switch responses were same consecutive percepts) in more than 3 experimental trials. Further, of the perceptual switches included in analyses, only those followed by a switch to the alternative percept were included.

In task 2, of the 19 subjects remaining, we took further quality control measures with regards to report and replay conditions. Specifically, for our analyses on the replay condition, we excluded replay trials for which the replay condition included $\pm 25\%$ replay reports than

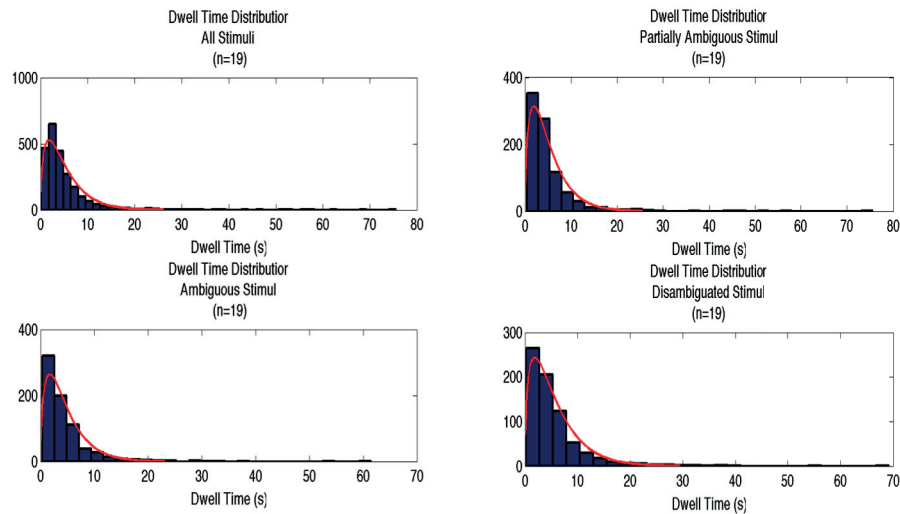


Figure 4.16 – Histogram of dwell times fit to a gamma distribution. The top left graph shows the distribution for all stimuli; the top right graph shows the distribution of disambiguated stimuli. The bottom two graphs show dwell times for partial and fully ambiguous stimuli from left to right, respectively. All dwell times appear to fit a gamma distribution.

switch reports in the previous report trial.

4.4 Neuroimaging Results - Perceptual Uncertainty

4.4.1 Imaging Analysis

Scans were preprocessed and analyzed using SPM 12. We first generated voxel displacement maps (VDM) and applied these to functional data. We then warped and realigned data to the mean image before apply a bias-field correction. Then data were co-registered to individual anatomical volumes before being segmented (6 class tissue probability maps) and normalized to the MNI152 template and smoothed to 8 mm. The resulting images were then used for analysis. Our data analysis plan follows the one used in the Card Game. We first conduct a whole-brain analysis, with a threshold of $p = .05$, (FWE corrected) before proceeding to a cluster level corrected whole brain analysis ($p = .001$) to determine if any voxels survive this threshold. We then perform a small volume correction with the same list of a priori regions as in the Card Game. Following this, we perform an non-parametric permutation tests using per Eklund et al. (Eklund et al., 2016) using the SnPM toolbox (<http://warwick.ac.uk/snmpm>).

4.4.2 General Linear Models

We employed model-based fMRI analysis to capture the variance of an event's average signal amplitude according to a given measure representing risk and surprise. As in the Card Game,

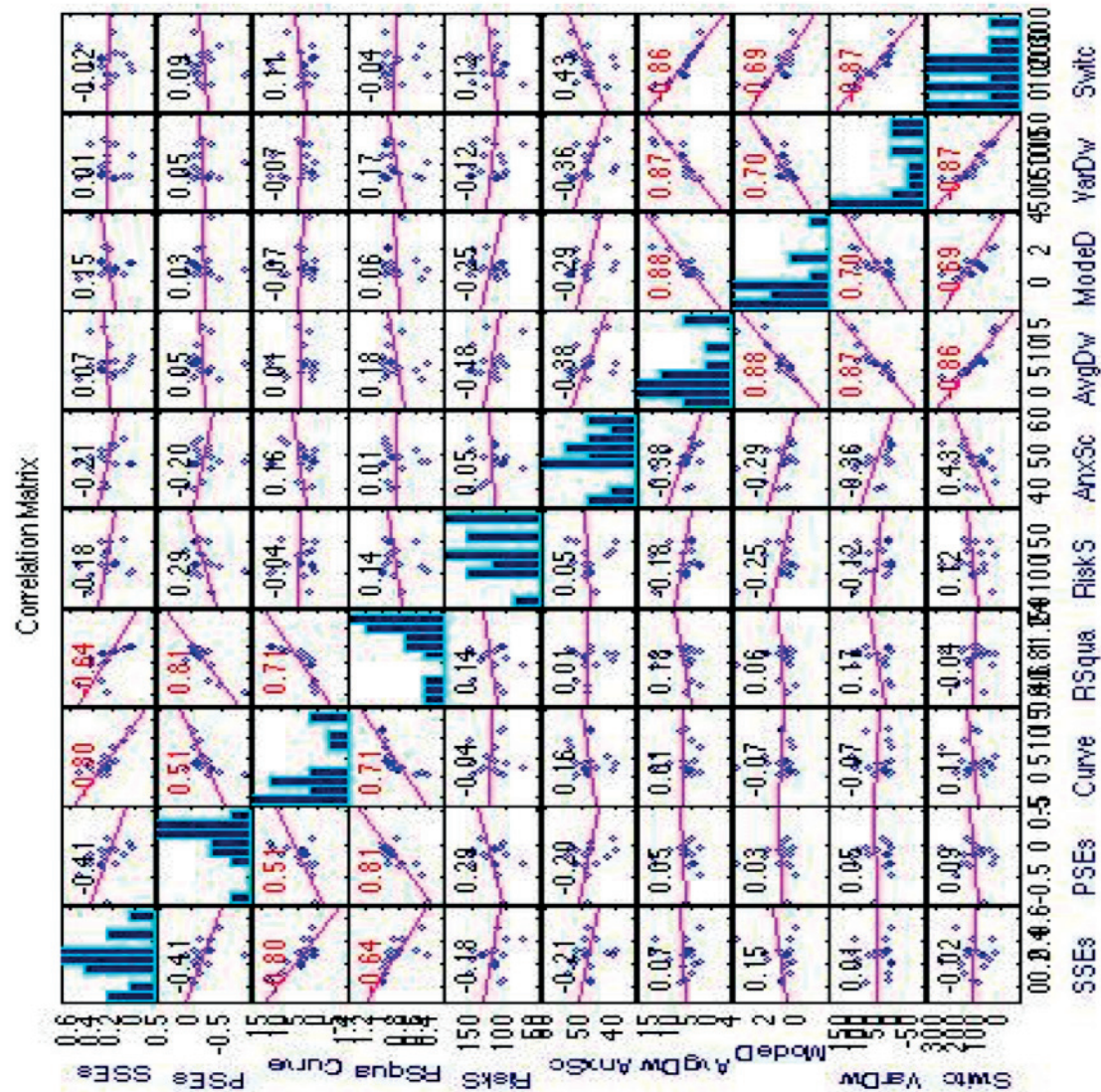


Figure 4.17 – Correlation Matrix for Psychometric Curve Parameters; Goodness of Fit Measures; STAI-T and DOSPERT measures; and summary statistics for dwell times (n = 17).

given the nature of the phenomena studied, we do not expect thresholded statistics to survive family-wise error corrections and thus plan to further scrutinize results with the use of small volume corrections with our list of 7 a priori regions of interest.

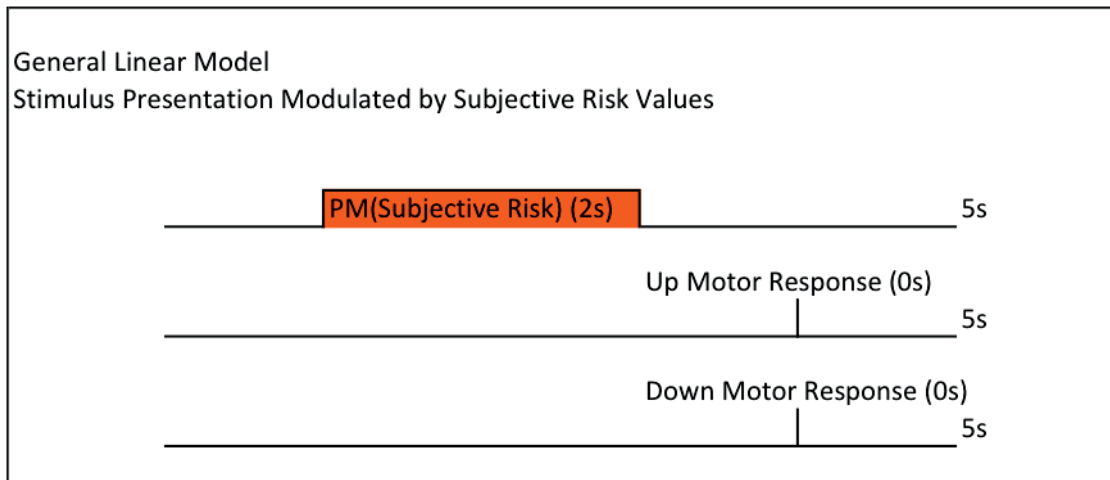


Figure 4.18 – General Linear Model Designs for Task 1

4.4.3 fMRI Results: Necker Cube Task 1

To better match the fMRI design of our financial uncertainty task, we constructed a GLM containing 3 regressors of interest. The first contained onsets for stimulus presentation, offset by 1 second, with a duration of 2 seconds, modulated by subjective risk values. These were obtained by pulling subjective bias values for each stimulus generated by fitting behavioral results to a sigmoidal curve, and squaring them. This regressor of interest was offset relative to the event (stimulus presentation) following results from Preuschoff et al (Preuschoff et al., 2006; Kuhnén & Knutson, 2005), in which risk processing was found to be delayed relative to stimulus onset. We also included two 0 duration regressors for the button press separated into response type (1 or -1 percept) and 6 motion regressors of no interest.

We performed a t-test on the parametric modulation of the first regressor to determine perceptual risk activation. Whole brain analyses as well as non-parametric permutation tests yielded no significant results.

We then applied SVCs to our preselected ROIs and find a significant cluster in the right anterior insula ($p = 0.038$ FWE, $T = 3.68$, $k = 14$) (Figure 4.19).

We also performed 2 t-tests to determine if an up response differed from a down response (Up > Down; Down > Up) and found no differences. A conjunction analysis of the responses regressors yielded activations in motor areas as expected.

4.4.4 fMRI Results: Necker Cube Task 2

We then set out to investigate perceptual uncertainty given both the task design and behavioral results. Recall that behavioral results in task 2 suggest that, while the manipulation of the stimulus yielded the expected responses (that is, subjective responses to ambiguity in task 2

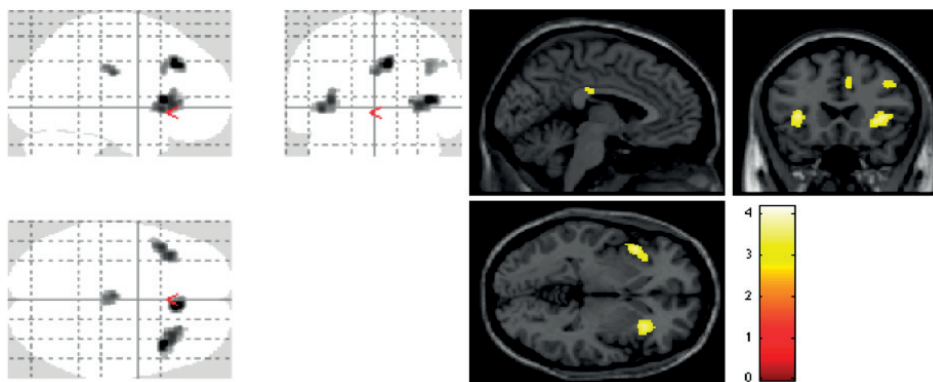


Figure 4.19 – Statistical maps overlaid on glass brain and canonical T1 image, for perceptual risk in task 1. The statistical map was thresholded at $p = .005$, uncorrected, for visualization purposes

reflected subjective responses to ambiguity in task 1 according to the response curve), it did so in a skewed manner, where up percepts were favored regardless of stimulus bias. Further, in considering the phenomenology of the perceptual switches, one has to account for the fact that the switch is surprising in and of itself, regardless of stimulus value. That is, the perceptual switch of a fully ambiguous cube should elicit surprise by virtue of its spontaneity, and this in a low level domain (visual perception). For the same reason, we cannot exclude the possibility of dwell time modulations on the surprise signal, again, regardless of stimulus bias. That is, given that a subject would expect her visual field to remain unchanged, the longer this stability lasts, the more surprising it should be when a switch occurs. Of course, it is possible that such a surprise would be conscious, while other regions may respond to shorter dwell times with more surprise; that is a system may exist in the brain that expects a perceptual switch. Finding neural correlates of regions implicated in shorter dwell times would lend credence to the attractor model, which presupposes an accumulation of noise over time. In task 2, we constructed 5 different GLMs to investigate uncertainty related to perceptual switches.

GLM 1: Report Vs Replay

In a first instance, we sought the differences in neural activation between true switch reports and replay presses. Two t-tests were conducted: Report > Replay and Replay > Report. Report regressors included switch onsets only for trials that matched subsequent replay responses to within 25%. All other trials were excluded.

In the first t-test, Report > Replay, no significant differences emerged at $p = 0.001$, cluster level corrected ($k = 25$ voxels). In the second t-test, Replay > Report, we find widespread bilateral activations at ($p = 0.001$, cluster level corrected, $k = 25$ voxels) notably in bilateral angular, middle frontal, and middle temporal gyri; right occipital fusiform gyrus; and left superior frontal gyrus (see Appendix B, Table B.2).

4.4. Neuroimaging Results - Perceptual Uncertainty

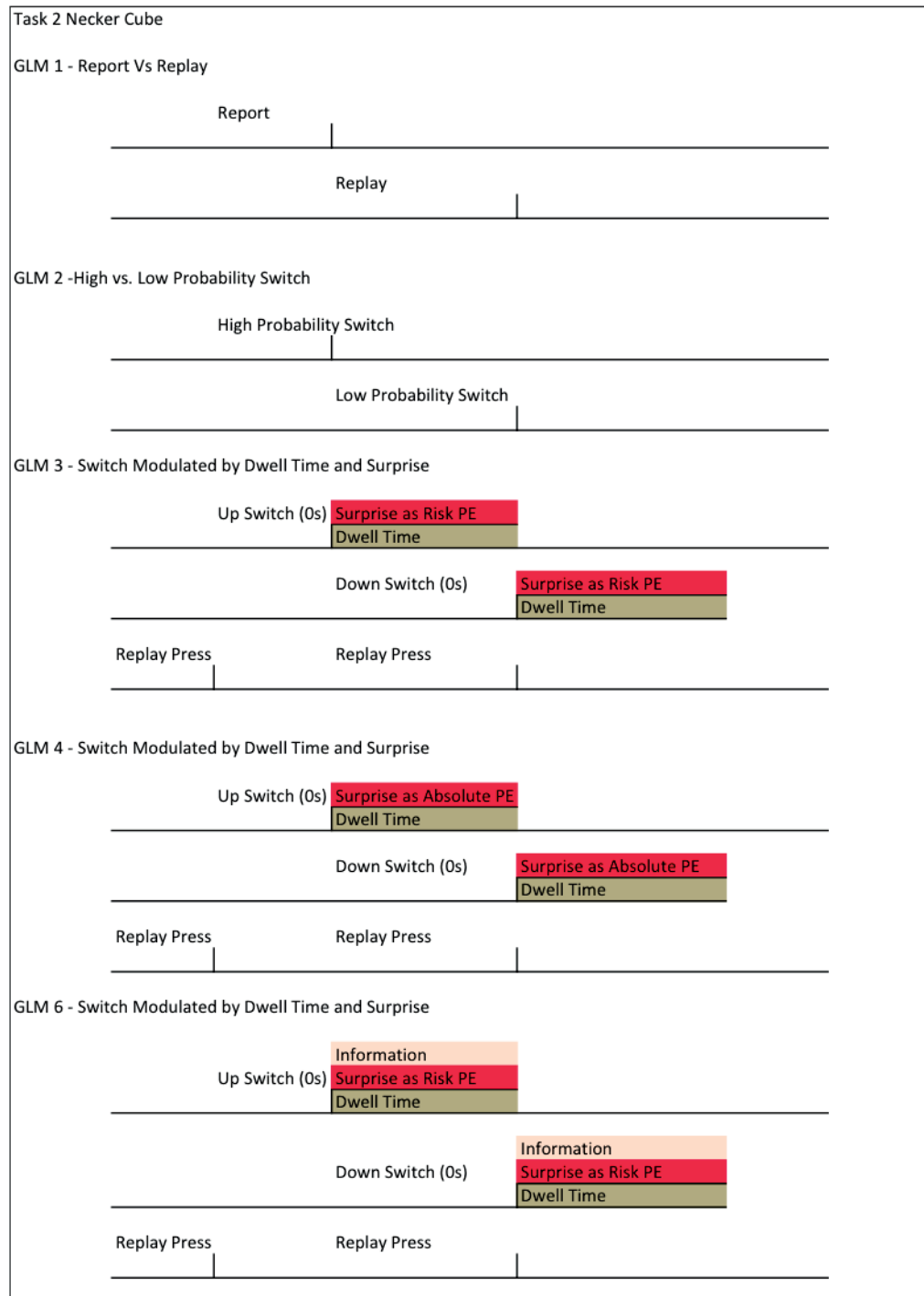


Figure 4.20 – Representative general linear model designs performed on Necker Cube task 2 fMRI data.

These results are in-line with previous findings that replay-related activity elicits more neural activation than ambiguous stimuli (Kornmeier & Bach, 2012). This is also in line with the

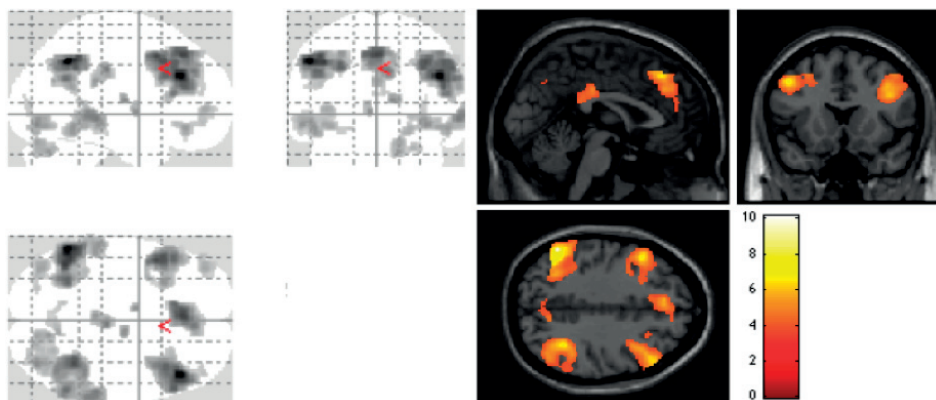


Figure 4.21 – Statistical maps for Replay > Report condition ($p=.001$, $k = 25$), overlaid on glass brain and canonical T1 image

notion that increased uncertainty may activate more neurons at a low activity rate while less uncertainty (such as in the disambiguated stimuli of the replay condition) will activate fewer neurons but more strongly (Vilares & Kording, 2011). Notably, activity is confined to regions that are not included in our list of a priori regions of interest in relation to uncertainty processing.

GLM 1: High versus Low Probability Perceptual States

Here we constructed a GLM composed of two regressors : Switch onsets for switches to high probability percepts (e.g. -1 for a stimulus biased at 20%) against switches to low probability percepts (e.g. 1 for a stimulus biased at 20%). Switch onsets for fully ambiguous stimuli (unbiased cube, 50% value) were not included in the regressors. Two t-tests were conducted: High Probability Switch > Low Probability Switch and Low Probability Switch > High Probability Switch. The aim here is to gain a picture of uncertainty across biases.

No results were found for High Probability Switch > Low Probability Switch. For Low Probability Switch > High Probability Switch, significant activation did not survive FWE cluster correction. Similarly, SVC analyses on our regions of interest did not yield significant results.

While the uncorrected statistical maps above hint at clusters in the ACC and right anterior insula for low probability states (a crude measure of most surprising stimuli), none of the analyses above yield significant results.

GLMs 3 and 4: Modulating Perceptual Switches by Surprise values

In the following GLMs, we separated switch onsets according to whether they indicated a switch to a from above state (1) or a from below state (-1) before modulating them with a surprise value. The only switch onsets included in the regressor for an 1 switch were those followed by a -1 switch, and the only onsets included in the 1 regressor were those followed by

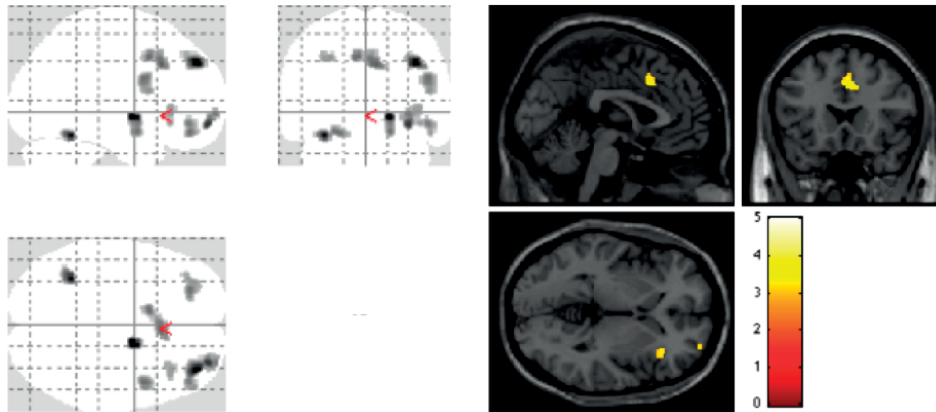


Figure 4.22 – Statistical Maps for Low Probability > High Probability (threshold used in image, $p = .005$, $k = 57$, for visualization purposes), overlaid on glass brain and canonical T1 image.

a -1 switch. We separate switches according to percept (and thus lose a degree of freedom) because the two are not equivalent. Further, we included 7 regressors of no interest (replay presses and 6 motion regressors). GLMs 3 and 4 differ in the surprise model used only: we perform an analyses on surprise as risk prediction error; as well as surprise as absolute prediction error, just as we did in the Financial Uncertainty task. Recall that these two models do not differ numerically but in their curves. Thus we modulate switch regressors by two values: a given switch's surprise value based on the percept and stimulus bias; and the previous dwell time of the switch - that is the amount of time spent in the previous perceptual state. We expect that the longer the dwell time, the higher the surprise will be, regardless of stimulus bias conditions. These GLMs are meant to capture the surprise signal inherent to the perceptual switch alone.

Up > Down

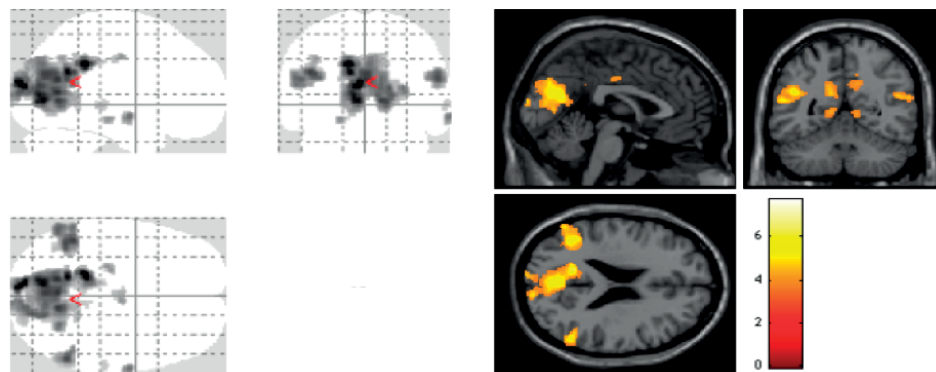


Figure 4.23 – Statistical Maps for Up > Down Switch contrast ($p = .001$, $k = 23$) in Necker Cube, Task 2 overlaid on glass brain and canonical T1 image.

The Down > Up switch t-test yielded no significant activation ($p = 0.001$, $k = 22$ voxels). These

Chapter 4. Are Perceptual Switches Surprising?

contrasts were conducted as controls, to ensure that different percept switches were equivalent. Thus, we did not expect any significant BOLD responses to emerge. However, we encounter an incidental finding in relation to the contrast. A pattern emerges for switches into the from-above percept. Recall that behavioral results indicate an overall bias for the from-above state. This bias may suggest a preferences for this state and be reflected in the BOLD response found. On the other hand, an incongruity in relation to the perceptual state may be captured by the BOLD response. We believe this explanation is unlikely. If the from-above state is the congruent state, based on the behavioral results, then we would expect the Down > Up contrast to reflect a BOLD response. In searching for a plausible explanation, given the regions implicated, notably the precuneus and the angular gyrus, we suggest that it is lying supine in the scanner while perceiving the cube to be from above that presents an incongruity. Both the angular gyrus and the precuneus have been implicated in body position and representation (Blanke et al., 2002; Lopez et al., 2008; Seghier 2013; Cavanna & Trimble, 2013).

Switch Contrast

We then performed a conjunction analysis by combining regressors for both kinds of switches to examine switch-related neural activity. As switches are reported via button press, this regressor is confounded by motor effects, thus we cannot conclude that statistical maps relate to a main effect of switch alone. Cluster level corrected ($p = 0.001$, $k = 37$) results show bilateral activity in the anterior insula, thalamus, supramarginal gyri as well as extensive activation patterns in the cerebellum and brainstem. In addition, clusters were found in right putamen and left nucleus accumbens (See Appendix B, Table B.3).

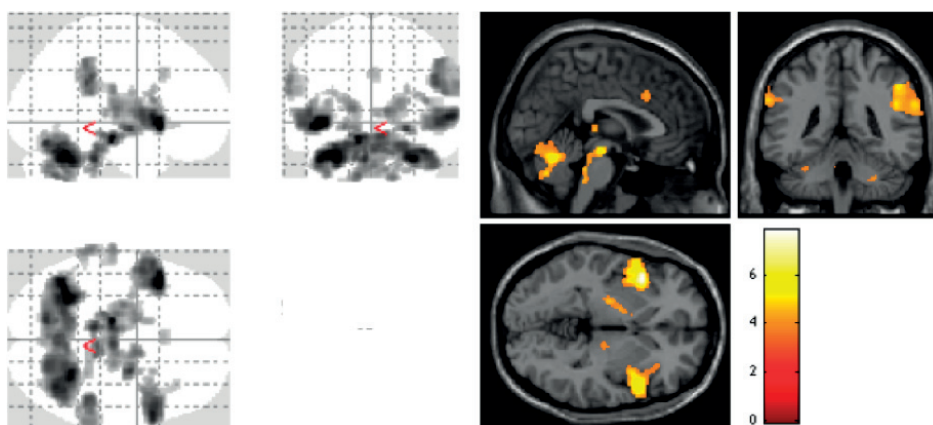


Figure 4.24 – Statistical maps for the Switch response contrast ($p = .001$, $k = 37$), overlaid on glass brain and canonical T1 image.

Switch Modulated by Surprise as Risk Prediction Error

Here we separated switch onsets according to whether they indicated a switch to a from above state (Up) or a from below state (Down) before modulating them with their surprise value.

4.4. Neuroimaging Results - Perceptual Uncertainty

In GLM 3, the surprise value was calculated by squaring the value of the stimulus $p(\text{Up})$, to obtain a quadratic function similar to that used in the financial uncertainty task whereby notably, a switch under full ambiguity should yield the least surprise and a switch under least ambiguity should yield the most.

$$Surprise_{up} = -4 * p_{up} + 8 * p_{up}^2 \quad Surprise_{down} = 4 - 12 * p_{up} + 8 * p_{up}^2 \quad (4.2)$$

Whole brain analysis yielded no significant clusters ($p=.001$, $k = 29$), nor did statistical non-parametric tests.

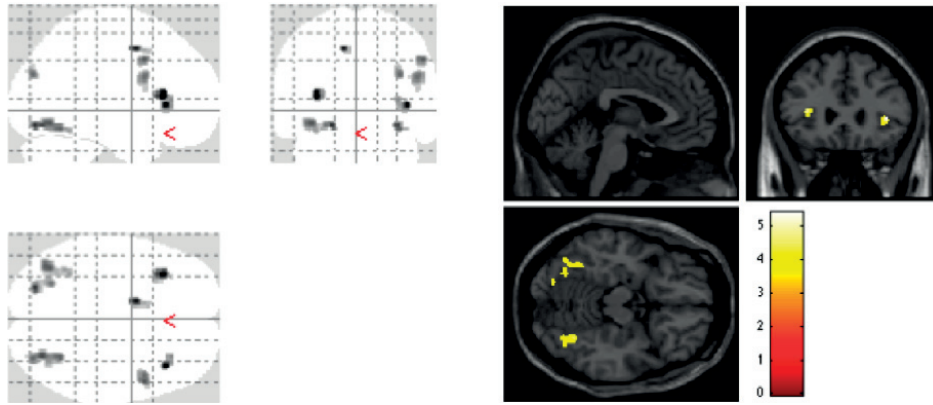


Figure 4.25 – Statistical maps of surprise as risk prediction error ($p = 0.001$, $k = 28$) overlaid on glass brain and canonical T1 image. While clusters in the bilateral insular cortex appear in the image, these are not significant in a whole brain analysis.

We then applied SVC correction to our list of a priori ROIs. A cluster ($k = 7$) in the right anterior insula yielded a p value of 0.05 (FWE). (See Appendix B, Table B.4). Results above show significant activation in the right insula both the cluster and peak levels, as well as in the left insula at peak levels.

Switch Modulated by Surprise as Absolute Prediction Error

In GLM 4, we modulate switch onsets by absolute prediction error, as we did in the Financial Uncertainty task. Prediction errors are calculated by subtracting an expected perceptual state from the switch state, denoted by either -1 or 1. We then take the absolute value of the resulting error, to emulate the absolute reward prediction error in our gambling task. Cluster level correction ($p = 0.001$, $k = 30$) yields no significant results; nor did statistical non-parametric tests. SVC analyses yield a significant cluster in the right anterior insula ($p = 0.039$, FWE; $k = 13$) (See Appendix B, Table B.5).

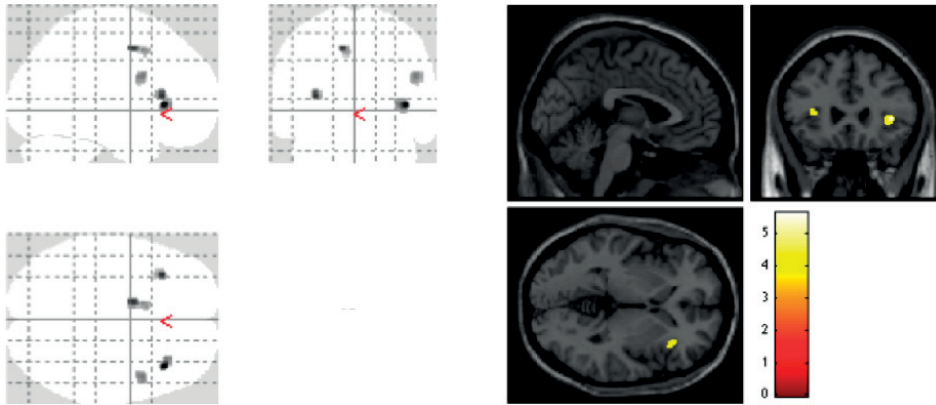


Figure 4.26 – Statistical maps of surprise as risk prediction error ($p = 0.001$, $k = 28$) overlaid on glass brain and canonical T1 image. While clusters in the bilateral insular cortex appear in the image, these are not significant in a whole brain analysis.

Switch Modulated by Previous Dwell Time

Here switch onset regressors were parametrically modulated by the switch reports preceding dwell time – that is the amount of time spent in the previous perceptual state. We hypothesize that the longer the dwell time, the higher the surprise will be, regardless of stimulus bias conditions. This test is meant to capture the surprise signal inherent to the perceptual switch alone, or a binary form of surprise, because we assume that an agent consciously expects an unchanging stimulus to remain static in her perception. Results of a t-test on combined parametrically modulated regressors of the switches reveal widespread activation at $p = .0001$, cluster corrected ($k = 12$). A high threshold was used because the resulting map appears noisy. Most, notably, the T-map shows significant activity in the dorsal striatum (putamen), one of our regions of interest. Regions surviving cluster-level correction include bilateral precuneus and middle occipital gyrus; thalamus; left temporal pole; left anterior insula; left middle temporal gyrus; and right mid-cingulate gyrus (See Appendix B, Table B.6).

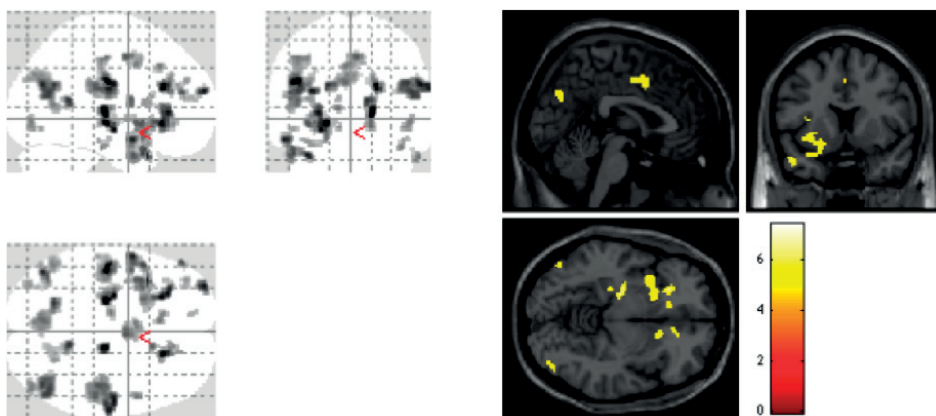


Figure 4.27 – Statistical map of dwell time modulation on the perceptual switch ($p = 0.0001$, $k = 12$), overlaid on glass brain and canonical T1 image.

We next moved on to an SVC analysis using our preselected ROIs. Results yield significant responses in the left anterior insula and right ventral striatum, as well as bilateral dorsal striatum (See Appendix B, Table B.7).

Effect of Surprise and Previous Dwell-Time

We next sought to combine the values associated with surprise given a stimulus bias condition and percept with the surprise associated with the previous dwell time. To do so, we performed a conjunction analysis on both parametrically modulated regressors. Whole brain cluster-corrected ($p = 0.001$, $k = 28$) results are found in bilateral occipital fusiform gyrus, as well as right middle occipital and right precentral gyrus (See Appendix B, Table B.8).

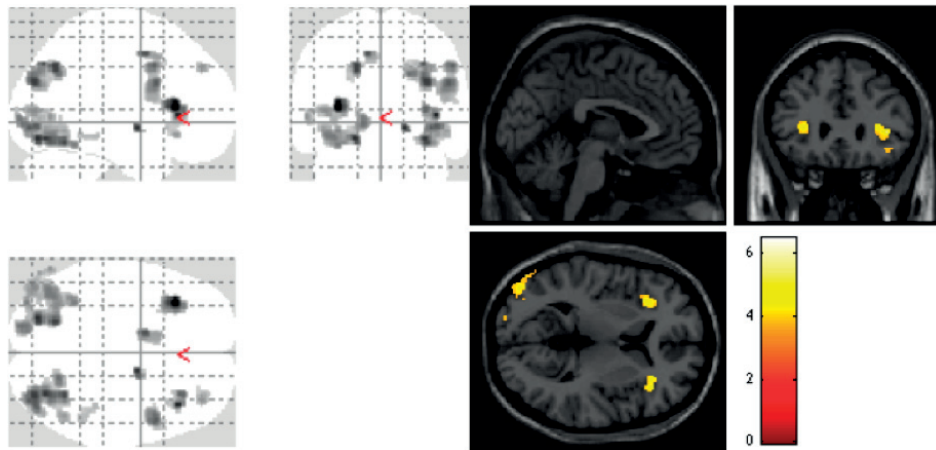


Figure 4.28 – Statistical map of dwell time modulation in conjunction with surprise modulation on the perceptual switch ($p = 0.0001$, $k = 12$), overlaid on glass brain and canonical T1 image.

As in the above contrasts of interest, we proceed to an SVC analysis of dwell time and surprise using our list of pre-defined ROIs which yields significant clusters in bilateral anterior insula (See Appendix B, Table B.9).

Effect of Information

We performed a group level t-test to determine if information plays a role in perceptual uncertainty. No significant voxels survived even at a low threshold ($p = 0.05$) suggesting that in an involuntary, perceptual context, information quantities do not play a role at the moment of a switch.

Inverse Dwell Time

Finally, we investigated the potential neural activity associated with shorter dwell times on the brain. Thus we parametrically modulated switch onsets by the inverse of their previous

dwelling times, instead of their dwell time. No results were found in performing a t-test on this parametrically modulated regressor.

4.5 Discussion

In the work detailed above, we employed a novel bi-stable perception task using the Necker Cube illusion to 1) probe the possibility that a spontaneous perceptual switch can be formalized as a prediction error; 2) to determine if the neural correlates of a perceptual prediction error match those found for a cognitive prediction error in the same subjects. Our fMRI results generally point to a common involvement of the insula in response to prediction errors in both financial and perceptual domains.

In a first instance, our behavioral results confirm a bias towards the "from above" percept in the perception of the Necker Cube, irrespective of stimulus conformation. In our first task, subjects classified more stimuli as viewed from above than "from below". Further, points of subjective equality were downshifted from truly (objectively) ambiguous Necker Cubes. In our second task, subjects spent significantly more time overall in "from above" perceptual states than in "from below" states. Further, first percepts in task 2 were reported as being "from above" significantly more frequently than "from below". These findings replicate a known feature of Necker Cube perception (Kornmeier et al., 2009; Sundareswara & Schrater, 2009; Kornmeier et al., 2017) though to our knowledge, it is not clear that an explanation for the bias has been elucidated. One could hypothesize that humans tend to view things from above, making this perceptual state, in Bayesian terms, a hyperprior; few navigate the environment with necks craned upward, although an interesting task would be to test a from-above bias in small children. Regardless, as mentioned, to our knowledge, no formal investigation of the bias has been performed. A recent study investigated the bias for from-above percepts in Aspergers' patients and found that the latter are not susceptible to the bias (Kornmeier et al., 2017). They argue that this discrepancy results from a lack of top-down control in individuals on the autism spectrum, though it is not clear how top-down control influences a percept bias per se. Indeed, autistic individuals often do not experience perceptual switches (Happé, 1996) and a Bayesian formulation of the reversal process suggests that autistic individuals may not hold priors with regards to ambiguous stimuli (Pellicano & Burr, 2012), a reasonable explanation for differences in illusion perception but not for the from-above bias.

The implication of the bias may have an impact on the dynamic modeling of Necker Cube switches. Namely, if perceptual alternations are viewed through the lens of an attractor model, then one must consider one attractor to be stronger than another. These results compromise our manipulation of the Necker Cube to a degree, as a cube biased towards a from-above percept is not equivalent to one biased towards a from-below percept. Nonetheless, applying subjective psychometric fits from task 1 controls for this bias. Psychometric curves for most subjects fit to a logistic function in task 1, as seen in fit parameters and points of subjective equality values. Further, in calculating a relative dwell time ratio per cube bias in task 2, we

find a positive linear relationship between the ratio and increasing bias towards from-above percepts. Thus the manipulation we perform on the Necker Cube persists above and beyond what appears to be a general preference towards from above percepts.

A significant question of interest regarding Necker Cube reversals concerns their dynamics. As mentioned in the introduction, initial views on perceptual reversals presupposed that percepts alternate in a regular, periodic fashion (Blake & Logothetis, 2002) in a model dubbed the oscillator model, however evidence has accumulated that perceptual switches occur in a stochastic manner, suggesting that switches result from random noise accumulation rising to some threshold before landing in an alternative perceptual state, a model dubbed the attractor model (Gigante et al., 2009). Fitting perceptual state dwell times to a gamma distribution provides evidence that dwell times are indeed random (Borsellini et al., 1972). In addition, we find dwell time distributions for stimuli at all 3 levels of ambiguity (low, medium, high) with scale parameters falling within the reported range in the literature (3-5, Gigante et al., 2009). This result supports our manipulation of the Necker Cube as we see that perceptual switches in the face of biased cubes behave nonetheless in a similar fashion to those for fully ambiguous cubes.

Perceptual switches are thought to be spontaneous and involuntary (Sterzer & Kleinschmidt, 2007;) however, there are questions on whether reversals are subject to volitional control (Hugrass & Cuthrie, 2012; van Ee et al., 2005) and a type of learning effect (Gigante et al., 2009; Pastukhov & Braun, 2008). The question has been studied within the context of probing the neural source of the switch: namely, does a perceptual switch arise from low-level brain areas (Polonsky et al., 2000; Parkonen et al., 2008; Pearson et al., 2007) or is it modulated by higher level regions? As seen above, there is evidence for both possibilities (Wang et al., 2013; Long & Toppino, 2004; Sterzer et al., 2009). Possible learning effects in perceptual switches indirectly implies a top-down influence of the process, and thus supports the Helmholtzian notion of an inferential process, while an automatic, memoryless experience of the switch would suggest a bottom-up process. Of course, when viewed from a hierarchical predictive coding account, such a dichotomy is not necessarily firm, as predictions and prediction errors are iterative processes where information is exchanged between low- and high-level areas in a bidirectional fashion (Rao & Ballard, 1998). We find no differences in dwell times across a trial in any ambiguity conditions, suggesting that perceptual reversals are memoryless, automatic, spontaneous and not subject to endogenous learning effects. We do not go as far as claiming that this provides evidence of a low-level process, for what makes the perceptual switch but a conscious experience of it? Our subjects must report on their switches, and thus must be able to recognize and declare their perceptual state, a process we believe is necessarily anchored to the high-level cortical domain. At the same time, our negative findings on dwell time drifts in our data support the assumptions we held in undertaking the task, namely that perceptual switches are spontaneous and involuntary and not subject to habituation or learning effects.

The asymmetry between from above and from below states introduces a confound in our results. While we account for that nuisance in separating up and down switches, and further

Chapter 4. Are Perceptual Switches Surprising?

verify that stimulus bias overcomes perceptual bias, the finding offers its own possible future line of investigating, notably because it is bolstered by neuroimaging results. In comparing switches into "from above" versus "from below" states, we find significant activity in the visual cortex but also in the angular gyrus and cuneus. Because we don't see significant activity in response to a "from below" perceptual switch, we conclude that switching to a "from above" percept may induce an incongruence in body awareness (Blanke & Mohr, 2005), and perceptual cues (Dyde et al., 2009). A subject lying down would expect to see things from below, as indicated by gravity and vestibular cues, yet the visual cue perceived runs contra to this expectation. The regions found in particular support this hypothesis; the angular gyrus is involved in spatial, body awareness and incongruence of the latter, for instance in out-of-body experiences (Farrer, 2008; Arzy et al., 2006). Given the behavioral preference of from-above percepts, one might expect "from-below" states to trigger increases in neural activity. This particular finding, while incidental, offers a means by which to investigate neural correlates of such a phenomenon.

Because we employed an fMRI paradigm, our task cannot capture perceptual switches alone. The perceptual switch as reported includes a motor response; our modulation of the switch is done to capture its degree of uncertainty. Thus we measured switch related BOLD responses by introducing replay trials as control conditions. Replay trials present disambiguated versions of the Necker Cube, presented at the same times as previously reported conditions. The replay condition is meant to capture BOLD responses related to viewing the cubes, deciding to report on a cube's conformation and related reaction times. Contrasts for report versus replay conditions show no significant activation for report > replay, but do yield widespread bilateral activations in angular, frontal and middle temporal gyrii in the replay > report condition. These regions do not overlap with those included in our a priori regions, the latter which were chosen for their particular role in decision-making under uncertainty. We did not find a significant perceptual switch specific response relative to replay responses in contrast to other studies. Using Lissajous figures as bistable stimuli, Weilhhammer and colleagues found report > replay BOLD responses in bilateral middle and inferior frontal gyrii, as well as in the right parietal regions (Weilhhammer et al., 2013) while Sterzer & Kleinschmidt found bilateral inferior cortical activity in report conditions using a motion-from-stability stimulus (Sterzer & Kleinschmidt, 2007). On the other hand, using a visual grating rivalry task, Sterzer & Rees found no significant differences in percent signal change between rivalry and replay conditions (Sterzer & Rees, 2008). Further, the manipulation we introduce in the experimental stimuli may blur differences between biased cubes in the report condition relative to the more heavily biased cubes in the replay condition. Our results for this contrast are, interestingly, in line with previous findings that suggest perceptual ambiguity dampens neuronal spiking activity (Emadi & Esteky, 2013). We nonetheless account for replay responses by including them as regressors in our analyses.

Other fMRI studies investigating bistable perception have generally found a fronto-parietal involvement for perceptual switches (Lumer et al., 1998; Sterzer et al., 2002; Sterzer & Kleinschmidt, 2007; Sterzer et al., 2009). Here, we do not focus on the perceptual switch per se.

only by how its modulated by uncertainty. Viewed through the lens of a predictive coding account, we first applied the notion of risk, borrowed from a neuroeconomic framework, to task 1. Specifically, we modulated a subject's viewing period by a (prediction) risk value, as this form of uncertainty occurs prior to the outcome of a decision. Thus in our view, the switch occurs at the outcome of some (involuntary, perceptual) decision. We cannot control for the possibility that a switch occurs during the 2 second window we consider. Nonetheless, we assume that this period represents a prediction window. Therefore, we modulate it by risk, or the variance of the expected "from-above" prediction. Using small volume correction, we succeed in finding a significant cluster in the right anterior insula using this formal account.

In task 2, we train our focus on the switch. We cannot adequately capture the prediction period in task 2 because the dwell times between switches vary a great deal; thus we cannot safely sample the BOLD response at a specific point in time. We assume the switch to be an outcome of an unseen decision-making process. Although one may argue that either a long dwell time may be rewarding in and of itself, or that a switch is rewarding because it reduces monotony and boredom, we generally assume that reward related variables do not affect our paradigm. In any case, reward is not of central interest to our research question, rather, uncertainty is. Thus we investigate the impact of switch induced surprise on the brain. We apply two similar accounts of surprise, risk prediction error, and absolute prediction error. As in the Card game, we find a specific role for the right anterior insula in response to surprise as absolute prediction error at the switch. These results are found with small volume correction analyses and do not survive whole-brain cluster level correction of statistical non-parametric tests. We propose that the BOLD response is not as robust in the perceptual paradigm for several reasons. 1) We modulate the switch by using subject's own bias values computed from their individual subjective fits; nonetheless, we cannot be certain that this subjectivity remains constant between task 1 and task 2. 2) We encounter a significant difficulty in controlling for errors in task 2, as we cannot identify errors of omission or commission in the experiment. We pre-process the data by only including switches that are followed by a switch to an alternative state, but nonetheless, switches input to our analyses cannot be falsified, and thus may be errors. 3) We only extract 5 values of surprise from task 2, whereas the financial uncertainty task had 9. Thus the Necker Cube task has comparatively less variance in its output, which may be reflected in the BOLD response.

In spite of these potential limitations, our results are remarkable because we lift a cognitive model of surprise and apply it to a perceptual paradigm, with all that the latter implies: spontaneity, automaticity, subconscious and of course a lack of deliberation or control. By using a predictive coding framework inspired by both an economic theorem (mean-variance theorem) as well as a reinforcement learning algorithm on a perceptual uncertainty task we obtain 1) any BOLD response at all; 2) a BOLD response in the same region (insula) as found in our financial uncertainty task using the same formal accounts on the same subjects.

We further scrutinize the role that the previous dwell time has on the experience of the perceptual switch and find striatal as well as insular involvement. Interestingly, the effect

Chapter 4. Are Perceptual Switches Surprising?

of the inverse of the dwell time does not yield any significant activity, which suggests that it is longer and not shorter dwell times which are more salient to the switch experience. This supports the idea that the switch is not expected. Given behavioral data examining a potential drift in dwell time with time on task, we can conclude that there is no learning or expectation suppression effect that is that a perceptual switch remains pertinent throughout the experiment. In combining previous dwell time (an estimated surprise value specific to the switch) and a computational account of stimulus-bound surprise, we find again bilateral anterior insula clusters.

Few studies have modeled perceptual switches as prediction errors. In a Necker Cube paradigm, Sundareswara & Schrater bias stimuli with flankers and apply a Markov Renewal Process framework to dwell times with the aim of characterizing switches as inferential processes (Sundareswara & Schrater, 2008). A more recent fMRI study applied a Bayesian framework to perceptual switches in a Lissajous figure paradigm, this time explicitly framing switches as prediction errors (Weilnhammer et al., 2017). While both the stimulus and formal account used differ from ours, Weilnhammer and colleagues nonetheless find a significant cluster in the right anterior insula with regards to switches as prediction errors, thus solidifying our results as well. Of note in our results, in relation to reinforcement learning, is the pointed absence of BOLD responses in the striatum. This null result may be due to the concomitant absence of reward, motivational and learning drives.

As seen above, our model-based analyses on risk and surprise in the Necker Cube firmly suggest a role for the right anterior insula in particular in the processing of perceptual uncertainty, which is in line with previous literature (Sterzer & Kleinschmidt, 2010). Indeed Sterzer & Kleinschmidt's review on the insula's role in perceptual switches is titled "often observed but barely understood", and does not address its plausible role in inferential processes. We propose that the insula emerges in perceptual uncertainty paradigms because the insula is uniquely tuned to uncertainty and thus reflects the nature of perception as an inferential process. This insight is invaluable because such a formal framework, regardless of which specific account or paradigm is used, can allow us to test populations that are not susceptible to illusions, such as autistic (Pellicano & Burr, 2012) and schizophrenic (Schmack et al., 2015) individuals

To conclude, in using a formal account borrowed from reinforcement learning and neuroeconomics, we find a BOLD response in the insula for switches cast as prediction errors. Our results support 1) high level cortical involvement in bistable perception and more generally support the notion of perception as an inferential process. In the future, a falsifiable perceptual uncertainty task, such as the dot-motion task, may further inform the inferential nature of perceptual uncertainty processing.

5 A Common Neural Representation of Uncertainty in the Insula

5.1 Introduction

In the previous chapter, we describe studies pertaining to the neural representation of uncertainty in two separate tasks, one objective and cognitive, and the other subjective and perceptual, in a same pool of subjects. We find common surprise-related BOLD responses in the insula for both tasks pertaining. In the following chapter, we attempt to directly compare these two tasks in a first instance. In a second instance, we perform a different analysis of the data sets with the aim of using a multi-variate pattern approach to determine if a more cogent picture of our research questions may emerge.

While neuroimaging has afforded us a distinctly powerful and galvanizing means by which to neurally investigate human behavior, difficult questions regarding the statistical reliability of results abound particularly in relation to the large number of statistical tests performed on each brain volume and the control of the concomitant false positive error rate (Woo et al., 2014). These concerns reached an apex recently that put into question a great deal of published results (Eklund et al., 2016). While pains were taken to address multiple comparison errors in previous chapters, here we opt to recast our data as a classification problem. Specifically, we seek to determine the neural patterns of activation related to specific levels of uncertainty and not necessarily a significant difference in BOLD signal between baseline and experimental condition. Such an analysis would allow us to target our research question from a different perspective and importantly may yield more insight into the neural representation of uncertainty than would standard hypothesis testing (Haxby, 2012). To that end, we first cast data from our two experiments into three levels of uncertainty (low, medium and high), and perform a standard general linear model analysis. We next investigate the parameter weights for each condition in regions of interest, to determine if they differ significantly from one another. Finally, we perform a multivariate pattern analysis to determine if 1) neural activity can accurately classify conditions according to their labels; 2) if the maps of such activity resemble one another across tasks (cognitive and perceptual).

5.2 Categorical Univariate Analyses

To further scrutinize the common neural activity found in our two tasks, we re-analyzed our fMRI data from Chapters 3 and 4, with the aim of directly comparing BOLD responses between the two tasks in a within subject design. We separated our fMRI data sets for the financial and perceptual uncertainty task into three categories, without modeling onsets. In the Card Game (Chapter 3), these categories refer to onsets for low, medium and high risk and high, medium and low surprise. We then did the same for the Necker Cube tasks, albeit with low, medium and high ambiguity of stimuli (which represents low, medium and high risk in task 1, and high, medium and low surprise in task2). We first perform standard parametric tests on our imaging analyses before examining different regressors parameter weights.

5.2.1 Financial Uncertainty : Categorical Analyses

We performed a categorical analysis by separating regressors at card 1 and card 2 according to three levels of risk and surprise respectively.

Financial Uncertainty - Main Effect of Risk

We performed an ANOVA across three different levels of risk (respectively, confidence) at the group level. Cluster level correction ($p=0.001$; $k=24$) yields a significant cluster notably in the right anterior insula, as well as bilateral inferior occipital gyri; left fusiform gyrus; and right angular gyrus. SVC analyses confirms bilateral insular engagement (See Appendix C, Table C.1). The amygdala, long known to modulate decision-making, particularly with respect to its affective component, makes its first appearance here in our study, though the cluster found for the region does not reach significance ($p=0.081$; $k=176$).

Financial Uncertainty, Risk at Card : Parameter Weights ANOVA

We then extracted percent signal changes for each level of risk in our 13 regions of interest (7 bilateral and the LC) and performed an ANOVA on the latter to determine if any significant differences emerged in pitting these coefficients against each other; and further to obtain a measure of confidence in the direction of the difference, if any.

Results of the ANOVA show significant differences in nearly all relevant regions, except for the ventral striatum. These differences appear to be driven by the lowest risk (highest confidence) condition; notably, high and medium risk levels do not appear to differ.

Financial Uncertainty, Surprise at Card 2 : - ANOVA

We perform the same mass univariate analysis for different categories of surprise (respectively, information), this time at card 2. Cluster level corrected results ($p=0.001$, $k=23$) show bilateral

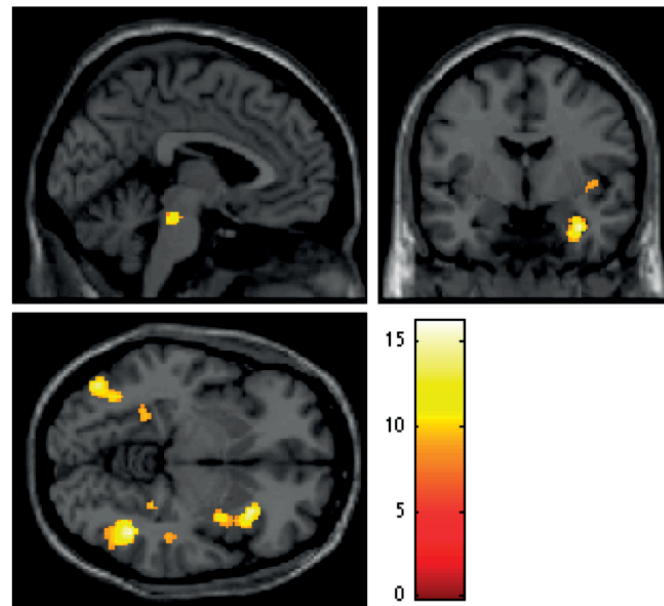


Figure 5.1 – Statistical map of main effect of risk (respectively, confidence) at card 1 ($p=0.001$; $k=24$) -2 -3 -6.

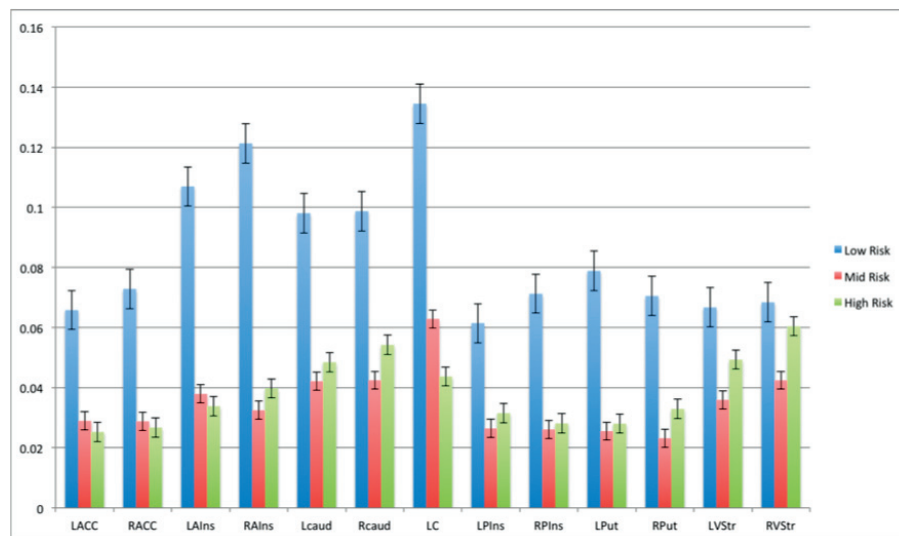


Figure 5.2 – Average percent signal changes in 7 ROIs according to risk level. Our model derived 5 different levels of risk in the task. Risk levels were divided into low (no risk); medium (the next two quantities of risk); and highest (the two highest quantities of risk).

anterior insula responses as well as significant clusters in the left supramarginal and left middle frontal gyrii; and right postcentral and middle temporal gyrii. SVC analyses on our list of a priori regions find significant bilateral BOLD responses in the ACC; anterior and posterior insulae; and caudate (See Appendix C, Table C.2).

Chapter 5. A Common Neural Representation of Uncertainty in the Insula

Table 5.1 – ANOVA results on differences between average percent signal changes across different levels of risk in 7 ROIs.

| Region | F | p |
|--------|---------|---------|
| LACC | 8.8268 | 0.0004* |
| RACC | 11.5628 | 0.0001* |
| LAIIns | 7.2490 | 0.0015* |
| RAIns | 8.3925 | 0.0006* |
| Lcaud | 5.0238 | 0.0096* |
| Rcaud | 4.5689 | 0.0142* |
| LC | 12.6222 | 0.0000* |
| LPIns | 2.4771 | 0.0925* |
| RPIns | 3.6342 | 0.0324* |
| LPut | 7.0930 | 0.0017* |
| RPut | 4.3033 | 0.0179* |
| LVStr | 1.4937 | 0.2328 |
| RVStr | 0.9247 | 0.4022 |

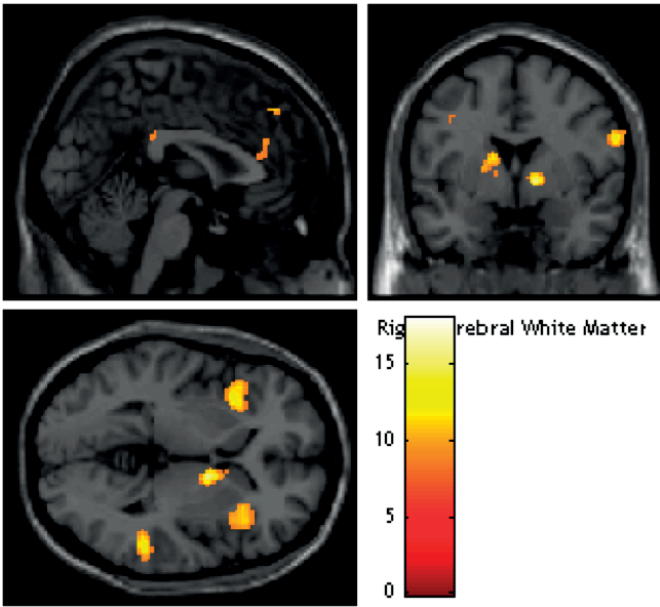


Figure 5.3 – Main effect of surprise (respectively, information) ($p=0.001$; $k=23$) overlaid on glass brain and canonical T1 image (1 0 1).

Financial Uncertainty, Surprise at Card 2: Parameter Weights ANOVA

We then extracted percent signal changes for each level of surprise in our 13 regions of interest (7 bilateral and the LC) and performed an ANOVA on the latter to determine if any significant differences emerged between surprise levels and to further gain a picture of any trends (linear or non) in the differences.

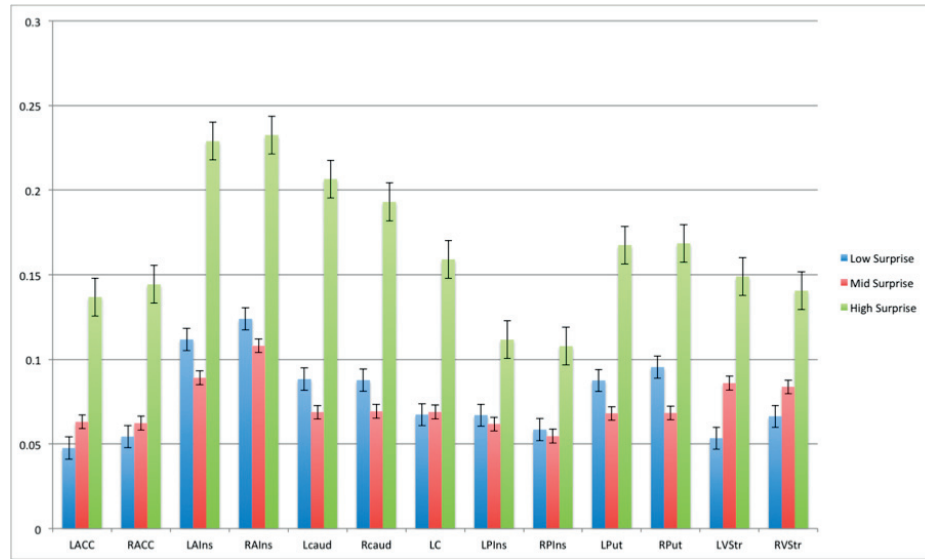


Figure 5.4 – Average percent signal changes in 7 ROIs according to surprise level. We derived 9 levels of surprise in our task. Here, we group the three lowest, three middle, and three highest levels of surprise to obtain 3 classes of the value.

We performed an ANOVA on the percent signal changes associated with each level of surprise (respectively, information) and find significant differences in all regions but for the left posterior insula.

Table 5.2 – ANOVA results on differences between average percent signal changes across different levels of surprise in 7 ROIs.

| Region | F | p |
|--------|---------|---------|
| LACC | 9.5771 | 0.0002* |
| RACC | 10.212 | 0.0002* |
| LAIns | 11.9647 | 0.0000* |
| RAIns | 8.9257 | 0.0004* |
| Lcaud | 13.0086 | 0.0000* |
| Rcaud | 9.3871 | 0.0003* |
| LC | 6.1698 | 0.0037* |
| LPIns | 2.7938 | 0.0692 |
| RPIns | 3.6986 | 0.0306* |
| LPut | 13.6017 | 0.0000* |
| RPut | 11.6956 | 0.0001* |
| LVStr | 7.2418 | 0.0015* |
| RVStr | 6.2546 | 0.0034* |

A cursory view of these values shows a pattern similar to that of risk at card 1; highest surprise (lowest information), drives the difference in means, while lowest and medium surprise do not appear to differ from one another. This result provides us with a sanity check on our

framework, as it mirrors the trend found for risk in card 1, sampled at a different timepoint, namely in that we assume highest risk to prompt the least surprise and lowest confidence to yield the most information.

5.2.2 Perceptual Uncertainty: Categorical Analyses

The categorical analysis of the perceptual uncertainty task is more fraught than that of the gambling task because of the subjective nature of the categories as well as the overall bias in percept towards the form above percept. Thus we perform two kinds of categorical analysis: one examining different levels of uncertainty (3 levels) and one examining the different levels of bias (towards the from above percept; 5 levels). We first performed a categorical analysis of uncertainty by separating trial types according to 3 levels of uncertainty.

Necker Cube Task 2 - Categorical Analyses

In task 2, we repeat the same analyses as we did for task 1, namely by looking at a main effect of ambiguity on neural activity as well as a main effect of stimulus bias towards the from above percept. To that end, we designed two general linear model designs.

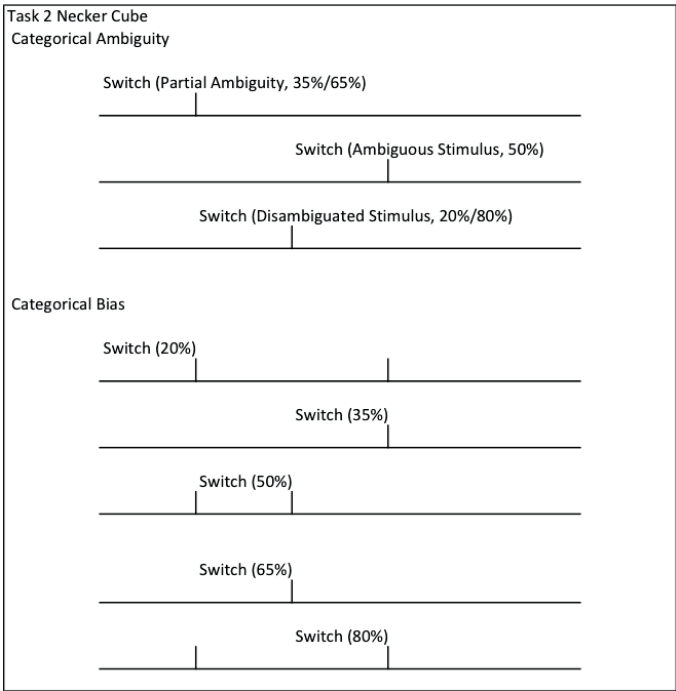


Figure 5.5 – Representative schemas of general linear model designs for the categorical analysis of the Necker Cube Task 2 fMRI data

Necker Cube Task 2 - Main Effect of Ambiguity

Similarly to Task 1, we performed a GLM that included 3 regressors of interest, in addition to 7 regressors of no interest (1 to account for replay presses and 6 to account for motion regressors). These regressors included switch onsets separated according to stimulus ambiguity: High (50%), Mid (35%/65%) and Low (20%/80%). We then performed an ANOVA to investigate a main effect of perceptual ambiguity on the brain.

Results show bilateral activation in the middle temporal gyrus as well as in the insula (anterior extending to posterior) and the putamen. In addition, significant clusters were found in the left inferior frontal gyrus, and right middle frontal gyrus. We then applied SVC to our list of ROIs, which yielded significant bilateral clusters in bilateral insulae and putamen (See Appendix C, Table C.3).

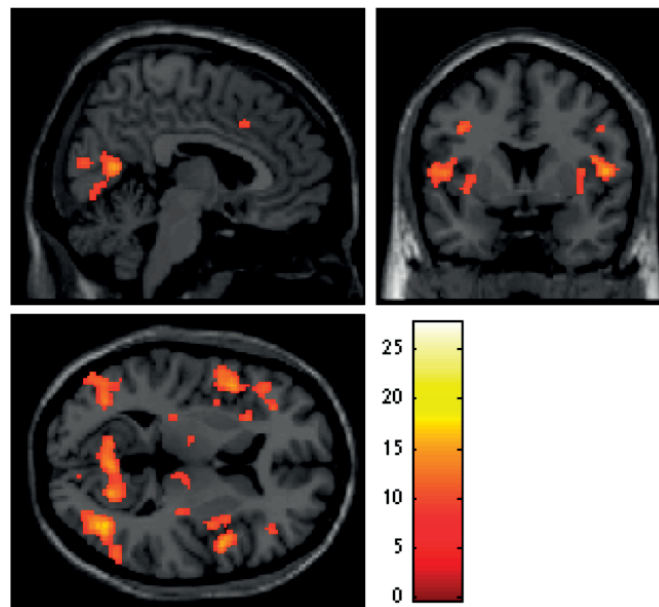


Figure 5.6 – Statistical map of the main effect of stimulus ambiguity on perceptual switches, Necker Cube Task 2 ($p = 0.001$, $k = 22$) (6 4 8).

As in the financial uncertainty task, we next extracted parameter weights to determine how percent signal changes vary across different levels of ambiguity in the perceptual uncertainty task.

Results above show that, once again, the insula plays a significant role in the representation of uncertainty. However, it must be noted that percent signal changes appear to show a u-shaped relationship to one another, where higher signal changes correspond to lowest and highest ambiguity. While this trend may, *prima facie*, appear counterintuitive, one could argue that the partially ambiguous stimulus may be the most ambiguous stimulus; in a sense, the most ambiguous stimulus is known to be perceptually unreliable, while the least ambiguous stimulus is known to be perceptually reliable, while the stimulus at the midpoint may not

Chapter 5. A Common Neural Representation of Uncertainty in the Insula

Table 5.3 – ANOVA results for the main effect of ambiguity on percent signal changes in the Necker Cube, Task 2, in 7 ROIs

| Region | F | p |
|--------|--------|---------|
| lACC | 1.4516 | 0.2432 |
| rACC | 1.2410 | 0.2972 |
| lAIns | 4.0229 | 0.0235* |
| rAIns | 3.1351 | 0.0515 |
| lCaud | 0.8038 | 0.4529 |
| rCaud | 3.6501 | 0.0326* |
| LC | 4.7389 | 0.0127* |
| lPIns | 4.7939 | 0.0121* |
| rPIns | 2.8169 | 0.0686 |
| lPut | 1.9092 | 0.1581 |
| rPut | 2.5875 | 0.0845 |
| lVStr | 1.4516 | 0.2432 |
| rVStr | 1.2410 | 0.2972 |

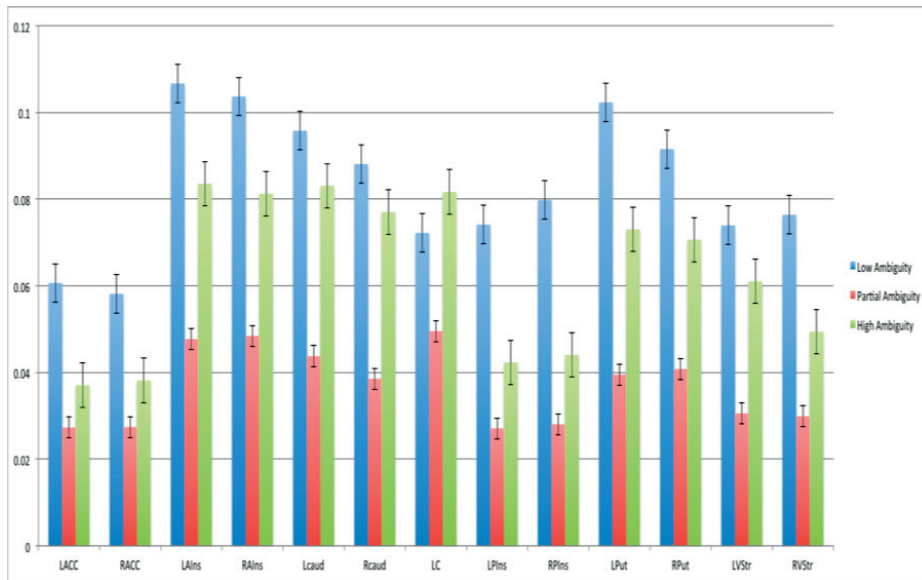


Figure 5.7 – Graphical representation of average percent signal changes across 3 levels of ambiguity in 7 ROIs.

occupy a clear category or expectation.

Necker Cube Task 2 - Main effect of 'from above' Bias

Here we sought to find differences in neural activity related to the bias level of switches. Thus we constructed a GLM with 5 regressors of interest, as well the 7 regressors of no interest (replay presses and motion regressors). Each of the 5 regressors included perceptual switch

5.3. Multivariate Pattern Analyses - Preliminary Analyses

reports for a specific bias condition related to the probability of viewing the cube from above. An ANOVA shows differences bilaterally in the anterior insula and in the anterior cingulate cortex ($p = 0.05$, FWE), as well as in the thalamus and cerebellum (See Appendix C, Table C.4).

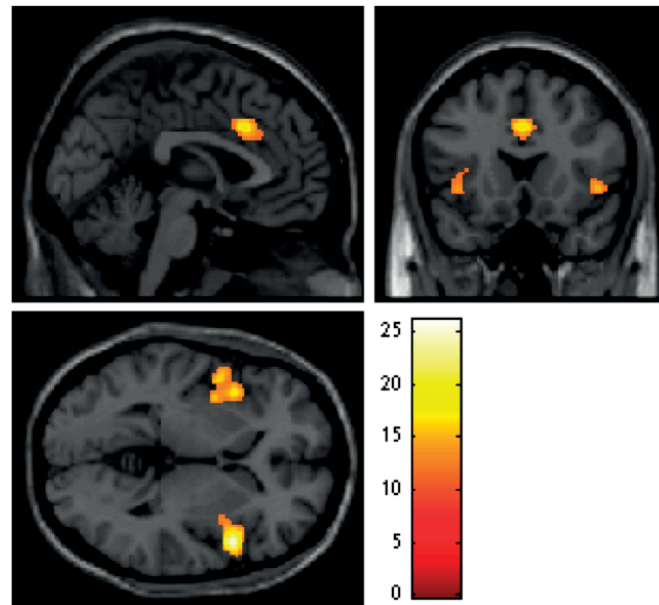


Figure 5.8 – Statistical map for the main effect of from above bias in Necker Cube, task 2 ($p=0.05$, FWE corrected) (0 17 1).

Prominent clusters in the bilateral insula appear here as well and this at a whole brain, FWE corrected threshold. We must conclude that the effect of viewing objects from above in the scanner induces a significant effect at the level of insula, above and beyond the effects of ambiguity. We nonetheless moved on to investigating a main effect of from above bias on percent signal changes in our preselected ROIs.

Results above suggest a strong effect of bias towards from above conformation, particularly in the insula (anterior and posterior) as well as in the right striatum (dorsal and ventral) and right anterior cingulate cortex and LC.

Taken together, results above confirm a predominant role of the anterior insula in encoding uncertainty across cognitive and perceptual domains. Nonetheless, the trend of percent signal changes across categories suggests that additional analyses should be performed to reinforce this relationship.

5.3 Multivariate Pattern Analyses - Preliminary Analyses

The General Linear Model represents the canonical means by which to analyse fMRI data (Friston, 1995). The model relies on a mass univariate approach. However, several authors are advocating for a multivariate approach because of the problems a univariate approach

Chapter 5. A Common Neural Representation of Uncertainty in the Insula

Table 5.4 – Main effect of from above bias on average percent signal changes in 7 ROIs

| Region | F | p |
|--------|---------|---------|
| lACC | 0.7587 | 0.5551 |
| rACC | 4.6925 | 0.0018* |
| lAIns | 7.0615 | 0.0001* |
| rAIns | 11.1467 | 0.0000* |
| lCaud | 1.5746 | 0.1885 |
| rCaud | 9.8983 | 0.0000* |
| LC | 8.2071 | 0.0000* |
| lPIns | 3.3749 | 0.0130* |
| rPIns | 4.0423 | 0.0048* |
| lPut | 1.9688 | 0.1066 |
| rPut | 4.9993 | 0.0011* |
| lVStr | 0.7587 | 0.5551 |
| rVStr | 4.6925 | 0.0018* |

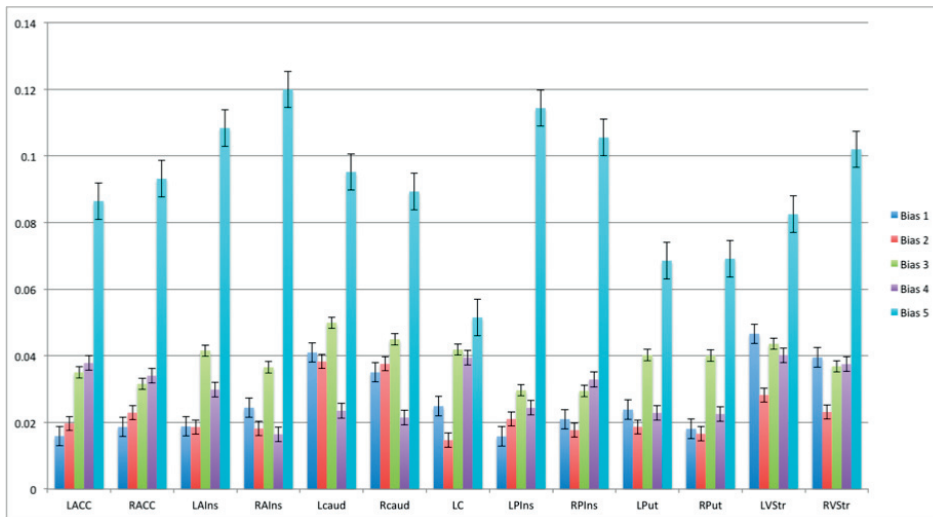


Figure 5.9 – Graphical representation of average percent signal change across subjects for varying levels of stimulus bias towards a from above percept in 7 ROIs.

offers (Poline & Brett, 2012). The univariate approach relies the independent testing of each voxel to determine significant changes in the BOLD signal in relation to an experimental event. However, the assumption of independence of a voxel does not conform to the assumption that voxels in a region, or subregion of the brain should not be independent. To date, the application of a smoothing kernel to brain data was one means by which to control for this local dependence (Friston, 1995). However, the advent of multi-variate methods that rely on the information content of patterns of activation afford us another means by which to investigate variations in brain activity in relation to an experimental event (Haynes, 2015).

Our primary goal in separating our fMRI datasets into categories as opposed to a parametric

design was to determine if multivariate analyses methods would succeed in identifying similar patterns of activation in classifying data according to level of uncertainty, namely 3 different levels of risk and surprise for the financial uncertainty task (respectively confidence and information); and ambiguity in the perceptual uncertainty task.

We employ the decoding toolbox (Hebart et al., 2015) and perform a linear classification technique that relies on a support vector machine with a fixed cost function, $c = 1$. Each individual subject's dataset is separated into a training set and a test set with a leave-one out cross validation test. In financial uncertainty, 2 of three sessions are used as training sets while the third was cast as a test set. For the perceptual uncertainty task, as there was one session only, we separated trials $i=1:k$ into two sessions, one for odd indices and the other for even ones. This creates a design where there are two sessions, one employed for training and the other for testing.

Results of this analysis yield accuracy values, where successful classification yields a value above chance (thus 33.33% for the three different classes of uncertainty) for each voxel, for each individual subject. In addition, the technique yields accuracy maps for the brain data.

5.3.1 Financial Uncertainty: Surprise (Information) Classification

We obtained accuracy maps using a linear classification algorithm for 21 subjects with a searchlight method. We then normalized all subjects' accuracy maps to the MNI template before smoothing normalized images to a 8 mm FWHM Gaussian filter. Finally, we performed a t-test on these images in a random effects model. It is important to note that this test is qualitatively different from standard group level analyses, because accuracy maps represent information content. Thus here, the t-test does not represent hypothesis testing, as a below chance result is meaningless (Allefeld et al., 2016). At best, the t-test we perform emulates a fixed effects analysis. Therefore, the maps below indicate that some subjects show an above chance classification accuracy in certain regions of the brain, however, we cannot make a population level inference about the maps.

The image below is representative only (Figure 5.10), as cluster level correction may not be an accurate means by which to represent information culled from this analysis. Using a cluster level corrected threshold ($p = 0.001$; $k = 41$), we find significant clusters in bilateral anterior insulae; left inferior temporal gyrus, transverse temporal gyrus, left calcarine and right superior occipital cortex (See Appendix C, Table C.5).

In addition to the insula and the anterior cingulate, results above show a cluster in the left putamen, similar to accuracy maps for risk. Further, in addition to regions we expect to emerge in this analysis, we find a cluster in the entorhinal cortex.

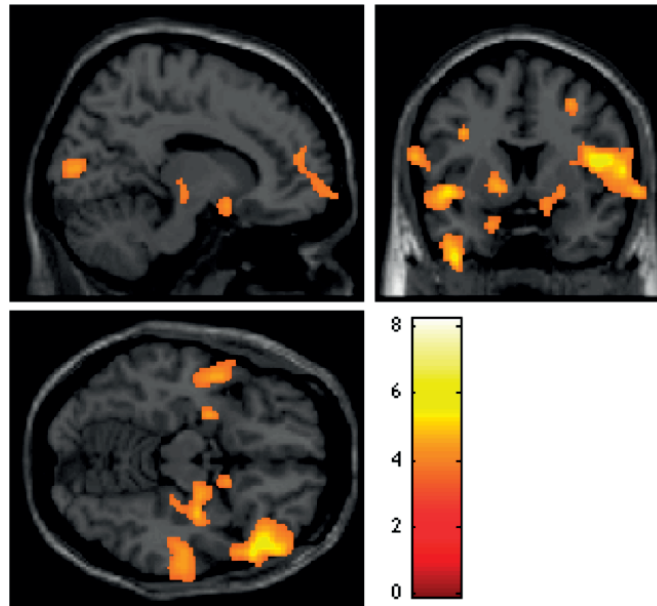


Figure 5.10 – Average accuracy maps for the classification of different levels of surprise categorization in the financial uncertainty task. The average represented here is the result of a fixed effects analysis and is not corrected for inter-individual variability. (27-16)

5.3.2 Perceptual Uncertainty: Surprise Classification

We obtained accuracy maps using a linear classification algorithm for 19 subjects with a searchlight method (Figure 5.11). As in the financial uncertainty task, we then normalized all subjects accuracy maps to the MNI template before smoothing normalized images to a 8 mm FWHM Gaussian filter. Finally, we performed a t-test on these images in a random effects model. We performed a nominal group level t-test but no significant clusters emerged. Nonetheless, we show group level maps below as a means to represent the information contained in the pattern found.

Accuracy maps for single-subject classification show a cluster in the left putamen, and left anterior insula, as we find in the financial uncertainty task, as well as a cluster in the left entorhinal cortex, as in the cognitive uncertainty task. In addition, here we find patterns of activity in the orbital gyrus, as region known to be implicated in decision-making and valuation (Padoa-Schioppa & Assad, 2006).

5.4 Parameter weights correlations

The results above and in previous chapters support common neural activations associated with uncertainty processing and surprise in cognitive and perceptual domains. To obtain a more granular picture of this commonality, we explore beta weight estimates for surprise linked to gamble outcomes and surprise linked to visual illusions within subjects.

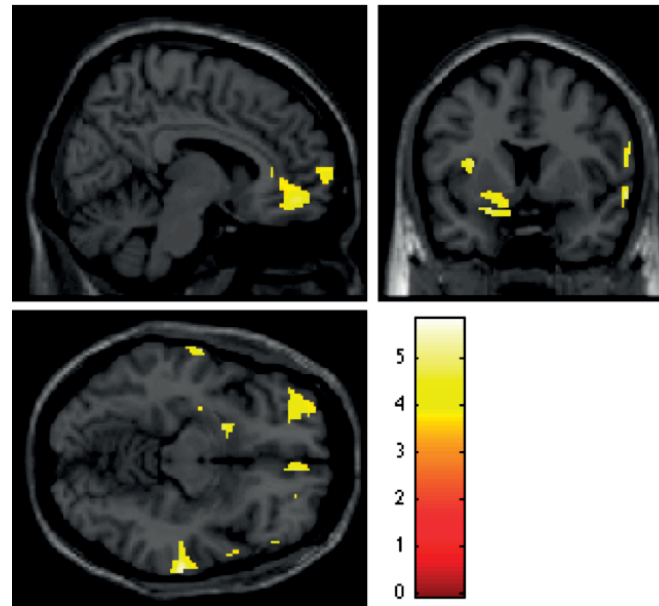


Figure 5.11 – Average accuracy maps for the classification of different levels of ambiguity categorization in the perceptual uncertainty task (2 7 -16). As in Figure 10, the average here does not take into account inter-subject variance. Therefore, these maps must be viewed with caution, as outliers may potentially drive the average.

We extract beta parameters for all subjects that successfully performed both tasks, leaving us with a sample size of 14. Beta parameter extraction is performed for ROIs that show both common activations in both tasks for surprise and that are included in our a priori lists, which leaves the bilateral anterior insula. We also include the LC as an especially relevant region for our research questions and reference region for this analysis, (region VI-VII of the cerebellum vermis). Finding a reference region for two tasks that engage audition and vision as well as decision-making is no small feat and indeed, cerebellar activity is seen in the Card Game however, we were guided in our choice by the following criteria: a wish to avoid cortical regions, as these tend to be less well defined; finding a low-level region; and choosing a mid-line region, to avoid laterality effects. While we see a linear relationship between insular values across tasks, none are significant (Figure 11).

5.5 Discussion

Herein, we conduct a series of analyses to corroborate our model-based findings on common neural representations of uncertainty across cognitive and perceptual domains. We opt for a categorical model of investigation, which implies a loss in variance in our data, with an aim to deploy multivariate pattern approaches, yet we still find a role for the insula in particular with regards to uncertainty processing when performing a standard univariate analysis on these unmodeled data.

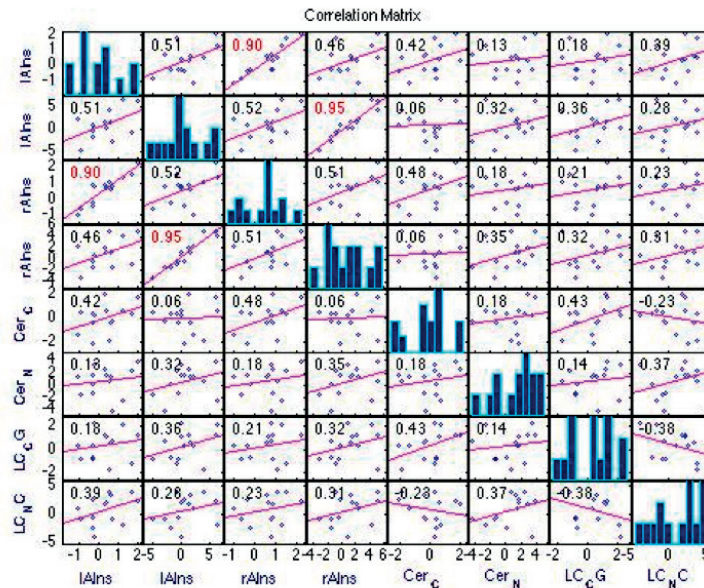


Figure 5.12 – Correlation matrix of group average beta weights for surprise in the financial uncertainty task (odd columns) and surprise in the perceptual uncertainty tasks (even columns) in bilateral insula, cerebellar vermis and locus coeruleus.

Table 5.5 – Correlation matrix of p values associated with regression of surprise coefficients from financial (CG) and perceptual (NC) uncertainty tasks in bilateral insula, cerebellar vermis and locus coeruleus in 14 subjects. Results in red show a strong inter-task relationship that does not reach significance

| | L Ains CG | L Ains NC | R Ains CG | R Ains NC | Cer CG | Cer NC | LC CG | LC NC |
|-----------|-----------|-----------|-----------|-----------|--------|--------|--------|--------|
| L Ains CG | 1.0000 | 0.0617 | 0.0000* | 0.1017 | 0.1365 | 0.6550 | 0.5378 | 0.1720 |
| L Ains NC | 0.0617 | 1.0000 | 0.0565 | 0.0000* | 0.8442 | 0.2695 | 0.2079 | 0.3365 |
| R Ains CG | 0.0000* | 0.0565 | 1.0000 | 0.0620 | 0.0860 | 0.5296 | 0.4680 | 0.4263 |
| R Ains NC | 0.1017 | 0.0000* | 0.0620 | 1.0000* | 0.8308 | 0.2164 | 0.2644 | 0.2794 |
| Cer CG | 0.1365 | 0.8442 | 0.0860 | 0.8308 | 1.0000 | 0.5341 | 0.1205 | 0.4338 |
| Cer NC | 0.6550 | 0.2695 | 0.5296 | 0.2164 | 0.5341 | 1.0000 | 0.6285 | 0.1913 |
| LC CG | 0.5378 | 0.2079 | 0.4680 | 0.2644 | 0.1205 | 0.6285 | 1.0000 | 0.1842 |
| LC NC | 0.1720 | 0.3365 | 0.4263 | 0.2794 | 0.4338 | 0.1913 | 0.1842 | 1.0000 |

We first tested for mean differences across three different levels of risk and three different levels of surprise. We find a main effect for both risk and surprise in the insular cortex, corroborating our findings in Chapter 3.

Parameter weights extraction in the financial uncertainty task support our assumed relationship between risk and surprise, notably that higher signal percent changes for low risk, a prediction variable, are mirrored by higher percent signal changes for high surprise, a prediction error variable. A significant main effect of risk is seen in all regions except for the ventral

striatum, however only the anterior cingulate cortex, LC (as assessed by the Keren atlas) and left anterior insula see a linear decrease in BOLD response with increased risk, though differences between medium and low risk do not appear important. This result segues well with the notion that increased risk, or a convergence towards ambiguity (Knightian uncertainty) decreases neural activity (Vilares & Kording, 2011). Our categorical univariate analysis on surprise also shows a significant main effect of surprise in all ROIs, except for the left posterior insula. This effect appears to be driven by high surprise in particular, as low and medium levels of surprise do not differ from one another. Nonetheless, BOLD responses increase linearly with surprise in bilateral anterior cingulate cortex, LC and ventral striatum. The ventral striatum's increase for surprise is telling, as we expect surprise to act as a learning signal, a process which is hypothesized to occur in the region in particular (Daw et al. 2011; Grahm et al., 2009; Wise et al., 2004; McClure et al., 2003; Shohamy, 2011).

We performed a similar categorical analysis on Task 2 of our perceptual study in Chapter 4. We constructed regressors composed of switches for low, medium and high levels of perceptual ambiguity (respectively, surprise). An F test probing a main effect of ambiguity finds a significant insular involvement, although the clusters seen span the posterior as well as the anterior insula, putting in question the commonality of the BOLD response found in the categorical analysis of perceptual uncertainty relative to that of financial uncertainty, at least with respect to the insula. To examine the relationship between the different levels of ambiguity beyond mere significant differences, we performed ANOVAs on percent signal changes in our 13 ROIs. We find significant differences in the left posterior insula, left anterior insula, right caudate and LC. In plotting these quantities however, we do not see a linear relationship emerging; we find lower BOLD response signal changes for partial ambiguity relative to high or low ambiguity. This pattern is unexpected; however, it may be that partially ambiguous stimulus is in reality more ambiguous than either fully ambiguous or disambiguated stimuli.

Finally, because of our findings in Chapter 4 in relation to the asymmetry between from-above and from-below percepts, we also probed a main effect of bias towards the from above perspective. We separated switch onsets according to which stimulus condition they occurred in (cubes biased from 20-80% towards the from-above perspective). Here, whole brain results thresholded at $p=0.05$ corrected for family wise error shows bilateral anterior insula activation, as well as in the anterior cingulate cortex. Together, these regions form the principal actors of the salience network (Uddin, 2015), the network we posit to be especially relevant to surprising events. An ANOVA on percent signal changes for different bias conditions in 13 ROIs shows a significant difference in the right anterior cingulate cortex, which scales linearly with increasing bias towards the from above perspective. Indeed, in all regions, percent signal changes for switches in the 80% condition are markedly higher than in all other conditions. This finding could be an artifact but we suggest otherwise. These results underscore our behavioral findings from Chapter 4. Specifically, we suggest that not only are switches in the 80% condition more surprising than in the 50% or 35%/65%, but they are more surprising than the 20% condition. We hypothesize that we hold both a strong prior for perceiving a disambiguated cube and that we hold a strong prior for perceiving a cube from above. Thus

Chapter 5. A Common Neural Representation of Uncertainty in the Insula

we weight a cube's unsigned ambiguity by our preferred, or most sensible, or most expected percept, which is the from-above percept. Switches in this condition may thus induce the strongest surprise, a possibility that is corroborated by our fMRI data.

Our primary aim in conducting this analysis was to perform a classification on our differing levels of uncertainty across tasks, to determine common and distinct neural patterns of activation in the different tasks. We reiterate here that MVPA techniques do not yield a statistical result in line with hypothesis testing; that is we are not testing to see if average patterns found are at below chance level. Instead, we seek an above chance classification accuracy, or odds of correctly predicting an out-of-sample datum to its true label. Thus, below-chance results are not directly applicable in this context, nor are group-level inferential statistics. Our results above are preliminary, as we may be better served by performing a representational similarity analysis, as well as permutation tests on our accuracy maps. Nonetheless, the goal of our MVPA analysis was to determine if the same regions found in our univariate analyses emerged here. The insula appears in both tasks' MVPA accuracy maps. Further, a cluster in the left putamen, as well as in the left entorhinal appear in the Card Game and Necker Cube tasks. Finally orbitofrontal regions emerge in the Necker Cube task. While we do not highlight this region in this thesis, the OFC has a prominent role the decision-making literature, specifically in relation to valuation. While the insula appears in both group accuracy maps for the Card Game as well as the Necker Cube Task, the remaining patterns of information content between the two tasks do not resemble one another. Thus, further analyses are warranted to probe the results above, with the aim of solidifying our findings in previous chapters.

Our results in this Chapter, as well as in Chapters 3 and 4, repeatedly find the insula involved in different kinds of uncertainty (risk and surprise) and different domains (perceptual and financial). BOLD responses in this region are significant in both model-based analyses (Chapters 3 and 4); categorical analyses (this Chapter). Further, classification of BOLD responses according to uncertainty levels finds insular involvement (this Chapter). As our studies were conducted in the same pool of subjects, we extracted beta weights for surprise in the Card Game and surprise in Task 2 of the Necker Cube for the 14 subjects who performed both tasks. We extracted beta weights in bilateral insula and the LC (Keren atlas), the former because it is both in our list of a priori regions and also because, as noted, it is found in all analyses related to uncertainty; and the latter because it was, at least initially, a prime region of interest with respect to uncertainty and surprise in particular. We also extract surprise-related beta weights for the cerebellum, as a reference region. We find a linear relationship for financial (cognitive surprise) as well as perceptual surprise, bilaterally, though the p values found do not reach significance ($p = 0.0617$ and $p = 0.0620$, in left and right insulae, respectively). We hypothesize that statistical significance ultimately eludes us due to the low sample size of $n = 14$. Tellingly, our reference region shows no correlation in surprise related beta weights across tasks, which shows that it is not a system-wide within subject correlation that we find in the insula.

The insula was introduced in Chapter 1 of this thesis but, unlike other regions reviewed in

the introduction, it has subsequently played a leading role in uncertainty processing in each of the following Chapters, including this one, across all forms of data analysis and in both experimental tasks. Thus, we must pose the following question: why the insula?

Studies using fMRI have found insula BOLD responses in relation to several phenomena. A considerable body of evidence implicates the insula in language (Ackermann & Riecker, 2004; Ackermann & Riecker, 2010; Ardila et al., 2014); auditory processing (Bamiou et al., 2003); interoception, or awareness of bodily signals (Craig & Craig, 2009; Craig, 2011); pain (Peyron et al., 2000); disgust (Wicker et al., 2003); addiction and drug craving (Naqvi & Bechara, 2009; Naqvi & Bechara, 2010; Garavan, 2010); gustatory function (Small, 2010); perception (Sterzer & Kleinschmidt, 2010); decision-making (Weller et al., 2009; Preuschoff et al., 2008; Singer et al., 2009) and emotion in general (Gasquoine, 2014). One could be forgiven for disregarding insular responses altogether and declaring the region a cortical junk drawer. An urge to do just that with regards to the insula likely comes from how significant results have been reported; namely, in neuroimaging studies, a specificity of function is often explicitly stated with regards to the insula in a particular study, which of course is challenged when a following unrelated study finds a similar BOLD response. We argue that rather than ignore insular responses in neuroimaging study, we must seek a commonality in the functions mentioned above. Indeed, an attempt at this has already been made, by casting the insula as an interoceptive center, a region that integrates bodily states into awareness (Craig & Craig, 2009). Interoception in itself already presents a specific, if broad category that can readily explain particularly salient emotions (love, fear, psychological pain) as well as sensory states (tasting, craving, disgust, pain). Awareness of said states can also explain the insula's putative role in language processing: awareness generally includes a declarative function, which is most readily translated into language production. Thus the interoception explanation covers most of the studies referenced above. But then what of the insula's role in uncertainty processing, decision-making, prediction errors and perception?

This facet of the insula has indeed been examined more recently (Klein et al., 2007; Klein et al., 2013), specifically the insula's role in error awareness. Thus the plot thickens with this idea, because it is not mere error detection that we attribute to the insula, but error in addition to an awareness of it that is crucial to the insula. While error is generically viewed as synonymous to a conflict or a mistake, from a predictive coding perspective, it is simply a difference between two states. For instance, in Chapter 3, we cast our information quantity as a confidence error, a quantity we believe does not engender conflict, but does encompass a difference. Klein and colleagues briefly allude to predictive coding to note that it is compatible with their account but do not examine the point in more detail.

In a related manner, the insula (and the anterior cingulate cortex) has emerged as a main component of the salience network in resting state fMRI states (Menon & Uddin, 2010). Indeed, one could argue that salience is an alternative to a neutral state - or the default mode network. Salience in itself implies error or deviation - or at the very minimum, a difference from a neutral or default state.

Chapter 5. A Common Neural Representation of Uncertainty in the Insula

Prior to the advent of neuroimaging, lesion cases provided the best means by which to assign function to location in neurological studies. In the case of the insula, lesion studies appear problematic, for few are isolated to the region (Gasquoine, 2014). Nevertheless, we review some evidence suggesting that insula dysfunction leads to an impaired awareness of prediction errors. For instance, it has been found that individuals with insula lesions quit smoking more easily (Naqvi et al., 2007). If one views craving as a homeostatic prediction error, one can see that a lack of awareness of the error may result in an inability to identify the bodily state of craving, leading to an ease in smoking cessation. Several studies report various language impairments following insular dysfunction (Ardila et al., 2014; Shuren, 1993). As mentioned, most insula lesions also affect neighboring language related locations, however, findings related to impaired language processing in patients does not contradict the region's putative role in awareness, as such a function implies the living of a declarative state. Indeed, most studies on the insula's role in language focus on speech production, a by-product of declarative capabilities.

Taken together, we hypothesize that the insula's involvement in uncertainty processing reflects its role in acting as a mediator between prediction errors (irrespective of origin) and declarative states. Thus, if a state can be named and it arises from a prediction error, we would expect insular involvement. Such a capacity could explain the studies reviewed above as well as our findings. For instance, pain can be named and should, at least in principle, result from a prediction error. The same can be said for disgust. However, prediction errors that are not expected to engender an awareness or declarative state, such as early trials in an implicit learning task, would not be expected to elicit insula activation.

Recall that all the uncertainty related phenomena that we study straddle both computational accounts as well as feeling states, including surprise, confidence and uncertainty. We propose that the insula applies a transfer function to these errors, whose output is a name (Figure 13).

The role of the insula we propose is based on the model proposed by Singer et al., 2009 and incorporates Craig's view of interoception as well as Klein's view of insula modulating error awareness. This function of the insula would be far from trivial. For instance, what happens if an individual's transfer function goes awry, perhaps mistaking love for fear? What is clear when one considers computational accounts of human function and behavior is that we do not assume that humans explicitly or consciously make these calculations. Our task in Chapter 3 is exceptional in that we assume adult subjects are capable of easily and explicitly computing, for instance, their probability of winning in a trial. It is an example of conscious inference, as opposed to Helmholtzian unconscious inference or Fristonian active inference. We concede that outside of the financial realm, it is difficult to find concrete examples where we expect individuals to explicitly compute. say, the variance of a given cue's value in day-to-day decisions. Such a process would be unrealistically taxing. We do assume however that the brain does make these computations, even if the conscious brain does not, just as an infant understands gravity, yet does not know it's related equation. In this sense, we propose that computations effective meaning are made explicit via the insula, which in essence transforms

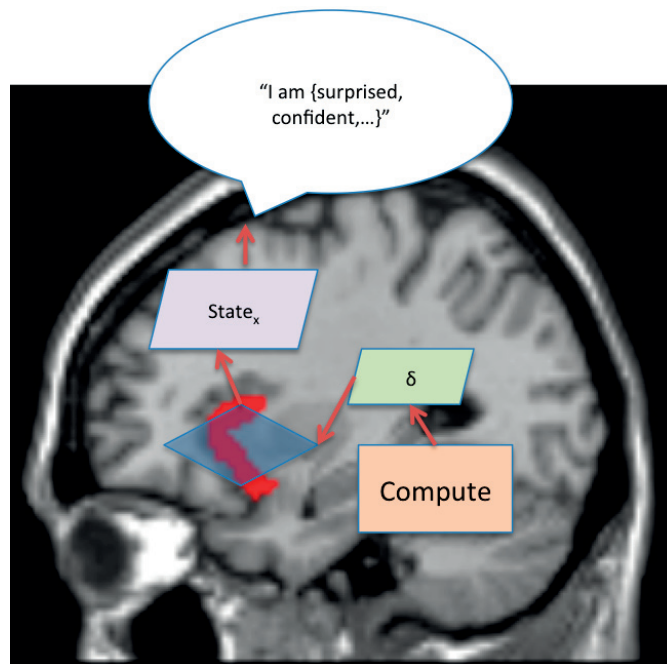


Figure 5.13 – A proposed model for insular function. We hypothesize that the insula integrates prediction errors arising from computations in lower levels of the brain, before applying a transfer function to the latter, to output a declarative state available to consciousness. Here the insula acts as a translator between the computational brain and the word.

values into easy to understand categories, such as "I am somewhat unsure" or "I am more informed", etc. just as it does for visceral input.

5.6 Conclusion

In this chapter, we sought to pull together results from our tasks in Chapters 3 and 4, conducted in the same pool of subjects, to directly compare uncertainty related BOLD responses in a perceptual task versus a cognitive one. We find the insula emerging as a uniquely relevant region to uncertainty across tasks. Because the insula has also been implicated in a wide variety of phenomena, we propose a model to integrate our view of the region's role, namely in low level signals, including prediction errors, before translating them into declarative states.

6 Conclusion

6.1

The work outlined above offers several insights into decision-making under uncertainty. We begin by considering the question of how humans address the problem of uncertainty, under the assumption that 1) uncertainty is a feature that permeates all our interactions: 2) the brain resolves this uncertainty. We thus work within the paradigm whereby we view the brain as an inference machine. From this starting point, we view human decision-making as an inferential process. We parsed the decision-making process into different components, namely its predictive components and its error components before applying probabilistic models that capture the hidden mechanisms we assume are at play when the human and its brain are faced with uncertainty.

6.2 Questions Posed

- In a first instance, to probe noradrenergic involvement in uncertainty processing, we performed a neuroimaging study to indirectly localize the locus coeruleus. In a first instance, we undertook to indirectly localize the locus coeruleus, the main site of noradrenergic neurons in the brain, using fMRI. Such a method of indirect localization did not prove fruitful. We then performed a pupillometry study using an auditory oddball task, to elicit a reliable cognitive pupil response, a putative proxy of noradrenergic activity. We then administered this task in an MRI scanner over three different sequences, to determine which would yield the best brainstem BOLD response. Results of our pilot studies ultimately guided us towards a different approach with respect to our main questions of interest; namely, we concluded that LC localization with fMRI methods is tenuous. Our subsequent studies thus focused on other regions in the brain to probe their role in uncertainty processing.
- Next, we aimed to disentangle the BOLD responses to three key variables related to decision-making under uncertainty. We first investigated the uncertainty predictions,

notably risk, as variance; and confidence, as precision. We next investigated surprise, as absolute prediction error, and information, as a confidence error. We find objective confidence engages the insula, when controlling for reward expectation. As confidence is a form of belief, this finding underscores the importance of belief in decision-making under uncertainty. With regards to uncertainty-related prediction errors, we find insular and anterior cingulate involvement that scales with surprise. Intriguingly, we also find a common insular activation for information, our confidence error, as well as distinct patterns in the cuneus and frontal cortex. Thus we disassociate the signals of these two quantities, surprise and information, in the brain and in so doing provide evidence that information is not always surprising, and that information is reflected in the BOLD responses, a novel finding.

- Finally, we administer a novel perceptual uncertainty task to the same subjects that performed the financial uncertainty task above. We induce uncertainty by presenting the Necker Cube illusion, with cubes biased towards one percept or another. By applying a predictive coding account to perceptual switches, we model the perceptual switch as a unsigned prediction error, or surprise. In so doing, we find bilateral insula involvement for perceptual surprise. This finding provides strong evidence of a dedicated role for the insula in general uncertainty processing, irrespective of domain or modality.

6.3 Impact of Findings

The work herein has borrowed concepts from several different fields, including neuroscience, philosophy, psychology, economics, signal processing and machine learning. It has concomitantly generated findings that impact three broad areas.

- The neural evidence of decision variables related to uncertainty, namely confidence, surprise and information allows us to enrich the framework within which we investigate neuropsychiatric illnesses. Namely, several of the latter involve subtle impairments in executive function, and decision-making in particular, but quantifying such disruptions can prove challenging. In Parkinson's disease, for example, performance on decision-making tasks is not always impaired (Ryterska et al., 2013). Schizophrenia, depression, anxiety and obsessive-compulsive disorder all similarly appear to engender decision-making impairment but the exact nature of the deficit remains elusive. Indeed, most tasks are meant to capture the operationalization of a given decision-variable without a computational formulation. By defining specific variables with a computational account, we can better differentiate those processes that may be uniquely affected by a disease state. In so doing, rehabilitative and therapeutic plans targeting specific dysfunctions can be refined. In Parkinsonians, evidence of decision-making dysfunction may be due to a dopaminergic overdose linked to dopamine therapy flooding intact striatal regions, and not to the disease itself (Aarts et al, 2012). Similarly, in obsessive-compulsive disorder, specific impairment in prediction error integration may be responsible for

aberrant behavioral responses, rather than a system-wide cognitive impairment. Finally, one can imagine that a system that cannot resolve the surprise signal - that is a patient that does not show surprise signals - would exhibit significant behavioral impairments that interfere with daily tasks. Thus, the differentiation of decision-variables above and beyond rewards and outcomes, as well as their computational definitions, formulated to capture their hidden nature, presents a powerful tool with which to address clinical questions.

- Currently, artificial intelligence algorithms look to reproduce human behavior. This endeavor necessitates a simplification of a complex system, which in turn elaborates simple and powerful models of cognitive processes in the human. By testing these models empirically, we can better assess whether the model is correct, at least in its presumption of imitating a human being. Empirical testing of these algorithms can thus better inform their suitability for a given task. In the case of learning and decision-making, one can see how a truly intelligent machine would need to know how to resolve uncertainty in its many forms. If we can create machines capable of integrating and employing, for instance, a surprise signal, or relying on its confidence estimates, we would be endowing said machine with a uniquely intelligent human capability. In other words, capturing the elusive states of confidence and surprise, as well as, potentially, their dynamics, adds to our understanding of how we can quantify the latter and thus refine decision-making processes in a machine.
- Finally, this work provides empirical evidence that we are inference machines. By showing common brain areas correlating with uncertainty encapsulated in two different domains, namely a high-level cognitive domain such as numerical cognition, and a low-level perceptual domain such as visual perception, we provide evidence supporting Hermann von Helmholtz' initial ideas on the brain's purpose in relation to its environment. The empirical findings in our imaging data therefore suggest our continual exchange with the environment (the environment not being limited to what is external to the physical body). That is, we engage in a constant process of approximation. A corollary to this conclusion is that, if our seemingly optimal functioning as a species is based on approximations, on what is good enough to move forward across a lifespan, it is because we are limited in our capacity to perceive and integrate. Thus, our empirical data support an epistemological idea: we cannot know the truth. Aside from the fact that we may not need to know the truth, providing empirical evidence suggesting that we cannot know what is true is a valuable, if sobering insight into our human behavior.

References

- Aarts, E., Helmich, R. C., Janssen, M. J. R., Oyen, W. J. G., Bloem, B. R., & Cools, R. (2012). Aberrant reward processing in Parkinson's disease is associated with dopamine cell loss. *NeuroImage*, 59(4), 3339–3346. <http://doi.org/10.1016/j.neuroimage.2011.11.073>
- Abler, B., Walter, H., Erk, S., Kammerer, H., & Spitzer, M. (2006). Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *NeuroImage*, 31(2), 790–795. <http://doi.org/10.1016/j.neuroimage.2006.01.001>
- Ackermann, H., & Riecker, A. (2004). The contribution of the insula to motor aspects of speech production: a review and a hypothesis. *Brain and language*, 89(2), 320–328.
- Ackermann, H., & Riecker, A. (2010). The contribution (s) of the insula to speech production: a review of the clinical and functional imaging literature. *Brain Structure and Function*, 214(5-6), 419–433.
- Allefeld, C., Görden, K., & Haynes, J. D. (2016). Valid population inference for information-based imaging: From the second-level t-test to prevalence inference. *NeuroImage*, 141, 378–392. <http://doi.org/10.1016/j.neuroimage.2016.07.040>
- Allman, J. M., Tetreault, N. A., Hakeem, A. Y., Manaye, K. F., Semendeferi, K., Erwin, J. M., ... Hof, P. R. (2011). The von Economo neurons in the frontoinsula and anterior cingulate cortex. *Annals of the New York Academy of Sciences*, 1225(1), 59–71. <http://doi.org/10.1111/j.1749-6632.2011.06011.x>
- Amiez, C., Joseph, J. P., & Procyk, E. (2006). Reward encoding in the monkey anterior cingulate cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, 16(7), 1040–55. <http://doi.org/10.1093/cercor/bhj046>
- Andrews, T. J., Schluppeck, D., Homfray, D., Matthews, P., & Blakemore, C. (2002). Activity in the fusiform gyrus predicts conscious perception of Rubin's vase-face illusion. *NeuroImage*, 17(2), 890–901. [http://doi.org/10.1016/S1053-8119\(02\)91243-7](http://doi.org/10.1016/S1053-8119(02)91243-7)
- Angela, J. Y., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46(4), 681–692.
- Ardila, A., Bernal, B., & Rosselli, M. (2014). Participation of the insula in language revisited: a meta-analytic connectivity study. *Journal of Neurolinguistics*, 29, 31–41.
- Astafiev, S. V., Snyder, A. Z., Shulman, G. L., & Corbetta, M. (2010). Comment on “Modafinil Shifts Human in the Suprachiasmatic Area,” 328(April), 1–3. <http://doi.org/10.1126/science.1177200>
- Aston-Jones, G., Rajkowski, J., & Cohen, J. (1999). Role of locus coeruleus in attention and behavioral flexibility. *Biological Psychiatry*, 46(9), 1309–20. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10560036>
- Azzena, G. B., Azzena, M. T., & Marini, R. (1974). Optokinetic nystagmus and the vestibular nuclei. *Experimental neurology*, 42(1), 158–168.

- Bach, D. R., & Dolan, R. J. (2012). Knowing how much you don't know: a neural organization of uncertainty estimates. *Nature Reviews Neuroscience*, 13(8), 572–586. <http://doi.org/10.1038/nrn3289>
- Baker, D. H., Karapanagiotidis, T., Coggan, D. D., Wailes-newson, K., & Smallwood, J. (2015). Brain networks underlying bistable perception Binocular rivalry (b) Necker cube Left eye Right eye Both eyes. *NeuroImage*, 119, 229–234. <http://doi.org/10.1016/j.neuroimage.2015.06.053>
- Baldi, P., & Itti, L. (2010). Of bits and wows : A Bayesian theory of surprise with applications to attention. *Neural Networks*, 23(5), 649–666. <http://doi.org/10.1016/j.neunet.2009.12.007>
- Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The role of the dorsal striatum in reward and decision-making. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 27(31), 8161–5. <http://doi.org/10.1523/JNEUROSCI.1554-07.2007>
- Bamiou, D. E., Musiek, F. E., & Luxon, L. M. (2003). The insula (Island of Reil) and its role in auditory processing: literature review. *Brain research reviews*, 42(2), 143-154.
- Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: a coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, 76, 412–27. <http://doi.org/10.1016/j.neuroimage.2013.02.063>
- Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, 47(1), 129-141.
- Bechara, a, Damasio, H., & Damasio, a R. (2000). Emotion, decision making and the orbitofrontal cortex. *Cerebral Cortex (New York, N.Y. : 1991)*, 10(3), 295–307. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/10731224>
- Bechara, A. (2004). The role of emotion in decision-making: evidence from neurological patients with orbitofrontal damage. *Brain and Cognition*, 55(1), 30–40. <http://doi.org/10.1016/j.bandc.2003.04.001>
- Beck, J. M., Ma, W. J., Kiani, R., Hanks, T., Churchland, A. K., Roitman, J., ... Pouget, A. (2008). Probabilistic Population Codes for Bayesian Decision Making. *Neuron*, 60(6), 1142–1152. <http://doi.org/10.1016/j.neuron.2008.09.021>
- Behrens, T. E. J., Hunt, L. T., Woolrich, M. W., & Rushworth, M. F. S. (2008). Associative learning of social value. *Nature*, 456(7219), 245–249. <http://doi.org/10.1038/nature07538>
- Bencomo, N., & Bencomo, N. (2014). A world full of surprises : Bayesian theory of surprise to quantify degrees of uncertainty A World Full of Surprises : Bayesian Theory of Surprise to Quantify Degrees of Uncertainty, (May). <http://doi.org/10.1145/2591062.2591118>
- Bendahan, S., Goette, L., Thoresen, J., Loued-Khenissi, L., Hollis, F., & Sandi, C. (2017). Acute stress alters individual risk taking in a time-dependent manner and leads to anti-social risk. *European Journal of Neuroscience*, 45(7), 877–885.

<http://doi.org/10.1111/ejn.13395>

- Bense, S., Janusch, B., Vucurevic, G., Bauermann, T., Schlindwein, P., Brandt, T., ... Dieterich, M. (2006). Brainstem and cerebellar fMRI-activation during horizontal and vertical optokinetic stimulation. *Experimental Brain Research*, 174(2), 312–323. <http://doi.org/10.1007/s00221-006-0464-0>
- Berridge, K. C. (2013). Computation of Reward Motivation, 35(7), 1124–1143. <http://doi.org/10.1111/j.1460-9568.2012.07990.x>.From
- Birn, R. M., Roeber, B. J., & Pollak, S. D. (2017). Early childhood stress exposure , reward pathways , and adult decision making, (Mid), 1–6. <http://doi.org/10.1073/pnas.1708791114>
- Bland, A. R., & Schaefer, A. (2012). Different varieties of uncertainty in human decision-making. *Frontiers in Neuroscience*, 6(JUN), 1–11. <http://doi.org/10.3389/fnins.2012.00085>
- Blanke, O., & Mohr, C. (2005). Hallucination of neurological origin Implications for neurocognitive mechanisms of corporeal awareness and self consciousness, 50, 184–199. <http://doi.org/10.1016/j.brainresrev.2005.05.008>
- Blanke, O., Ortigue, S., Landis, T., & Seeck, M. (2002). Neuropsychology: stimulating illusory own-body perceptions. *Nature*, 419(6904), 269.
- Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. *Artificial Intelligence*, 97(1–2), 245–271. [http://doi.org/10.1016/S0004-3702\(97\)00063-5](http://doi.org/10.1016/S0004-3702(97)00063-5)
- Borsellino, A., De Marco, A., Allazetta, A., Rinesi, S., & Bartolini, B. (1972). Reversal time distribution in the perception of visual ambiguous stimuli. *Kybernetik*, 10(3), 139–144.
- Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*. US: American Psychological Association. <http://doi.org/10.1037/0033-295X.108.3.624>
- Bouret, S., & Sara, S. J. (2005). Network reset: a simplified overarching theory of locus coeruleus noradrenaline function. *Trends in Neurosciences*, 28(11), 574–82. <http://doi.org/10.1016/j.tins.2005.09.002>
- Brand, M., Recknor, E. C., Grabenhorst, F., & Bechara, A. (2007). Decisions under ambiguity and decisions under risk: correlations with executive functions and comparisons of two different gambling tasks with implicit and explicit rules. *Journal of Clinical and Experimental Neuropsychology*, 29(1), 86–99. <http://doi.org/10.1080/13803390500507196>
- Brascamp, J. W., Ee, R. Van, Noest, J., Jacobs, R. H. A. H., & Berg, A. V. Van Den. (2017). The time course of binocular rivalry reveals a fundamental role of noise, (2006), 1244–1256.
- Brascamp, J. W., Klink, P. C., & Levitt, W. J. M. (2015). The “laws” of binocular rivalry: 50 years of Levelt’s propositions. *Vision Research*, 109(Part A), 20–37. <http://doi.org/10.1016/j.visres.2015.02.019>

- Brascamp, J. W., van Ee, R., Pestman, W. R., & van den Berg, A. V. (2005). Distributions of alternation rates in various forms of bistable perception. *Journal of Vision*, 5(4), 1. <http://doi.org/10.1167/5.4.1>
- Braun, J., & Mattia, M. (2010). NeuroImage Attractors and noise : Twin drivers of decisions and multistability. *NeuroImage*, 52(3), 740–751. <http://doi.org/10.1016/j.neuroimage.2009.12.126>
- Britz, J., Pitts, M. A., & Michel, C. M. (2011). Right parietal brain activity precedes perceptual alternation during binocular rivalry. *Human brain mapping*, 32(9), 1432–1442.
- Brooks, J. C., Faull, O. K., Pattinson, K. T., & Jenkinson, M. (2013). Physiological noise in brainstem fMRI. *Frontiers in human neuroscience*, 7.
- Brown, J. W. (2005). Learned Predictions of Error Likelihood in the Anterior Cingulate Cortex. *Science*, 307(5712), 1118–1121. <http://doi.org/10.1126/science.1105783>
- Browning, M., Behrens, T. E., Jocham, G., O'Reilly, J. X., & Bishop, S. J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature Neuroscience*, 18(4), 590–596. <http://doi.org/10.1038/nn.3961>
- Bryden, D. W., Johnson, E. E., Tobia, S. C., Kashtelyan, V., & Roesch, M. R. (2011). Attention for learning signals in anterior cingulate cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 31(50), 18266–74. <http://doi.org/10.1523/JNEUROSCI.4715-11.2011>
- Buckthrought, A., Jessula, S., & Mendola, J. D. (2011). Bistable percepts in the brain: fMRI contrasts monocular pattern rivalry and binocular rivalry. *PLoS ONE*, 6(5), 24–26. <http://doi.org/10.1371/journal.pone.0020367>
- Buelow, M. T., & Suhr, J. A. (2009). Construct validity of the Iowa gambling task. *Neuropsychology Review*, 19(1), 102–114. <http://doi.org/10.1007/s11065-009-9083-4>
- Buelow, M. T., Frakey, L. L., Grace, J., & Friedman, J. H. (2017). The Contribution of Apathy and Increased Learning Trials to Risky Decision-Making in Parkinson ' s Disease, 29(August 2013), 100–109. <http://doi.org/10.1093/arclin/act065>
- Buettner, U. W., & Büttner, U. (1979). Vestibular nuclei activity in the alert monkey during suppression of vestibular and optokinetic nystagmus. *Experimental Brain Research*, 37(3), 581–593.
- Burr, D. (2004). Eye movements: Keeping vision stable. *Current Biology*, 14(5), 195–197. <http://doi.org/10.1016/j.cub.2004.02.020>
- Cahill, L., Prins, B., Weber, M., & McGaugh, J. L. (1994). β -Adrenergic activation and memory for emotional events. *Nature*, 371(6499), 702–704. <http://doi.org/10.1038/371702a0>
- Cavanagh, J. F., Zambrano-Vazquez, L., & Allen, J. J. B. (2012). Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology*, 49(2), 220–238. <http://doi.org/10.1111/j.1469-8986.2011.01293.x>

- and the economics of overconfidence. *Journal of Economic Behavior and Organization*, 61(3), 453–470. <http://doi.org/10.1016/j.jebo.2004.10.010>
- Chamba, G., Weissmann, D., Rousset, C., Renaud, B., & Pujol, J. F. (1991). Distribution of alpha-1 and alpha-2 binding sites in the rat locus coeruleus. *Brain research bulletin*, 26(2), 185-193.
- Charpentier, C. J., Aylward, J., Roiser, J. P., & Robinson, O. J. (2017). Enhanced Risk Aversion, But Not Loss Aversion, in Unmedicated Pathological Anxiety. *Biological Psychiatry*, 81(12), 1014–1022. <http://doi.org/10.1016/j.biopsych.2016.12.010>
- Chua, E. F., Schacter, D. L., Rand-Giovannetti, E., & Sperling, R. A. (2006). Understanding metamemory: neural correlates of the cognitive process and subjective level of confidence in recognition memory. *Neuroimage*, 29(4), 1150-1160.
- Clark, L., Bechara, a, Damasio, H., Aitken, M. R. F., Sahakian, B. J., & Robbins, T. W. (2008). Differential effects of insular and ventromedial prefrontal cortex lesions on risky decision-making. *Brain : A Journal of Neurology*, 131(Pt 5), 1311–22. <http://doi.org/10.1093/brain/awn066>
- Cont, R. (2006). Model uncertainty and its impact on the pricing of derivative instruments. *Mathematical Finance*, 16(3), 519–547. <http://doi.org/10.1111/j.1467-9965.2006.00281.x>
- Cools, R., Barker, R. A., Sahakian, B. J., & Robbins, T. W. (2001). Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. *Cerebral cortex*, 11(12), 1136-1143.
- Corlett, P. R., Murray, G. K., Honey, G. D., Aitken, M. R. F., Shanks, D. R., Robbins, T. W., ... Fletcher, P. C. (2007). Disrupted prediction-error signal in psychosis: Evidence for an associative account of delusions. *Brain*, 130(9), 2387–2400.
- Cornelissen, F. W., Peters, E. M., & Palmer, J. (2002). The Eyelink Toolbox: eye tracking with MATLAB and the Psychophysics Toolbox. *Behavior Research Methods, Instruments, & Computers*, 34(4), 613-617.
- Craig, A. D. (2011). Significance of the insula for the evolution of human awareness of feelings from the body. *Annals of the New York Academy of Sciences*, 1225, 72-82.
- Craig, A. D. (Bud). (2009). How do you feel—now? The anterior insula and human awareness. *Nature Reviews Neuroscience*. Craig, A. D. (Bud): Atkinson Research Laboratory, Barrow Neurological Institute, Phoenix, AZ, US, 85013, bcraig@chw.edu: Nature Publishing Group. <http://doi.org/10.1038/nrn2555>
- Craig, A. D., & Craig, A. D. (2009). How do you feel--now? The anterior insula and human awareness. *Nature reviews neuroscience*, 10(1).
- Critchley, H. D., Wiens, S., Rotshtein, P., Öhman, A., & Dolan, R. J. (2004). Neural systems supporting interoceptive awareness. *Nature Neuroscience*, 7(2), 189–195. <http://doi.org/10.1038/nrn1176>
- Crottaz-Herbette, S., & Menon, V. (2006). Where and When the Anterior Cingulate Cortex Modulates Attentional Response: Combined fMRI and ERP Evidence. *Journal of*

Cognitive Neuroscience, 18(5), 766–780. <http://doi.org/10.1162/jocn.2006.18.5.766>

- d'Acremont, M., & Bossaerts, P. (2008). Neurobiological studies of risk assessment: A comparison of expected utility and mean-variance approaches. *Cognitive, Affective, & Behavioral Neuroscience*, 8(4), 363–374. <http://doi.org/10.3758/CABN.8.4.363>
- d'Acremont, M., Lu, Z.-L., Li, X., Van der Linden, M., & Bechara, A. (2009). Neural correlates of risk prediction error during reinforcement learning in humans. *NeuroImage*, 47(4), 1929–1939. <http://doi.org/10.1016/j.neuroimage.2009.04.096>
- Damasio, A. R. (1994). Descartes' error: Emotion, rationality and the human brain. *New York: Putnam*, 352.
- Dave, C., & Wolfe, K. W. (2003). On confirmation bias and deviations from bayesian updating. *Internet access: http://www.peel.pitt.edu/esa2003/papers/wolfe_confirmationbias.pdf*. [Prieiga per Internetą 2011 02 24].
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6), 1204–1215. <http://doi.org/10.1016/j.neuron.2011.02.027>
- Dayan, P. (1998). A hierarchical model of binocular rivalry. *Neural Computation*, 10(5), 1119–1135.
- De Martino, B., Fleming, S. M., Garrett, N., & Dolan, R. J. (2013). Confidence in value-based choice. *Nat Neurosci*, 16(1), 105–110. <http://doi.org/10.1038/nn.3279>
- Delazer, M., Sinz, H., Zamarian, L., Stockner, H., Seppi, K., Wenning, G. K., ... Poewe, W. (2009). Decision making under risk and under ambiguity in Parkinson's disease. *Neuropsychologia*, 47(8–9), 1901–1908. <http://doi.org/10.1016/j.neuropsychologia.2009.02.034>
- Dieter, K. C., Brascamp, J., Tadin, D., & Blake, R. (2016). Does visual attention drive the dynamics of bistable perception ? *Attention, Perception, & Psychophysics*, (May), 1861–1873. <http://doi.org/10.3758/s13414-016-1143-2>
- Ding, L., & Gold, J. I. (2012). Separate, Causal Roles of the Caudate in Saccadic Choice and Execution in a Perceptual Decision Task. *Neuron*, 75(5), 865–874. <http://doi.org/10.1016/j.neuron.2012.07.021>
- Doherty, J. O., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2014). Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning, 452(2004). <http://doi.org/10.1126/science.1094285>
- Doherty, J. P. O., Hampton, A., & Kim, H. (2007). Model-Based fMRI and Its Application to Reward Learning, 53, 35–53. <http://doi.org/10.1196/annals.1390.022>
- Downar, J., Crawley, a P., Mikulis, D. J., & Davis, K. D. (2001). The effect of task relevance on the cortical response to changes in visual and auditory stimuli: an event-related fMRI study. *NeuroImage*, 14(6), 1256–1267. <http://doi.org/10.1006/nimg.2001.0946>
- 146 Doya, K. (2008). Modulators of decision making. *Nature Neuroscience*, 11(4), 410–416.

<http://doi.org/10.1038/nn2077>

Dyde, R. T., Jenkin, Æ. M. R., Jenkin, H. L., Zacher, Æ. J. E., & Harris, L. R. (2009). The effect of altered gravity states on the perception of orientation, 647–660. <http://doi.org/10.1007/s00221-009-1741-5>

Einhäuser, W., Martin, K. A., & König, P. (2004). Are switches in perception of the Necker cube related to eye position?. *European Journal of Neuroscience*, 20(10), 2811–2818.

Einhäuser, W., Stout, J., Koch, C., & Carter, O. (2008). Pupil dilation reflects perceptual selection and predicts subsequent stability in perceptual rivalry. *Proceedings of the National Academy of Sciences*, 105(5), 1704–1709.

Engelmann, J. B., & Tamir, D. (2009). Individual differences in risk preference predict neural responses during financial decision-making. *Brain Research*, 1290, 28–51. <http://doi.org/10.1016/j.brainres.2009.06.078>

Ernst, M., & Paulus, M. P. (2005). Neurobiology of decision making: A selective review from a neurocognitive and clinical perspective. *Biological Psychiatry*, 58(8), 597–604. <http://doi.org/10.1016/j.biopsych.2005.06.004>

Euteneuer, F., Schaefer, F., Stuermer, R., Boucsein, W., Timmermann, L., Barbe, M. T., ... Kalbe, E. (2009). Dissociation of decision-making under ambiguity and decision-making under risk in patients with Parkinson's disease: A neuropsychological and psychophysiological study. *Neuropsychologia*, 47(13), 2882–2890. <http://doi.org/10.1016/j.neuropsychologia.2009.06.014>

Euteneuer, F., Schaefer, F., Stuermer, R., Boucsein, W., Timmermann, L., Barbe, M. T., ... & Kalbe, E. (2009). Dissociation of decision-making under ambiguity and decision-making under risk in patients with Parkinson's disease: a neuropsychological and psychophysiological study. *Neuropsychologia*, 47(13), 2882–2890.

Faraji, M., Preuschoff, K., & Gerstner, W. (2016). Balancing New Against Old Information: The Role of Surprise in Learning, 1–53. http://doi.org/10.1162/neco_a_01025

Feldman, H., & Friston, K. J. (2010). Attention, Uncertainty, and Free-Energy. *Frontiers in Human Neuroscience*, 4(December), 1–23. <http://doi.org/10.3389/fnhum.2010.00215>

Fellows, L. K., & Farah, M. J. (2005). Different underlying impairments in decision-making following ventromedial and dorsolateral frontal lobe damage in humans. *Cerebral Cortex*, 15(1), 58–63. <http://doi.org/10.1093/cercor/bhh108>

Fernandes, P., Regala, J., Correia, F., & Gonçalves-Ferreira, A. J. (2012). The human locus coeruleus 3-D stereotactic anatomy. *Surgical and Radiologic Anatomy*, 34(10), 879–885. <http://doi.org/10.1007/s00276-012-0979-y>

Finger, S., & Wade, N. J. (2002). The neuroscience of Helmholtz and the theories of Johannes Müller. Part 1: Nerve cell structure, vitalism, and the nerve impulse. *Journal of the History of the Neurosciences*, 11(2), 136–55. <http://doi.org/10.1076/jhin.11.2.136.15190>

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete Coding of Reward Dopamine

Neurons. *Science*, 299(March), 1898–1902. <http://doi.org/10.1126/science.1077349>

Fouragnan, E., Queirazza, F., Retzler, C., Mullinger, K. J., & Philiastides, M. G. (2017). Spatiotemporal neural characterization of prediction error valence and surprise during reward learning in humans. *Scientific Reports*, 7(1), 1–18.

Frässle, S., Sommer, J., Jansen, A., Naber, M., & Einhäuser, W. (2014). Binocular rivalry: frontal activity relates to introspection and action but not to perception. *Journal of Neuroscience*, 34(5), 1738-1747.

Friston, K. (1995). Analysis of fMRI Time-Series Revisited. *NeuroImage*.
<http://doi.org/10.1006/nimg.1995.1007>

Friston, K. (2010). The free-energy principle: A unified brain theory? *Nature Reviews Neuroscience*, 11(2), 127–138. <http://doi.org/10.1038/nrn2787>

Friston, K. J., Holmes, A. P., Poline, J. B., Grasby, P. J., Williams, S. C. R., Frackowiak, R. S., & Turner, R. (1995). Analysis of fMRI time-series revisited. *Neuroimage*, 2(1), 45-53.

Friston, K., Adams, R. A., Perrinet, L., & Breakspear, M. (2012). Perceptions as hypotheses: Saccades as experiments. *Frontiers in Psychology*, 3(MAY).
<http://doi.org/10.3389/fpsyg.2012.00151>

Garavan, H. (2010). Insula and drug cravings. *Brain structure and function*, 214(5-6), 593-601.

Naqvi, N. H., & Bechara, A. (2009). The hidden island of addiction: the insula. *Trends in neurosciences*, 32(1), 56-67.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: A review of underlying mechanisms. *Clinical Neurophysiology*, 120(3), 453–463. <http://doi.org/10.1016/j.clinph.2008.11.029>

Garrido, M. I., Teng, C. L. J., Taylor, J. A., Rowe, E. G., & Mattingley, J. B. (2016). Surprise responses in the human brain demonstrate statistical learning under high concurrent cognitive demand. *Npj Science of Learning*, 1(1), 16006.
<http://doi.org/10.1038/npjscilearn.2016.6>

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. *Neuroscience and Biobehavioral Reviews*, 37(7), 1297–1310. <http://doi.org/10.1016/j.neubiorev.2013.03.023>

Gasquoine, P. G. (2014). Contributions of the insula to cognition and emotion. *Neuropsychology review*, 24(2), 77-87.

Gigante, G., Mattia, M., Braun, J., & Del Giudice, P. (2009). Bistable perception modeled as competing stochastic integrations at two levels. *PLoS computational biology*, 5(7), e1000430

Gilovich, T., & Griffl, D. (2002). Introduction – Heuristics and Biases: Then and Now. In *Heuristics and Biases: The Psychology of Intuitive Judgment* (pp. 1–11).
<http://doi.org/10.1017/CBO9780511808098>

148 Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., & Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus

- function. *Cognitive, Affective & Behavioral Neuroscience*, 10(2), 252–69.
<http://doi.org/10.3758/CABN.10.2.252>
- Gläscher, J., Daw, N., Dayan, P., & O’Doherty, J. P. (2010). States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, 66(4), 585–595.
<http://doi.org/10.1016/j.neuron.2010.04.016>
- Glavin, G. B. (1985). Stress and brain noradrenaline: A review. *Neuroscience and Biobehavioral Reviews*, 9(2), 233–243. [http://doi.org/10.1016/0149-7634\(85\)90048-X](http://doi.org/10.1016/0149-7634(85)90048-X)
- Gleichgerrcht, E., Ibáñez, A., Roca, M., Torralva, T., & Manes, F. (2010). Decision-making cognition in neurodegenerative diseases. *Nature Reviews. Neurology*, 6(11), 611–23.
<http://doi.org/10.1038/nrneurol.2010.148>
- Glimcher, P. W. (2004). Neuroeconomics: The Consilience of Brain and Decision. *Science*, 306(5695), 447–452. <http://doi.org/10.1126/science.1102566>
- Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, 108(Supplement 3), 15647–15654.
- Glover, G. H., Li, T. Q., & Ress, D. (2000). Image-based method for retrospective correction of physiological motion effects in fMRI: RETROICOR. *Magnetic resonance in medicine*, 44(1), 162–167.
- Goette, L., Bendahan, S., Thoresen, J., Hollis, F., & Sandi, C. (2014). Stress pulls us apart : Trait anxiety modulates self-confidence under stress. *Psychoneuroendocrinology*, 1–9.
<http://doi.org/10.1016/j.psyneuen.2015.01.019>
- Goñi, J., Aznárez-Sanado, M., Arrondo, G., Fernández-Seara, M., Loayza, F. R., Heukamp, F. H., & Pastor, M. A. (2011). The neural substrate and functional integration of uncertainty in decision making: An information theory approach. *PLoS ONE*, 6(3).
<http://doi.org/10.1371/journal.pone.0017408>
- Gregory, R. L. (1973). Eye and brain : the psychology of seeing. *British Medical Journal*, 4(5893), 682. <http://doi.org/10.1136/bmj.4.5893.682-b>
- Gregory, R. L. (1997). Knowledge in perception and illusion. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 352(1358), 1121–1127.
- Grahn, J. A., Parkinson, J. A., & Owen, A. M. (2009). The role of the basal ganglia in learning and memory: neuropsychological studies. *Behavioural brain research*, 199(1), 53–60.
- Happé, F. G. (1996). Studying weak central coherence at low levels: children with autism do not succumb to visual illusions. A research note. *Journal of Child Psychology and Psychiatry*, 37(7), 873–877.
- Hare, T. A., O’Doherty, J., Camerer, C. F., Schultz, W., & Rangel, A. (2008). Dissociating the Role of the Orbitofrontal Cortex and the Striatum in the Computation of Goal Values and Prediction Errors. *Journal of Neuroscience*, 28(22), 5623–5630.

<http://doi.org/10.1523/JNEUROSCI.1309-08.2008>

Hartley, C. A., & Phelps, E. A. (2012). Anxiety and decision-making. *Biological psychiatry*, 72(2), 113-118.

Harvey, A. K., Pattinson, K. T., Brooks, J. C., Mayhew, S. D., Jenkinson, M., & Wise, R. G. (2008). Brainstem functional magnetic resonance imaging: disentangling signal from physiological noise. *Journal of Magnetic Resonance Imaging*, 28(6), 1337-1344.

Hatfield, G. (2002). Perception as Unconscious Inference BT - Perception and the Physical World: Psychological and Philosophical Issue in Perception. *Perception and the Physical World: Psychological and Philosophical Issue in Perception*. (5), 115-143. <http://doi.org/10.1002/0470013427.ch5>

Haxby, J. V. (2012). Multivariate pattern analysis of fMRI : Parcellating abstract from concrete representations. *Neuroimage*, 62(2), 852-855. <http://doi.org/10.1016/j.neuroimage.2012.03.016>.Multivariate

Hayden, B. Y., Heilbronner, S. R., Pearson, J. M., & Michael, L. (2011). NIH Public Access, 31(11), 4178-4187. <http://doi.org/10.1523/JNEUROSCI.4652-10.2011>.Surprise

Haynes, J. D. (2015). A Primer on Pattern-Based Approaches to fMRI: Principles, Pitfalls, and Perspectives. *Neuron*, 87(2), 257-270. <http://doi.org/10.1016/j.neuron.2015.05.025>

Haynes, J., & Rees, G. (2006). Decoding mental states from brain activity in humans, 7(July), 523-534. <http://doi.org/10.1038/nrn1931>

Hebart, M. N., GÃ¶rgen, K., & Haynes, J.-D. (2015). The Decoding Toolbox (TDT): a versatile software package for multivariate analyses of functional imaging data. *Frontiers in Neuroinformatics*, 8(January), 1-18. <http://doi.org/10.3389/fninf.2014.00088>

Hebart, M. N., Schriever, Y., Donner, T. H., & Haynes, J. D. (2016). The Relationship between Perceptual Decision Variables and Confidence in the Human Brain. *Cerebral Cortex*, 26(1), 118-130. <http://doi.org/10.1093/cercor/bhu181>

Heekeren, H. R., Marrett, S., & Ungerleider, L. G. (2008). The neural systems that mediate human perceptual decision making, 9(june). <http://doi.org/10.1038/nrn2374>

Hesselmann, G., Kell, C. A., Eger, E., & Kleinschmidt, A. (2008). Spontaneous local variations in ongoing neural activity bias perceptual decisions. *Proceedings of the National Academy of Sciences*, 105(31), 10984-10989. <http://doi.org/10.1073/pnas.0712043105>

Hoffmann KP, Distler C, Erickson R, Mader W (1988). Physiological and anatomical identification of the nucleus of the optic tract and dorsal terminal nucleus of the accessory optic tract in monkeys. *Exp Brain Res* 69:635-644

Hohwy, J., Roepstorff, A., & Friston, K. (2008). Predictive coding explains binocular rivalry: An epistemological review. *Cognition*, 108(3), 687-701.

Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, 1(4), 304-9.

<http://doi.org/10.1038/1124>

- Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G. H., & Cohen, J. D. (2004). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience*, 7(5), 497–498. <http://doi.org/10.1038/n1238>
- Hooper, C. J., Luciana, M., Conklin, H. M., & Yarger, R. S. (2004). Adolescents' performance on the Iowa Gambling Task: implications for the development of decision making and ventromedial prefrontal cortex. *Developmental psychology*, 40(6), 1148.
- Howard, I. P. (1996). Alhazen's Neglected Discoveries of Visual Phenomena. *Perception*, 25(10), 1203–1217. <http://doi.org/10.1068/p251203>
- Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science (New York, N.Y.)*, 310(5754), 1680–1683. <http://doi.org/10.1126/science.1115327>
- Huettel, S. a, Stowe, C. J., Gordon, E. M., Warner, B. T., & Platt, M. L. (2006). Neural signatures of economic preferences for risk and ambiguity. *Neuron*, 49(5), 765–75. <http://doi.org/10.1016/j.neuron.2006.01.024>
- Hugrass, L., & Crewther, D. (2012). Willpower and conscious percept: volitional switching in binocular rivalry. *PloS one*, 7(4), e35963.
- Huk, A. C., & Shadlen, M. N. (2005). Neural Activity in Macaque Parietal Cortex Reflects Temporal Integration of Visual Motion Signals during Perceptual Decision Making, 25(45), 10420–10436. <http://doi.org/10.1523/JNEUROSCI.4684-04.2005>
- Ide, J. S., Shenoy, P., Yu, A. J., & Li, C. R. (2013). Bayesian Prediction and Evaluation in the Anterior Cingulate Cortex, 33(5), 2039–2047. <http://doi.org/10.1523/JNEUROSCI.2201-12.2013>
- Iglesias, S., Mathys, C., Brodersen, K. H., Kasper, L., Piccirelli, M., denOuden, H. E. M., & Stephan, K. E. (2013). Hierarchical Prediction Errors in Midbrain and Basal Forebrain during Sensory Learning. *Neuron*, 80(2), 519–530. <http://doi.org/10.1016/j.neuron.2013.09.009>
- Intaite, M., Koivisto, M., & Castelo-Branco, M. (2014). Event-related potential responses to perceptual reversals are modulated by working memory load. *Neuropsychologia*, 56(1), 428–438. <http://doi.org/10.1016/j.neuropsychologia.2014.02.016>
- Intaite, M., Noreika, V., & Šoli, A. (2013). Interaction of bottom-up and top-down processes in the perception of ambiguous figures, 89, 24–31. <http://doi.org/10.1016/j.visres.2013.06.011>
- Itti, L., & Baldi, P. (2005). Bayesian Surprise Attracts Human Attention. *Vision Research*, 49(Nips), 1295–1306. <http://doi.org/10.1016/j.visres.2008.09.007>
- Jepma, M., & Nieuwenhuis, S. (2011). Pupil diameter predicts changes in the exploration-exploitation trade-off: evidence for the adaptive gain theory. *Journal of Cognitive*

- Neuroscience*, 23(7), 1587–96. <http://doi.org/10.1162/jocn.2010.21548>
- Jones, C. L., Minati, L., Harrison, N. A., Ward, J., & Critchley, H. D. (2011). Under pressure: Response urgency modulates striatal and insula activity during decision-making under risk. *PLoS ONE*, 6(6). <http://doi.org/10.1371/journal.pone.0020942>
- Jones, C. L., Ward, J., & Critchley, H. D. (2010). The neuropsychological impact of insular cortex lesions. *Journal of Neurology, Neurosurgery & Psychiatry*, 81(6), 611–618.
- Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between Pupil Diameter and Neuronal Activity in the Locus Coeruleus , Colliculi , and Cingulate Cortex. *Neuron*, 89(1), 221–234. <http://doi.org/10.1016/j.neuron.2015.11.028>
- Joshi, S., Li, Y., Kalwani, R. M., & Gold, J. I. (2016). Relationships between pupil diameter and neuronal activity in the locus coeruleus, colliculi, and cingulate cortex. *Neuron*, 89(1), 221–234.
- Kahneman, D. (2014). Thinking Fast and Slow. *Igarss 2014*, (1), 1–5. <http://doi.org/10.1007/s13398-014-0173-7.2>
- Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, 47(2), 263–292.
- Kanai, R., Bahrami, B., & Rees, G. (2010). Human parietal cortex structure predicts individual differences in perceptual rivalry. *Current biology*, 20(18), 1626–1630.
- Kanai, R., Bahrami, B., & Rees, G. (2010). Report Human Parietal Cortex Structure Predicts Individual Differences in Perceptual Rivalry. *Current Biology*, 20(18), 1626–1630. <http://doi.org/10.1016/j.cub.2010.07.027>
- Kandasamy, N., Hardy, B., Page, L., Schaffner, M., Graggaber, J., Powlson, A. S., ... Coates, J. (2013). Cortisol shifts financial risk preferences. <http://doi.org/10.1073/pnas.1317908111>
- Kawamoto, A. H., & Anderson, J. A. (1985). A neural network model of multistable perception. *Acta Psychologica*, 59(1), 35–65.
- Kehagia, A. a., Barker, R. a., & Robbins, T. W. (2012). Cognitive impairment in Parkinson's disease: the dual syndrome hypothesis. *Neurodegenerative Diseases*, 11(2), 79–92. <http://doi.org/10.1159/000341998>
- Kennerley, S. W., Behrens, T. E. J., & Wallis, J. D. (2011). Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. *Nature Neuroscience*, 14(12), 1581–9. <http://doi.org/10.1038/nn.2961>
- Kennerley, S. W., Walton, M. E., Behrens, T. E. J., Buckley, M. J., & Rushworth, M. F. S. (2006). Optimal decision making and the anterior cingulate cortex. *Nature Neuroscience*, 9(7), 940–947. <http://doi.org/10.1038/nn1724>
- Kepecs, A., Uchida, N., Zariwala, H. a., & Mainen, Z. F. (2008). Neural correlates, computation and behavioural impact of decision confidence. *Nature*, 455(7210), 227–231. <http://doi.org/10.1038/nature07200>

- Keren, N. I., Lozar, C. T., Harris, K. C., Morgan, P. S., & Eckert, M. a. (2009). In vivo mapping of the human locus coeruleus. *NeuroImage*, 47(4), 1261–7.
<http://doi.org/10.1016/j.neuroimage.2009.06.012>
- Kiani, R., & Shadlen, M. N. (2009). Representation of Confidence Associated with a Decision by Neurons in the Parietal Cortex. *Science*, 324(5928), 759–764.
<http://doi.org/10.1126/science.1169405>
- Klein, T. A., Endrass, T., Kathmann, N., Neumann, J., von Cramon, D. Y., & Ullsperger, M. (2007). Neural correlates of error awareness. *Neuroimage*, 34(4), 1774–1781.
- Klein, T. A., Ullsperger, M., & Danielmeier, C. (2013). Error awareness and the insula: links to neurological and psychiatric diseases. *Frontiers in human neuroscience*, 7, 14.
- Kleinschmidt, A., Bichel, C., Zeki, S., & Frackowiak, R. S. J. (1998). Human brain activity during spontaneously reversing perception of ambiguous ® gures, (October).
- Kloosterman, N. A., Meindertsma, T., Loon, A. M. Van, Lamme, V. A. F., Yoram, S., & Donner, T. H. (2015). Pupil size tracks perceptual content and surprise, 41(August 2014), 1068–1078. <http://doi.org/10.1111/ejn.12859>
- Kluger, D. S., & Schubotz, R. I. (2017). Strategic adaptation to non-reward prediction error qualities and irreducible uncertainty in fMRI. *Cortex*, 97, 32–48.
<http://doi.org/10.1016/j.cortex.2017.09.017>
- Knight, F., Of, R., & Classics, E. (1921). Risk, Uncertainty and Profit. *Climate Change 2013 - The Physical Science Basis*, XXXI, 1–30.
<http://doi.org/10.1017/CBO9781107415324.004>
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*.
<http://doi.org/10.1016/j.tins.2004.10.007>
- Kolling, N., Behrens, T. E. J., Wittmann, M. K., & Rushworth, M. F. S. (2016). Multiple signals in anterior cingulate cortex. *Current Opinion in Neurobiology*, 37, 36–43.
<http://doi.org/10.1016/j.conb.2015.12.007>
- Kondo, H. M., & Kashino, M. (2007). Neural mechanisms of auditory awareness underlying verbal transformations. *Human Brain Mapping Journal*, 36(1), 123–130.
<http://doi.org/10.1016/j.neuroimage.2007.02.024>
- Kopp, B., & Lange, F. (2013). Electrophysiological indicators of surprise and entropy in dynamic task-switching environments. *Frontiers in Human Neuroscience*, 7.
<http://doi.org/10.3389/fnhum.2013.00300>
- Korn, C. W., & Bach, D. R. (2016). A solid frame for the window on cognition : Modeling event- related pupil responses, 16(February), 1–16.
- Kornmeier, J., Hein, C. M., & Bach, M. (2009). Multistable perception: when bottom-up and top-down coincide. *Brain and cognition*, 69(1), 138–147.
- Kornmeier, J., Wörner, R., Riedel, A., & van Elst, L. T. (2017). A different view on the Necker cube—Differences in multistable perception dynamics between Asperger and non-

Asperger observers. *PloS one*, 12(12), e0189197.

Krain, A. L., Wilson, A. M., Arbuckle, R., Castellanos, F. X., & Milham, M. P. (2006). Distinct neural mechanisms of risk and ambiguity: a meta-analysis of decision-making. *NeuroImage*, 32(1), 477–84. <http://doi.org/10.1016/j.neuroimage.2006.02.047>

KROLL, Y., LEVY, H., & MARKOWITZ, H. M. (1984). Mean-Variance Versus Direct Utility Maximization. *The Journal of Finance*, 39(1), 47–61. <http://doi.org/10.1111/j.1540-6261.1984.tb03859.x>

Krug, K., Brunskill, E., Scarna, A., Goodwin, G. M., & Parker, A. J. (2008). Perceptual switch rates with ambiguous structure-from-motion figures in bipolar disorder, (May), 1839–1848. <http://doi.org/10.1098/rspb.2008.0043>

Kudlicka, A., Clare, L., & Hindle, J. V. (2011). Executive functions in Parkinson's disease: systematic review and meta-analysis. *Movement Disorders : Official Journal of the Movement Disorder Society*, 26(13), 2305–15. <http://doi.org/10.1002/mds.23868>

Kuhnen, C. M., & Knutson, B. (2005). The neural basis of financial risk taking. *Neuron*, 47(5), 763–770. <http://doi.org/10.1016/j.neuron.2005.08.008>

Lak, A., Nomoto, K., Keramati, M., Sakagami, M., & Kepecs, A. (2017). Midbrain Dopamine Neurons Signal Belief in Choice Accuracy during a Perceptual Decision. *Current Biology*, 27(6), 821–832. <http://doi.org/10.1016/j.cub.2017.02.026>

Lauriola, M., & Levin, I. P. (2001). Relating individual differences in Attitude toward Ambiguity to risky choices. *Journal of Behavioral Decision Making*, 14(2), 107–122. <http://doi.org/10.1002/bdm.368>

Lawson, R. P., Rees, G., & Friston, K. J. (2014). An aberrant precision account of autism. *Frontiers in Human Neuroscience*, 8(May), 1–10. <http://doi.org/10.3389/fnhum.2014.00302>

Lebreton, M., Abitbol, R., Daunizeau, J., & Pessiglione, M. (2015). Automatic integration of confidence in the brain valuation signal. *Nature Neuroscience*, 18(8), 1159–1167. <http://doi.org/10.1038/nn.4064>

Lenggenhager, B., Mouthon, M., & Blanke, O. (2009). Spatial aspects of bodily self-consciousness. *Consciousness and Cognition*, 18(1), 110–117. <http://doi.org/10.1016/j.concog.2008.11.003>

Leopold, D. A., & Logothetis, N. K. (1999). Multistable phenomena : changing views in perception, 6613(July), 254–264.

Lestienne, R. (2001). Spike timing, synchronization and information processing on the sensory side of the central nervous system. *Progress in Neurobiology*, 65(6), 545–591. [http://doi.org/10.1016/S0301-0082\(01\)00019-3](http://doi.org/10.1016/S0301-0082(01)00019-3)

Logothetis, N. K. (2008). What we can do and what we cannot do with fMRI. *Nature*, 453(7197), 869–878. <http://doi.org/10.1038/nature06976>

154 Logothetis, N. K., Auguth, M., Oeltermann, A., Pauls, J., & Trinath, T. (2001). A neurophysiological investigation of the basis of the BOLD signal in fMRI. *Nature*,

- 412(6843), 150–157. Retrieved from <http://www.nature.com/nature/journal/v412/n6843/full/412150a0.html>
- Lopez, C., Halje, P., & Blanke, O. (2008). Body ownership and embodiment: vestibular and multisensory mechanisms. *Neurophysiologie Clinique/Clinical Neurophysiology*, 38(3), 149–161.
- Lorini, E., & Castelfranchi, C. (2006). the Unexpected Aspects of Surprise. *International Journal of Pattern Recognition and Artificial Intelligence*, 20(6), 817–833. <http://doi.org/10.1142/S0218001406004983>
- Loughlin, S. E., Foote, S. L., & Grzanna, R. (1986). Efferent projections of nucleus locus coeruleus: morphologic subpopulations have different efferent targets. *Neuroscience*, 18(2), 307–319.
- Ma, W. J., & Jazayeri, M. (2014). Neural Coding of Uncertainty and Probability. *Annual Review of Neuroscience*, 37, 205–220. <http://doi.org/10.1146/annurev-neuro-071013-014017>
- Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, 14(2), 154–162. <http://doi.org/10.1038/nn.2723>
- Mars, R. B., Debener, S., Gladwin, T. E., Harrison, L. M., Haggard, P., Rothwell, J. C., & Bestmann, S. (2008). Trial-by-Trial Fluctuations in the Event-Related Electroencephalogram Reflect Dynamic Changes in the Degree of Surprise. *Journal of Neuroscience*, 28(47), 12539–12545. <http://doi.org/10.1523/JNEUROSCI.2925-08.2008>
- Mars, R. B., Shea, N. J., Kolling, N., & Rushworth, M. F. S. (2012). Model-based analyses: Promises, pitfalls, and example applications to the study of cognitive control. *Quarterly Journal of Experimental Psychology*, 65(2), 252–267. <http://doi.org/10.1080/17470211003668272>
- Martin, K. A. C., Ko, P., & Einha, W. (2004). Are switches in perception of the Necker cube related to eye position ?, 20(August), 2811–2818. <http://doi.org/10.1111/j.1460-9568.2004.03722.x>
- McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, 38(2), 339–346. [http://doi.org/10.1016/S0896-6273\(03\)00154-5](http://doi.org/10.1016/S0896-6273(03)00154-5)
- McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron*, 38(2), 339–346.
- Medford, N., & Critchley, H. D. (2010). Conjoint activity of anterior insular and anterior cingulate cortex: awareness and response. *Brain Structure and Function*, 214(5–6), 535–549.
- Megumi, F., Bahrami, B., Kanai, R., & Rees, G. (2015). Brain activity dynamics in human parietal regions during spontaneous switches in bistable perception. *NeuroImage*, 107, 190–197.

- Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function*, 214(5–6), 655–667. <http://doi.org/10.1007/s00429-010-0262-0>
- Meredith, G. M. (1967). Some attributive dimensions of reversibility phenomena and their relationship to rigidity and anxiety. *Perceptual and Motor Skills*, 24(3, PT. 1), 843–849.
- Meyniel, F., Sigman, M., & Mainen, Z. F. (2015). Perspective Confidence as Bayesian Probability : From Neural Origins to Behavior. *Neuron*, 88(1), 78–92. <http://doi.org/10.1016/j.neuron.2015.09.039>
- Michelon, P., Snyder, A. Z., Buckner, R. L., McAvoy, M., & Zacks, J. M. (2003). Neural correlates of incongruous visual information: An event-related fMRI study. *NeuroImage*, 19(4), 1612–1626. [http://doi.org/10.1016/S1053-8119\(03\)00111-3](http://doi.org/10.1016/S1053-8119(03)00111-3)
- Miltner, W. H. R., Braun, C. H., & Coles, M. G. H. (1997). Event-Related Brain Potentials Following Incorrect Feedback in a Time-Estimation Task: Evidence for a “Generic” Neural System for Error Detection. *Journal of Cognitive Neuroscience*, 9(6), 788–798. <http://doi.org/10.1162/jocn.1997.9.6.788>
- Minzenberg, M. J., & Carter, C. S. (2008). Modafinil: A review of neurochemical actions and effects on cognition. *Neuropsychopharmacology*, 33(7), 1477–1502. <http://doi.org/10.1038/sj.npp.1301534>
- Minzenberg, M. J., Watrous, A. J., Yoon, J. H., Ursu, S., & Carter, C. S. (2008). Modafinil Shifts Human Locus Coeruleus to Low-Tonic, High-Phasic Activity During Functional MRI. *Science*, 322(5908), 1700–1702. <http://doi.org/10.1126/science.1164908>
- Monsell, S. (2003). Task switching. *Trends in Cognitive Sciences*. [http://doi.org/10.1016/S1364-6613\(03\)00028-7](http://doi.org/10.1016/S1364-6613(03)00028-7)
- Monti, M. (2011). Statistical Analysis of fMRI Time-Series: A Critical Review of the GLM Approach. *Frontiers in Human Neuroscience*, 5(March), 1–13. <http://doi.org/10.3389/fnhum.2011.00028>
- Moradi, F., & Heeger, D. J. (2009). Inter-ocular contrast normalization in human visual cortex. *Journal of Vision*, 9(3)(3109), 1–223. <http://doi.org/10.1167/9.3.13>
- Moreno-Bote, R., Rinzel, J., & Rubin, N. (2007). Noise-Induced Alternations in an Attractor Network Model of Perceptual Bistability. *Journal of Neurophysiology*, 98(3), 1125–1139. <http://doi.org/10.1152/jn.00116.2007>
- Muhammed, K., Manohar, S., Ben Yehuda, M., Chong, T. T. J., Tofaris, G., Lennox, G., ... Husain, M. (2016). Reward sensitivity deficits modulated by dopamine are associated with apathy in Parkinson’s disease. *Brain*, 139(10), 2706–2721. <http://doi.org/10.1093/brain/aww188>
- Murphy, P. R., O’Connell, R. G., O’Sullivan, M., Robertson, I. H., & Balsters, J. H. (2014). Pupil diameter covaries with BOLD activity in human locus coeruleus. *Human Brain*

- Mapping*, 35(8), 4140–4154. <http://doi.org/10.1002/hbm.22466>
- Murphy, P. R., Robertson, I. H., Balsters, J. H., & O’connell, R. G. (2011). Pupillometry and P3 index the locus coeruleus-noradrenergic arousal function in humans. *Psychophysiology*, 48(11), 1532–43. <http://doi.org/10.1111/j.1469-8986.2011.01226.x>
- Mustari MJ, Fuchs AF, Kaneko CRS, Robinson RF (1994) Anatomical connections of the primate pretectal nucleus of the optic tract. *J Comp Neurol* 349:111–128
- Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, 69(6), 1204-1215.
- Nagai, T., Satoh, K., Imamoto, K., & Maeda, T. (1981). Divergent projections of catecholamine neurons of the locus coeruleus as revealed by fluorescent retrograde double labeling technique. *Neuroscience letters*, 23(2), 117-123.
- Naqvi, N. H., & Bechara, A. (2010). The insula and drug addiction: an interoceptive view of pleasure, urges, and decision-making. *Brain Structure and Function*, 214(5-6), 435-450.
- Naqvi, N. H., Rudrauf, D., Damasio, H., & Bechara, A. (2007). Damage to the insula disrupts addiction to cigarette smoking. *Science*, 315(5811), 531-534.
- Nassar, M. R., Rumsey, K. M., Wilson, R. C., Parikh, K., Heasly, B., & Gold, J. I. (2012). Rational regulation of learning dynamics by pupil-linked arousal systems. *Nature neuroscience*, 15(7), 1040.
- Nassar, M. R., Wilson, R. C., Heasly, B., & Gold, J. I. (2010). An Approximately Bayesian Delta-Rule Model Explains the Dynamics of Belief Updating in a Changing Environment, 30(37), 12366–12378. <http://doi.org/10.1523/JNEUROSCI.0822-10.2010>
- Necker, LA (1832) Observations on some remarkable Optical Phænomena seen in Switzerland; and on an Optical Phenomenon which occurs on viewing a Figure of a Crystal or geometric Solid. *The London and Edinburgh Philosophical Magazine and Journal of Science (3rd Series)* 1, No 5, 329-337.
- Neuroscience, H., Kornmeier, J., & Bach, M. (2012). Ambiguous figures – what happens in the brain when perception changes but not the stimulus, 6(March), 1–23. <http://doi.org/10.3389/fnhum.2012.00051>
- Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, 2(2), 175–220. <http://doi.org/10.1037/1089-2680.2.2.175>
- Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus--norepinephrine system. *Psychological bulletin*, 131(4), 510.
- Nieuwenhuys, R. (2012). *The insular cortex. A review. Progress in Brain Research* (1st ed., Vol. 195). Elsevier B.V. <http://doi.org/10.1016/B978-0-444-53860-4.00007-6>
- Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, 191(3), 507–520. <http://doi.org/10.1007/s00213-006-0502-4>

- Niv, Y., Edlund, J. A., Dayan, P., & O'Doherty, J. P. (2012). Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. *Journal of Neuroscience*, 32(2), 551–562. <http://doi.org/10.1523/JNEUROSCI.5498-10.2012>
- Noest, A. J., van Ee, R., Nijs, M. M., & van Wezel, R. J. A. (2007). Percept-choice sequences driven by interrupted ambiguous stimuli: A low-level neural model. *Journal of Vision*, 7(8), 10. <http://doi.org/10.1167/7.8.10>
- O'Neill, M., & Schultz, W. (2013). Risk Prediction Error Coding in Orbitofrontal Neurons. *Journal of Neuroscience*, 33(40), 15810–15814. <http://doi.org/10.1523/JNEUROSCI.4236-12.2013>
- O'Reilly, J. X., Schuffelgen, U., Cuell, S. F., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proceedings of the National Academy of Sciences*, 110(38), E3660–E3669. <http://doi.org/10.1073/pnas.1305373110>
- Okuda, J., Fujii, T., Ohtake, H., Tsukiura, T., Tanji, K., Suzuki, K., ... & Yamadori, A. (2003). Thinking of the future and past: The roles of the frontal pole and the medial temporal lobes. *Neuroimage*, 19(4), 1369-1380.
- Oman, C. M., Howard, I. P., Smith, T., Beall, A. C., Natapoff, A., Zacher, J. E., ... Jenkin, H. L. (n.d.). The Role of Visual Cues in Microgravity Spatial Orientation Authors, 69–81.
- Padoa-Schioppa, C., & Assad, J. A. (2006). Neurons in the orbitofrontal cortex encode economic value. *Nature*, 441(7090), 223-226.
- Park, J. W., Bhimani, R. V., & Park, J. (2017). Noradrenergic modulation of dopamine transmission evoked by electrical stimulation of the locus coeruleus in the rat brain. *ACS chemical neuroscience*, 8(9), 1913-1924.
- Parkkonen, L., Andersson, J., Hämäläinen, M., & Hari, R. (2008). Early visual brain areas reflect the percept of an ambiguous scene. *Proceedings of the National Academy of Sciences*, 105(51), 20500-20504.
- Parr, T., & Friston, K. J. (2017). The active construction of the visual world. *Neuropsychologia*, 104(July), 92–101. <http://doi.org/10.1016/j.neuropsychologia.2017.08.003>
- Parris, B. A., Kuhn, G., Mizon, G. A., Benattayallah, A., & Hodgson, T. L. (2009). Imaging the impossible: An fMRI study of impossible causal relationships in magic tricks. *NeuroImage*, 45(3), 1033–1039. <http://doi.org/10.1016/j.neuroimage.2008.12.036>
- Pastukhov, A., & Braun, J. (2008). A short-term memory of multi-stable perception. *Journal of Vision*, 8(13), 7.
- Payzan-LeNestour, E., & Bossaerts, P. (2012). Do not bet on the unknown versus try to find out more: estimation uncertainty and “unexpected uncertainty” both modulate exploration. *Frontiers in neuroscience*, 6, 150.
- Payzan-lenestour, E., Dunne, S., Bossaerts, P., Doherty, J. P. O., & O'Doherty, J. P. (2013). The neural representation of unexpected uncertainty during value-based decision making. *Neuron*, 79(1), 191–201. <http://doi.org/10.1016/j.neuron.2013.04.037>

- Pearson, J., & Blake, R. (2007). The effects of transcranial magnetic stimulation on visual rivalry. *Journal of Vision*, 7(2007), 1–11. <http://doi.org/10.1167/7.7.2.Introduction>
- Pearson, J., Tadin, D., & Blake, R. (2007). The effects of transcranial magnetic stimulation on visual rivalry. *Journal of vision*, 7(7), 2-2.
- Pelillo, M. (2014). Alhazen and the nearest neighbor rule. *Pattern Recognition Letters*, 38(1), 34–37. <http://doi.org/10.1016/j.patrec.2013.10.022>
- Pellicano, E., & Burr, D. (2012). When the world becomes “too real”: a Bayesian explanation of autistic perception. *Trends in Cognitive Sciences*, 16(10), 504–10. <http://doi.org/10.1016/j.tics.2012.08.009>
- Pertl, M.-T., Benke, T., Zamarian, L., & Delazer, M. (2015). Decision Making and Ratio Processing in Patients with Mild Cognitive Impairment. *Journal of Alzheimer's Disease : JAD*, 48(3), 765–779. <http://doi.org/10.3233/JAD-150291>
- Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, 442(7106), 1042–1045. <http://doi.org/10.1038/nature05051>
- Petersen, R. C., Caracciolo, B., Brayne, C., Gauthier, S., Jelic, V., & Fratiglioni, L. (2014). Mild cognitive impairment: A concept in evolution. *Journal of Internal Medicine*, 275(3), 214–228. <http://doi.org/10.1111/joim.12190>
- Peyron, R., Laurent, B., & Garcia-Larrea, L. (2000). Functional imaging of brain responses to pain. A review and meta-analysis (2000). *Neurophysiologie Clinique/Clinical Neurophysiology*, 30(5), 263-288.
- Philastides, M. G., Biele, G., Vavatzanidis, N., Kazzer, P., & Heekeren, H. R. (2010). NeuroImage Temporal dynamics of prediction error processing during reward-based decision making. *NeuroImage*. <http://doi.org/10.1016/j.neuroimage.2010.05.052>
- Picard, F., & Friston, K. (2014). Predictions, perception, and a sense of self. *Neurology*, 1–7. <http://doi.org/10.1212/WNL.0000000000000798>
- Platt, M. L., & Huettel, S. a. (2008). Risky business: the neuroeconomics of decision making under uncertainty. *Nature Neuroscience*, 11(4), 398–403. <http://doi.org/10.1038/nn2062>
- Pochon, J.-B., Riis, J., Sanfey, A. G., Nystrom, L. E., & Cohen, J. D. (2008). Functional Imaging of Decision Conflict. *Journal of Neuroscience*, 28(13), 3468–3473. <http://doi.org/10.1523/JNEUROSCI.4195-07.2008>
- Poletti, M., Frosini, D., Lucetti, C., Del Dotto, P., Ceravolo, R., & Bonuccelli, U. (2010). Decision making in de novo Parkinson's disease. *Movement Disorders : Official Journal of the Movement Disorder Society*, 25(10), 1432–6. <http://doi.org/10.1002/mds.23098>
- Poline, J. B., & Brett, M. (2012). The general linear model and fMRI: Does love last forever? *NeuroImage*, 62(2), 871–880. <http://doi.org/10.1016/j.neuroimage.2012.01.133>
- Polonsky, A., Blake, R., Braun, J., & Heeger, D. J. (2000). Neuronal activity in human primary visual cortex correlates with perception during binocular rivalry. *Nature neuroscience*, 3(11), 1153.

- Pouget, A., Drugowitsch, J., & Kepecs, A. (2016). Confidence and certainty: Distinct probabilistic quantities for different goals. *Nature Neuroscience*.
<http://doi.org/10.1038/nn.4240>
- Preuschoff, K., Bossaerts, P., & Quartz, S. R. (2006). Neural differentiation of expected reward and risk in human subcortical structures. *Neuron*, 51(3), 381–90.
<http://doi.org/10.1016/j.neuron.2006.06.024>
- Preuschoff, K., Hart, B. M., & Einhauser, W. (2011). Pupil dilation signals surprise: Evidence for noradrenaline's role in decision making. *Frontiers in neuroscience*, 5, 115.
- Preuschoff, K., Quartz, S. R., & Bossaerts, P. (2008). Human insula activation reflects risk prediction errors as well as risk. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 28(11), 2745–52.
<http://doi.org/10.1523/JNEUROSCI.4286-07.2008>
- Proedrou, K., Nouretdinov, I., Vovk, V., & Gammerman, A. (2002). Transductive confidence machines for pattern recognition. *Machine Learning: Ecml 2002*, 2430(January), 381–390. http://doi.org/10.1007/3-540-36755-1_32
- Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1994). Locus coeruleus activity in monkey: Phasic and tonic changes are associated with altered vigilance. *Brain Research Bulletin*, 35(5–6), 607–616. [http://doi.org/10.1016/0361-9230\(94\)90175-9](http://doi.org/10.1016/0361-9230(94)90175-9)
- Rajkowski, J., Kubiak, P., & Aston-Jones, G. (1994). Locus coeruleus activity in monkey: phasic and tonic changes are associated with altered vigilance. *Brain research bulletin*, 35(5–6), 607–616.
- Ramachandran, S., Meyer, T., & Olson, C. R. (2017). Prediction suppression and surprise enhancement in monkey inferotemporal cortex. *Journal of Neurophysiology*, 118(1), 374–382. <http://doi.org/10.1152/jn.00136.2017>
- Rao, R. P. N., & Ballard, D. H. (1999). Predictive coding in the visual cortex : a functional interpretation of some extra-classical receptive-field effects, 79–87.
- Rees, G. (2007). Neural correlates of the contents of visual awareness in humans. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1481), 877–886. <http://doi.org/10.1098/rstb.2007.2094>
- Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. *Science*, 306(5695), 443–7.
<http://doi.org/10.1126/science.1100301>
- Rimmele, U., Lackovic, S. F., Tobe, R. H., Leventhal, B. L., & Phelps, E. A. (2016). Beta-adrenergic blockade at memory encoding, but not retrieval, decreases the subjective sense of recollection. *Journal of cognitive neuroscience*, 28(6), 895–907.
- Robbins, T. W. (1984). Cortical noradrenaline, attention and arousal. *Psychological medicine*, 14(1), 13–21.
- Rolls, E. T., Grabenhorst, F., & Deco, G. (2010). Decision-Making, Errors, and Confidence in the Brain. *Journal of Neurophysiology*, 104(5), 2359–2374.

<http://doi.org/10.1152/jn.00571.2010>

- Rommelfanger, K. S., & Weinshenker, D. (2007). Norepinephrine: The redheaded stepchild of Parkinson's disease. *Biochemical Pharmacology*, 74(2), 177–90.
<http://doi.org/10.1016/j.bcp.2007.01.036>
- Roy, M., Shohamy, D., Daw, N., Jepma, M., Wimmer, G. E., & Wager, T. D. (2014). Representation of aversive prediction errors in the human periaqueductal gray. *Nature neuroscience*, 17(11), 1607.
- Ryterska, A., Jahanshahi, M., & Osman, M. (2013). What are people with Parkinson's disease really impaired on when it comes to making decisions? A meta-analysis of the evidence. *Neuroscience and Biobehavioral Reviews*, 37(10 Pt 2), 2836–46.
<http://doi.org/10.1016/j.neubiorev.2013.10.005>
- Sadacca, B. F., Wikenheiser, A. M., & Schoenbaum, G. (2017). Toward a theoretical role for tonic norepinephrine in the orbitofrontal cortex in facilitating flexible learning. *Neuroscience*, 345, 124–129. <http://doi.org/10.1016/j.neuroscience.2016.04.017>
- Samanez-Larkin, G. R., Kuhnen, C. M., Yoo, D. J., & Knutson, B. (2010). Variability in Nucleus Accumbens Activity Mediates Age-Related Suboptimal Financial Risk Taking. *Journal of Neuroscience*, 30(4), 1426–1434. <http://doi.org/10.1523/JNEUROSCI.4902-09.2010>
- Sanders, J. I., Hangya, B., & Kepecs, A. (2016). Signatures of a Statistical Computation in the Human Sense of Confidence. *Neuron*, 90(3), 499–506.
<http://doi.org/10.1016/j.neuron.2016.03.025>
- Sanders, J. I., Kepecs, A., & Harbor, C. S. (2017). confidence, 90(3), 499–506.
<http://doi.org/10.1016/j.neuron.2016.03.025.Signatures>
- Sara, S. J. (2009). The locus coeruleus and noradrenergic modulation of cognition. *Nature Reviews. Neuroscience*, 10(3), 211–223. <http://doi.org/10.1038/nrn2573>
- Sara, S. J. (2015). Locus Coeruleus in time with the making of memories. *Current Opinion in Neurobiology*, 35, 87–94. <http://doi.org/10.1016/j.conb.2015.07.004>
- Sara, S. J., & Bouret, S. (2012). Orienting and Reorienting: The Locus Coeruleus Mediates Cognition through Arousal. *Neuron*, 76(1), 130–141.
<http://doi.org/10.1016/j.neuron.2012.09.011>
- Sara, S. J., & Herve-Minvielle, A. (1995). Inhibitory influence of frontal cortex on locus coeruleus neurons. *Proceedings of the National Academy of Sciences*, 92(13), 6032–6036. <http://doi.org/10.1073/pnas.92.13.6032>
- Sarinopoulos, I., Grupe, D. W., Mackiewicz, K. L., Herrington, J. D., Lor, M., Steege, E. E., & Nitschke, J. B. (2010). Uncertainty during anticipation modulates neural responses to aversion in human insula and amygdala. *Cerebral Cortex*, 20(4), 929–940.
<http://doi.org/10.1093/cercor/bhp155>
- Sasaki, M., Shibata, E., Tohyama, K., Takahashi, J., Otsuka, K., Tsuchiya, K., ... Sakai, A. (2006). Neuromelanin magnetic resonance imaging of locus ceruleus and substantia nigra in Parkinson's disease. *NeuroReport*, 17(11), 1215–1218.

<http://doi.org/10.1097/01.wnr.0000227984.84927.a7>

- Schiffer, A. M., Ahlheim, C., Wurm, M. F., & Schubotz, R. I. (2012). Surprised at all the entropy: Hippocampal, caudate and midbrain contributions to learning from prediction errors. *PLoS ONE*, 7(5). <http://doi.org/10.1371/journal.pone.0036445>
- Schmack, K., Schnack, A., Priller, J., & Sterzer, P. (2015). Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. *Schizophrenia Research: Cognition*, 2(2), 72–77. <http://doi.org/10.1016/j.scog.2015.03.005>
- Schmack, K., Schnack, A., Priller, J., & Sterzer, P. (2015). Perceptual instability in schizophrenia: Probing predictive coding accounts of delusions with ambiguous stimuli. *Schizophrenia Research: Cognition*, 2(2), 72–77.
- Schonberg, T., O'Doherty, J. P., Joel, D., Inzelberg, R., Segev, Y., & Daw, N. D. (2010). Selective impairment of prediction error signaling in human dorsolateral but not ventral striatum in Parkinson's disease patients: evidence from a model-based fMRI study. *NeuroImage*, 49(1), 772–81. <http://doi.org/10.1016/j.neuroimage.2009.08.011>
- Schrater, P. R. (2017). Perceptual multistability predicted by search model for Bayesian decisions Rashmi Sundareswara, 8(2008), 1–19. <http://doi.org/10.1167/8.5.12.Introduction>
- Schultz, W. (1997). Dopamine neurons and their role in reward mechanisms. *Current Opinion in Neurobiology*. [http://doi.org/10.1016/S0959-4388\(97\)80007-4](http://doi.org/10.1016/S0959-4388(97)80007-4)
- Schultz, W. (2000). Multiple reward signals in the brain. *Nature reviews neuroscience*, 1(3), 199.
- Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, 275(5306), 1593–1599. <http://doi.org/10.1126/science.275.5306.1593>
- Schwartenbeck, P., Fitzgerald, T. H. B., Mathys, C., Dolan, R., Kronbichler, M., & Friston, K. (2015). Evidence for surprise minimization over value maximization in choice behavior. *Scientific Reports*, 5, 1–14. <http://doi.org/10.1038/srep16575>
- Scocchia, L., Valsecchi, M., & Triesch, J. (2014). Top-down influences on ambiguous perception: the role of stable and transient states of the observer. *Frontiers in Human Neuroscience*, 8(December), 1–18. <http://doi.org/10.3389/fnhum.2014.00979>
- Seligman, M. E. (1972). Learned helplessness. *Annual Review of Medicine*, 207–412. <http://doi.org/10.1146/annurev.me.23.020172.002203>
- Shackman, A. J., Salomons, T. V., Slagter, H. A., Andrew, S., Winter, J. J., & Davidson, R. J. (2011). NIH Public Access. *Cortex*, 12(3), 154–167. <http://doi.org/10.1038/nrn2994>
- Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal*, 27(July 1928), 379–423. <http://doi.org/10.1145/584091.584093>
- Sharot, T., Velasquez, C. M., & Dolan, R. J. (2010). Do Decisions Shape Preference? *Psychological Science*, 21(9), 1231–1235. <http://doi.org/10.1177/0956797610379235>

- Shenhav, A., Botvinick, M. M., & Cohen, J. D. (2013). The expected value of control: An integrative theory of anterior cingulate cortex function. *Neuron*, 79(2), 217–240. <http://doi.org/10.1016/j.neuron.2013.07.007>
- Shohamy, D. (2011). Learning and motivation in the human striatum. *Current opinion in neurobiology*, 21(3), 408-414.
- Shpiro, A., Moreno-Bote, R., Rubin, N., & Rinzel, J. (2009). Balance between noise and adaptation in competition models of perceptual bistability. *Journal of computational neuroscience*, 27(1), 37.
- Shuren, J. (1993). Insula and aphasia. *Journal of neurology*, 240(4), 216-218.
- Silvetti, M., Seurinck, R., & Verguts, T. (2013). Value and prediction error estimation account for volatility effects in ACC: A model-based fMRI study. *Cortex*, 49(6), 1627–1635. <http://doi.org/10.1016/j.cortex.2012.05.008>
- Singer, T., Critchley, H. D., & Preuschoff, K. (2009). A common role of insula in feelings, empathy and uncertainty. *Trends in Cognitive Sciences*, 13(8), 334–40. <http://doi.org/10.1016/j.tics.2009.05.001>
- Sinz, H., Zamarian, L., Benke, T., Wenning, G. K., & Delazer, M. (2008). Impact of ambiguity and risk on decision making in mild Alzheimer's disease. *Neuropsychologia*, 46(7), 2043–2055. <http://doi.org/10.1016/j.neuropsychologia.2008.02.002>
- Small, D. M. (2010). Taste representation in the human insula. *Brain Structure and Function*, 214(5-6), 551-561.
- Smith, A. (2000). *Adam Smith 1759. Villa Borsig Workshop Series 2000*. <http://doi.org/10.5040/9781501301308>
- Sommer, J., Jansen, A., Naber, M., & Einha, W. (2014). Binocular Rivalry : Frontal Activity Relates to Introspection, 34(5), 1738–1747. <http://doi.org/10.1523/JNEUROSCI.4403-13.2014>
- Sterzer, P., & Kleinschmidt, A. (2007). A neural basis for inference in perceptual ambiguity, 104(1).
- Sterzer, P., & Kleinschmidt, A. (2007). A neural basis for inference in perceptual ambiguity. *Proceedings of the National Academy of Sciences*, 104(1), 323-328.
- Sterzer, P., & Kleinschmidt, A. (2010). Anterior insula activations in perceptual paradigms: often observed but barely understood. *Brain Structure and Function*, 214(5-6), 611-622.
- Sterzer, P., & Rees, G. (2008). A neural basis for percept stabilization in binocular rivalry. *Journal of cognitive neuroscience*, 20(3), 389-399.
- Sterzer, P., Kleinschmidt, A., & Rees, G. (2009). The neural bases of multistable perception, (310). <http://doi.org/10.1016/j.tics.2009.04.006>
- Strange, B. A., Duggins, A., Penny, W., Dolan, R. J., & Friston, K. J. (2005). Information theory, novelty and hippocampal responses: Unpredicted or unpredictable? *Neural*

- Networks*, 18(3), 225–230. <http://doi.org/10.1016/j.neunet.2004.12.004>
- Strüber, D., & Herrmann, C. S. (2002). MEG alpha activity decrease reflects destabilization of multistable percepts. *Cognitive Brain Research*, 14(3), 370–382. [http://doi.org/10.1016/S0926-6410\(02\)00139-8](http://doi.org/10.1016/S0926-6410(02)00139-8)
- Strüber, D., & Stadler, M. (1999). Differences in top - Down influences on the reversal rate of different categories of reversible figures. *Perception*, 28(10), 1185–1196. <http://doi.org/10.1068/p2973>
- Summerfield, C., Trittschuh, E. H., Monti, J. M., Mesulam, M. M., & Egner, T. (2008). Neural repetition suppression reflects fulfilled perceptual expectations. *Nature Neuroscience*, 11(9), 1004–1006. <http://doi.org/10.1038/nn.2163>
- Sundaeswara, R., & Schrater, P. R. (2008). Perceptual multistability predicted by search model for Bayesian decisions. *Journal of vision*, 8(5), 12-12.
- Sundaeswara, R., Schrater, P. R., & Sundaeswara, R. (2008). Perceptual multistability predicted by search model for Bayesian decisions Rashmi Sundaeswara. *Journal of Vision*, 8(5), 1–19. <http://doi.org/10.1167/8.5.12.Introduction>
- Sutton, R. S., & Barto, A. G. (1998). Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*, 9(5), 1054–1054. <http://doi.org/10.1109/TNN.1998.712192>
- Sylvia, T., Guy, K., Sarah, M., Julian, S., Neil, L., Miles, H., & Cowen, P. J. (2013). Beta adrenergic blockade reduces utilitarian judgement. *Biological Psychology*, 92(2), 323–328. <http://doi.org/10.1016/j.biopsycho.2012.09.005>
- Taleb, N. N. (2001). Fooled by Randomness: The Hidden Role of Chance in Life and in the Markets. *The Hidden Role of Chance in Life and in the Markets*. <http://doi.org/10.1080/14459795.2010.528786>
- Timmermann, C., Spriggs, M. J., Kaelen, M., Leech, R., Nutt, D. J., Moran, R. J., ... & Muthukumaraswamy, S. D. (2017). LSD modulates effective connectivity and neural adaptation mechanisms in an auditory oddball paradigm. *Neuropharmacology*.
- Tobler, P. N., O'Doherty, J. P., Dolan, R. J., & Schultz, W. (2006). Reward Value Coding Distinct From Risk Attitude-Related Uncertainty Coding in Human Reward Systems. *Journal of Neurophysiology*, 97(2), 1621–1632. <http://doi.org/10.1152/jn.00745.2006>
- Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron*, 21(4), 753–759. [http://doi.org/10.1016/S0896-6273\(00\)80592-9](http://doi.org/10.1016/S0896-6273(00)80592-9)
- Troje, N. F., & McAdam, M. (2010). The viewing-from-above bias and the silhouette illusion. *i-Perception*, 1(3), 143-148.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., ... & Joliot, M. (2002). Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage*, 15(1), 273-289.

- Uddin, L. Q. (2015). Salience processing and insular cortical function and dysfunction. *Nature Reviews Neuroscience*, 16(1), 55.
- Vaghi, M. M., Vértés, P. E., Kitzbichler, M. G., Apergis-Schoute, A. M., van der Flier, F. E., Fineberg, N. A., ... Robbins, T. W. (2017). Specific Frontostriatal Circuits for Impaired Cognitive Flexibility and Goal-Directed Planning in Obsessive-Compulsive Disorder: Evidence From Resting-State Functional Connectivity. *Biological Psychiatry*, 81(8), 708–717. <http://doi.org/10.1016/j.biopsych.2016.08.009>
- Vaillancourt, D. E., Schonfeld, D., Kwak, Y., Bohnen, N. I., & Seidler, R. (2013). Dopamine overdose hypothesis: evidence and clinical implications. *Movement Disorders*, 28(14), 1920-1929.
- Van Den Hout, M., & Kindt, M. (2003). Phenomenological validity of an OCD-memory model and the remember/know distinction. *Behaviour Research and Therapy*, 41(3), 369–378. [http://doi.org/10.1016/S0005-7967\(02\)00097-9](http://doi.org/10.1016/S0005-7967(02)00097-9)
- Van Eimeren, T., Ballanger, B., Pellecchia, G., Miyasaki, J. M., Lang, A. E., & Strafella, A. P. (2009). Dopamine agonists diminish value sensitivity of the orbitofrontal cortex: a trigger for pathological gambling in Parkinson's disease?. *Neuropsychopharmacology*, 34(13), 2758.
- Venkatraman, V., & Huettel, S. A. (2012). Strategic control in decision-making under uncertainty. *European Journal of Neuroscience*, 35(7), 1075–1082. <http://doi.org/10.1111/j.1460-9568.2012.08009.x>
- Vilares, I., & Kording, K. (2011). Bayesian models: The structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, 1224(1), 22–39. <http://doi.org/10.1111/j.1749-6632.2011.05965.x>
- Volz, K. G., Schubotz, R. I., & Cramon, D. Y. Von. (2003). Predicting events of varying probability : uncertainty investigated by fMRI, 19, 271–280. [http://doi.org/10.1016/S1053-8119\(03\)00122-8](http://doi.org/10.1016/S1053-8119(03)00122-8)
- Von Helmholtz, H. (1925). Physiological Optics. *Uspekhi Fizicheskikh Nauk*, III(10), 1193–1213. <http://doi.org/10.1007/978-3-540-39053-4>
- Von Neumann, J., & Morgenstern, O. (1944). Theory of Games and Economic Behavior. *Princeton University Press*, 625. <http://doi.org/10.1177/1468795X06065810>
- Voon, V., Hassan, K., Zurowski, M., Duff-Canning, S., De Souza, M., Fox, S., ... & Miyasaki, J. (2006). Prospective prevalence of pathologic gambling and medication association in Parkinson disease. *Neurology*, 66(11), 1750-1752.
- Voorn, P., Vanderschuren, L. J. M. J., Groenewegen, H. J., Robbins, T. W., & Pennartz, C. M. a. (2004). Putting a spin on the dorsal-ventral divide of the striatum. *Trends in Neurosciences*, 27(8), 468–74. <http://doi.org/10.1016/j.tins.2004.06.006>
- Wagatsuma, A., Okuyama, T., Sun, C., Smith, L. M., Abe, K., & Tonegawa, S. (2018). Locus coeruleus input to hippocampal CA3 drives single-trial learning of a novel context. *Proceedings of the National Academy of Sciences*, 115(2), E310-E316.
- Wang, M., Arteaga, D., & He, B. J. (2013). Brain mechanisms for simple perception and bistable perception, 2013. <http://doi.org/10.1073/pnas.1221945110/->

- Wang, M., Arteaga, D., & He, B. J. (2013). Brain mechanisms for simple perception and bistable perception. *Proceedings of the National Academy of Sciences*, 110(35), E3350-E3359.
- Warren, R. M., & Gregory, R. L. (1958). An auditory analogue of the visual reversible figure. *The American journal of psychology*.
- Watanabe, T., Masuda, N., Megumi, F., Kanai, R., & Rees, G. (2014). Energy landscape and dynamics of brain activity during human bistable perception. *Nature Communications*, 5, 1–11. <http://doi.org/10.1038/ncomms5765>
- Weber, E. U., Shafir, S., & Blais, A. R. (2004). Predicting risk sensitivity in humans and lower animals: risk as variance or coefficient of variation. *Psychological review*, 111(2), 430.
- Weilnhammer, V. A., Ludwig, K., Hesselmann, G., & Sterzer, P. (2013). Frontoparietal cortex mediates perceptual transitions in bistable perception. *Journal of Neuroscience*, 33(40), 16009-16015.
- Weilnhammer, V. A., Ludwig, K., Sterzer, P., & Hesselmann, G. (2014). Revisiting the Lissajous figure as a tool to study bistable perception. *Vision Research*, 98, 107–112. <http://doi.org/10.1016/j.visres.2014.03.013>
- Weilnhammer, V., Stuke, H., Hesselmann, G., Sterzer, P., & Schmack, K. (2017). A predictive coding account of bistable perception - a model-based fMRI study, 1–21.
- Weller, J. A., Levin, I. P., Shiv, B., & Bechara, A. (2009). The effects of insula damage on decision-making for risky gains and losses. *Social Neuroscience*, 4(4), 347–358. <http://doi.org/10.1080/17470910902934400>
- Weller, J. A., Levin, I. P., Shiv, B., & Bechara, A. (2009). The effects of insula damage on decision-making for risky gains and losses. *Social neuroscience*, 4(4), 347-358.
- Werner, N. S., Schweitzer, N., Meindl, T., Duschek, S., Kambeitz, J., & Schandry, R. (2013). Interoceptive awareness moderates neural activity during decision-making. *Biological Psychology*, 94(3), 498–506. <http://doi.org/10.1016/j.biopsycho.2013.09.002>
- Wessel, J. R., & Ridderinkhof, K. R. (2010). Conscious perception of errors and its relation to the anterior insula, 629–643. <http://doi.org/10.1007/s00429-010-0261-1>
- White, T. P., Engen, N. H., Sørensen, S., Overgaard, M., & Shergill, S. S. (2014). Uncertainty and confidence from the triple-network perspective: Voxel-based meta-analyses. *Brain and cognition*, 85, 191-200.
- Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My insula: the common neural basis of seeing and feeling disgust. *Neuron*, 40(3), 655-664.
- Wiebking, C., de Greck, M., Duncan, N. W., Tempelmann, C., Bajbouj, M., & Northoff, G. (2015). Interoception in insula subregions as a possible state marker for depression—an exploratory fMRI study investigating healthy, depressed and remitted participants. *Frontiers in Behavioral Neuroscience*, 9(April).

<http://doi.org/10.3389/fnbeh.2015.00082>

- Wilke, M., Logothetis, N. K., & Leopold, D. A. (2006). Local field potential reflects perceptual suppression in monkey visual cortex. *Proceedings of the National Academy of Sciences*, 103(46), 17507–17512. <http://doi.org/10.1073/pnas.0604673103>
- Wise, R. A. (2004). Dopamine, learning and motivation. *Nature reviews neuroscience*, 5(6), 483.
- Wunderlich, K., Smittenaar, P., & Dolan, R. J. (2012). Dopamine enhances model-based over model-free choice behavior. *Neuron*, 75(3), 418–24. <http://doi.org/10.1016/j.neuron.2012.03.042>
- Xue, G., Lu, Z., Levin, I. P., & Bechara, A. (2010). The impact of prior risk experiences on subsequent risky decision-making: The role of the insula. *NeuroImage*, 50(2), 709–716. <http://doi.org/10.1016/j.neuroimage.2009.12.097>
- Yamasaki, M., & Takeuchi, T. (2017). Locus Coeruleus and Dopamine-Dependent Memory Consolidation. *Neural plasticity*, 2017.
- Yoshida, W., Seymour, B., Koltzenburg, M., & Dolan, R. J. (2013). Uncertainty increases pain: evidence for a novel mechanism of pain modulation involving the periaqueductal gray. *Journal of Neuroscience*, 33(13), 5638–5646.
- Yu, A. J., & Dayan, P. (2005). Uncertainty, neuromodulation, and attention. *Neuron*, 46(4), 681–92. <http://doi.org/10.1016/j.neuron.2005.04.026>
- Yu, A., & Dayan, P. (2002). Expected and Unexpected Uncertainty: ACh and NE in the Neocortex. <http://doi.org/citeulike-article-id:496920>
- Zaki, J., Davis, J. I., & Ochsner, K. N. (2012). Overlapping activity in anterior insula during interoception and emotional experience. *NeuroImage*, 62(1), 493–499.
- Zaretskaya, N., Thielscher, A., Logothetis, N. K., & Bartels, A. (2010). Disrupting parietal function prolongs dominance durations in binocular rivalry. *Current biology*, 20(23), 2106–2111.
- Zhou, Y. H., Gao, J. B., White, K. D., Merk, I., & Yao, K. (2004). Perceptual dominance time distributions in multistable visual perception, 263, 256–263. <http://doi.org/10.1007/s00422-004-0472-8>

Appendix A

| Expected Reward $p = .001$; $k = 27$; SVC Analysis | | | | | | | | | | | |
|--|-------------|---------|---------|-------------|------|--------|--------|----|----|-----------|-----------|
| cluster | cluster | cluster | cluster | peak | peak | peak | peak | | | | |
| p(FWE-corr) | p(FDR-corr) | equivk | p(unc) | p(FWE-corr) | T | equivZ | p(unc) | x | y | z {mm} | |
| 0.026 | 0.877 | 18 | 0.405 | 0.019 | 4.21 | 3.52 | 0 | 10 | 14 | -6 | R Caud |
| 0.038 | 0.521 | 11 | 0.521 | 0.022 | 4.21 | 3.52 | 0 | 14 | 10 | -10 | R Put |
| 0.004 | 0.265 | 32 | 0.265 | 0.003 | 4.31 | 3.58 | 0 | 12 | 12 | -8 | R VStr |
| | | | | 0.006 | 4.03 | 3.4 | 0 | 10 | 16 | -6 | |

Table A.1 SVC analyses results for the expected reward contrast on 13 ROIS

RPE Contrast $p = .001$; $k = 27$

| cluster | cluster | cluster | cluster | peak | peak | peak | peak | | | | |
|-------------|-------------|---------|---------|-------------|------|--------|--------|-----|-----|-----------|------------------------------|
| p(FWE-corr) | p(FDR-corr) | equivk | p(unc) | p(FWE-corr) | T | equivZ | p(unc) | x | y | z {mm} | |
| 0 | 0 | 5108 | 0 | 0.023 | 6.96 | 4.9 | 0 | -20 | 34 | 42 | Left Superior Frontal Gyrus |
| 0 | 0 | 756 | 0 | 0.05 | 6.48 | 4.7 | 0 | -40 | 36 | -10 | Left Lateral Orbital Gyrus |
| 0 | 0 | 1746 | 0 | 0.07 | 6.28 | 4.61 | 0 | -60 | -48 | -6 | Left Middle Temporal Gyrus |
| 0 | 0 | 1478 | 0 | 0.095 | 6.09 | 4.53 | 0 | 42 | 34 | 12 | Right Inferior Frontal Gyrus |
| 0.007 | 0.008 | 362 | 0.001 | 0.213 | 5.58 | 4.28 | 0 | -32 | -10 | -4 | Left Putamen |
| 0.006 | 0.007 | 382 | 0.001 | 0.337 | 5.27 | 4.12 | 0 | 14 | 12 | -2 | Right Caudate (Accumbens) |
| 0.055 | 0.042 | 214 | 0.008 | 0.762 | 4.54 | 3.72 | 0 | -2 | -60 | 22 | Precuneus |
| 0.044 | 0.041 | 229 | 0.006 | 0.796 | 4.48 | 3.69 | 0 | -48 | -62 | 24 | Left Angular Gyrus |
| 0.057 | 0.042 | 211 | 0.008 | 0.797 | 4.48 | 3.68 | 0 | 16 | 60 | 20 | Right Superior Frontal Gyrus |
| 0.066 | 0.044 | 201 | 0.009 | 0.835 | 4.41 | 3.64 | 0 | 2 | -8 | 34 | Right Mid-cingulate Gyrus |

Table A.2 Table of results for activity related to the reward prediction error.

| Confidence, p =0.001; k = 31 | | | | | | | | | | | |
|------------------------------|---------------------|------------|------------|---------------------|----------|------------|------------|---------|---------|-----------|---------------------------------------|
| cluster | cluster | cluster | cluster | peak | peak | peak | peak | | | | |
| p(FW E- corr) | p(FD R- corr) | equiv k | p(unc) | p(FW E- corr) | T | equiv Z | p(unc) | x | y | z {mm} | |
| 0.026 | 0.04 | 300 | 0.004 | 0.086 | 6.0 6 | 4.51 | 0 | 34 | 36 | 12 | Right Inferior Frontal Gyrus |
| 0.004 | 0.009 | 477 | 0.001 | 0.107 | 5.9 2 | 4.45 | 0 | 64 | - 48 | 26 | Right Angular Gyrus |
| 0.005 | 0.009 | 454 | 0.001 | 0.213 | 5.4 9 | 4.24 | 0 | - 44 | 46 | 8 | Left Middle Frontal Gyrus |
| 0.001 | 0.003 | 681 | 0 | 0.241 | 5.4 1 | 4.2 | 0 | - 36 | - 56 | 34 | Left Angular Gyrus |
| 0.038 | 0.042 | 270 | 0.006 | 0.368 | 5.1 1 | 4.04 | 0 | 32 | 22 | -6 | Right Anterior Insula |
| 0.041 | 0.042 | 263 | 0.007 | 0.586 | 4.7 4 | 3.83 | 0 | - 16 | 30 | 30 | Left Superior frontal Gyrus |

Confidence SVC Analyses

| cluster | cluster | cluster | cluster | peak | peak | peak | peak | | | | |
|---------------------|---------------------|------------|------------|---------------------|------|------------|------------|----|----|---------------|-----------|
| p(FW E- corr) | p(FD R- corr) | equiv k | p(unc) | p(FW E- corr) | T | equiv Z | p(unc) | x | y | z {mm } | |
| 0.002 | 0.027 | 162 | 0.027 | 0.004 | 5.11 | 4.04 | 0 | 32 | 22 | -6 | R Ains |

Table A.3 Whole brain analyses (p=0.001, cluster level corrected) and SVC analyses of the confidence contrast.

Confidence - Statistical Non Parametric Tests - $p = 0.05$ FWE corrected

| Cluster Level | Voxel Level | | | | | |
|---------------|--------------|----------|-----|-----|----|-------------------------|
| k | P(FWE -corr) | Pseudo-t | x | y | z | |
| 67 | 0.0067 | 5.66 | 58 | -54 | 28 | r Angular Gyrus |
| 24 | 0.0082 | 5.58 | 46 | -30 | -2 | r Middle Temporal Gyrus |
| 6 | 0.0175 | 5.28 | -34 | -56 | 36 | l Angular Gyrus |
| 9 | 0.0196 | 5.24 | -50 | -30 | -6 | L Middle Temporal Gyrus |
| 9 | 0.0292 | 5.05 | 32 | 24 | -6 | r Ant. Insula |
| 6 | 0.0332 | 4.99 | -42 | 46 | 6 | l middle frontal Gyrus |
| 5 | 0.0347 | 4.97 | -56 | -50 | 38 | l supramarginal Gyrus |

Table A.4 Results of statistical non-parametric tests for the confidence contrast (FWE corrected at $p = 0.05$).

Regions implicated include frontal cortical gyrii; anterior insula and importantly middle temporal gyrii.

ABS RPE, $p=0.001$; $k=34$

| cluster | cluster | cluster | cluster | peak | peak | peak | peak | | | | |
|-------------|-------------|---------|---------|-------------|------|--------|--------|-----|-----|-----------|--------------------------------------|
| p(FWE-corr) | p(FDR-corr) | equivk | p(unc) | p(FWE-corr) | T | equivZ | p(unc) | x | y | z {mm} | |
| 0 | 0 | 3900 | 0 | 0.01 | 7.34 | 5.05 | 0 | -28 | 20 | -2 | L Ains |
| | | | | 0.032 | 6.6 | 4.75 | 0 | -20 | -6 | 8 | Thalamus |
| | | | | 0.066 | 6.17 | 4.56 | 0 | -40 | 34 | 12 | Left Inferior Frontal Gyrus |
| 0 | 0 | 2538 | 0 | 0.03 | 6.63 | 4.77 | 0 | 40 | 22 | -10 | R Ains |
| | | | | 0.037 | 6.52 | 4.72 | 0 | 32 | 24 | -12 | R Posterior Orbital Gyrus |
| 0 | 0 | 2045 | 0 | 0.037 | 6.52 | 4.72 | 0 | -62 | -46 | 22 | L Supramar ginal Gyrus |
| 0 | 0 | 1003 | 0 | 0.056 | 6.26 | 4.61 | 0 | 60 | -40 | -2 | R Middle Temporal Gyrus |
| 0.001 | 0.001 | 674 | 0 | 0.1 | 5.91 | 4.45 | 0 | 14 | 10 | 4 | R Caudate |
| 0 | 0 | 1586 | 0 | 0.194 | 5.5 | 4.24 | 0 | 6 | 28 | 16 | R ACC |

ABS RPE SVC

| cluster | cluster | cluster | cluster | peak | peak | peak | peak | | | | |
|-------------|-------------|---------|---------|-------------|------|--------|--------|-----|----|-----------|--------|
| p(FWE-corr) | p(FDR-corr) | equivk | p(unc) | p(FWE-corr) | T | equivZ | p(unc) | x | y | z {mm} | |
| 0 | 0.004 | 333 | 0.004 | 0.004 | 5.14 | 4.06 | 0 | 0 | 30 | 14 | L ACC |
| | | | | 0.008 | 4.83 | 3.89 | 0 | -10 | 24 | 26 | |
| | | | | 0.032 | 4.08 | 3.44 | 0 | -2 | 42 | 14 | |
| | | | | 0.035 | 4.03 | 3.4 | 0 | -8 | 36 | 14 | |
| 0.001 | 0.033 | 215 | 0.017 | 0.002 | 5.5 | 4.24 | 0 | 6 | 28 | 16 | R ACC |
| | | | | 0.024 | 4.16 | 3.49 | 0 | 6 | 24 | 32 | |
| 0 | 0.004 | 347 | 0.004 | 0 | 6.96 | 4.9 | 0 | -28 | 22 | -2 | L AIns |
| | | | | 0.005 | 4.99 | 3.97 | 0 | -36 | 10 | 4 | |
| | | | | 0.043 | 3.84 | 3.29 | 0.001 | -28 | 14 | -18 | |
| 0 | 0.002 | 428 | 0.002 | 0 | 6.4 | 4.67 | 0 | 30 | 22 | -10 | R Ains |
| | | | | 0.001 | 6.11 | 4.54 | 0 | 40 | 20 | -10 | |
| | | | | 0.001 | 5.78 | 4.38 | 0 | 34 | 20 | 0 | |
| 0.003 | 0.106 | 130 | 0.053 | 0.001 | 5.88 | 4.43 | 0 | -14 | 12 | 6 | L Caud |
| | | | | 0.002 | 5.34 | 4.16 | 0 | -16 | 4 | 14 | |

| | | | | | | | | | | | |
|-------|-------|-----|-------|-------|------|------|-------|-----|----|-----|--------|
| | | | | 0.003 | 5.11 | 4.04 | 0 | -18 | 10 | 12 | |
| 0.045 | 0.834 | 2 | 0.834 | 0.031 | 3.87 | 3.3 | 0 | -8 | 6 | 2 | |
| 0.002 | 0.032 | 165 | 0.032 | 0.001 | 5.73 | 4.36 | 0 | 12 | 12 | 4 | R Caud |
| | | | | 0.001 | 5.54 | 4.27 | 0 | 18 | 8 | 14 | |
| 0.008 | 0.184 | 76 | 0.129 | 0.001 | 5.83 | 4.4 | 0 | -22 | -4 | 8 | L Put |
| | | | | 0.008 | 4.68 | 3.8 | 0 | -30 | 0 | -2 | |
| 0.012 | 0.184 | 57 | 0.184 | 0.014 | 4.39 | 3.63 | 0 | -26 | 14 | -2 | |
| | | | | 0.024 | 4.1 | 3.45 | 0 | -16 | 8 | -10 | |
| | | | | 0.036 | 3.86 | 3.3 | 0 | -20 | 10 | -2 | |
| 0.013 | 0.892 | 52 | 0.203 | 0.011 | 4.52 | 3.71 | 0 | 24 | -6 | 6 | R Put |
| | | | | 0.028 | 3.99 | 3.38 | 0 | 20 | 6 | 2 | |
| 0.035 | 0.892 | 12 | 0.549 | 0.025 | 4.05 | 3.42 | 0 | 28 | 12 | 2 | |
| | | | | 0.032 | 3.93 | 3.34 | 0 | 30 | 8 | 2 | |
| | | | | 0.04 | 3.8 | 3.26 | 0.001 | 26 | 16 | 2 | |
| 0.047 | 0.892 | 4 | 0.75 | 0.027 | 4.02 | 3.4 | 0 | 26 | 14 | -8 | |

Table A.5 Whole brain cluster level corrected results of surprise as absolute reward prediction error; the lower half of the table shows results of SVC analyses for the surprise contrast on 6 (12 bilateral) ROIs.

Surprise SnPM

| k | pFWE-Corr | pFDR-corr | Pseudo-t | puncorr | x | y | z | |
|-----|-----------|-----------|----------|---------|-----|-----|-----|---------------------------|
| 594 | 0.0012 | 0.0035 | 6.34 | 0.0001 | 32 | 22 | -10 | R Posterior Orbital Gyrus |
| | 0.0013 | 0.0035 | 6.28 | 0.0001 | 44 | 20 | -10 | |
| | 0.0131 | 0.0035 | 5.31 | 0.0001 | 38 | 14 | 0 | |
| 350 | 0.0017 | 0.0035 | 6.21 | 0.0001 | -62 | -44 | 24 | L Supramarginal Gyrus |
| | 0.0056 | 0.0035 | 5.71 | 0.0001 | -58 | -52 | 6 | |
| 320 | 0.002 | 0.0035 | 6.09 | 0.0001 | -30 | 22 | -4 | L Ains |
| 191 | 0.0026 | 0.0035 | 6 | 0.0001 | -42 | 34 | 12 | L Inferior Frontal Gyrus |
| 207 | 0.0046 | 0.0035 | 5.8 | 0.0001 | 56 | -34 | -2 | R Superior Temporal Gyrus |
| | 0.0059 | 0.0035 | 5.66 | 0.0001 | 50 | -28 | -4 | |
| 25 | 0.0073 | 0.0035 | 5.61 | 0.0001 | -20 | -4 | 8 | L Caud |
| 40 | 0.0092 | 0.0035 | 5.48 | 0.0001 | 2 | 28 | 14 | R ACC |
| 149 | 0.0092 | 0.0035 | 5.47 | 0.0001 | -64 | -36 | -6 | L Middle Temporal Gyrus |
| | 0.0093 | 0.0035 | 5.47 | 0.0001 | -52 | -30 | -4 | |
| 18 | 0.0232 | 0.0044 | 5.05 | 0.0002 | 54 | -44 | 12 | L Superior Temporal Gyrus |
| 13 | 0.0266 | 0.0044 | 4.98 | 0.0002 | 18 | 8 | 12 | R Caud |
| 3 | 0.0319 | 0.0035 | 4.91 | 0.0001 | 0 | 22 | 42 | Supplementary Motor Gyrus |
| 14 | 0.0327 | 0.0035 | 4.91 | 0.0001 | -38 | 14 | 28 | L Middle Frontal Gyrus |
| 2 | 0.0347 | 0.0044 | 4.88 | 0.0002 | -10 | 26 | 26 | L ACC |

Table A.6 Results of non-parametric statistical tests. Bilateral anterior cingulate cortex involvement remains as does activity in left caudate and left anterior insula.

Information; p =0.001; k=31

| cluster | cluster | cluster | cluster | peak | peak | peak | peak | | | | |
|-----------------|-----------------|------------|------------|-----------------|----------|------------|------------|---------|---------|---------------|---------------------------------------|
| p(FWE -corr) | p(FDR -corr) | equiv k | p(unc) | p(FWE -corr) | T | equiv Z | p(unc) | x | y | z {mm } | |
| 0 | 0 | 2984 | 0 | 0.004 | 8.0 3 | 5.31 | 0 | - 46 | - 36 | 40 | Left Supramargin al Gyrus |
| 0 | 0 | 4088 | 0 | 0.062 | 6.2 7 | 4.61 | 0 | 24 | - 70 | 24 | Right Middle Occipital Gyrus |
| 0 | 0 | 844 | 0 | 0.105 | 5.9 4 | 4.46 | 0 | 24 | - 62 | -26 | Left Middle Occipital Gyrus |
| 0 | 0 | 1838 | 0 | 0.16 | 5.6 8 | 4.33 | 0 | 22 | 4 | 16 | Right Putamen |
| 0 | 0 | 769 | 0 | 0.229 | 5.4 5 | 4.22 | 0 | - 32 | 38 | 18 | Left Middle Frontal Gyrus |
| 0.009 | 0.009 | 390 | 0.001 | 0.35 | 5.1 6 | 4.07 | 0 | 24 | 60 | 4 | Right Superior Frontal Gyrus |

Information, SVC analysis

| cluster | cluster | cluster | cluster | peak | peak | peak | peak | | | | |
|-----------------|-----------------|------------|------------|-----------------|------|------------|------------|---------|----|---------------|---------|
| p(FWE -corr) | p(FDR -corr) | equiv k | p(unc) | p(FWE -corr) | T | equiv Z | p(unc) | x | y | z {mm } | |
| 0.006 | 0.081 | 93 | 0.081 | 0.002 | 5.42 | 4.2 | 0 | - 28 | 16 | 10 | A Ins L |
| | | | | 0.028 | 4.12 | 3.46 | 0 | - 38 | -2 | 8 | |
| | | | | 0.028 | 4.11 | 3.46 | 0 | - 34 | 2 | 8 | |
| | | | | 0.029 | 4.09 | 3.44 | 0 | - 42 | -4 | 8 | |
| 0.011 | 0.257 | 62 | 0.147 | 0.003 | 5.33 | 4.16 | 0 | 32 | 18 | 10 | Ains R |
| | | | | 0.005 | 5.08 | 4.02 | 0 | 30 | 22 | 8 | |
| | | | | 0.017 | 4.37 | 3.62 | 0 | 34 | 4 | 8 | |
| 0.02 | 0.257 | 37 | 0.257 | 0.007 | 4.87 | 3.91 | 0 | 34 | -8 | 12 | |

| | | | | | | | | | | | |
|-------|-------|-----|-------|-------|------|------|-------|---------|---------|----|--------|
| 0.045 | 0.775 | 3 | 0.775 | 0.041 | 3.74 | 3.22 | 0.001 | - 18 | 0 | 18 | Caud L |
| 0.04 | 0.667 | 6 | 0.667 | 0.009 | 4.58 | 3.74 | 0 | 20 | 4 | 18 | Caud R |
| 0.021 | 0.667 | 26 | 0.341 | 0.015 | 4.31 | 3.58 | 0 | 16 | 18 | 12 | |
| 0.003 | 0.063 | 107 | 0.063 | 0.002 | 5.15 | 4.06 | 0 | 38 | - 12 | 12 | Pins L |
| 0.011 | 0.228 | 42 | 0.228 | 0.018 | 4.09 | 3.45 | 0 | - 38 | - 12 | 8 | Pins R |
| 0.005 | 0.075 | 97 | 0.075 | 0.007 | 4.81 | 3.87 | 0 | - 24 | 12 | 6 | Put L |
| | | | | 0.009 | 4.65 | 3.78 | 0 | - 26 | 2 | 8 | |
| 0.004 | 0.166 | 115 | 0.055 | 0.003 | 5.21 | 4.1 | 0 | 32 | 2 | 2 | Put R |
| | | | | 0.004 | 5.04 | 4 | 0 | 26 | 6 | 10 | |
| | | | | 0.004 | 5.04 | 4 | 0 | 26 | 2 | 12 | |
| | | | | 0.016 | 4.33 | 3.6 | 0 | 28 | -6 | 12 | |

Table A.7 Results of information contrast, sampled after Card 2; The upper part of the table shows results of a whole brain analysis ($p = 0.001$, cluster level corrected $k = 31$) and the bottom half shows SVC analyses results on 6 ROIs.

Information, SnPM, $p=0.05$ FWE

| k | pFWE-Corr | pFDR-corr | Pseudo-t | puncorr | x | y | z | Region |
|-----|-----------|-----------|----------|---------|-----|-----|-----|--------------------------|
| 147 | 0.0001 | 0.0034 | 7.49 | 0.0001 | -44 | -34 | 40 | L Supramarginal Gyrus |
| 72 | 0.0053 | 0.0034 | 5.54 | 0.0001 | -12 | -72 | 26 | Left Cuneus |
| 40 | 0.0073 | 0.0034 | 5.44 | 0.0001 | 32 | 18 | 10 | R Ains |
| 69 | 0.0171 | 0.0043 | 5.17 | 0.0002 | 20 | -70 | 24 | Right Cuneus |
| | 0.0189 | 0.0034 | 5.11 | 0.0001 | 14 | -74 | 30 | |
| 19 | 0.0184 | 0.0034 | 5.13 | 0.0001 | 32 | -76 | 20 | R Superior Occ. Gyrus |
| 16 | 0.0211 | 0.0034 | 5.07 | 0.0001 | -52 | -8 | 24 | R Postcentral Gyrus |
| 5 | 0.0315 | 0.0043 | 4.94 | 0.0002 | -18 | -52 | -14 | Cerebellum |
| 5 | 0.0382 | 0.0034 | 4.87 | 0.0001 | 26 | 60 | 4 | R Superior Frontal Gyrus |
| 9 | 0.0389 | 0.0034 | 4.86 | 0.0001 | 26 | 2 | 10 | R Putamen |
| 9 | 0.04 | 0.0034 | 4.85 | 0.0001 | -54 | -22 | 12 | L Central Operculum |
| 3 | 0.0421 | 0.0043 | 4.82 | 0.0002 | -16 | -68 | 6 | L Calcarine cortex |
| 2 | 0.0462 | 0.0034 | 4.78 | 0.0001 | -8 | -74 | -10 | L Lingual Gyrus |
| 3 | 0.0465 | 0.0034 | 4.77 | 0.0001 | 24 | -60 | -22 | Cerebellum |

Table A.8 Results of non-parametric statistical tests for information. In addition to a cluster in the right anterior insula, we find a bilateral cluster in the cuneus, as well as clusters in the frontal lobe and cerebellum.

Appendix B

Task 1 t-Test $-p=001$; $k = 31$; SVC

| cluster | cluster | cluster | peak | peak | peak | | | | |
|--------------|-----------|--------------|--------------|-------------|--------------|-----------|-----------|-----------|--------------|
| p(FWE-corr) | equivk | p(unc) | p(FWE-corr) | T | p(unc) | x | y | z {mm} | |
| 0.038 | 14 | 0.168 | 0.064 | 3.68 | 0.001 | 34 | 24 | 0 | rAIns |

Table B.1 Significant SVC results for perceptual risk in Task 1 in the Right Anterior Insula

Task 2 -GLM1 -t-test - Replay > Report, $p=0.001$, $k=25$

| cluster | cluster | cluster | peak | peak | peak | | | null | Region |
|-------------|---------|---------|-------------|-------|--------|-----|-----|--------|------------------------|
| p(FWE-corr) | equivk | p(unc) | p(FWE-corr) | T | p(unc) | x | y | z {mm} | |
| 0 | 969 | 0 | 0.001 | 10.14 | 0 | -54 | -56 | 42 | Angular Gyrus L |
| | | | 0.028 | 7.3 | 0 | -42 | -58 | 40 | |
| 0 | 1850 | 0 | 0.002 | 9.24 | 0 | 46 | 30 | 30 | Mid Frontal Gyrus R |
| 0 | 871 | 0 | 0.041 | 7.04 | 0 | -2 | 32 | 46 | Superior Frontal Gyrus |
| 0 | 1100 | 0 | 0.045 | 6.98 | 0 | -46 | 12 | 44 | Mid Frontal Gyrus L |
| 0 | 1290 | 0 | 0.055 | 6.84 | 0 | 36 | -52 | 40 | Angular Gyrus R |
| 0.001 | 518 | 0 | 0.327 | 5.59 | 0 | 60 | -34 | -6 | MTG R |
| 0 | 594 | 0 | 0.518 | 5.21 | 0 | 36 | -80 | -10 | Occ. Fusiform Gyrus R |
| 0.038 | 238 | 0.004 | 0.638 | 5 | 0 | -58 | -32 | -10 | MTG L |

Table B.2 Whole Brain Analyses for Replay > Report condition ($p=.001$, $k=25$)

Task 2 -T-Test - Up > Down, p =.001, k= 23

| cluster | cluster | cluster | peak | peak | peak | | | nu ll | |
|-------------|---------|---------|-------------|------|--------|-----|-----|----------|------------------------|
| p(FWE-corr) | equivk | p(unc) | p(FWE-corr) | T | p(unc) | x | y | z | Region |
| 0 | 5418 | 0 | 0.017 | 7.71 | 0 | -8 | -92 | 16 | L Occipital Pole |
| | | | 0.023 | 7.53 | 0 | -16 | -42 | 34 | Precuneus |
| | | | 0.049 | 6.99 | 0 | -14 | -56 | 26 | Precuneus |
| 0 | 706 | 0 | 0.25 | 5.86 | 0 | -46 | -52 | 26 | R Angular Gyrus |
| 0.021 | 255 | 0.002 | 0.289 | 5.75 | 0 | 56 | -56 | 22 | L Angular gyrus |

Table B.3 Whole Brain results for contrast Up > Down switches (p =.001, k = 23), in the Necker Cube, Task 2.

Task 2 –T-Test - Switch, p =.001, k=37

| cluster | cluster | peak | peak | peak | peak | | | null | |
|---------------------|---------|-----------------|-----------------|------|--------|-----|-----|--------|-----------------------------|
| p(FW E- corr) | equivk | p(FWE- corr) | p(FDR- corr) | T | p(unc) | x | y | z {mm} | Regio n |
| 0 | 1119 | 0.009 | 0.085 | 7.82 | 0 | -44 | 16 | -4 | L AIns |
| 0 | 4231 | 0.013 | 0.085 | 7.6 | 0 | -32 | -56 | -26 | Cerebellum |
| | | 0.021 | 0.085 | 7.25 | 0 | 30 | -54 | -30 | |
| | | 0.029 | 0.085 | 7.03 | 0 | -40 | -60 | -28 | |
| 0 | 1605 | 0.026 | 0.085 | 7.1 | 0 | -12 | -22 | -14 | Ventral DC |
| 0 | 947 | 0.057 | 0.125 | 6.58 | 0 | 58 | 10 | 2 | L AIns |
| 0 | 868 | 0.162 | 0.13 | 5.87 | 0 | 48 | -40 | 28 | R Supramarginal Gyrus |
| 0.003 | 603 | 0.366 | 0.206 | 5.27 | 0 | 10 | -2 | 6 | Thalamus |
| 0.032 | 333 | 0.45 | 0.221 | 5.1 | 0 | -66 | -38 | 30 | R Supramarginal Gyrus |

Table B.4 Whole brain results for switch response (p =.001, k = 37).

SVC Analysis - Switch Modulated by Surprise and Previous Dwell Time

| cluster | cluster | cluster | peak | peak | peak | | | | |
|-------------|---------|---------|-------------|------|--------|-----|----|-----------|--------|
| p(FWE-corr) | equivk | p(unc) | p(FWE-corr) | T | p(unc) | x | y | z {mm} | |
| 0.059 | 4 | 0.721 | 0.047 | 3.93 | 0 | -30 | 18 | 10 | AIns L |
| 0.05 | 7 | 0.621 | 0.039 | 4.04 | 0 | 38 | 24 | 2 | Ains R |

Table B.5 SVC results for switch onsets modulated by surprise as risk prediction error.

NC Task 2 - Abs RPE -p =0.001, k = 30

| cluster | cluster | cluster | cluster | peak | peak | peak | peak | peak | | | | |
|-------------|-------------|---------|---------|-------------|-------------|------|--------|--------|----|----|--------|-----------|
| p(FWE-corr) | p(FDR-corr) | equivk | p(unc) | p(FWE-corr) | p(FDR-corr) | T | equivZ | p(unc) | x | y | z {mm} | |
| 0.039 | 0.498 | 13 | 0.498 | 0.022 | 0.282 | 4.34 | 3.54 | 0 | 36 | 26 | 2 | Ains R |

Table B.6 SVC analysis of surprise as absolute prediction error in Necker Cube task 2.

Whole Brain Analysis - Switch Modulated by Previous Dwell Time; $p=.0001$; $k=12$

| cluster | cluster | peak | peak | peak | | | null | |
|-------------|---------|-------------|------|--------|-----|-----|--------|--------------------------------|
| p(FWE-corr) | equivk | p(FWE-corr) | T | p(unc) | x | y | z {mm} | Region |
| 0.025 | 80 | 0.024 | 7.37 | 0 | 16 | -22 | 4 | Thalamus |
| 0 | 371 | 0.024 | 7.37 | 0 | -46 | -20 | 28 | L Central Operculum |
| 0.001 | 180 | 0.026 | 7.33 | 0 | -22 | -20 | -4 | L Ventral DC |
| | | 0.042 | 6.98 | 0 | -26 | -14 | -10 | |
| 0 | 494 | 0.053 | 6.83 | 0 | -24 | 26 | -6 | L AIns |
| 0 | 468 | 0.073 | 6.62 | 0 | 56 | -14 | 24 | R Parietal Operculum |
| 0 | 231 | 0.082 | 6.53 | 0 | -38 | 14 | -30 | L Temporal Pole |
| 0.042 | 65 | 0.088 | 6.49 | 0 | -44 | -72 | 24 | L Middle Occipital Gyrus |
| 0.023 | 82 | 0.115 | 6.3 | 0 | 12 | -50 | 12 | Precuneus |
| 0 | 272 | 0.118 | 6.29 | 0 | -20 | -62 | 24 | Precuneus |
| 0 | 278 | 0.13 | 6.22 | 0 | 50 | -68 | 16 | R Middle Occipital Gyrus |
| 0.002 | 172 | 0.207 | 5.89 | 0 | 4 | -2 | 46 | R Mid Cingulate Gyrus |
| 0.037 | 69 | 0.362 | 5.48 | 0 | -48 | -62 | 10 | L Middle Temporal Gyrus |

Table B.7 Whole brain analysis for switch onset modulated by previous dwell time ($p=.0001$, $k=12$)

SVC Analysis - Switch Modulated by Previous Dwell Time; $p = .0001$; $k = 12$

| cluster | cluster | peak | peak | peak | | | | |
|-------------|---------|-------------|------|--------|-----|-----|-----------|--------|
| p(FWE-corr) | equivk | p(FWE-corr) | T | p(unc) | x | y | z {mm} | |
| 0.003 | 16 | 0.001 | 6.25 | 0 | -28 | 28 | 4 | Ains L |
| 0.001 | 40 | 0.002 | 5.95 | 0 | -38 | 8 | -6 | |
| 0.007 | 6 | 0.005 | 5.24 | 0 | -32 | 12 | 10 | |
| 0.006 | 8 | 0.009 | 4.93 | 0 | -32 | 14 | -20 | |
| 0.008 | 1 | 0.008 | 4.8 | 0 | -14 | 14 | -4 | Caud L |
| 0.004 | 9 | 0.002 | 5.66 | 0 | 14 | 18 | -6 | Caud R |
| 0.001 | 42 | 0.001 | 6.26 | 0 | -28 | -14 | -8 | Put L |
| 0 | 67 | 0.003 | 5.56 | 0 | -22 | 12 | -10 | |
| | | 0.003 | 5.5 | 0 | -22 | 18 | -2 | |
| | | 0.007 | 4.98 | 0 | -28 | 10 | -2 | |
| 0.008 | 3 | 0.008 | 4.89 | 0 | 18 | 18 | -8 | Put R |
| 0.001 | 7 | 0 | 5.82 | 0 | 14 | 18 | -8 | VStr R |

Table B.8 SVC analysis results for switch onset modulated by previous dwell time

Whole Brain Analysis - Switch Modulated by Surprise and Previous Dwell Time; $p = .001$; $k = 28$

| cluster | cluster | peak | peak | peak | | | null | |
|-------------|---------|-------------|------|--------|-----|-----|--------|----------------------------|
| p(FWE-corr) | equivk | p(FWE-corr) | T | p(unc) | x | y | z {mm} | Region |
| 0 | 738 | 0.394 | 5.38 | 0 | -22 | -80 | -14 | L Occipital Fusiform |
| 0.003 | 460 | 0.451 | 5.27 | 0 | 32 | -74 | -14 | R Occipital Fusiform |
| 0.011 | 351 | 0.457 | 5.25 | 0 | 34 | -82 | 26 | R Middle Occipital |
| 0.014 | 332 | 0.672 | 4.88 | 0 | 50 | 10 | 26 | R Precentral Gyrus |

Table B.9 Whole Brain analysis for switch onset modulated by previous dwell time and surprise ($p = .0001$, $k = 28$)

SVC Analysis - Switch Modulated by Surprise and Previous Dwell Time

| cluster | cluster | peak | peak | peak | | | | |
|-------------|---------|-------------|------|--------|-----|----|--------|--------|
| p(FWE-corr) | equivk | p(FWE-corr) | T | p(unc) | x | y | z {mm} | AIns L |
| 0.032 | 20 | 0.02 | 4.45 | 0 | -30 | 18 | 10 | |
| | | 0.027 | 4.27 | 0 | -32 | 22 | 6 | |
| 0.02 | 36 | 0.035 | 4.09 | 0 | 30 | 26 | 8 | Ains R |
| | | 0.045 | 3.96 | 0 | 38 | 24 | 0 | |

Table B.10 SVC analysis of switch onset modulated by both previous dwell time and surprise

Appendix C

Categorical Risk Card Game - $p=0.001$; $k=24$

| cluster | cluster | cluster | peak | peak | peak | peak | | | | |
|-------------|-------------|---------|-------------|-------|--------|--------|-----|-----|-----------|---|
| p(FWE-corr) | p(FDR-corr) | equivk | p(FWE-corr) | F | equivZ | p(unc) | x | y | z {mm} | |
| 0 | 0 | 977 | 0.177 | 15.73 | 4.29 | 0 | 48 | -74 | 8 | Right Inferior occipital gyrus |
| 0.041 | 0.047 | 219 | 0.185 | 15.62 | 4.27 | 0 | -30 | -38 | -16 | Left Fusiform Gyrus |
| 0 | 0 | 626 | 0.199 | 15.45 | 4.25 | 0 | -40 | -64 | -12 | Left Inferior occipital gyrus |
| 0.007 | 0.01 | 338 | 0.277 | 14.64 | 4.15 | 0 | 32 | 22 | -6 | R Anterior Insula |
| 0 | 0 | 763 | 0.312 | 14.34 | 4.11 | 0 | 60 | -56 | 26 | R Angular Gyrus |
| 0.081 | 0.079 | 176 | 0.145 | 16.2 | 4.34 | 0 | 34 | -6 | -28 | R Amygdala |

Categorical Risk Card Game SVC Analyses

| cluster | cluster | cluster | peak | peak | peak | peak | | | | |
|-------------|-------------|---------|-------------|-------|--------|--------|-----|----|-----------|--------|
| p(FWE-corr) | p(FDR-corr) | equivk | p(FWE-corr) | F | equivZ | p(unc) | x | y | z {mm} | |
| 0.03 | 0.334 | 21 | 0.032 | 10.11 | 3.45 | 0 | -28 | 24 | 2 | Ains L |
| 0.001 | 0.009 | 186 | 0.003 | 14.64 | 4.15 | 0 | 32 | 22 | -6 | Ains R |
| | | | 0.017 | 11.23 | 3.64 | 0 | 44 | 0 | 2 | |
| | | | 0.024 | 10.59 | 3.54 | 0 | 38 | 4 | -6 | |
| 0.009 | 0.149 | 48 | 0.015 | 10.58 | 3.53 | 0 | 38 | 0 | -4 | Pins L |
| | | | 0.026 | 9.67 | 3.37 | 0 | 40 | -8 | 2 | |

Table C.1 Results of ANOVA on main effect of risk (respectively, confidence) at card 1, separated according to low, medium and high risk. The top table represents whole brain analyses, while the lower portion of the table shows SVC analyses.

Categorical Surprise Card Game - $p=0.001$; $k=23$

| cluster | cluster | cluster | peak | peak | peak | peak | | | | |
|-------------|-------------|---------|-------------|-------|--------|--------|-----|-----|---------------|--------------------------------|
| p(FWE-corr) | p(FDR-corr) | equivk | p(FWE-corr) | F | equivZ | p(unc) | x | y | z {mm } | |
| 0 | 0 | 675 | 0.072 | 17.94 | 4.55 | 0 | -60 | -48 | 22 | Left Supramarginal gyrus |
| 0.008 | 0.009 | 319 | 0.11 | 16.95 | 4.43 | 0 | 62 | -10 | 30 | Right Postcentral gyrus |
| 0.021 | 0.02 | 255 | 0.277 | 14.72 | 4.16 | 0 | 50 | -28 | -6 | Right Middle Temporal Gyrus |
| 0.001 | 0.003 | 447 | 0.499 | 13.15 | 3.94 | 0 | -34 | 18 | -2 | Left Anterior Insula |
| 0.003 | 0.004 | 393 | 0.521 | 13.02 | 3.92 | 0 | 44 | 22 | -10 | Right Anterior Insula |
| 0.028 | 0.022 | 235 | 0.623 | 12.44 | 3.83 | 0 | -46 | 12 | 30 | Left Middle Frontal Gyrus |

Categorical Surprise Card Game - SVC

| cluster | cluster | cluster | peak | peak | peak | peak | | | | |
|-------------|-------------|---------|-------------|-------|--------|--------|-----|-----|---------------|---------|
| p(FWE-corr) | p(FDR-corr) | equivk | p(FWE-corr) | F | equivZ | p(unc) | x | y | z {mm } | |
| 0.033 | 0.624 | 22 | 0.056 | 9.34 | 3.31 | 0 | 0 | 36 | 24 | ACC L |
| 0.04 | 0.44 | 13 | 0.054 | 9.15 | 3.27 | 0.001 | 2 | 36 | 24 | ACC R |
| 0.001 | 0.014 | 192 | 0.006 | 13.15 | 3.94 | 0 | -34 | 18 | -2 | A Ins L |
| 0.03 | 0.323 | 21 | 0.02 | 11.05 | 3.61 | 0 | -28 | 14 | -18 | |
| 0.001 | 0.008 | 184 | 0.008 | 12.65 | 3.87 | 0 | 40 | 22 | -4 | A Ins R |
| | | | 0.011 | 12.14 | 3.79 | 0 | 42 | 20 | -8 | |
| | | | 0.017 | 11.26 | 3.65 | 0 | 32 | 22 | -2 | |
| 0.013 | 0.374 | 38 | 0.005 | 12.91 | 3.9 | 0 | -14 | 2 | 14 | Caud L |
| 0.011 | 0.153 | 45 | 0.008 | 12.13 | 3.79 | 0 | 12 | 10 | 4 | Caud R |
| 0.039 | 0.648 | 5 | 0.033 | 9.28 | 3.3 | 0 | 36 | -10 | 16 | P Ins L |
| 0.016 | 0.272 | 26 | 0.009 | 11.5 | 3.69 | 0 | -40 | -14 | 10 | P Ins R |

Table C.2 Results of ANOVA on main effect of surprise (respectively, information) at card 2, separated according to low, medium and high surprise. The top table represents whole brain analyses, while the lower portion of the table shows SVC analyses.

Task 2 - ANOVA - Ambiguity - $p=.001$, $k = 22$ - cluster

| cluster | cluster | peak | peak | peak | | | null | Region |
|-------------|---------|-------------|-------|--------|-----|-----|--------|-----------------------------|
| p(FWE-corr) | equivk | p(FWE-corr) | F | p(unc) | x | y | z {mm} | |
| 0 | 1289 | 0.003 | 27.65 | 0 | 50 | -60 | 0 | Right Middle Temporal gyrus |
| 0 | 987 | 0.03 | 21.46 | 0 | -54 | -58 | 0 | Left Middle Temporal Gyrus |
| 0.005 | 348 | 0.088 | 18.71 | 0 | 14 | 16 | 36 | Supplementary Motor Cortex |
| 0 | 837 | 0.122 | 17.87 | 0 | 4 | -62 | 12 | Right Calcarine Cortex |
| 0.007 | 320 | 0.128 | 17.75 | 0 | 40 | -26 | 0 | Right Anterior Insula |
| 0 | 539 | 0.152 | 17.31 | 0 | -34 | -6 | -2 | Left Posterior Insula |
| 0 | 825 | 0.189 | 16.75 | 0 | -44 | 14 | 16 | Left Inferior Frontal Gyrus |
| 0.016 | 269 | 0.299 | 15.53 | 0 | 52 | 4 | 10 | Right Anterior Insula |
| 0.016 | 268 | 0.334 | 15.22 | 0 | 44 | 42 | 26 | Right Middle Frontal Gyrus |

Task 2 - ANOVA - Ambiguity – SVC Analysis

| cluster | cluster | peak | peak | peak | | | null | |
|-------------|---------|-------------|-------|--------|-----|-----|--------|--------|
| p(FWE-corr) | equivk | p(FWE-corr) | F | p(unc) | x | y | z {mm} | |
| 0.025 | 26 | 0.017 | 11.76 | 0 | -38 | -2 | 0 | lAins |
| 0.049 | 9 | 0.034 | 10.36 | 0 | -28 | 20 | 8 | |
| 0.008 | 64 | 0.019 | 11.45 | 0 | 40 | -2 | 8 | rAins |
| | | 0.027 | 10.73 | 0 | 36 | 6 | 8 | |
| 0.022 | 17 | 0.012 | 11.44 | 0 | 34 | -16 | 4 | lPIins |
| 0.02 | 20 | 0.016 | 10.97 | 0 | 38 | -6 | 0 | |
| | | 0.037 | 9.32 | 0.001 | 38 | -10 | -2 | |
| 0.005 | 67 | 0.001 | 16.16 | 0 | -38 | -12 | -4 | rPIins |
| 0.004 | 91 | 0.005 | 13.89 | 0 | -32 | -6 | -2 | lPut |
| | | 0.007 | 13.22 | 0 | -30 | -16 | -6 | |
| 0.02 | 28 | 0.021 | 10.99 | 0 | 32 | -14 | 4 | rPut |
| | | 0.044 | 9.59 | 0 | 30 | -18 | 6 | |

Table C.3 Whole brain analysis of main effect of ambiguity on switch response in Necker Cube, Task 2 (p=0.001; k =22) SVC results of main effect of ambiguity in the Necker Cube, Task 2 on 7 a priori ROIs and subsequent SVC analysis.

Task 2 - ANOVA - Bias - $p=0.05$, FWE

| cluster | cluster | peak | peak | peak | | | null | |
|-------------|---------|-------------|-------|--------|-----|-----|-----------|------------------------|
| p(FWE-corr) | equiv | p(FWE-corr) | F | p(unc) | x | y | z {mm} | |
| 0 | 316 | 0 | 26.12 | 0 | 36 | -52 | -28 | Cerebellum |
| | | 0.001 | 16.68 | 0 | 26 | -60 | -20 | |
| 0 | 433 | 0 | 25.39 | 0 | 52 | 10 | 2 | R AIns |
| | | 0.014 | 12.82 | 0 | 36 | 8 | 8 | |
| 0 | 478 | 0 | 21.13 | 0 | 64 | -38 | 24 | R Supramarginal Gyrus |
| | | 0 | 18.56 | 0 | 56 | -38 | 28 | |
| 0 | 522 | 0 | 20.97 | 0 | -50 | 2 | 4 | L AIns |
| | | 0 | 18.53 | 0 | -48 | -2 | 12 | |
| | | 0 | 17.23 | 0 | -40 | 12 | 0 | |
| 0 | 91 | 0 | 20.43 | 0 | -18 | -68 | -46 | Cerebellum |
| 0 | 368 | 0 | 18.75 | 0 | -36 | -60 | -26 | Cerebellum |
| | | 0 | 18.53 | 0 | -40 | -58 | -34 | |
| | | 0.002 | 15.02 | 0 | -26 | -58 | -28 | |
| 0 | 284 | 0 | 18.43 | 0 | 0 | 18 | 36 | Middle Cingulate Gyrus |
| | | 0.002 | 15.31 | 0 | -10 | 32 | 32 | |
| 0.001 | 41 | 0 | 17.12 | 0 | 18 | -64 | -48 | Cerebellum |
| 0 | 96 | 0.004 | 14.26 | 0 | 12 | -14 | 8 | Thalamus |
| 0.001 | 39 | 0.005 | 13.99 | 0 | -14 | -18 | 8 | Thalamus |
| 0.003 | 23 | 0.008 | 13.46 | 0 | 36 | 44 | 36 | R Middle Frontal Gyrus |
| | | | | | | | | |

Table C.4 Whole brain analysis results for the main effect of stimulus bias (bias towards “from above”), $p=0.05$, FWE ,corrected

| cluster | cluster | cluster | cluster | peak | peak | peak | peak | | | | |
|-------------|-------------|---------|---------|-------------|------|--------|--------|-----|-----|------|------------------------------|
| p(FWE-corr) | p(FDR-corr) | equivk | p(unc) | p(FWE-corr) | T | equivZ | p(unc) | x | y | z | |
| | | | | | | | | | | {mm} | |
| 0 | 0 | 12651 | 0 | 0.01 | 8.23 | 5.31 | 0 | 64 | -4 | 12 | R. Anterior Insula |
| 0.014 | 0.008 | 572 | 0.001 | 0.028 | 7.53 | 5.06 | 0 | -36 | -4 | -44 | L. Inf. Temporal Gyrus |
| 0 | 0 | 5317 | 0 | 0.076 | 6.88 | 4.82 | 0 | -42 | -34 | 4 | L. Transverse Temporal Gyrus |
| 0.001 | 0 | 1024 | 0 | 0.292 | 5.96 | 4.42 | 0 | -14 | -80 | 2 | L. Calcarine Cortex |
| 0 | 0 | 1399 | 0 | 0.515 | 5.51 | 4.21 | 0 | 8 | 36 | 36 | Anterior Cingulate Cortex |
| 0.071 | 0.033 | 385 | 0.005 | 0.653 | 5.28 | 4.09 | 0 | 24 | -72 | 40 | R. Superior Occipital Cortex |
| 0.093 | 0.037 | 355 | 0.006 | 0.908 | 4.79 | 3.83 | 0 | -64 | 10 | 14 | L. Anterior Insula |

Table C.5 Table of results showing clusters emerging from a group level analysis on accuracy maps for the classification of data into correct categories of surprise (respectively, information) levels.



Apathy and noradrenaline: silent partners to mild cognitive impairment in Parkinson's disease?

Leyla Loued-Khenissi^a and Kerstin Preuschoff^b

Purpose of review

Mild cognitive impairment (MCI) is a comorbid factor in Parkinson's disease. The aim of this review is to examine the recent neuroimaging findings in the search for Parkinson's disease MCI (PD-MCI) biomarkers to gain insight on whether MCI and specific cognitive deficits in Parkinson's disease implicate striatal dopamine or another system.

Recent findings

The evidence implicates a diffuse pathophysiology in PD-MCI rather than acute dopaminergic involvement. On the one hand, performance in specific cognitive domains, notably in set-shifting and learning, appears to vary with dopaminergic status. On the other hand, motivational states in Parkinson's disease along with their behavioral and physiological indices suggest a noradrenergic contribution to cognitive deficits in Parkinson's disease. Finally, Parkinson's disease's pattern of neurodegeneration offers an avenue for continued research in nigrostriatal dopamine's role in distinct behaviors, as well as the specification of dorsal and ventral striatal functions.

Summary

The search for PD-MCI biomarkers has employed an array of neuroimaging techniques, but still yields divergent findings. This may be due in part to MCI's broad definition, encompassing heterogeneous cognitive domains, only some of which are affected in Parkinson's disease. Most domains falling under the MCI umbrella include fronto-dependent executive functions, whereas others, notably learning, rely on the basal ganglia. Given the deterioration of the nigrostriatal dopaminergic system in Parkinson's disease, it has been the prime target of PD-MCI investigation. By testing well defined cognitive deficits in Parkinson's disease, distinct functions can be attributed to specific neural systems, overcoming conflicting results on PD-MCI. Apart from dopamine, other systems such as the neurovascular or noradrenergic systems are affected in Parkinson's disease. These factors may be at the basis of specific facets of PD-MCI for which dopaminergic involvement has not been conclusive. Finally, the impact of both dopaminergic and noradrenergic deficiency on motivational states in Parkinson's disease is examined in light of a plausible link between apathy and cognitive deficits.

Keywords

apathy, learning, neuroimaging, noradrenaline, Parkinson's

INTRODUCTION

Mild cognitive impairment (MCI) refers to cognitive decline that does not meet the clinical criteria for dementia. MCI is a widely reported comorbid factor in Parkinson's disease [1]. Whereas MCI can predict dementia in Parkinson's disease [2], MCI assessment accuracy based on cognitive batteries is relatively poor [3,4]. As such, neuroimaging techniques are now being used to identify its neural signature. MCI is a profile that arises in many populations, including the aged and Alzheimer's patients. Since its cause is unknown, it is unclear whether the same mechanism prompts its emergence in different diseases [5]. Further, MCI incorporates deficits across

heterogeneous cognitive domains [6], most related to fronto-dependent executive function [7], but at times inclusive of learning processes [8,9]. Studies

^aBrain Mind Institute, Laboratory of Behavioral Genetics – LGC, EPFL SV BMI LGC, AAB 2 01 (Bâtiment AAB), Lausanne and ^bGeneva Finance Research Institute (GFRI) & Interfaculty Center for Affective Sciences (CISA), University of Geneva, Geneva, Switzerland

Correspondence to Leyla Loued-Khenissi, Brain Mind Institute, Laboratory of Behavioral Genetics – LGC, EPFL SV BMI LGC, AAB 2 01 (Bâtiment AAB), Station 19, CH-1015 Lausanne, Switzerland. Tel: +41 216939693; e-mail: leyla.loued-khenissi@epfl.ch

Curr Opin Neurol 2015, 28:344–350

DOI:10.1097/WCO.0000000000000218

KEY POINTS

- Neuroimaging research supports a diffuse neural marker for PD-MCI with a neurovascular basis emerging as a strong candidate in its cause.
- Though difficult to image, the locus coeruleus noradrenaline complex, given its widespread cortical projections, chemical link to dopamine, and marked deterioration in Parkinson's disease, should be investigated as a strong contributor to Parkinson's disease behavioral impairments.
- The widespread emergence of apathy in Parkinson's disease, supported by behavioral and EEG markers, should be investigated in relation to a dopaminergic or noradrenergic neural basis and known cognitive profiles in Parkinson's disease.

on Parkinson's disease MCI (PD-MCI) have been inconclusive with regards to the domains affected and dopaminergic involvement. PD-MCI is thought to be a consequence of cortical dopaminergic changes in Parkinson's disease arising from compromised fronto-striatal circuits, notably the mesocortical and nigrostriatal loops (see Fig. 1) [10]. However, evidence of changes in prefrontal dopamine is equivocal [8,11,12]. Since Parkinson's disease

is marked by nigrostriatal dopaminergic loss, basal-ganglia-dependent learning processes have been studied extensively [13,14], with a particular focus on the striatum. It is generally thought that the ventral and dorsal striatum play distinct functional roles, which are only partially understood to date. The dopamine overdose hypothesis may explain observed selective impairment in Parkinson's disease patients on dopaminergic replacement therapy (PDON) relative to unmedicated patients (PDOFF) [15]. In early Parkinson's disease, the dorsal striatum displays extensive degeneration, whereas the ventral striatum remains preserved. Dopaminergic medication relieves dorso-related motor symptoms, but may overdose a functional ventral striatum, prompting selective behavioral impairments such as impulse control disorders (ICDs). PDOFF populations thus offer a window into dorsal striatum-dependent functions. Further questions regarding cognitive deficits converge on recent recognition that apathy is a common symptom in early Parkinson's disease [16]. Questions on apathy's behavioral impact and its neural basis remain open. Dopamine has long been the focus of Parkinson's disease research; however, disease characteristics extend beyond the dopaminergic system, suggesting other factors may drive observed deficits.

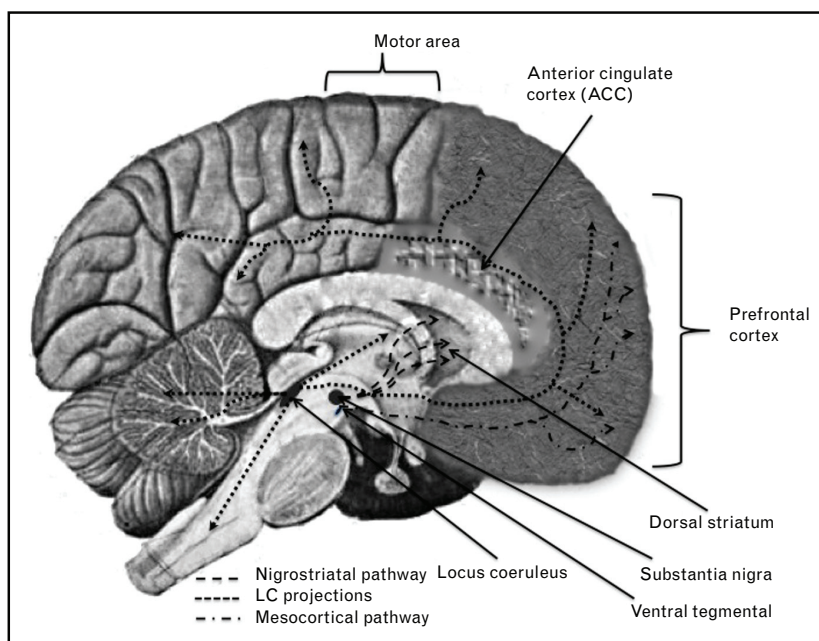


FIGURE 1. An overview of brain regions implicated in MCI, Parkinson's disease, and the locus coeruleus noradrenergic system. MCI test batteries primarily include executive function, which is traditionally linked to the prefrontal cortex (as well as the anterior cingulate cortex). PD is characterized by a damaged nigrostriatal pathway that starts in the substantia nigra and projects to the dorsal striatum (dashed lines). Noradrenergic projections start from the locus coeruleus and project out to the cortex and the cerebellum (dotted lines). LC, locus coeruleus; MCI, mild cognitive impairment; PD, Parkinson's disease.

NEUROIMAGING MILD COGNITIVE IMPAIRMENT

The search for biomarkers of PD-MCI has employed various neuroimaging measures including functional, structural, and diffusion measures. Although this endeavor has yielded a number of potential biomarkers, the evidence has simultaneously generated a less ordered view of PD-MCI signatures and causes. Cognitive scores in Parkinson's disease patients correlate with dorso-fronto parietal connectivity; inhibited subcortical primary sensory activation; and preserved nigrostriatal pathways in resting-state functional magnetic resonance imaging (fMRI), but not with presynaptic dopaminergic uptake [17]. Atrophy in various cortical regions is associated with neuropsychiatric symptoms [18], as well as MCI in some studies [19,20,21,22,23], though others found no such differences in PD-MCI compared to Parkinson's disease without MCI [24,25]. The expected effect of MCI on the subcortical regions is even less clear, though hippocampal atrophy was found to predict conversion to PD-MCI and to dementia from PD-MCI in a longitudinal study [26]. Research has also investigated white matter differences, which can indicate neurovascular abnormalities [27]. White matter hyperintensities were found to predict cognitive decline [24], and several recent studies reported white matter abnormalities in PD-MCI [26,28,29]. Early Parkinson's disease patients specifically show evidence of atherosclerosis alongside white matter hyperintensities – factors that lead to microvascular injury and possible cognitive decline [30]. Interestingly, both orthostatic and prandial hypotension is a sign of noradrenergic disturbance [13,31], a neurotransmitter which is affected early in Parkinson's disease [32]. The evidence suggests PD-MCI's neural footprint remains difficult to delineate even with various imaging measures, though neurovascular abnormalities emerge as strong causal candidates. Neurovascular differences indeed correlate with MCI in other patient populations [33–36]. The studies above do not show a distinct link between dopamine and PD-MCI, but they do yield an array of diffuse neural correlates, which may reflect the fuzzy nature of MCI's behavioral characterization.

COGNITIVE FLEXIBILITY IN PARKINSON'S DISEASE

Parkinson's disease patients display executive dysfunction, but evidence on specific domains affected remains murky [37,38]. One persistent finding is set-shifting impairment in Parkinson's disease patients [1,39]. Cognitive flexibility appears to rely on the

dorsal striatum [40] and medication response correlates with improved task switching in Parkinson's disease, further supporting the dorsal striatum's role in cognitive flexibility [41]. One fMRI study in PD-OFF patients found no impairment in set-shifting, but did reveal atypical task-related activation in the cortex, suggesting compensatory anomalous cortical activity inhibits behavioral impairment [42]. Previous studies produced conflicting results on medication's remedial effects on set-shifting impairment, but the studies above support striatal dopamine's role in cognitive flexibility, as well as a cortical up-regulation in early stages of the disease, perhaps masking striatal deficiencies.

LEARNING DEFICITS IN PARKINSON'S DISEASE

Reinforcement learning has been extensively studied in Parkinson's disease [43–45] to support models cast within a basal ganglia dopaminergic framework. When controlling for medication effects, studies reveal deficits in learning from trial-by-trial feedback [46], a hallmark of implicit learning [47]. Indeed, a meta-analysis found Parkinson's disease patients to be significantly impaired in implicit learning across 27 studies using the serial reaction time task [48]. While implicit learning is thought to depend on the basal ganglia, explicit, declarative learning relies on the hippocampus and medial temporal lobe [49]. The interplay between the two systems has yet to be defined [50], but a selective impairment in Parkinson's disease would suggest implicit learning occurs in the dorsal striatum. Most tasks measuring one type of learning versus another rely on both mechanisms [51], but recent evidence suggests explicit and implicit learning can be dissociated by manipulating a task's feedback structure (delayed versus discrete) [52]. An [¹¹C] raclopride PET study showed striatal (accumbens) D2 release accompanied learning from discrete feedback in a probabilistic classification task [53]. Further, learning from delayed feedback activates the hippocampus, whereas learning from immediate feedback engages the striatum [48]. A study investigated competing learning mechanisms in Parkinson's disease, with two initial tasks that tested novel tool features (explicit) and novel tool skill (implicit), and a follow-up task that assessed both learning acquisitions 3 weeks later. Patients did not differ from controls in either the initial learning session or on knowledge of novel tool attributes in the follow-up session; however, the Parkinson's disease group did not retain skilled tool use [54]. Two more recent studies highlight differences in retention for Parkinson's disease patients. An initial test

of sequence learning was not affected in Parkinson's disease, though patient retention a week later was [55]. Patients tested on an implicit learning sequence task performed as well as healthy controls in a first block, but not in a second block [56]. Further, no differences were found in an implicit learning task of semantic categorization between healthy controls and Parkinson's disease patients [57,58]. These divergent findings call into question the impairment of implicit learning in Parkinson's disease, as well as its dependence on the dorsal striatum. It has been posited that dorsal striatal dopaminergic signals are necessary for performance, or action-selection, rather than learning *per se* [14,43,59,60]. These two roles may be specific to distinct striatal regions, but action-selection is often used to determine learning. Thus, recent studies have examined the functional dissociation of the dorsal and ventral striatum in relation to learning acquisition (or memory encoding) and action-selection (or memory retrieval). An fMRI study in the healthy controls investigated stimulus-response learning with feedback, followed by a session that assessed how well associations were learned. Activation in the ventral striatum was confined to the learning session, whereas activation in the dorsal striatum emerged in the second session, where associations had already been learned and the task demand was appropriate response selection [61]. A novel fMRI study dissociated dopamine's roles in anticipation and reward to determine whether placebo would be as effective as dopaminergic replacement therapy in Parkinson's disease reward learning. Both placebo and medication groups exhibited learning signals in the ventral striatum [62]. Vo *et al.* [63], in 2014, found PDOFF patients learned stimulus-response associations as well as controls, whereas PDON patients were impaired. Further, PDOFF patients outperformed controls and PDON patients, supporting the hypothesis that cortical D1 is up-regulated in Parkinson's disease [64,65]. The studies listed above support the dorsal striatum's role in action-selection, but a recent case study of a patient suffering bilateral damage to the dorsal striatum showed specific impairment in learning stimulus values and not action values [66]. The evidence suggests a different frame within which to study functions specific to the ventral and dorsal striatum. Notably, learning's dependence on the ventral striatum and action-selection's reliance on the dorsal striatum merit closer scrutiny in future studies.

APATHY

Apathy is a common, early symptom in Parkinson's disease that predicts MCI and dementia [67]. Apathy

may significantly impact processes requiring motivation, such as action-selection and cognitive task performance, if not cognition itself. While apathy's neural correlates remain unknown, the search for a neural mechanism of Parkinson's disease apathy focuses on the dopaminergic system. Compared to healthy and Parkinson's disease controls, apathetic patients showed a reduction in left limbic striatal and frontal connectivities in resting-state fMRI, though apathy scores showed no correlation with structural differences [68^{***}]. An fMRI study [69] examined dopaminergic medication effects during an emotional Stroop task in PDON and PDOFF patients, and found that when presented with negative Stroop stimuli, PDOFF patients had higher apathy scores, decreased fear recognition, and reduced anterior cingulate cortex (ACC) activation. While ACC activation was recovered with medication, it is interesting to note that the cingulate receives projections from the locus coeruleus noradrenaline (LC-NE) system [70]. Though Parkinson's disease-related apathy is an early symptom, it can also emerge following deep brain stimulation (DBS) implantation as a suspected consequence of dopaminergic medication washout. Increased apathy after DBS correlated with reduced right ventral striatal activity in a PET study [71]. Further, dopamine-resistant apathy correlated with nucleus accumbens atrophy [72[†]]. Like MCI, apathy in Parkinson's disease is primarily assessed via psychometric scale [73], but electroencephalogram (EEG) studies have yielded compelling behavioral and physiological consequences of Parkinson's disease apathy. An event related potential (ERP) study measured feedback-related negativity (FRN) in response to gains and losses. Apathetic patients showed a reduced difference between FRN for losses and FRN for gains when compared to Parkinson's disease patients and healthy controls [74]. In a similar vein, an EEG study examined differences in ERPs between Parkinson's disease patients and healthy controls during the Iowa Gambling Task – a task of decision-making under ambiguity [75]. ERP for gains differed from ERP for losses in the healthy controls, as expected, but no differences emerged in Parkinson's disease patients. Another study reported a blunted P3 signal in apathetic PDOFF patients [76] (a P3 signal arises upon encounter of a salient stimulus). Furthermore, Parkinson's disease patients did not display the Von Restorff effect, where novelty enhances stimulus recall [77[†]]. In the same study, the P3 signal was larger for novel stimuli in healthy controls relative to patients, irrespective of medication status, implicating a nondopaminergic system. A potential candidate is the noradrenergic system whose activation has been linked to the P3 signal via pupillometry

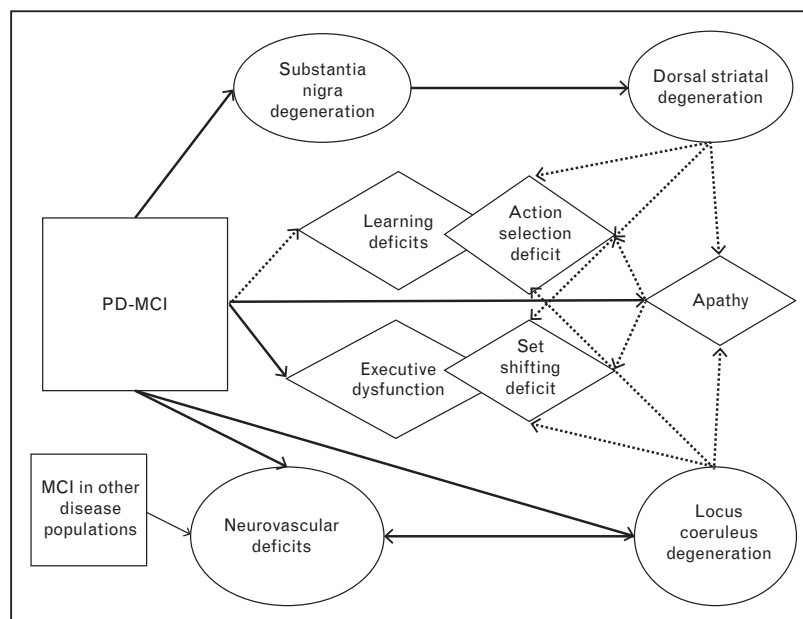


FIGURE 2. An overview of recent findings related to PD-MCI. Circles represent known neural correlates; diamonds represent putative behavioral symptoms. Solid lines indicate known relationships while dashed lines represent possible links between factors. MCI, mild cognitive impairment; PD, Parkinson's disease.

studies [78,79]. And while the neural correlates to apathy above implicate the ventral striatum, it should be noted that the region receives projections from the LC in addition to its dopaminergic projections. The dearth in research on apathy's link to observed cognitive deficits provides an avenue of investigation into the motivational factors of cognitive performance.

THE NORADRENERGIC SYSTEM

Parkinson's disease research has centered on the dopaminergic system; however, many of the observed cognitive deficits may also be linked to a pathological noradrenergic system in Parkinson's disease patients. Post mortem analysis of Parkinsonian brains reveals Lewy body accumulation in the LC [80], as well as a reduction in frontal norepinephrine and serotonin, but not dopamine [81,82]. LC degeneration precedes nigrostriatal neural loss [32]. Dopamine and noradrenaline are both tyrosine-derived catecholamines; their interaction may be of particular interest [83], given that the LC-NE system has widespread cortical projections (Fig. 2) [84]; noradrenaline may protect against dopaminergic deficiency [85]; and noradrenaline modulates dopaminergic activation [86]. Indeed, recent research in learning and decision-making has already moved beyond the bounds of the basal ganglia to scrutinize LC-NE's contribution to these functions [87,88]. As such, there is now compelling

evidence that LC-NE degeneration in Parkinson's disease may contribute to PD-MCI [89]. Specifically, cognitive inflexibility in early Parkinson's disease could reflect early dysfunction of the LC-NE system [90]. Adaptive gain theory [91] describes LC neurons' dual firing modes: a phasic mode that signals exploitation, and a tonic mode that prompts exploration. A compromised LC-NE system could lead to decreased tonic noradrenergic transmission, inhibiting flexibility and enhancing perseveration [90]. A dysfunctional LC-NE system could further prevent patients from registering salient signals demanding action, which may explain action-selection deficits and contribute to Parkinson's disease-related apathy. Neuroimaging evidence of LC-NE involvement in Parkinson's disease has been sparse to date, due to the difficulty inherent in imaging a small, brainstem region [92], but among the many neurotransmitter systems affected in Parkinson's disease [93,94], noradrenaline's characteristics stand out as markedly relevant to the study of cognitive function.

CONCLUSION

Mild cognitive impairment in Parkinson's disease is not confined to dopaminergic deficits *per se*, behooving us to consider nondopaminergic mechanisms for its emergence. Two lines of investigation merit closer future inspection: the role apathy plays in observed behavioral deficits and the role of the LC-NE's

influence on learning, apathy, and distinct measures of MCI in a Parkinson's disease model.

Acknowledgements

None.

Financial support and sponsorship

Funding: This study was supported in part by the Swiss National Science Fund.

Conflicts of interest

There are no conflicts of interest.

REFERENCES AND RECOMMENDED READING

Papers of particular interest, published within the annual period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Kudlicka A, Clare L, Hindle JV. Executive functions in Parkinson's disease: systematic review and meta-analysis. *Movement Disord* 2011; 26:2305–2315.
2. Pedersen KF, Larsen JP, Tysnes OB, Alves G. Prognosis of mild cognitive impairment in early Parkinson disease: the Norwegian ParkWest study. *J Am Med Assoc Neurol* 2013; 70:580–586.
3. Chou KL, Lenhart A, Koeppe RA, Bohnen NI. Abnormal MoCA and normal range MMSE scores in Parkinson disease without dementia: cognitive and neurochemical correlates. *Parkinsonism Relat Disord* 2014; 20:1076–1080.
4. Hu M, Szeewczyk-Królikowski K, Tomlinson P, et al. Predictors of cognitive impairment in an early stage Parkinson's disease cohort. *Movement Disord* 2014; 29:351–359.
5. Petersen RC, Caracciolo B, Brayne C, et al. Mild cognitive impairment: a concept in evolution. *J Intern Med* 2014; 275:214–228.
6. Barker RA, Williams-Gray CH. Mild cognitive impairment and Parkinson's disease: something to remember. *J Parkinson's Dis* 2014; 4:651–656.
7. Alvarez JA, Emory E. Executive function and the frontal lobes: a meta-analytic review. *Neuropsychol Rev* 2006; 16:17–42.
8. Robbins TW, Cools R. Cognitive deficits in Parkinson's disease: a cognitive neuroscience perspective. *Movement Disord* 2014; 29:597–607.
9. Kehagia AA, Barker RA, Robbins TW. Neuropsychological and clinical heterogeneity of cognitive impairment and dementia in patients with Parkinson's disease. *Lancet Neurol* 2010; 9:1200–1213.
10. Pellecchia MT, Picillo M, Santangelo G, et al. Cognitive performances and DAT imaging in early Parkinson's disease with mild cognitive impairment: a preliminary study. *Acta Neurol Scand* 2015; 131:275–281.
11. Cropley VL, Fujita M, Bara-Jimenez W, et al. Preand postsynaptic dopamine imaging and its relation with frontostriatal cognitive function in Parkinson disease: PET studies with [11 C] NNC 112 and [18 F] FDOPA. *Psychiatry Res Neuroimag* 2008; 163:171–182.
12. Sawamoto N, Piccini P, Hotton G, et al. Cognitive deficits and striato-frontal dopamine release in Parkinson's disease. *Brain* 2008; 131:1294–1302.
13. Lewis SJ, Pavese N, Rivero-Bosch M, et al. Brain monoamine systems in multiple system atrophy: a positron emission tomography study. *Neurobiol Dis* 2012; 46:130–136.
14. Eisenegger C, Naef M, Linssen A, et al. Role of dopamine D2 receptors in human reinforcement learning. *Neuropsychopharmacology* 2014; 39:2366–2375.
15. Cools R. Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease. *Neurosci Biobehav Rev* 2006; 30:1–23.
16. Santangelo G, Vitale C, Trojano L, et al. Relationship between apathy and cognitive dysfunctions in de novo untreated Parkinson's disease: a prospective longitudinal study. *Eur J Neurol* 2015; 22:253–260.
17. Lebedev AV, Westman E, Simmons A, et al. Large-scale resting state network correlates of cognitive impairment in Parkinson's disease and related dopaminergic deficits. *Front Syst Neurosci* 2014; 8:45.
18. O'Callaghan C, Shine JM, Lewis SJG, Hornberger M. Neuropsychiatric symptoms in Parkinson's disease: fronto-striatal atrophy contributions. *Parkinsonism Relat Disord* 2014; 20:867–872.
19. Noh SW, Han YH, Mun CW, et al. Analysis among cognitive profiles and gray matter volume in newly diagnosed Parkinson's disease with mild cognitive impairment. *J Neurol Sci* 2014; 347 (1–2):210–213.
20. Mak E, Zhou J, Tan LC, et al. Cognitive deficits in mild Parkinson's disease are associated with distinct areas of grey matter atrophy. *J Neurol Neurosurg Psychiatry* 2014; 85:576–580.
21. Segura B, Baggio HC, Marti MJ, et al. Cortical thinning associated with mild cognitive impairment in Parkinson's disease. *Mov Disord* 2014; 29:1495–1503.
- MCI correlated with thinning of the tempoparietal cortex.
22. Koshimori Y, Segura B, Christopher L, et al. Imaging changes associated with cognitive abnormalities in Parkinson's disease. *Brain Struct Funct* 2014. [Epub ahead of print]
- MCI correlated with thinning of the left dorsolateral superior frontal gyrus.
23. Hanganu A, Bedetti C, Degroot C, et al. Mild cognitive impairment is linked with faster rate of cortical thinning in patients with Parkinson's disease longitudinally. *Brain* 2014; 137:1120–1129.
24. Agosta F, Canu E, Stefanova E, et al. Mild cognitive impairment in Parkinson's disease is associated with a distributed pattern of brain white matter damage. *Hum Brain Mapp* 2014; 35:1921–1929.
25. Yarnall AJ, Breen DP, Duncan GW, et al. Characterizing mild cognitive impairment in incident Parkinson disease: the ICICLE-PD study. *Neurology* 2014; 82:308–316.
26. Kandiah N, Zainal NH, Narasimhalu K, et al. Hippocampal volume and white matter disease in the prediction of dementia in Parkinson's disease. *Parkinsonism Relat Disord* 2014; 20:1203–1208.
27. Chutinet A, Rost NS. White matter disease as a biomarker for long-term cerebrovascular disease and dementia. *Curr Treatment Options Cardiovasc Med* 2014; 16:1–12.
28. Sunwoo MK, Jeon S, Ham JH, et al. The burden of white matter hyperintensities is a predictor of progressive mild cognitive impairment in patients with Parkinson's disease. *Eur J Neurol* 2014; 21:922–950.
29. Zheng Z, Shemmassian S, Wijekoon C, et al. DTI correlates of distinct cognitive impairments in Parkinson's disease. *Hum Brain Mapp* 2014; 35:1325–1333.
30. Kim JS, Oh YS, Lee KS, et al. Carotid artery thickening and neurocirculatory abnormalities in de novo Parkinson disease. *J Neural Transm* 2014; 121:1259–1268.
- Parkinson's disease patients showed supine hypertension as well as orthostatic hypotension, which are indicators of atherosclerosis.
31. Kaufmann H, Goldstein DS. Autonomic dysfunction in Parkinson disease. *Handb Clin Neurol* 2013; 117:259–278.
32. Zarow C, Lyness SA, Mortimer JA, Chui HC. Neuronal loss is greater in the locus coeruleus than nucleus basalis and substantia nigra in Alzheimer and Parkinson diseases. *Arch Neurol* 2003; 60:337–341.
33. O'Brien T, Erkinjuntti J, Reisberg T, et al. Vascular cognitive impairment. *Lancet Neurol* 2003; 2:89–98.
34. Papma JM, de Groot M, de Koning I, et al. Cerebral small vessel disease affects white matter microstructure in mild cognitive impairment. *Hum Brain Mapp* 2014; 35:2836–2851.
35. Boespflug EL, Eliassen J, Welge J, Krikorian R. Associative learning and regional white matter deficits in mild cognitive impairment. *J Alzheimers Dis* 2014; 41:421–430.
36. Chuang YF, Eldreth D, Erickson KI, et al. Cardiovascular risks and brain function: a functional magnetic resonance imaging study of executive function in older adults. *Neurobiol Aging* 2014; 35:1396–1403.
37. Dirnberger G, Jahanshahi M. Executive dysfunction in Parkinson's disease: a review. *J Neuropsychol* 2013; 7:193–224.
38. Kehagia AA, Barker RA, Robbins TW. Cognitive impairment in Parkinson's disease: the dual syndrome hypothesis. *Neurodegener Dis* 2013; 11:79–92.
39. Nagano-Saito A, Habak C, Mejia-Constain B, et al. Effect of mild cognitive impairment on the patterns of neural activity in early Parkinson's disease. *Neurobiol Aging* 2014; 35:223–231.
40. MacDonald AA, Seergobin KN, Tamjeedi R, et al. Examining dorsal striatum in cognitive effort using Parkinson's disease and fMRI. *Ann Clin Translat Neurol* 2014; 1:390–400.
- The team had previously found cognitive flexibility deficits relate to the dorsal striatum, but results were challenged on the basis that cognitive flexibility was conflated with cognitive effort. The team repeated the experiment, this time testing cognitive effort alone.
41. Aarts E, Nusslein AA, Smittenaar P, et al. Greater striatal responses to medication in Parkinson's disease are associated with better task-switching but worse reward performance. *Neuropsychologia* 2014; 62:390–397.
42. Gerrits NJ, van der Werf YD, Verhoef KM, et al. Compensatory fronto-parietal hyperactivation during set-shifting in unmedicated patients with Parkinson's disease. *Neuropsychologia* 2015; 68:107–116.
43. Shiner T, Seymour B, Wunderlich K, et al. Dopamine and performance in a reinforcement learning task: evidence from Parkinson's disease. *Brain* 2012; 135:1871–1883.
44. Maia TV, Frank MJ. From reinforcement learning models to psychiatric and neurological disorders. *Nature Neurosci* 2011; 14:154–162.
45. Frank MJ, Seeberger LC, O'Reilly RC. By carrot or by stick: cognitive reinforcement learning in Parkinsonism. *Science* 2004; 306:1940–1943.
46. Ryterska A, Jahanshahi M, Osman M. What are people with Parkinson's disease really impaired on when it comes to making decisions? A meta-analysis of the evidence. *Neurosci Biobehav Rev* 2013; 37:2836–2846.

47. Maddox WT, Ashby FG, Bohil CJ. Delayed feedback effects on rule-based and information-integration category learning. *J Exp Psychol Learn Memory Cogn* 2003; 29:650–662.
 48. Clark GM, Lum JA, Ullman MT. A meta-analysis and meta-regression of serial reaction time task performance in Parkinson's disease. *Neuropsychology* 2014; 28:945–958.
 49. Foerde K, Race E, Verfaellie M, Shohamy D. A role for the medial temporal lobe in feedback-driven learning: evidence from amnesia. *J Neurosci* 2013; 33:5698–5704.
 50. Doll BB, Shohamy D, Daw ND. Multiple memory systems as substrates for multiple decision systems. *Neurobiol Learn Memory* 2014; 117:4–13.
 51. Foerde K, Shohamy D. The role of the basal ganglia in learning and memory: insight from Parkinson's disease. *Neurobiol Learn Memory* 2011; 96:624–636.
 52. Smith JD, Boomer J, Zakrzewski AC, *et al.* Deferred feedback sharply dissociates implicit and explicit category learning. *Psychol Sci* 2013; 25:447–457.
 53. Wilkinson L, Tai YF, Lin CS, *et al.* Probabilistic classification learning with corrective feedback is associated with in vivo striatal dopamine release in the ventral striatum, while learning without feedback is not. *Hum Brain Mapp* 2014; 35:5106–5115.
- [¹¹C]Raclopride is a radioligand that binds preferentially to striatal D2 receptors. Rather than approximate dopaminergic activity via BOLD measures or other proxies, D2 activity can be directly assessed here and thus offers compelling evidence of learning confined to the ventral striatum.
54. Roy S, Park NW, Roy EA, Almeida QJ. Interaction of memory systems during acquisition of tool knowledge and skills in Parkinson's disease. *Neuropsychologia* 2015; 66:55–66.
- The task required patients learn tool attributes (declarative, explicit memory dependent); tool use (implicit, procedural learning); and display skilled tool use (assumed to require both types of learning) 3 weeks later. No group differences emerged with regards to tool attribute knowledge, suggesting patients are specifically impaired in implicit learning.
55. Smits-Bandstra S, Gracco V. Retention of implicit sequence learning in persons who stutter and persons with Parkinson's disease. *J Mot Behav* 2014; 4:1–18.
 56. Gamble KR, Cummings TJ Jr, Lo SE, Ghosh PT, *et al.* Implicit sequence learning in people with Parkinson's disease. *Front Hum Neurosci* 2014; 8:563.
 57. Arroyo-Anlló EM, Ingrand P, Neau JP, Gil R. Procedural learning of semantic categorization in Parkinson's disease. *J Alzheimers Dis* 2014; 45:205–216.
 58. Filoteo JV, Maddox WT. Procedural-based category learning in patients with Parkinson's disease: impact of category number and category continuity. *Front Syst Neurosci* 2014; 8:14.
 59. Leventhal DK, Stoetznr CR, Abraham R, *et al.* Dissociable effects of dopamine on learning and performance within sensorimotor striatum. *Basal Ganglia* 2014; 4:43–54.
 60. Smittenaar P, Chase HW, Aarts E, *et al.* Decomposing effects of dopaminergic medication in Parkinson's disease on probabilistic action selection: learning or performance? *Eur J Neurosci* 2012; 35:1144–1151.
 61. Hiebert NM, Vo A, Hampshire A, *et al.* Striatum in stimulus-response learning via feedback and in decision making. *Neuroimage* 2014; 101:448–457.
 62. Schmidt L, Braun EK, Wager TD, Shohamy D. Mind matters: placebo enhances reward learning in Parkinson's disease. *Nat Neurosci* 2014; 17:1793–1797.
 63. Vo A, Hiebert NM, Seergobin KN, *et al.* Dopaminergic medication impairs feedback-based stimulus-response learning but not response selection in Parkinson's disease. *Front Hum Neurosci* 2014; 8:784.
 64. Cools R, Miyakawa A, Sheridan M, D'Esposito M. Enhanced frontal function in Parkinson's disease. *Brain* 2010; 133:225–233.
 65. Nagano-Saito A, Kato T, Arahata Y, *et al.* Cognitive and motor-related regions in Parkinson's disease: FDOPA and FDG PET studies. *Neuroimage* 2004; 22:553–561.
 66. Vo K, Rutledge RB, Chatterjee A, Kable JW. Dorsal striatum is necessary for stimulus-value but not action-value learning in humans. *Brain* 2014; 137:3129–3135.
 67. Martínez-Horta S, Pagonabarraga J, Fernández de Bobadilla R, *et al.* Apathy in Parkinson's disease: more than just executive dysfunction. *J Int Neuropsychol Soc* 2013; 19:571–582.
 68. Baggio HC, Segura B, Garrido-Millán JL, *et al.* Resting-state frontostriatal functional connectivity in Parkinson's disease-related apathy. *Mov Disord* 2015; 30:671–679.
- This study performed volumetric and shape analyses on subcortical regions and voxel based morphometry on fronto-subcortical regions.
69. Fleury V, Cousin E, Czernecki V, *et al.* Dopaminergic modulation of emotional conflict in Parkinson's disease. *Front Aging Neurosci* 2014; 6:164.
 70. Tervo DG, Proskurin M, Manakov M, *et al.* Behavioral variability through stochastic choice and its gating by anterior cingulate cortex. *Cell* 2014; 159:21–32.
 71. Robert GH, Le Jeune F, Lozachmeur C, *et al.* Preoperative factors of apathy in subthalamic stimulated Parkinson disease A PET study. *Neurology* 2014; 83:1620–1626.
 72. Carrière N, Besson P, Dujardin K, *et al.* Apathy in Parkinson's disease is associated with nucleus accumbens atrophy: a magnetic resonance imaging shape analysis. *Mov Disord* 2014; 29:897–903.
- Other neuroimaging data collected included shape analysis, DTI, and cortical thickness measures none of which correlated with apathy.
73. Santangelo G, Barone P, Cuoco S, *et al.* Apathy in untreated, de novo patients with Parkinson's disease: validation study of Apathy Evaluation Scale. *J Neurol* 2014; 261:2319–2328.
 74. Martínez-Horta S, Riba J, de Bobadilla RF, *et al.* Apathy in Parkinson's disease: neurophysiological evidence of impaired incentive processing. *J Neurosci* 2014; 34:5918–5926.
 75. Mapelli D, Di Rosa E, Cavalletti M, *et al.* Decision and dopaminergic system: an ERPs study of Iowa gambling task in Parkinson's disease. *Front Psychol* 2014; 5:684.
 76. Mathis S, Neau JP, Pluchon C, *et al.* Apathy in Parkinson's disease: an electrophysiological study. *Neurol Res Int* 2014; 2014:290513.
 77. Schomaker J, Berendse HW, Foncke EM, *et al.* Novelty processing and memory formation in Parkinson's disease. *Neuropsychologia* 2014; 62:124–136.
- Healthy controls exhibited the Von Restorff effect, where novelty modulates memory. At encoding, words were either presented in novel or standard font. While healthy controls had a higher rate of recall for words in the novel font, Parkinson's disease patients had a higher rate of recall for words in the standard font.
78. Nieuwenhuis S, Aston-Jones G, Cohen JD. Decision making, the P3, and the locus coeruleus: norepinephrine system. *Psychol Bull* 2005; 131:510–532.
 79. Murphy PR, Robertson IH, Balsters JH, O'Connell RG. Pupillometry and P3 index the locus coeruleus–noradrenergic arousal function in humans. *Psychophysiology* 2011; 48:1532–1543.
 80. Rommelfanger KS, Weinshenker D. Norepinephrine: the redheaded stepchild of Parkinson's disease. *Biochem Pharmacol* 2007; 74:177–190.
 81. Nayyar T, Bubser M, Ferguson MC, *et al.* Cortical serotonin and norepinephrine denervation in parkinsonism: preferential loss of the beaded serotonin innervation. *Eur J Neurosci* 2009; 30:207–216.
 82. Goldstein DS, Sullivan P, Holmes C, *et al.* Catechols in postmortem brain of patients with Parkinson disease. *Eur J Neurol* 2011; 18:703–710.
 83. Pietrajtis K, Sara SJ, Logothetis NK, Eschenko O. Spike timing among neurons recorded simultaneously in locus coeruleus, ventral tegmental area and frontal cortex during somatosensory stimulation: LC leads! In 7th Forum of European Neuroscience, 3–7 July 2010; Amsterdam, The Netherlands.
 84. Espay AJ, LeWitt PA, Kaufmann H. Norepinephrine deficiency in Parkinson's disease: the case for noradrenergic enhancement. *Mov Disord* 2014; 29:1710–1719.
 85. Szot P, Franklin A, Sikkema C, *et al.* Sequential loss of LC noradrenergic and dopaminergic neurons results in a correlation of dopaminergic neuronal number to striatal dopamine concentration. *Front Pharmacol* 2012; 3:184.
 86. Szot P, Knight L, Franklin A, *et al.* Lesioning noradrenergic neurons of the locus coeruleus in C57Bl/6 mice with unilateral 6-hydroxydopamine injection, to assess molecular, electrophysiological and biochemical changes in noradrenergic signaling. *Neuroscience* 2012; 216:143–157.
 87. Preuschoff K, Marius't Hart B, Einhäuser W. Pupil dilation signals surprise: evidence for noradrenaline's role in decision making. *Front Neurosci* 2011; 5:115.
 88. Sara SJ, Bouret S. Orienting and reorienting: the locus coeruleus mediates cognition through arousal. *Neuron* 2012; 76:130–141.
 89. Del Tredici K, Braak H. Dysfunction of the locus coeruleus–norepinephrine system and related circuitry in Parkinson's disease-related dementia. *J Neurol Neurosurg Psychiatry* 2013; 84:774–783.
 90. Vazey EM, Aston-Jones G. The emerging role of norepinephrine in cognitive dysfunctions of Parkinson's disease. *Front Behav Neurosci* 2012; 6:48.
 91. Aston-Jones G, Cohen JD. An integrative theory of locus coeruleus–norepinephrine function: adaptive gain and optimal performance. *Annu Rev Neurosci* 2005; 28:403–450.
 92. Chen X, Huddleston DE, Langley J, *et al.* Simultaneous imaging of locus coeruleus and substantia nigra with a quantitative neuromelanin MRI approach. *Magn Reson Imag* 2014; 32:1301–1306.
 93. Nagatsu T, Sawada M. Biochemistry of postmortem brains in Parkinson's disease: historical overview and future prospects. Vienna: Springer; 2007. pp. 113–120.
 94. Halliday GM, Leverenz JB, Schneider JS, Adler CH. The neurobiological basis of cognitive impairment in Parkinson's disease. *Mov Disord* 2014; 29:634–650.

Curriculum Vitae

LEYLA LOUED-KHENISSI

Nationality: USA, Tunisia **Languages:** French, English, Arabic

Education

- 2014-2018** **Ecole Polytechnique Fédérale de Lausanne**
Brain Mind Institute
Doctoral Assistant, Neuroscience (EDNE)
Thesis Title: “Belief as a Wise Wager : The Neural Representation of Uncertainty, Surprise and Confidence Across Cognitive and Perceptual Domains ”
- 2007-2010 Open University, Milton Keynes, United Kingdom
M.Sc. Research Methods in Psychology
Dissertation Title: “Could Envy Lie Behind Egalitarian Motives?”
- 1999-2003 Concordia University, Montreal, Canada
B.A. Psychology
- 1993-1996 Lycee Rochambeau, Bethesda, Maryland
Section S
American High School Diploma

Research Experience

- 2012 *Research Internship, Brain Mind Institute, EPFL, PI: Dr. Carmen Sandi.* Investigation of social hierarchy formation and socially relevant emotions in humans using eye tracking. Effects of stress on economic behavior.
- 2007 *Research Internship, Department of Neurology USZ, PI: Dr. Peter Brugger* Investigation of Dermo-optical perception, a study replication of an experiment originally conducted in 1966
- 2004-2005 *Research Associate, UCLA, Laboratory of Neuromolecular Imaging. PI: Prof. E. D. London* Neuroimaging of methamphetamine abuse, PMS and OCD using FDG. Duties included running PET scans, acquiring, preprocessing and analyzing image data, using MRICro, SPM99, SPM2, and MedX. Studies focused on smokers, methamphetamine abusers, PMDD sufferers and OCD patients.
- 2001-2002 *Data Analyst, Functional Brain Mapping Division, New York State*209

Psychiatric Institute. PI: Mark Laruelle. Neuroimaging using PET and some SPECT in substance abuse, autism, schizophrenia and OCD. Duties included pre-processing and analysis of image data, using SPM and MedX. Image data acquired using C11 ligands; data analyzed included NNC, DASB, MCN and RAC scans.

Writing and Editing Experience

- Freelance editor, English Editing Experts, May 2007-present
- Freelance contractor, American Journal Experts, April 2012-December 2012
- Volunteer with the Multidisciplinary Association for Psychedelic Studies
- Relief International, Intern, Santa Monica, CA, December 2003-May 2004
- Intern, Marijuana Policy Project, Capitol Hill, Washington DC August-September 2003

Technical Skills

- Operating user, 3T Siemens MRI scanner, LREN, CHUV, Lausanne
- MSWord, Excel and HTML; Unix Operating systems; Matlab, SPSS, C++, Adobe Photoshop, E-Prime and InDesign ; PsychoPhysicsToolbox ; MRICron
- Eyelink Eyetracker ; BioPac Systems

Honors and Awards

- Invited Speaker, Alpine Brain Imaging Meeting, Champéry, Switzerland, January 8, 2018
- Invited Speaker, Nestle Research Center, Lausanne, September 18, 2017
- Invited Speaker, Apéro(Neuro)Science, BioScience Network Lausanne, Lausanne, March 14, 2017
- Invited speaker, Human Brain Project Summit, Florence, Italy 2016
- First Prize Winner, Science Writing Competition, Human Brain Project Summit, Florence, Italy, 2016
- Short-list, Richard Casement Internship for Science writing, Economist, 2004
- AP Scholar Award, June 1996
- English Merit Award, June 1996

Leadership Skills

- Life Sciences Faculty Council, September 2016-December 2017

Courses Taught During Doctoral Studies

- Neuroscience III (Prof. Carmen Sandi)
- Algebre Lineaire (Prof. Kathryn Hess-Bellwald)
 - _Guest Lecture : Application de l'algebre linéaire en IRMf

Additional Courses Taken

- CTI Business Course – Spring 2016
- Project Management Course – Fall 2016
- ²¹⁰·Python and Advanced Python – Fall 2016

Publications

Bendahan, S., Goette, L., Thoresen, J., Loued-Khenissi, L., Hollis, F., & Sandi, C. (2016). Acute stress alters individual risk taking in a time-dependent manner and leads to anti-social risk. *European Journal of Neuroscience*.

Loued-Khenissi, L., & Preuschoff, K. (2015). Apathy and noradrenaline: silent partners to mild cognitive impairment in Parkinson's disease?. *Current opinion in neurology*, 28(4), 344-350.

[Narendran R](#), [Frankle WG](#), [Keefe R](#), [Gil R](#), [Martinez D](#), [Slifstein M](#), [Kegeles LS](#), [Talbot PS](#), [Huang Y](#), [Hwang DR](#), [Khenissi L](#), [Cooper TB](#), [Laruelle M](#), [Abi-Dargham A](#). Altered prefrontal dopaminergic function in chronic recreational ketamine users. [Am. J. Psychiatry](#) 2005 Dec;162(12):2352-9.

Acknowledgements/Abstracts

Goette, L., Bendahan, S., Thoresen, J., Hollis, F., & Sandi, C. (2015). Stress pulls us apart: Anxiety leads to differences in competitive confidence under stress. *Psychoneuroendocrinology*, 54, 115-123.

Martinez, Diana; Slifstein, Mark; Broft, Allegra; Mawlawi, Osama; Hwang, Dah-Ren; Huang, Yiyun; Cooper, Thomas; Kegeles, Lawrence; Zarahn, Eric; Abi-Dargham, Anissa; Haber, Suzanne N.; Laruelle M. Imaging Human Mesolimbic Dopamine Transmission With Positron Emission Tomography. Part II: Amphetamine-Induced Dopamine Release in the Functional Subdivisions of the Striatum *J. Cereb. Blood Flow Metab.* 2003 23, 285-300.

H. Blair Simpson, Ilise Lombardo, Mark Slifstein, Henry Yiyun Huang, Dah-Ren Hwang, Anissa Abi-Dargham, Michael R. Liebowitz and Marc Laruelle Serotonin transporters in obsessive-compulsive disorder: a positron emission tomography study with [¹¹C]McN 5652 [Biological Psychiatry Volume 54, Issue 12](#) , 15 December 2003, Pages 1414-1421

ED London, SM Berman, BT Voytek, SL Simon, J Monterosso, JA Geaga, MS Hong, KM Hayashi, PM Thompson, MA Mandelkern, AL Brody, RA Rawson and W Ling: Cerebral metabolic dysfunction and impaired vigilance in recently abstinent methamphetamine abusers. *Biol. Psych.*, submitted 10/15/04; returned 12/01/04 for resubmission by 03/30/05.

Poster Presentations

1. Imaging the Locus Coeruleus: An fMRI Sequence Comparison Study

Leyla Loued-Khenissi ,Vasiliki Liakoni, Remi Castella, Antoine Lutti, Ferhat Kherif, Bogdan Draganski, Kerstin Preuschoff, Swiss Society for Neuroscience Annual Meeting, Lausanne, January, 2016

2. The Neural Correlates of Surprise, L. Loued-Khenissi, Antoine Lutti, K. Preuschoff, Alpine Brain Imaging Meeting, Champéry, January 2016

3. Imaging the Neural Correlates of Surprise, Leyla Loued-Khenissi, Vasiliki Liakoni, Antoine Lutti, Bogdan Draganski, Ferath Kherif, Kerstin Preuschoff
Organization for Human Brain Mapping (OHBM) Annual Meeting, Geneva, June 2016

4. The brain on bistable illusions, L. Loued-Khenissi, A. Pfeuffer, W. Einhauser, K. Preuschoff, Federation of European Neuroscience Societies (FENS) Forum, Copenhagen, July 2016

5. Common Neural Correlates of Uncertainty in Perceptual and Cognitive Decision-Making, Leyla Loued-Khenissi, A. Pfeuffer, W. Einhauser, Kerstin Preuschoff Society for Neuroeconomics Annual meeting, Berlin, August 2016

6. Common Neural Correlates of Uncertainty in Perceptual and Cognitive Decision-Making L. Loued-Khenissi, A. Pfeuffer, W. Einhauser, K. Preuschoff, Alpine Brain Imaging Meeting, Champéry, January 2017

7. Neural Correlates of Surprise and Entropy in a Financial Decision-Making Task L. Loued-Khenissi, K. Preuschoff, Neural Implementation of Learning Symposium, EPFL, November, 2017

