
Integrating Information Into Beliefs: Good and Bad News

Neil Garrett

Experimental Psychology

University College London

Dissertation submitted for the degree of

Doctor of Philosophy

of

University College London

June 2016

Declaration

I, Neil Garrett, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

June, 2016

Abstract

When integrating new information into our beliefs, an important factor is valence: whether a piece of news is good or bad. Evidence suggests that for self-relevant beliefs, separate rules and mechanisms underlie integration of these different types of news with good news being integrated to a greater degree than bad news. This asymmetry results in a positive bias. In this thesis, I present 4 studies that explore the boundaries of this asymmetry.

In study 1, I explore how the asymmetry is altered in clinical depression. I show that depressed patients update beliefs regarding their future in response to bad news to a greater degree than healthy controls resulting in unbiased updating. fMRI results suggest that this increased capacity to integrate bad news is the result of greater responsiveness in the receipt of undesirable information. These findings suggest that a positive state of mental health is linked to biased processing of information that supports positively skewed views of the future. In study 2, I examine whether the asymmetry varies in response to changes in the environment. Two separate experiments show that information integration in response to bad news (but not good news) is enhanced in threatening environments. These findings suggest that biased processing of information is not set in stone, but flexibly changes in response to the environment. In study 3, I adapt the update bias paradigm to investigate if a positive update bias exists under an alternative method of classifying good and bad news. A positive update bias remains under this alternative classification. In addition, under both the original and this alternative method of classification, updating is shown to correspond more closely to a rational Bayesian agent for good compared to bad news. These findings suggest that the update bias is robust to variations in classification schemes and analysis. In study 4, I explore whether an asymmetry in updating exists for positive as well as for negative life

events. I show that participants update their beliefs to a greater extent when receiving good news compared to bad news, regardless of whether the information concerns a positive or a negative life event. These findings suggest that the bias is not a phenomenon specific to negative life events and is robust to variations in task stimuli. Taken together with the findings of study 3, the results make a strong case for a true optimistic asymmetry in belief updating.

The four studies presented in this thesis expand our knowledge of how individuals integrate information about the world, characterize the boundaries of asymmetric information integration, examine the neural mechanisms that support it and its responsiveness to environmental threat.

Acknowledgements

The course of most PhDs are invariably described as a series of ups and downs. Whilst this one has certainly had its high points, the lows have gladly been sparse and infrequent. A large part of this has been down to the existence of alcohol of course. But a comparably large factor has been the people I've encountered and worked with along the way.

The Affective Brain Lab has been a fantastic place to learn and attempt to develop. I will always be grateful to everyone that passed through its doors whilst I was amongst its ranks. This began with who I like to refer to as the "founding crew" comprising myself, Caroline, Christina and Tali. Soon followed a second generation in which Steph, Andreas, Kaustubh and Seb came into the fold. And then, before you knew it, a third generation was in town for the party. "3G" saw Donal, Eleanor and Filip join and this is the generation in which I leave, making me a granddaddy of sorts. Across these different eras, everyone has been a thoroughly good egg, great to have around and immensely generous with their time, giving guidance and feedback in abundance whenever it's been asked. Out of everyone though I have to say a special thank you to Tali who was kind enough to take me on as a student way back yonder and has been a fantastic supervisor at each and every step of my PhD.

Finally, I should give at least a bit of praise to *Office of Man*, the setting in which much of this PhD has been conducted. It's been rowdy, smelled at times and seen many a PhD student come and go whilst I've been stationed there. A lot of the time I've had little idea what anyone was talking about (flipper cells?) but it's been a lot of fun interacting with everyone that's come its way and sharing experiences despite often working in very different neuroscience fields. Keep in touch folks.

Contents

Declaration	2
Abstract	3
Acknowledgements	5
Contents	6
List of Figures	12
List of Tables	13
Contributions	14
Publications during the PhD	15
1 General Introduction	
1.1 Information integration and biased beliefs.....	16
1.2 Update Bias Task and information integration.....	19
1.3 Underlying mechanism of biased updating.....	21
1.3.1 Memory, stimulus properties and skewed priors.....	21
1.3.2 Reinforcement learning.....	22
1.3.3 Asymmetry in learning.....	25
1.4 Boundaries of asymmetric updating.....	26
1.4.1 Outcomes open to interpretation.....	26
1.4.2 Motivation.....	27
1.5 Neural representation of estimation errors.....	30
1.6 An adaptive function?	33
1.7 Information integration across development.....	36
1.8 Outstanding questions and aims of thesis.....	38
1.9 Outline of thesis.....	39
2 Experimental Methods	

2.1	Update bias task.....	40
2.1.1	Task and stimuli.....	40
2.1.2	Differences in design between chapters.....	44
2.1.3	Behavioural analysis.....	45
2.1.3.1	Update and information integration.....	45
2.1.3.2	Control variables.....	47
2.1.4	Base rate distribution and estimation errors.....	48
2.1.5	Power analysis.....	53
2.2	Physiological Measures.....	54
2.2.1	fMRI; Chapter 3.....	54
2.2.2	Skin conductance and cortisol; Chapter 4.....	58
2.3	Mini-International Neuropsychiatric Interview; Chapter 3.....	60
2.4	Mood/Personality Questionnaires.....	60
2.4.1	Beck Depression Inventory; Chapters 3-6.....	60
2.4.2	State Trait Anxiety Inventory; Chapters 3 and 4.....	61
2.4.3	Life Orientation Test Revised; Chapters 3-6.....	61
3	Neural Substrates of Unbiased Belief Updating in Depression	
3.1	Introduction.....	63
3.2	Materials and Methods.....	64
3.2.1	Participants.....	64
3.2.2	Procedure.....	66
3.2.3	Behavioural Task.....	66
3.2.4	Behavioural Analysis.....	67
3.2.5	MRI scanning.....	68
3.2.6	fMRI data analysis.....	68
3.3	Results	
3.3.1	Unbiased updating in MDD, biased updating in controls.....	70
3.3.2	Control variables.....	71
3.3.2.1	Memory.....	72
3.3.2.2	Life event ratings.....	72

3.3.2.3	Task factors.....	73
3.3.2.4	Framing.....	74
3.3.3	Unbiased updating in MDD explained by adequate use and neural tracking of negative estimation errors	
3.3.3.1	Learning parameters.....	75
3.3.3.2	fMRI data.....	76
3.4	Discussion.....	80
4	Updating Beliefs Under Threat	
4.1	Introduction.....	84
4.2	Experiment I.....	86
4.2.1	Materials and Methods.....	86
4.2.1.1	Participants.....	86
4.2.1.2	Manipulation Procedure.....	86
4.2.1.3	Manipulation Check.....	87
4.2.1.4	Behavioural Task.....	88
4.2.1.5	Statistical Analysis.....	89
4.2.2	Results.....	91
4.3	Experiment II.....	99
4.3.1	Materials and Methods.....	99
4.3.1.1	Participants.....	99
4.3.1.2	Task, stimuli and control variables.....	100
4.3.2	Results.....	101
4.4	Discussion.....	104
5	How Robust is the Optimistic Update Bias for Estimating Self-risk and Population Base Rates?	
5.1	Introduction.....	106
5.2	Materials and Methods.....	109
5.2.1	Participants.....	109
5.2.2	Task and Procedure.....	109

5.2.2.1	Stimuli.....	109
5.2.2.2	Procedure.....	109
5.2.3	Data Analysis.....	110
5.2.3.1	Classification, Updates and Bayesian comparison...	110
5.2.3.2	Desirability Ratings.....	115
5.3	Results.....	115
5.3.1	Do participants update in a Bayesian manner?.....	115
5.3.2	Is updating beliefs regarding personal risk and base rates biased?.....	117
5.3.2.1	Updating of personal risk.....	118
5.3.2.2	Updating of base rates.....	118
5.3.2.3	Control variables.....	119
5.3.3	Do estimation errors underlie desirability and update?.....	123
5.3.3.1	Updating and estimation errors.....	123
5.3.3.2	Desirability and estimation errors.....	126
5.3.3.3	Effects of question order and frame.....	126
5.4	Discussion.....	128
6	Optimistic Update for Positive Life Events? An Unbiased Test	
6.1	Introduction.....	131
6.1.1	Obtaining statistics for positive events.....	132
6.2	Materials and Methods.....	133
6.2.1	Construction of stimuli.....	133
6.2.1.1	Participants.....	133
6.2.1.2	Task.....	134
6.2.1.3	Analysis and event selection.....	135
6.2.2	Update Bias Task.....	135
6.2.2.1	Participants.....	135
6.2.2.2	Analysis.....	139
6.3	Results.....	141
6.4	Discussion.....	142

7	General Discussion	
7.1	Summary and limitations of experimental investigations.....	144
7.1.1	Chapter 3: Neural Substrates of Unbiased Belief Updating in Depression.....	145
7.1.1.1	Summary.....	145
7.1.1.2	Limitations.....	145
7.1.2	Chapter 4: Updating Beliefs Under Threat.....	146
7.1.2.1	Summary.....	146
7.1.2.2	Limitations.....	146
7.1.3	Chapter 5: How Robust is the Optimistic Update Bias for Estimating Self-risk and Population Base Rates?.....	147
7.1.3.1	Summary.....	147
7.1.3.2	Limitations.....	149
7.1.4	Chapter 6: Optimistic Update for Positive Life Events? An Unbiased Test.....	149
7.1.4.1	Summary.....	149
7.1.4.2	Limitations.....	150
7.2	Key questions for future research.....	151
7.2.1	The interplay between anxiety, biased updating and depression.....	151
7.2.2	Neurological mechanism by which stress alters information Integration.....	154
7.2.3	Understanding the process of likelihood ratio formation.....	155
7.2.4	The role of motivation in asymmetric updating.....	157
7.2.5	Environmental volatility.....	159
7.2.6	Bias modification.....	160
7.2.7	The role of genetics in belief updating.....	161
7.2.8	The role of attention in belief updating.....	162
7.3	Conclusion.....	164

List of Abbreviations	166
Bibliography	167

List of Figures

1.1	The Update Bias Task	19
1.2	Updating, Learning Parameters and Estimation Errors	32
1.3	Information integration over the lifespan.	37
2.1	The Update Bias Task.	41
2.2	Histograms of base rate distributions.	49
2.3	Simulation Results.	51
3.1	Timeline of task.	67
3.2	Updating in MDD and controls.	71
3.3	Learning parameters and BDI scores.	76
3.4	Brain activity tracking estimation errors.	78
3.5	Brain activity tracking estimation errors in rIFG.	79
4.1	Manipulation check.	92
4.2	Bias in learning parameters.	94
4.3	Learning parameters for bad news and change in stress.	98
4.4	Learning parameters and stress (firefighters).	103
5.1	Paradigm.	108
5.2	Bayesian update and participant update.	116
5.3	Update for self risk and base rates.	123
5.4	Regression coefficients predicting update from estimation errors.	125
5.5	Order and frame effects self risk updates.	127
5.6	Order and frame effects, base rate updates.	128
6.1	Task for stimuli construction.	134
6.2	Update bias task.	136
6.3	Update for positive and negative life events.	142

List of Tables

2.1	Stimuli used in the Update Bias Task.....	43
3.1	Demographic and clinical characteristics of participants.....	65
3.2	Control Variables.....	74
4.1	Task related variables	95
4.2	Parameter estimates of covariates in Experiment I.....	99
4.3	Parameters estimates of covariates in Experiment II.....	103
5.1	Control variables	121
6.1	Stimuli used in study.....	137
6.2	Control variables.....	139
6.3	Categorisation of events.....	140

Contributions

The work reported in this thesis is entirely my own unless otherwise indicated. All chapters have benefited from guidance and advice from my supervisor, Dr. Tali Sharot.

The Update Bias Task used in Chapters 3, 4, 5 and 6 was adopted from a study originally conducted by Sharot and colleagues (Sharot et al., 2011) and used in a number of subsequent studies (Chowdhury et al., 2014; Korn et al., 2013; Moutsiana et al., 2013, 2015; Sharot et al., 2012a, 2012b).

Chapter 3 was the result of a collaboration between myself, Dr. Chris Korn at the University of Zürich (previously at University College London), Dr. Paul Faulkner and Prof. Jon Roiser (both at the Institute of Cognitive Neuroscience at the time of the study). Chris collected the behavioural and fMRI data of the first 3 participants in this study. I collected the remainder of data for this study. Paul and Jon assisted with participant recruitment and administering the Mini International Neuropsychiatric Inventory.

Data in Chapter 4, Experiment 1 was partly based on Ana Maria Gonzalez's MSc Cognitive Neuroscience project at the ICN. This was co-supervised by myself and Dr. Tali Sharot. Ana carried out the behavioural data collection under my supervision. This experiment was conducted in collaboration with Liat Levita at the University of Sheffield. Liat advised on collection and analysis of skin conductance data.

Publications during the PhD

Chapter 1 is partially composed from a review article published in *Trends in Cognitive Science* (Sharot and Garrett, 2016). Chapter 5, with the exception of the Bayesian analysis, has been published in *PLOS One* (Garrett and Sharot, 2014). Chapter 3 has been published in *Frontiers in Human Neuroscience* (Garrett et al., 2014).

Chapter 1

General Introduction

1.1 Information integration and biased beliefs

How do individuals integrate new information into beliefs about themselves and their future? The fact that beliefs of these kind show a tendency to be biased in a positive direction suggests that valence is an important factor. When individuals consider their skills and abilities, for instance, most exhibit a *superiority illusion* (Brown, 1986; Hoorens, 1993; Kruger and Dunning, 1999; Meyer, 1975; Svenson, 1981), believing they are better and more skilled than the majority. When thinking about one's future, the majority of people exhibit *unrealistic optimism* (Armor and Taylor, 2002; Baker and Emery, 1993; Lovallo and Kahneman, 2003; Puri and Robinson, 2007; Sharot, 2011; Weinstein, 1980). This means that they underestimate the likelihood of negative life events and overestimate the likelihood of positive events occurring. One of the first studies to demonstrate this phenomenon presented college students with different positive (for instance: receiving an award, living past 80) and negative (for instance: having a heart

attack, divorced a few years after marriage) life events and asked the students to estimate their own likelihood relative to the average of their peers (Weinstein, 1980). The students rated themselves as more likely than average to experience positive events (for example 13% more likely to receive an award) and less likely than average to experience negative events (for example, 49% less likely to become divorced) (Weinstein, 1980).

Intuitively it may seem odd that beliefs are positively biased in this way. After all, in daily life, we encounter information and experiences that ought to cause beliefs to reflect an accurate representation of the world in which we live. The fact that beliefs are positively biased suggests that information is integrated to a greater degree when information represents good news compared to when it represents bad news. Indeed, whilst theories from economics (Von Neumann and Morgenstern, 1953), machine learning (Bishop, 2006; Russell and Norvig, 1995) and psychology (Körding and Wolpert, 2006; Maslow, 1950) conceive of humans as rational unbiased integrators of information, experimental evidence suggests that this is not the case when it comes to self-relevant beliefs. In these instances, evidence suggests that bad news is often underweighted relative to good news. This is observed when participants are presented with information concerning their character traits (Korn et al., 2012), attractiveness (Eil and Rao, 2011), intellect (Eil and Rao, 2011; Mobius et al., 2012), financial prospects (Wiswall and Zafar, 2015) and likelihood of experiencing aversive life events in the future (Sharot et al., 2011). In each of these instances, beliefs are updated to a greater degree when this information suggests a shift to a more desirable set of beliefs (i.e. good news) relative to when this information suggests a shift to a less desirable set of beliefs (i.e. bad news).

In one study (Eil and Rao, 2011), groups of participants undertook a series of brief “dates” with one another and then rated each other on an attractiveness scale. Ratings were used to compile a ranking of the most attractive to the least attractive in the group. Each participant was then asked to plot a histogram that represented their perceived likelihood of being rated at each position in the ranking from most attractive in the group (position 1) to the least attractive in the group (position 10). On each trial, participants were then told whether their overall rank was above or below another randomly selected anonymous participant. Hence participants could get good news (someone else in the group was rated less attractive than them) or bad news (someone else in the group was rated more attractive than them). Participants were then asked to redraw their belief histogram. Comparing participants’ posteriors to those of a Bayesian agent showed that following good news, subjects updated their beliefs in a normative, Bayesian, manner (R^2 , accounting for subjects’ posterior with a Bayesian posterior, approximately 0.8), but less so following receipt of bad news (R^2 approximately 0.5).

In another study (Mobius et al., 2012), participants ($n > 2000$) undertook an IQ test and then estimated the likelihood that they were in the top half of performers. After each estimate they received a noisy signal (correct with 75% probability) informing them whether they were in the top or bottom half of performers. They then re-estimated the likelihood that they were in the top half of performers. Participants’ posteriors were more similar to that of a Bayesian agent following receipt of good news than bad news, (although in general their updates were more conservative than a Bayesian).

1.2 Update Bias Task and information integration

Another type of belief updating in which asymmetry in information integration is observed concerns beliefs about the likelihood of future life events. This has been quantified using the Update Bias Task (UBT) (**Figure 1.1** and **Experimental Methods, Chapter 2**). The basic premise of the task is to elicit people's beliefs about the likelihood of different events occurring to them, provide them with information that is either better or worse than expected and then examine how these beliefs change in light of this information.

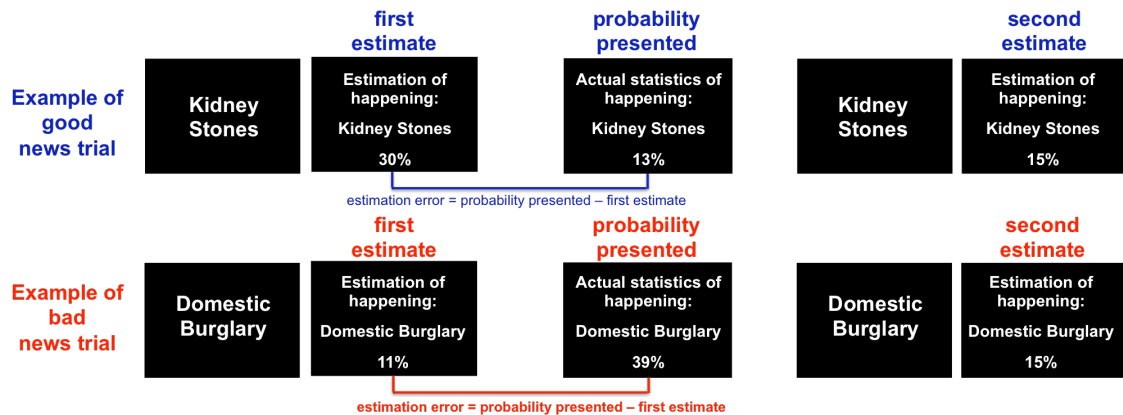


Figure 1.1: The Update Bias Task. Participants are presented with approximately 80 different aversive life events and are asked to estimate their likelihood of experiencing each event. They are then presented with the average likelihood of the event occurring to someone like them and are asked to re-estimate their likelihood. This is used to test and quantify asymmetric belief updating. Displayed (top) is an example of a good news trial in which the average likelihood is lower than the participant's first estimate and (bottom) an example of a bad news trial in which the average likelihood is higher than the participant's first estimate.

In the UBT, participants estimate the likelihood of experiencing different aversive life events in the future and are then presented with the average likelihood of these events happening to someone from the same socio-economic environment as them. Trials are partitioned into trials where participants receive *good news* in which the average presented is lower than their first estimate and *bad news* in which the average likelihood is higher

than their initial estimate. In a second session, participants see all aversive life events again and are asked to re-estimate their own likelihood of experiencing each event in the future. By comparing participant's first and second estimates, the extent to which participants update their estimates following presentation of the information provided (i.e. the change from first to second estimates) can be quantified and compared for good news trials and bad news trials. The finding from a range of studies (Chowdhury et al., 2014; Garrett et al., 2014; Garrett and Sharot, 2014; Korn et al., 2013; Kuzmanovic et al., 2015, 2016; Sharot et al., 2011, 2012a, 2012b) is that updating is asymmetric, being greater for good news compared to bad news (**Figure 1.2a**).

This task has also been adapted to examine how social information is integrated and whether valence also plays a role in the context of updating beliefs about one's abilities and character traits (Korn et al., 2012). Participants played a game of monopoly in groups of 5 and then privately rated one another according to different trait adjectives (e.g., polite, pedantic, reliable, selfish). Then a version of the UBT was administered in which participants rated themselves on each trait and saw the group's mean rating of them on this trait. As in the original UBT (Sharot et al., 2011), participants can get good (group gives them a higher rating on a positive trait or a lower rating on a negative trait) or bad (group gives them a lower rating on a positive trait or a higher rating on a negative trait) news. In a second session, participants then rated themselves again on each trait. The results showed that participants changed their ratings more when receiving good news compared to when in receipt of bad news (Korn et al., 2012).

1.3 Underlying mechanism of biased updating

1.3.1 Memory, stimulus properties and skewed priors

The asymmetric updating pattern observed in the UBT (**Figure 1.2a**) cannot be accounted for by memory differences. In the original study (Sharot et al., 2011), this was examined by asking participants to try and recall each of the probabilities they were presented with during the UBT, immediately after completing the task. There were no significant differences in memory errors (calculated as the absolute difference between participants recall of probabilities and the actual probabilities) for good news trials and bad news trials (Sharot et al., 2011). This was tested by calculating mean memory errors for good news and bad news trials for each participant and comparing these two sets of errors using paired sample ttests.

The pattern of asymmetric updating could also not be explained by differences in subjective properties of the stimuli, specifically: emotional arousal, familiarity, vividness, extent of negative valance and past experience. This was tested by having participants rate each life event according to each of these properties on a 6 point likert scale after completion of the UBT. Difference scores (calculated as the difference between the average rating for good news trials and average rating for bad news trials) were calculated for each participant for each of the subjective properties. When these 5 sets of difference scores were entered as covariates in a repeated measures ANOVA comparing belief updating for good news and bad news trials, the difference in updating between good and bad news remained significant (Sharot et al., 2011). Note that asymmetric updating also cannot be explained as the result of differential processing of high and low numbers as

this was controlled for by asking participants to estimate their likelihood of encountering adverse events on half of the trials and to estimate their likelihood of *not* encountering the adverse events on the other half of the trials (Sharot et al., 2011).

One possibility though is that biased updating is the result of positively skewed priors held by participants. Care is taken to control for the *mean* prior of participants in the UBT (by adding average first estimate as a covariate in the analysis) but the *distribution* of priors this mean is generated from cannot be inferred from the task so it is not possible to know whether these are positively skewed. However, studies that have elicited a full distribution of participants' prior beliefs such as those described above (Eil and Rao, 2011; Mobius et al., 2012) show that their posteriors diverge significantly from what would be expected from a Bayesian agent following bad news, but converge with Bayesian posteriors following good news. Here, although participants are shown to place full weights on their priors, they still underweight bad news relative to good news even when incentivized to report accurate beliefs.

1.3.2 Reinforcement learning

Reinforcement learning originated in the field of machine learning (Sutton and Barto, 1998) to describe a class of problems in which, over multiple trials, an agent is tasked with learning which action it should take (for instance, which of two buttons to press, with each button leading to rewards with differing probabilities) in order to achieve a goal (for instance, maximising rewards).

One simple mechanism (sometimes referred to as a type of “model free” learning) by which agents are able to learn in these types of contexts is to assign values to each

available action and iteratively scale these values up and down according to how well they correspond to actual outcomes that ensue when a specific action is selected. “Prediction Errors” (β) quantify this difference between expected and actual rewards (Schultz et al., 1997; Sutton and Barto, 1998). Formally, on a specific trial (t), a prediction error (β) for an action, A , selected on that trial, can be expressed as:

$$\beta(t) = r(t) - V_A(t)$$

Where, $r(t)$ represents the reward delivered on the current trial t and $V_A(t)$ the expected value of action A .

Prediction errors are therefore positive in instances in which rewards are greater than expected (for instance, the delivery of a reward where none was expected) and negative in instances in which rewards are less than expected (for instance, the absence of a reward where a reward was expected). Prediction errors are then used to increment or decrement values assigned to actions, according to a learning rate α . The value of a particular action, A , could be adjusted according to:

$$V_A(t+1) = V_A(t) + \alpha * \beta(t)$$

Hence, the greater the learning rate (α) is, the more that values assigned to actions are altered following receipt of feedback (e.g., reward/no reward) on each trial. Where multiple actions are possible (e.g., participant needs to choose between actions A and B), the values of each action are compared at the time of choice and a choice determined stochastically (e.g., via a softmax function) such that actions with higher values are more likely to be selected.

This simple learning mechanism (e.g., McClure et al., 2003; Pessiglione et al., 2006; Frank et al., 2007) and more complex versions of it (e.g., Collins and Frank, 2014)

has been shown to perform well at simulating trial by trial learning of humans and animals and to be biologically plausible with phasic firing of midbrain dopamine neurons shown to encode prediction errors (Schultz et al., 1997).

Learning in the UBT differs in a number of ways to learning in the type of tasks reinforcement learning models typically capture. Two of the key differences are that in the UBT, participants encounter each stimulus (life event) just once during the learning phase and do not receive any kind of tangible feedback - such as a reward or punishment - following their responses (simply information which can be interpreted in a number of different ways). This is different to reinforcement learning tasks where participants typically see the same stimuli multiple times and receive clear feedback (such as a reward or the absence of a reward) on each trial.

Despite these differences, adapting the reinforcement learning framework to learning in the UBT has been shown to perform well in terms of capturing how individuals update beliefs regarding their future and differences in learning following the receipt of good and bad news. In particular, a key feature of model free reinforcement learning that can be adopted to model learning in the UBT, is that learning takes place by means of prediction errors which quantify the difference between expectations (anticipating that implementing a certain action - such as a button press in response to a cue - will lead to an outcome such as a monetary reward) and the actual outcome that transpires (Schultz et al., 1997; Sutton and Barto, 1998). An analogous mechanism has been suggested to underwrite learning in the UBT (Garrett et al., 2014; Moutsiana et al., 2013; Sharot et al., 2011). Specifically, that *estimation errors* (the difference between

initial estimates and information participants are provided with on each trial, **Figure 1.1**) drive participant learning.

1.3.3 Asymmetry in learning

Reinforcement learning models which use a single learning rate and do not distinguish between positive prediction errors (generated for instance when receiving a reward after expecting nothing or after expecting something unpleasant as an outcome) and negative prediction errors (generated for instance when receiving an electric shock after expecting nothing or after expecting something pleasant such as a reward as an outcome) often perform worse at accounting for data compared to models that use two learning rates – one for positive prediction errors and one for negative prediction errors. This is despite penalising the addition of an extra parameter (Frank et al., 2007; Niv et al., 2012; Gershman, 2015; Lefebvre et al., 2016). When two learning parameters¹ – one for good news (α_G) and one for bad news (α_B) – are used to predict the degree of belief change (i.e. update) in the UBT, good news learning parameters (α_G) are revealed to be larger than bad news learning parameters (α_B) (**Figure 1.2b**), consistent with the asymmetric pattern of updating observed in the task (**Figure 1.2a**). The implication is that estimation errors generated in the receipt of good news are assigned more weight and better integrated towards belief revision than estimation errors generated in the receipt of bad news (Sharot et al., 2011).

¹ Throughout the thesis, to avoid confusion I try to use learning *parameters* to denote the weights assigned to good news/bad news estimation errors towards updating beliefs in the UBT. Learning *rates* is used to refer to weights assigned to negative/positive prediction errors from feedback received in reinforcement learning tasks.

Some of the reinforcement learning studies that apply two learning rates similarly find that learning rates for positive prediction errors are greater than learning rates for negative prediction errors (Frank et al., 2007; Lefebvre et al., 2016). However, this is not consistently found and in some instances the reverse asymmetry has been found, i.e. learning rates for negative prediction errors are greater than those for positive prediction errors (Gershman, 2015; Niv et al., 2012). Attempting to characterise the conditions and boundaries under which asymmetric learning is observed may help understand this discrepancy.

1.4 Boundaries of asymmetric updating

Positively biased asymmetric integration of information is most likely to be observed when two conditions are met: information is ambiguous and beliefs are motivated.

1.4.1 Outcomes open to interpretation

Information received in the UBT is unconstrained and able to be interpreted in different ways. Individuals can distrust the source of the information or downplay why it is that a particular instance of bad news ought not to apply to them. For example, on learning that one is more likely to have a heart attack than previously thought, individuals can bring to mind reasons why this need not require a large belief adjustment (e.g. they keep fit, maintain a healthy diet). It is of note that less of a positivity bias is observed when individuals update views about their intelligence compared to updating views about attractiveness (Eil and Rao, 2011). This might be because intelligence (garnered from an IQ test in this study) is more objective than attractiveness hence there is less space for

manoeuvring in terms of constructing ways that this information could be suspect (and therefore discounted). Similarly when one encounters a positive or negative outcome (as in reinforcement learning tasks) such as a financial reward, or an electric shock, the form of feedback is explicit and cannot be interpreted as anything other than a reward or a shock.

1.4.2 Motivation

Beliefs are thought to have value in and of themselves, in the sense that they are ‘consumed’, with positive beliefs eliciting positive feelings and thus having positive utility, and negative beliefs eliciting negative feelings and having negative utility (Brunnermeier and Parker, 2004; Loewenstein, 2006). People are thus motivated to maintain a positive optimistic view of themselves, their circumstances, and their future, and to disregard negative information. By contrast, when there is no intrinsic or external advantage for holding a belief, or when the advantage is relatively small, an asymmetry in updating may be less apparent.

A recent study (Cahill, 2015) supports the notion that motivation is a key requirement for biased belief updating to take place. Participants spun a gambling wheel on each trial and were told that this could be ‘tuned’ to one of two modes: *positive mode* in which the wheel pays out rewards more frequently than it incurs losses; or *negative mode* in which the wheel incurs losses more often than it pays out rewards. After each spin of the wheel, participants received feedback in the form of a reward or a loss. Following feedback they were asked to indicate which of the two modes they thought the wheel was tuned to. Note that both of the above constraints are met: (1) information is

ambiguous in relation to the belief. Although the outcome (reward or loss) is unambiguous, participants can get a reward when the machine is tuned to negative mode and a loss when tuned to positive mode; (2) there is a motivation to maintain the belief that the wheel is in positive mode (and move away from the belief that it is in negative mode) since participants presumably want to believe that they are going to receive rewards more often than they incur losses during the experiment. The results from this study show that learning from feedback is biased in a positive direction. Specifically, comparing beliefs for the same amount of evidence suggestive of being in positive mode (e.g. a loss followed by two reward outcomes) to being in negative mode (e.g. a reward followed by two loss outcomes), participants were more inclined to report that they thought the wheel was in positive mode than negative mode. This suggests that a bias towards positive information is present if the relevant conditions (ambiguity and motivation) are met.

A subsequent study has also reported asymmetric learning in a reinforcement learning task (Lefebvre et al., 2016). In this study, participants were instructed to choose between pairs of abstract visual cues and probabilistically received feedback (reward or loss/reward omission), learning over the course of the experiment which cue was the better option in each pair. Across participants, a model comprising 2 learning rates (one for positive prediction errors and one for negative prediction errors) provided a better fit to participant's choices than a model with 1 learning rate and learning rates in the 2 learning rate model were greater for positive prediction errors than for negative prediction errors (consistent with Frank et al., 2007). In other words, when participants received feedback that was better than expected, they updated beliefs to a greater extent relative to when they received feedback that was worse than expected. The extent to which a model with two

learning rates was a better fit to participant's choices relative to a model with one learning rate was also shown to correlate positively with trait optimism.

The motivation for asymmetric learning in the context of high level abstract beliefs (such as whether we are in a positive mode, or whether we are likely to be robbed in the future) is consistent with the idea of positive beliefs generating utility (Brunnermeier and Parker, 2004; Loewenstein, 2006). A similar motivation may be a factor behind the learning asymmetry found in some reinforcement learning paradigms such as the one described above (Lefebvre et al., 2016) and others (Frank et al., 2007). Specifically, if individuals derive utility from the belief that they are going to receive rewards from certain cues and disutility from the belief that they are going to incur losses from certain cues, learning to a greater degree following positive feedback than negative feedback (as captured by differences in the magnitude of learning rates) would help to generate beliefs which produce these utility gains for individuals. Interestingly however, in these reinforcement learning contexts, an asymmetry in learning has been found less consistently (e.g., Gershman, 2015; Niv et al., 2012) suggesting that other motivations and factors may play a role in determining the relative magnitude of learning rates.

One of these factors may be the timeframe over which beliefs are held. In the UBT, beliefs concern the occurrence of significant life events in one's lifetime. These beliefs are therefore of high importance to the participant and their importance and relevance continue to exist outside the context of the UBT experiment. In reinforcement learning paradigms however, beliefs usually concern abstract cues (such as fractal pictures, symbols or one arm bandits). Beliefs about them (such as whether they are likely to pay out small rewards or incur small financial losses) are most likely of less importance

to the participant, in part because they have no significance once the experiment has finished. The relative advantage therefore to holding positively biased beliefs in reinforcement learning studies is often likely to be small relative to in the UBT where beliefs concern more impactful stimuli and pertain over a much longer timeframe.

A second factor that is likely to play a role in reinforcement learning paradigms but less so in the UBT is the existence of loss aversion (Kahneman and Tversky, 1984, 1992). Individuals fear losses more than they derive pleasure from gains. As a result, the disutility incurred from losing £1 is greater than the utility generated from gaining £1. In reinforcement learning tasks which often involve learning to associate some cues with gains and others with losses, loss aversion ought to cause a bias towards learning to a greater degree from stimuli which incur losses over those that pay gains (since for the same amount lost or gained, subjectively the loss will be greater). These effects could interact with or dominate any effect of a positive bias that arises out of utility generated from holding beliefs about future rewards and losses.

Given the presence of these factors and others, it is likely that positive biases where they are found in reinforcement learning processes (Lefebvre et al., 2016; Frank et al., 2007) are characterised by different boundaries than those that arise in the context of high level beliefs such as those in the UBT.

1.5 Neural representation of estimation errors

Functional magnetic resonance imaging (fMRI) has been used to investigate areas of the brain which track estimation errors from good news and bad news (Sharot et al., 2011). Consistent with the behavioural data suggesting that learning from good news and bad

news is better explained by separate learning parameters, Blood Oxygenated Level Dependant (BOLD) response has been shown to correlate with the magnitude of estimation errors in separate parts of the brain. Specifically, BOLD response in the left Inferior Frontal Gyrus (IFG) and Medial Frontal Cortex (MFC) correlates positively with estimation errors generated from good news (**Figure 1.2c**) whereas BOLD response in the right IFG correlates negatively with estimation errors generated from bad news (**Figure 1.2d**). Differences in the extent to which BOLD response correlates with estimation errors in the right IFG correlated with the extent to which there was a bias in belief updating with greater tracking of negative estimation errors corresponding to less biased updating.

The fact that these fMRI results are lateralised may shed light on the common and the separate underlying cognitive processes engaged in when individuals integrate good and bad news respectively. Both the left and right IFG have been suggested to play a role in encoding “Bayesian surprise” – the extent to which information deviates from a Bayesian posterior in a weather prediction task (d’Acremont et al., 2013) – as well as error monitoring (Mitchell et al., 2009), reversal learning (Greening et al., 2011) and the encoding of risk prediction errors (d’Acremont et al., 2009). However the right IFG specifically has been suggested to play a key role in inhibition (Aron et al., 2004) - such as task switching in the Wisconsin Card Sorting Task (Konishi et al., 1998; Aron et al., 2003) or withholding responses in a Go/No Go task (Konishi et al., 1998) - as well as risk aversion (Christopoulos et al., 2009) and processing punishment information (O’Doherty et al., 2001).

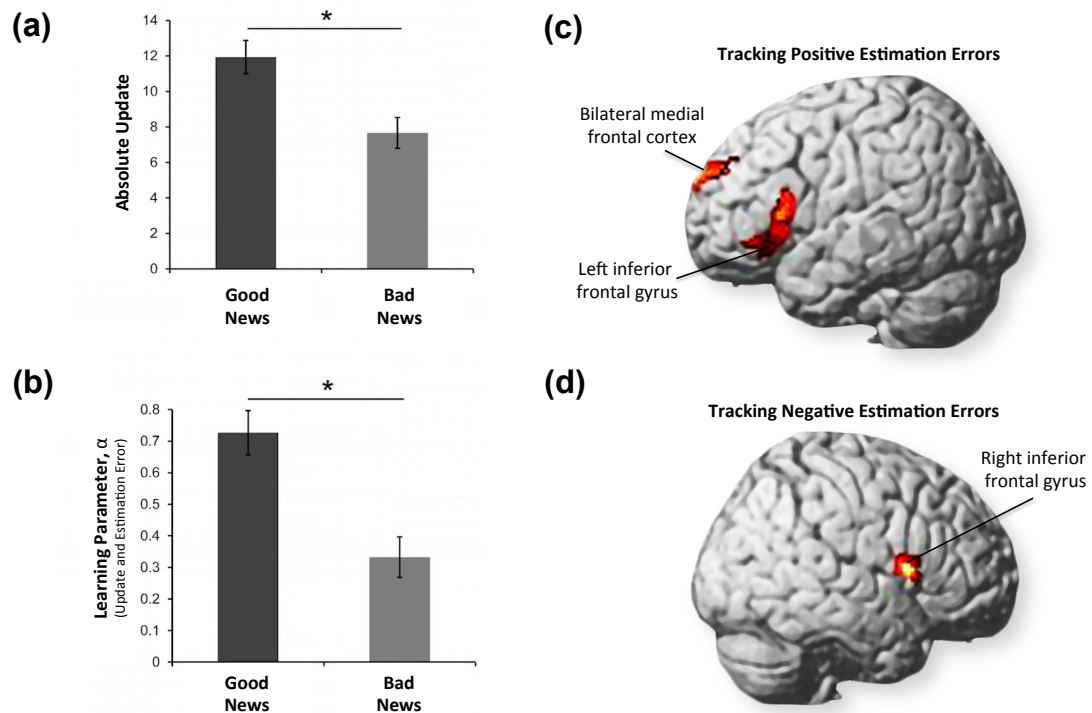


Figure 1.2: (a) In the UBT, participants adjust their beliefs to a greater extent when they receive good news (i.e. that a negative event is less likely to occur than expected) compared to when they receive bad news (i.e., that a negative event is more likely to occur than expected). (b) Learning parameters for good news (α_G) are greater than learning parameters for bad news (α_B). Learning parameters (α) quantify the degree to which estimation errors (the difference between the first estimate and the information given, see Figure 1.1) correspond to change in beliefs (derived from the equation: second estimate = first estimate + α (estimation error)). (c) Estimation errors for good news correlate with blood oxygenation level-dependent (BOLD) signal in the left inferior frontal gyrus (IFG) and the medial frontal cortex (MFG). (d) Estimation errors for bad news correlate negatively with BOLD response in the right IFG. Figure adapted from (Sharot et al., 2011).

Recently, Diffusion Tensor Imaging (DTI) has been used to further investigate the neural mechanisms that underlie biased belief updating (Moutsiana et al., 2015). Results showed that white matter connectivity (measured as the number of white matter tracts) between the left IFG and a number of sub cortical brain regions (specifically: left amygdala, putamen, pallidum, hippocampus and thalamus) as well as the left insula correlated with asymmetric belief updating with greater connectivity between the left IFG and each region associated with a greater bias in updating. This was due to a combination

of both greater updating in response to good news and reduced updating in response to bad news each correlating with the strength of white matter connectivity. This study suggests that asymmetric updating is not due to only frontal lobe function but an interaction between frontal regions and sub cortical structures modulating value and emotion (Baxter and Murray, 2002; Chang, 2001; LeDoux, 2000; Phelps, 2006). Consistent with this, applying Transcranial Magnetic Stimulation (TMS) to the left IFG abolishes the update bias (Sharot et al., 2012a).

1.6 An adaptive function?

An important question concerns why it is that humans integrate information in this asymmetric way as there seem to be obvious disadvantages to ignoring negative information such as underestimating risks. Asymmetric belief updating is suggested to be a mechanism which gives rise to positively biased beliefs (Sharot and Garrett, 2016) such as over-optimism and over-confidence. These beliefs have been blamed for a host of disasters such as war (Johnson, 2009), overly aggressive medical decisions (Paling, 2003), ill-preparedness in the face of natural catastrophe (Paton, 2003) and financial collapse (Shefrin, 2009). Moreover, positively biased views of the self can lead to error and cost, as shown for overconfident CEOs (for review, see Dunning et al., 2004) and overconfident stock traders (Barber and Odean, 1999).

Yet, set against these disadvantages there are also a number of advantages to holding positive expectations. These include advantages to physical health suggested by a number of epidemiological studies. For example, individuals (n=95,000) with positive expectations have been shown to be less likely to develop coronary heart disease over an 8

year period (Tindle et al., 2009). Another study (Scheier et al., 1999) reported that patients who had developed heart disease and undergone coronary artery surgery were less likely to be re-hospitalized in the 6 months following surgery if they were of an optimistic disposition, measured with the Life Orientation Test (a similar finding has been reported by Tindle et al., 2012). Other studies have found that lung (Novotny et al., 2010) and head/neck (Allison, 2003) cancer patients had better survival rates if they were optimistic, all else being equal and that optimism protects older adults against stroke (Kim et al., 2011). The reason that positive expectations are thought to be associated with better physical health is suggested to be twofold (Carver and Scheier, 2014). Firstly, optimists are more inclined to proactively engage in measures to safeguard their health such as refraining from smoking and taking regular exercise (Gitlay et al., 2007; Steptoe et al., 2006). Secondly, positive expectations are believed to influence emotional responses to distressing experiences. Optimism has been linked to lower cortisol response to stress for instance (Jobin et al., 2013). This reduction of stress and anxiety can lead to better physical health over time (Taylor and Brown, 1988; Taylor et al., 2000). In addition to the health advantages positive expectations has been shown to impart, they are also believed to be advantageous in terms of increasing motivation (Carver and Scheier, 2014) including improving productivity (Galasso and Simcoe, 2011), persistence (Solberg et al., 2009) and encouraging the prioritisation and pursuit of goals (Geers et al., 2009).

Given that there are seemingly disadvantageous and advantageous aspects to positive expectations, one possibility for their survival is that the advantageous aspects outweigh the disadvantageous (McKay and Dennett, 2010). Another possibility however, suggested by an evolutionary computational model (Johnson and Fowler, 2011), is that

populations have evolved over time to be overconfident as this maximises individual fitness. This is because individuals successfully claim resources from stronger less confident rivals and are resistant to rivals trying to claim resources that they are strong enough to defend. Importantly, overconfidence only evolves as an equilibrium in this model when 2 conditions are met: (1) there is uncertainty about one another's capability in attaining resources; (2) resources attainable from conflicts need to be sufficiently greater than costs incurred in conflicts. Interestingly, when this second assumption is not met and attainable resources are lower than the costs of conflict, the dynamics of the model play out such that a combination of unbiased and underconfident individuals evolves as the equilibrium. This raises the suggestion that when costs in the environment increase relative to rewards, a reduction of positive expectations ought to be observed. This idea is tested in **Chapter 4** where I examine whether positively biased updating remains when an individual's environment becomes more threatening.

Animal studies investigating positive expectations consistently show that in environments that are comfortable and/or rich in resources, positive expectations are present but these expectations are absent in environments which are unpleasant, threatening and/or depleted. This has been observed in pigs (Douglas et al., 2012), chicks (Salmeto et al., 2011), rodents (Harding et al., 2004) and starlings (Matheson et al., 2008). Studies typically have animals associate a response to a cue (such as a sound, visual cue or odour) with an appetitive outcome and an alternate response to a contrasting cue with a less appetitive or aversive outcome. A novel intermediate cue is then introduced and it is observed whether the animal's response is suggestive that the better or worse outcome is expected.

In one study (Matheson et al., 2008), starlings learned stimulus (visual cue of a specific duration) response (pecking one of two buttons) associations to two different visual cues (e.g. short, 2 seconds and long, 10 seconds). A correct response to one of the visual cues led to a high reward (immediate food) and an alternate response to the other visual cue led to a low reward (delayed food). Incorrect responses were not rewarded. After stimulus-response associations had been learned, an intermediate visual cue (e.g. medium, 6 seconds) was introduced. Animal responses to this cue indicated whether they expected this to lead to the high reward or low reward. Starlings housed in comfortable environments (large cages with natural branches and water baths) showed a higher tendency to expect intermediary cues to lead to high rewards relative to starlings housed in unenriched environments (smaller cages with unpredictable cleaning times, water baths 50% of the time). Other studies have reported similar results in different species using variations of this paradigm, including the use of one reward and one aversive outcome rather than two appetitive outcomes of differing magnitudes. All show that animals are more likely to interpret ambiguous stimuli as indicative of a previously learned positive outcome if living in a relatively stress-free environment.

1.7 Information integration across development

A number of studies using the UBT have been undertaken across a range of ages from adolescents (Moutsiana et al., 2013) through to old age (Chowdhury et al., 2014). When data from studies is combined, learning parameters for good news (α_G) are consistently higher than learning parameters for bad news (α_B). However, an interesting feature of learning parameters for bad news (α_B) is that they assume an inverse U shape over the

lifespan, being lowest in young and old and highest in middle age. This is not the result of general alterations in learning since learning parameters for good news (α_G) remain flat over the lifespan (**Figure 1.3**). In other words, this variation in learning seems selective to bad news specifically.

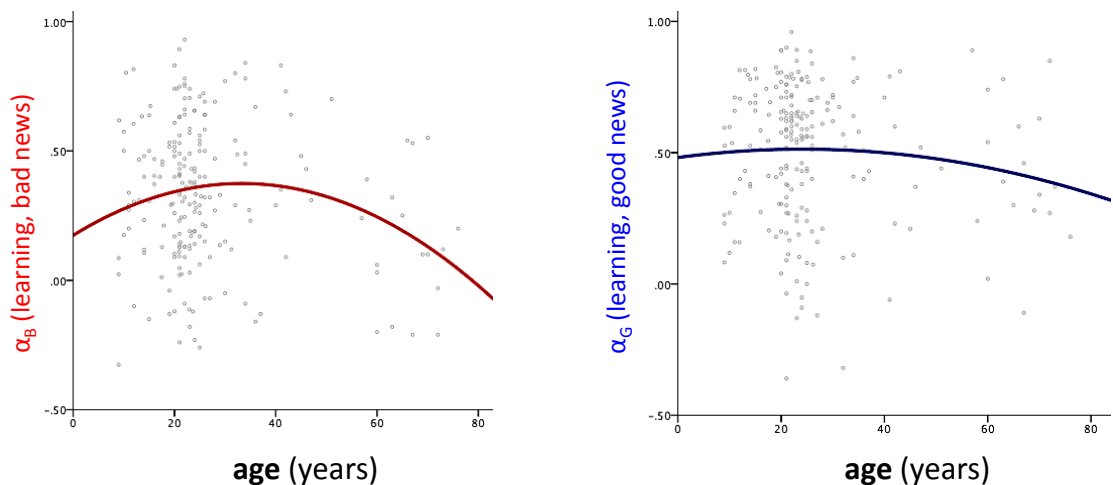


Figure 1.3: How Information Integration Alters with Age. Learning from bad news (α_B) changes with age following an inverse-U shape (quadratic) function; it is low in children and teenagers, peaks around middle age, and then starts to decline again. This pattern is not observed for learning of good news (α_G), which remains relatively stable over the lifetime. Figure includes data from 4 studies (Sharot et al., 2011; Moutsiana et al., 2013; Chowdhury et al., 2014; Garrett et al., 2014).

The neurobiological mechanism(s) giving rise to this variation in learning parameters with age is not known however. They could be the result of alterations in dopamine function or connectivity between brain regions, both of which have been associated with mediating biased updating in previous studies (Sharot et al., 2012b; Moutsiana et al., 2015) and shown to vary with age (Volkow et al., 1996; Fjell et al., 2016). However, this variation could also be accounted for by other neurological mechanisms and there may be separate mechanisms operating at different stages of the lifespan rather than a single mechanism or combinations of mechanisms at play. Another possibility is that they arise out of non-physiological environmental factors.

1.8 Outstanding questions and aims of thesis

The aims of this thesis are twofold. The first aim is to examine variability in belief updating. Since biased belief updating is suggested to maintain positive self-views (Sharot and Garrett, 2016) and positive self-views are often absent in depression, I investigate in **Chapter 3** if individuals with depression show less biased belief updating. In **Chapter 4**, I examine whether biased belief updating is a fixed human tendency or state dependant, allowing for fluctuations when individuals are placed in a threatening environment.

The second aim is to examine the robustness of asymmetric belief updating. To date, the majority of studies that have examined belief updating have done so in the context of beliefs about one's *own* likelihood of encountering *negative* life events in the future (Chowdhury et al., 2014; Garrett and Sharot, 2014; Garrett et al., 2014; Korn et al., 2013; Kuzmanovic et al., 2015; Moutsiana et al., 2013, 2015; Sharot et al., 2011, 2012a, 2012b). On the one hand, biased updating might be specific to processing information about one's own likelihood (as it is this which individuals are motivated to maintain positive views of). However, if individuals integrate information about base rates when considering their own susceptibility to different life events, a bias for updating base rates could also exist for base rates and contribute to the maintenance of positive self views. In **Chapter 5**, I examine whether asymmetric belief updating also exists for updating base rates, whether base rates are integrated into revised beliefs in a Bayesian manner and whether asymmetric effects exist under different classifications of good and bad news. Finally in **Chapter 6**, I investigate whether a bias in updating exists for positive as well as negative life events.

1.9 Outline of the thesis

Chapter 2 details the experimental techniques used in the studies reported in Chapters 3-6. This includes detail on the UBT, questionnaires administered and physiological measures used.

In **Chapter 3**, I investigate differences in updating between depressed patients and controls and use fMRI to relate these differences to alterations in information processing. Since depressed patients have been shown to be less optimistic than controls it was hypothesised that they would exhibit less asymmetry in information integration with this asymmetry corresponding to differences in estimation error tracking in the brain.

In **Chapter 4**, I report the result of two studies investigating how stress alters information integration.

Chapter 5 details the results of a study in which I extend the UBT so that participants also provide estimates of the base rate for each event. This allows alternate ways of classifying trials into good and bad news. Under each classification scheme, it is then investigated the degree to which updating for good and bad news corresponds to that of a Bayesian agent and whether a bias in updating exists for base rates (as well as for own likelihoods).

Chapter 6 details the results of an online study in which I examine if biased belief updating exists for positive as well as negative life events.

In **Chapter 7**, I provide a summary of the main findings of each experimental chapter. I also outline limitations of the studies presented and suggest some directions for future research.

Chapter 2

Experimental Methods

2.1 Update bias task

2.1.1 Task and stimuli

The Update Bias Task (UBT) has been used in a number of studies (Chowdhury et al., 2014; Garrett and Sharot, 2014; Garrett et al., 2014; Korn et al., 2013; Kuzmanovic et al., 2015; Moutsiana et al., 2013, 2015; Sharot et al., 2011, 2012a, 2012b) to investigate valence effects of belief updating. It is depicted in **Figure 2.1**.

On each trial, one of 80 different stimuli is presented on screen. Each of these is an adverse life event such as burglary, kidney stones (see **Table 2.1** for list of stimuli used). Participants are asked to imagine the event happening to them in the future and are then asked to estimate how likely the event is to happen to them in the future. Participants are then shown the base rate of the event in a demographically similar population to themselves. In a second session, immediately after the first, participants are asked again to provide estimates of their likelihood of encountering the same events so that it can be assessed how they update their estimations, following the information they are provided.

For each adverse life event, the base rate of that event occurring at least once to a person living in the same socio-cultural environment as the participants was determined from online resources (Office for National Statistics, Eurostat, PubMed). Very rare, or very common, events are not included; all events probabilities lie between 10% and 70%. To ensure that the range of possible overestimation is equal to the range of possible underestimation, participants are told that the range of probabilities lie between 3% and 77%. If the participant fails to submit a response in the first session or second session that trial is excluded from all consequent analyses.

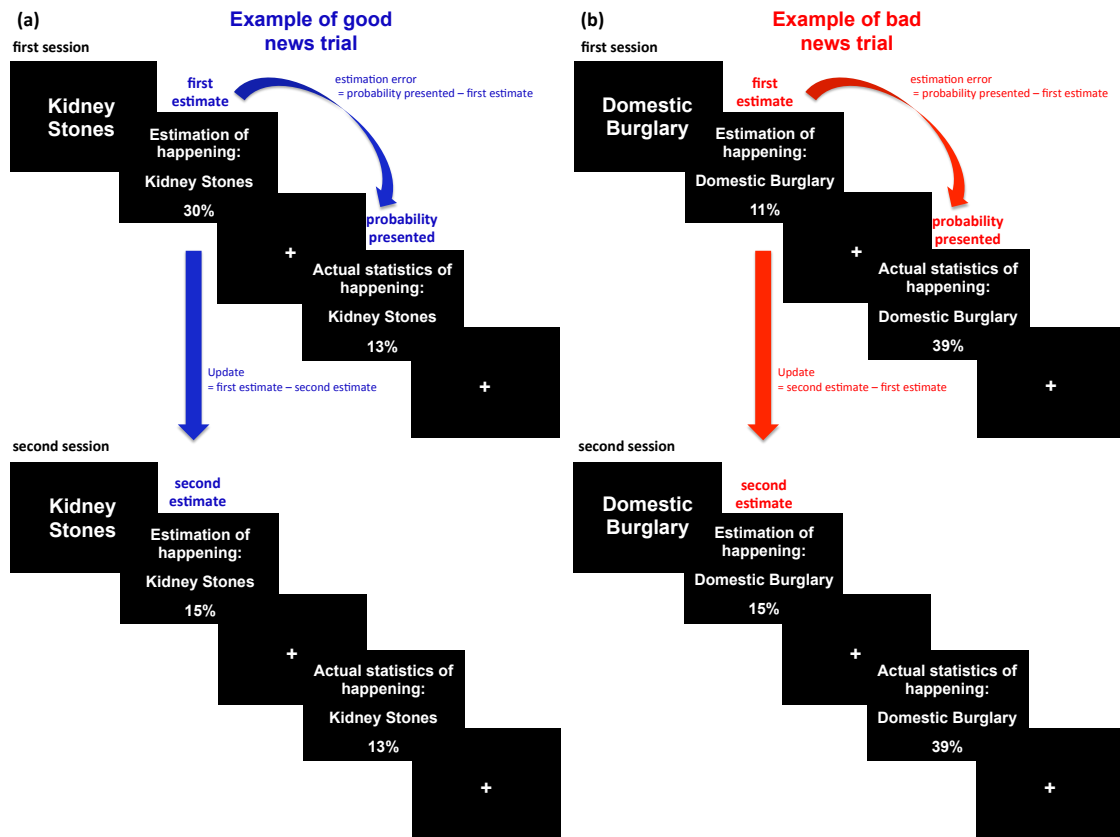


Figure 2.1: The Update Bias Task. Participants are presented with 80 different life events (see Table 2.1) and are asked to estimate their likelihood of experiencing each event. They are then presented with the base rate (average likelihood) of the event occurring to someone like them. In a second session they are asked to re-estimate their likelihood. Trials are categorized into good news and bad news according to whether the base rate presented in the first session is below (a, good news) or above (b, bad news) participants' first estimate.

A feature of this design is that participants often receive good news when the base rate presented to them is a low number and bad news when the base rate is a high number. To control for the possibility that differences in updating therefore could reflect differences in processing of high and low numbers, on half the trials participants estimate the likelihood of the event *happening* to them in the future and on the other half of trials, participants estimate the likelihood of the event *not happening* to them in the future. Framing estimations in these two ways ensures that differential updating cannot be attributed to differential processing of high and low numbers. Furthermore, under such framing half the trials are conceptually presented as negative events (i.e. divorce) and half as positive events (i.e. – never divorce).

LIST A	LIST B
age related blindness	back pain
alcoholism	abnormal heart rhythm
appendicitis	Alzheimer's disease
being cheated by husband/wife	anxiety disorder
being convicted of crime	arteries hardening (narrowing of blood vessels)
bone fracture	artificial joint
car stolen	asthma
card fraud	autoimmune disease
chronic high blood pressure	being fired
chronic ringing sound in ear (tinnitus)	bicycle theft
death before 60	blood clot in vein
death before 80	cancer_(of_digestive_system/lung/prostate/breast/skin)
dementia	computer crash with loss of important data
depression	death before 70
diabetes (type 2)	death by infection
disease of spinal cord	divorce
domestic burglary	epilepsy
drug abuse	eye cataract (clouding of the lens of the eye)
fraud when buying something on the internet	having a stroke
gallbladder stones	having fleas/lice
genital warts	hepatitis A or B
gluten intolerance	herpes

heart failure	hospital stay longer than three weeks
hernia (rupture of internal tissue wall)	kidney stones
house vandalised	limb amputation
household accident	liver disease
infertility	migraine
irritable bowel syndrome (disorder of the gut)	osteoporosis (reduced bone density)
knee osteoarthritis (causing knee pain and swelling)	severe injury due to accident (traffic or house)
miss a flight	severe insomnia
more than £30,000 debts	severe teeth problems when old
mouse/rat in house	sexual dysfunction
obesity	skin burn
Parkinson's disease	theft from person
restless legs syndrome	ulcer
serious hearing problems	victim of bullying at work (nonphysical)
sport related accident	victim of mugging
theft from vehicle	victim of violence at home
victim of violence by stranger	victim of violence by acquaintance
witness a traumatising accident	victim of violence with need to go to A&E

Table 2.1: List of stimuli used in the UBT. In the full version of the task, two lists of 40 events are used. Participants provide first and second estimates for one of the lists (sessions one and two) followed by first and second estimates for the other list (sessions 3 and 4). One of the lists is presented in the framing “estimation of happening to you?” whilst the other list is presented in the framing “estimation of NOT happening to you?”.

After completing the UBT, participants rate stimuli on prior experience [for the question “Has this event happened to you before?” the responses ranged from 1 (never) to 6 (very often)], familiarity [for the question “Regardless if this event has happened to you before, how familiar do you feel it is to you from TV, friends, movies, and so on?” the responses ranged from 1 (not at all familiar) to 6 (very familiar)] and negativity [for the question “How negative would this event be for you?” the responses ranged from 1 (not negative at all) to 6 (very negative)]. To test memory for the information presented, participants are asked to provide the actual probability previously presented of each event.

2.1.2 Differences in design between chapters

In **Chapter 3**, I use a version of the UBT which is exactly as described above only participants complete the task whilst in an fMRI scanner. Therefore responses are made by the participant via an MRI compatible button box. Due to this box only having 8 buttons, this meant that in the happen sessions, participants could only enter the digits 0 through to 7 (no 8s or 9s) in the happen sessions and could only enter the digits 2 to 9 in the not happen sessions (no 0s or 1s).

In **Chapter 4**, Experiments I and II each use a version of the UBT task in which only 40 events were used (each participant was presented with all events from one of the 2 lists of events in **Table 2.1**, the lists used were counterbalanced over participants). Since only 1 list of events was used for each participants and framing has not been shown to affect results previously (Sharot et al., 2011), all events were presented in the happen framing for these two studies. Furthermore, the information (base rates) was not redisplayed in the second session when participants provided their 2nd estimates. In Experiment II, due to time constraints, participants were only asked to provide memory estimates and subjective ratings for half of the events.

In **Chapter 5**, I use a version of the UBT which is as described above only participants estimate both the likelihood of each event occurring to themselves and how likely they think the event is to occur on average. In addition, after being presented with the information (base rate) in the first session, participants rate how desirable/undesirable they find this information.

In **Chapter 6** I use an online version of the UBT in which both positive and negative life events are used. This uses an alternative list of events to the negative life

events displayed in **Table 2.1**. In this version, everyday life events are used (see **Chapter 6, Table 6.1** for events used in this study) and participants rate how likely the event is to occur in the forthcoming month (rather than in their lifetime). Participants rated how negative or positive life events were and events were classified as positive and negative life events for each participant based on these ratings. In this version, since 1 list of 54 events was used for each participant (as opposed to 2 lists of 40 events each) and framing had not been shown to affect results previously (Sharot et al., 2011), for all events participants estimated the likelihood of the event *happening* to them in the future; they did not estimate the likelihood of the event *not happening* to them in the future. Also, participants were only shown base rates in the first session (they did not see these again in the second session when providing their 2nd estimates). Participants were allowed to enter responses between 5% and 95% in this version. Subjective ratings for Familiarity, Vividness and Emotional Arousal were collected from 75% of participants in this study. Ratings for Past Experience, Negativity and Memory Recall was collected from all participants.

2.1.3 Behavioural analysis

2.1.3.1 Update and information integration

Behavioural analysis is conducted as in previous studies (Chowdhury et al., 2013; Korn et al., 2013; Moutsiana et al., 2013; Sharot et al., 2011; Sharot et al., 2012a; Sharot et al., 2012b) using IBM SPSS statistics (version 19). All actual (base rates) and estimated percentages in ‘not happen’ sessions are transformed into the corresponding numbers of the ‘happen’ sessions by subtracting the respective number from 100. For each participant, trials are classified according to whether the participant initially overestimated or

underestimated the probability of the life event relative to the base rate presented. Specifically, if their initial estimate was lower than the base rate, this information would be categorised as “bad news”. If their initial estimate was higher than the base rate, this information would be categorised as “good news”². Trials in which the initial estimate was equal to the base rate are excluded from subsequent analyses since these cannot be categorized into either condition.

For each event, in each session, an estimation error term is calculated as the absolute difference between the participant’s estimate and the corresponding base rate

(Figure 2.1):

$$\textit{Estimation Error} = | \textit{First Estimate} - \textit{Base Rate} |$$

Update is the difference between participants first and second estimates but this difference is calculated depending on whether the trial is categorised as bad news:

$$\textit{Update (Good News)} = \textit{First Estimate} - \textit{Second Estimate}$$

$$\textit{Update (Bad News)} = \textit{Second Estimate} - \textit{First Estimate}$$

Thus, positive updates indicate a change toward the probability presented and negative updates a change away from the probability presented.

To explore the relationship between estimation errors and update, for each participant, two linear regressions are conducted entering estimation errors as independent

² Note that in **Chapter 3** which uses both positive and negative life events, this classification only applies for the negative life events. For positive life events, trials are classified in the opposite way. i.e. if initial estimate was lower than the average presented, this information would be categorised as “good news” and if initial estimate was higher than the average presented, this information would be categorised as “bad news”.

measures and updates as dependent measures – one for trials in which participants received good news and one for trials in which participants received bad news.

$$\text{Update (Good News)} = \alpha_G (\text{Estimation Error})$$

$$\text{Update (Bad News)} = \alpha_B (\text{Estimation Error})$$

This generates separate learning parameters for good news (α_G) and for bad news (α_B) for each participant. Learning parameters can then be evaluated at the group level (for example, to see if learning parameters are greater for good news than for bad news by conducting a paired sample ttest) and used as an individual difference measure to relate to other variables of interest (for example to correlate with depression symptoms in **Chapter 3**, or with stress scores in **Chapter 4**).

2.1.3.2 Control variables

Memory errors are calculated as the absolute difference between the base rate and the participants' recollection of that statistic:

$$\text{Memory Error} = | \text{Base Rate} - \text{Recollection of Probability Presented} |$$

Mean memory errors are then calculated for good news and bad news for each participant to examine whether there are differences in memory recall for the information provided in the UBT (Sharot et al., 2011). Similarly, mean rating scores for each rating that participants provide for each event (emotional arousal, familiarity, negativity, past experience and vividness) are calculated for good and bad news events and these compared to examine if differences exist on any of the dimensions. Note, that there is no need to correct for multiple comparisons when examining this since the aim of this analysis is to identify potential confounding factors (i.e. by not applying a correction such as Bonferroni, the analyses are more sensitive).

It is also important to control for participant priors, to rule out the possibility that valence effects are the result of the starting position of participants beliefs. For example, if one believes in general that negative life events are unlikely to happen to them (low prior) they may get (and expect to get) bad news more often in the task (and vice versa if one has a high prior). So valence effects found in the task could be attributed to other explanations (such as information expectation) if care is not taken to control for priors. This can be done by adding participants' mean first estimate as a covariate in the relevant analysis.

In addition, it is important to control for differences in estimation errors (see section 2.1.4 below).

2.1.4 Base rate distribution and estimation errors

When investigating biases in belief updating it is important to control for estimation errors. Running simulations that fail to ensure this create spurious biases in updating unless one controls for *estimation errors*, see below. Note that the purpose of these simulations is to demonstrate that skewing the distribution of base rates can artificially produce a bias in updating. It is not an attempt to try to identify and/or capture the precise cognitive processes participants engage in when undertaking the task. In the simulations I present here I invoke very basic assumptions about human behaviour, the idea being that if these nonetheless reveal updating patterns similar to human participants there is a confound in the task which needs to be addressed.

Consider four lists of base rates for life events; the first is skewed towards the bottom end (i.e. the majority of base rates are rare and fall below the midpoint of 50%), (**Figure 2.2a**); the second is skewed towards the top end (i.e. majority of base rates are

common and fall above the midpoint of 50%, (**Figure 2.2b**); the third and fourth are normally distributed around 50%, (**Figure 2.2c** and **2.2d**, this is the actual set used in the study reported in **Chapter 6**).

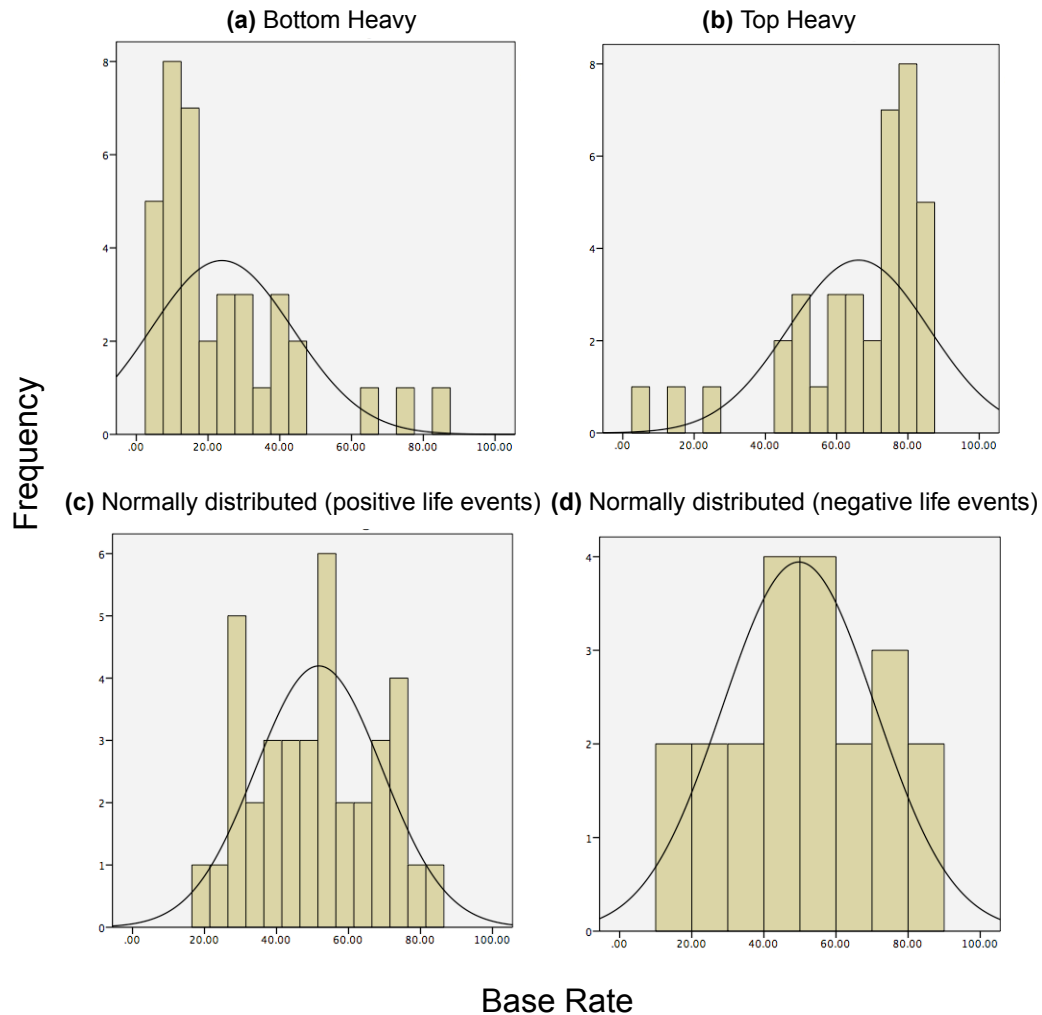


Figure 2.2: Histograms of base rates distributions used in simulations. Simulations were run using different distributions of base rates to examine the conditions under which an update bias is artificially produced. **(a)** Base rates are skewed towards the bottom end (i.e. the majority of base rates are rare). **(b)** Base rates are skewed towards the top end (i.e. majority of base rates are common). **(c, d)** Base rates used in the study reported in Chapter 6 which are normally distributed around a mean of 50% both for **(c)** positive life events and **(d)** negative life events.

For each simulation I randomly generate a *first estimate* on every trial for each virtual participant. This will be a random integer drawn from a uniform discrete

distribution between 5 and 95. I then “present” the simulator with the actual base rate for the event on that trial (*information*) and generate a *second estimate* - a random integer between the first estimate and the information (also drawn from a uniform distribution). For example, for the question “how likely are you to go out of town for leisure in the upcoming month” the simulator may randomly select 10% (first estimate), it will then observe a base rate of 36% (information) and adjust its answer to a random number between 10 and 36, let’s say 30% (second estimate). Thus, update on this trial would be 20 (update is calculated such that positive numbers always indicate a move towards the base rate).

I run 1000 such simulations (i.e., “experiments”) for each set of base rates (**Figure 2.2**). If the data produced by a simulation results in biased updating, this will indicate that the bias is due to a statistical artefact (the result of the mathematical constraints of the task) and not to an asymmetry in human learning. If, however, the simulation produces no bias in updating, but a bias is observed for human data, this would suggest the bias is due to asymmetric learning not to a statistical artefact.

The simulations clearly show that when base rates are skewed, an artificial bias in belief updating is observed (**Figure 2.3a** and **2.3b**), but not when the base rates are normally distributed (**Figure 2.3c**). Interestingly, when an artificial bias is detected it is observed in opposite directions for positive and negative life events creating a distinct “flip”. Specifically, when base rates are skewed towards the bottom end of the scale (rare events, **Figure 2.2a**), the simulation shows greater update for bad news than good news for positive life events (significant difference in 100% of the simulations run, **Figure 2.3a**), while for negative life events update for good news is greater than bad news

(significant difference in 100% of simulations, **Figure 2.3a**). However, when the base rates are skewed towards the top end of the scale (common events, **Figure 2.2b**), the opposite flip is observed (**Figure 2.3b**); larger update for good news than bad news for positive life events (significant difference in 100% of cases) and the opposite pattern for negative life events (**Figure 2.3b**, significant difference in 100% of cases). Finally, when the base rates are *normally* distributed the simulation does not reveal asymmetric updating (**Figure 2.3c**, significant difference in only 5% of cases for positive and only 5% of cases for negative life events).

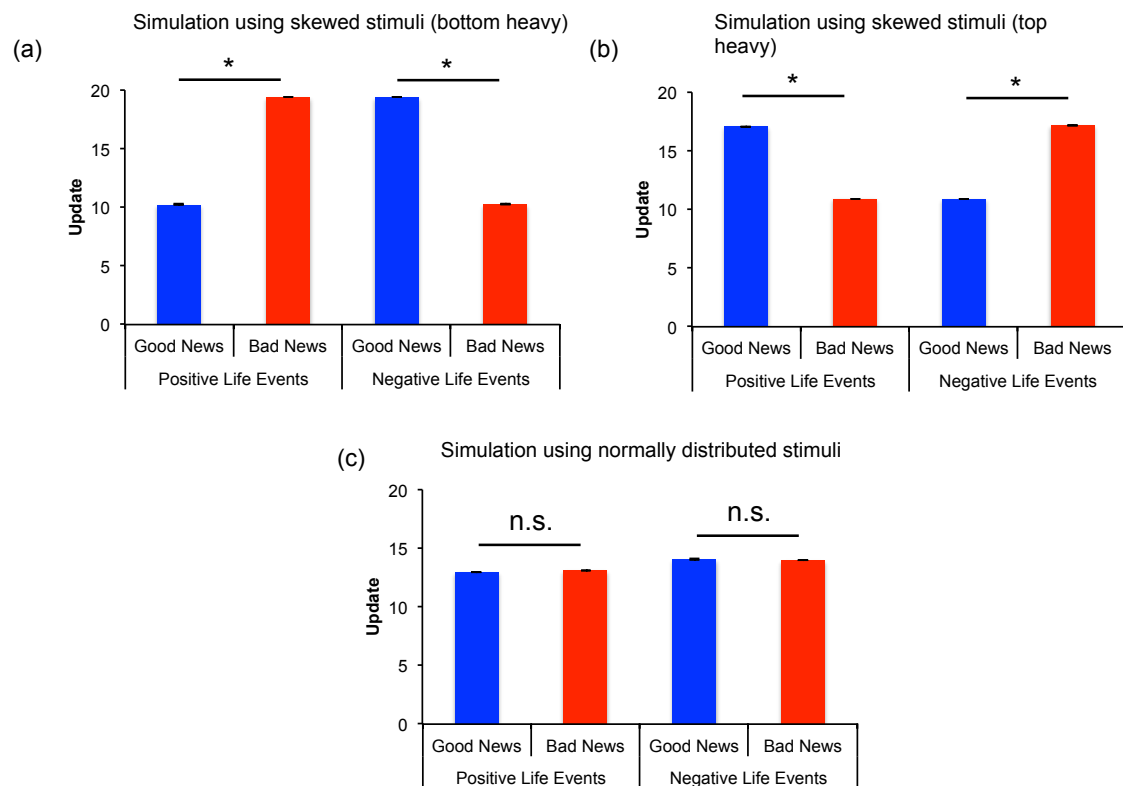


Figure 2.3: Simulations Results. For each group of base rates portrayed in Figure 2.2 I ran simulations to examine patterns of updating. When base rates were skewed, an artificial update bias was revealed, which flips for positive and negative life events (these artificial biases are abolished when controlling for estimation errors). Simulation for **(a)** base rates that are skewed towards the bottom end of the scale (rare events) and **(b)** base rates that are skewed towards the top end of the scale (common events). **(c)** When base rates are normally distributed there is no artificial bias in updating. *Error bars represent SEM, * $p < 0.05$, two-tailed paired sample t-test*

Why is an artificial bias produced for stimuli with a skewed distribution? If most base rates are skewed towards large numbers then on average there is more room to alter estimates when the first estimate is smaller than the base rate than when it is larger. In other words the difference between the first estimate and the information given (i.e. the *estimation error*) will be larger when subjects receive good news for positive events and when they receive bad news for negative events. Thus, updates will be greater when information is “good” for positive events and “bad” for negative events, and vice versa when base rates are skewed towards low numbers. This statistical artefact, however, can be corrected by controlling for “estimation errors”. If in the simulations above estimation errors are controlled for, no bias is observed for any set of base rates. This was tested by running an additional 20 simulations (10 for positive life events, 10 for negative life events, 20 participants per simulation) for each set of skewed base rates. For each set of simulated data I then conducted a repeated measures ANOVA with valence (good news/bad news) as a factor and entering the difference in estimation errors between good news and bad news as a covariate. When base rates were skewed towards the bottom end of the scale (**Figure 2.2a**) a valence effect between good news and bad news was not significant in 19 (95%) of these simulations. When base rates were skewed towards the top end of the scale (**Figure 2.2b**) a valence effect between good and bad news was no longer significant in all (100%) of these simulations.

Thus, to avoid false conclusions, researchers should either use normally distributed base rates with a mean at the midpoint of the scale and/or control for estimation errors, or both. In the case of the studies presented in this thesis, the distribution of base rates of the events used in Chapters 3, 4 and 5 is positively skewed (albeit less extreme than the base

rate distributions used in the simulations above, **Figure 2.2a**). Estimation errors are controlled for in each of these studies therefore. In Chapter 6, the events used are normally distributed (they are the actual base rates used in the simulations for normally distributed stimuli, **Figure 2.2c** and **Figure 2.2d**) and in addition I control for estimation errors.

2.1.5 Power analysis

The effect size index (d_z) for comparing update for bad and good news (using a paired sample t-test), calculated from the original study that used the UBT (Sharot et al, 2011), is estimated to be $d_z = 1.00$ (Mean Update [standard deviation]: bad news: 7.66[3.77]; good news: 11.92[4.08]). Using this effect size, to detect a difference in updating between good news and bad news requires a sample size of $n = 10$ to achieve a power level of 80% ($\beta = 0.2$) and a false discovery rate of 5% ($\alpha = 0.05$). All studies reported in this thesis use a sample size of $n > 10$. The effect size index for detecting differences in updating (from bad news) between participants (taken from Sharot et al., 2012a) was estimated to be $d_z = 1.18$. Using this effect size to detect a difference between two groups requires a sample size of $n = 13$ per group to achieve a power level of 80% ($\beta = 0.2$) and a false discovery rate of 5% ($\alpha = 0.05$). All studies reported in this thesis that examine group differences in updating (**Chapter 3** and **Chapter 4**) use a sample size of $n > 13$ per group. Power calculations were undertaken using G*Power (<http://www.gpower.hhu.de/en>).

2.2 Physiological measures

2.2.1 fMRI; Chapter 3

Functional Magnetic Resonance Imaging (fMRI) is a non-invasive technique which detects the oxygen concentration of blood in different parts of the brain. Since the onset of neural activity necessitates a change in this concentration, changes in Blood Oxygen Level Dependant (BOLD) signal can be used to infer changes in neural activity over time. This has been validated by studies which have simultaneously recorded neural (electrophysiological recordings) and BOLD (using fMRI) signal in primates (Logothetis et al, 2001; Goense and Logothetis, 2008).

fMRI takes advantage of the fact that oxygenated blood (which is diamagnetic) has different magnetic properties than deoxygenated blood (which is paramagnetic). As a result, the rate at which protons present in haemoglobin (an oxygen transporting protein in red blood cells) realign with the magnetic field depends on the ratio of oxygenated to deoxygenated red blood cells. Importantly, this ratio changes during and directly following brain activity as active neurons mobilise oxygen (Ogawa et al., 1992). Specifically, there will be an initial *increase* in deoxygenated blood as neurons use up available oxygen sources. This will be followed by a surge in oxygenated blood as haemoglobin replenishes oxygen that has been depleted during neural activity. This increase in oxygenated blood and corresponding reduction of deoxygenated blood can be detected as an increase in the MRI signal (Ogawa et al., 1992), which usually peaks 5-6 seconds after onset of the neural activity itself.

Before functional scans can be analysed, a number of pre-processing steps are carried out. There are a number of these steps available (http://www.fil.ion.ucl.ac.uk/spm/doc/spm5_manual.pdf), however only the steps applied in **Chapter 3** are covered here. The first pre-processing step is to make adjustments to images to account for head movement of participants whilst in the fMRI scanner. Such movements effectively mean that different neurons occupy a voxel space previously occupied by other neurons (in a nearby space). To account for this, realignment uses the first image in the sequence as a reference image and applies a rigid-body affine transformation. However, since this procedure does not completely correct for movement artefacts (Andersson et al, 2001), two other approaches can be applied to further reduce noise in the signal related to movement artefacts. The first is to apply unwarping to account for interactions between inhomogeneities (distortions) in the magnetic field and head movement. The second is to include the 6 movement parameters estimated from the realignment procedure as covariates in the design matrix (see below). Both of these additional corrections were used in the analysis reported in **Chapter 3**.

Owing to heterogeneity in the size of human brains, to compare images across different participants in a study and allow anatomical comparisons to be made between different studies (for instance, that the same area of the brain is engaged during a specific cognitive process in different studies), images need to be converted into a common space. Normalization is the process by which each image is warped into a standard space based on a specified template. In the analysis reported here (**Chapter 3**) the Montreal Neurological Institute (MNI) reference brain was used.

The final pre-processing step is to spatially smooth the images by applying a filter. This improves the signal to noise ratio by removing high frequency information (i.e. small scale changes in the image). In the analysis reported in **Chapter 3**, this was done by convolving the images with Gaussian kernels with full-width at half-maximum (FWHM) of 8mm.

There are a wide variety of ways to analyse fMRI data depending on the nature of the research question being asked. The approach taken in **Chapter 3** is to conduct a mass univariate approach. Simplifying, this essentially conducts a separate multiple regression for each and every voxel in the brain. In each of these regressions, the time course of the BOLD response at a particular voxel is the dependent variable³. Variables specified by the experimenter constitute the independent variable(s), input as a design matrix. In the experiment reported in **Chapter 3**, independent variables of interest comprise two types: (1) categorical regressors which specify when events of specific types occurred (for instance when good news was presented, when bad news was presented) (2) parametric regressors which modulate a categorical regressor with a quantity of experimental interest such that the regressor scales according to that quantity, allowing one to examine how BOLD response covaries with it.

For each voxel, beta parameters along with their corresponding F statistic (i.e. “usefulness” in accounting for variance in the BOLD response) are estimated for each independent variable included in the linear model using restricted maximum likelihood. These statistics can then be viewed in the form of a statistical map, highlighting for instance, voxels with high t statistics for a particular model parameter.

³ Since observations in the time series are not independent (i.e. there is often correlation in the BOLD response from one scan to another), a correction is applied to ensure sphericity (Glaser, 2004).

To reliably interpret the significance of these results however it is necessary to correct for the large number of comparisons made that comes out of running a regression for each and every voxel. One means of doing this is to apply a Bonferroni correction according to the number of voxels. Whilst this stringently guards against Type I errors, it also runs the risk of incurring Type II errors (i.e. failing to detect significant effects that are present in the data), particularly since voxels close to one another will often have very similar BOLD response profiles. Hence, less conservative approaches are often adopted.

One of these is to abandon looking at the significance of betas across the whole brain and instead restrict the search area to a specific anatomical region, defined *a priori* (for instance on theoretical grounds or from regions identified in previous comparable studies). This means that a correction is only applied for the number of voxels in the specified region (so called small volume correction). In the analysis presented in **Chapter 3** this was a viable option as a similar fMRI study had previously been run using the UBT (Sharot et al., 2011) and identified the right Inferior Frontal Gyrus as a region tracking bad news.

Another approach is to correct for the number of clusters (i.e. groups of voxels) rather than the number of voxels. However, in order to do this there needs to be some initial significance threshold set in order to generate clusters in the first instance. Setting a p value such as $p < 0.001$ acts as a filter on the map of voxel by voxel t statistics by only including voxels that meet this threshold of significance and correcting for the number of clusters that form at this threshold. In the analysis in **Chapter 3**, this was also adopted as an approach with activity reported cluster level corrected with the cluster defining threshold used set at $p < 0.001$.

2.2.2 Skin conductance and cortisol; Chapter 4

Exposure to stress causes alterations to brain function via two systems: the hypothalamic pituitary adrenal (HPA) axis and the sympathetic component of the autonomic nervous system (de Kloet et al., 2011; Lupien and McEwen, 1997).

The former causes the secretion of glucocorticoids from the adrenal cortex which acts to mobilise energy resources by increasing the glucose content of blood. The predominant glucocorticoid in humans is cortisol and this can be measured from saliva as well as blood and urine (Kirschbaum and Hellhamer, 1994). The process by which cortisol is secreted in response to a stressor is relatively gradual; peak cortisol response usually occurs 30-40 minutes after onset of a stressor and returns to baseline (pre-stressor levels) around 1 hour after the stressor has ended (Dickerson and Kemney, 2004). In Experiment I reported in **Chapter 4**, saliva samples were taken at 4 different time intervals across the experiment to measure cortisol response over an extended period.

Changes to the sympathetic component causes the release of catecholamines adrenaline and noradrenaline. Unlike the HPA, this autonomic component is fast acting, occurring immediately after onset of the stressor and returning to baseline around 10 minutes after the stressor has ended (Kirschbaum et al., 1993). The release of catecholamines results in a number of physiological responses including increase in heart rate, blood pressure and electrodermal responses. Skin conductance is one specific type of electrodermal activity which can be measured by applying an external electrical current at a constant voltage level to the skin. Sweat from eccrine sweat glands changes the conductivity of the skin and is regulated by the sympathetic component of the autonomic nervous system originating in the hypothalamus. Hence changes in skin conductance

(from sweating) are believed to reflect general changes in the autonomic nervous system due to psychological processes such as arousal (Figner and Murphy, 2011).

Analysis of skin conductance can be approached in one of two ways depending on what the experimenter is attempting to tie conductance responses to. One approach is to look at Skin Conductance *Level* (SCL). This examines mean conductivity over a specified time interval (such as 30 seconds or an experimental block) providing a measure of general arousal levels (i.e. tonic activity) during this period. This is useful for instance if one wants to see if participants are in a higher state of arousal in one condition than another (by comparing SCLs between conditions), or to ascertain if a stress manipulation has been effective (e.g. by comparing SCL pre and post manipulation). Another approach is to look at Skin Conductance *Response* (SCR). This examines the phasic response of skin conductance; variation in the signal over a short time scale (usually a few seconds). This is useful if one wants to look at how skin conductance is modulated on a trial by trial basis such as in response to specific types of stimuli. In this way a clumsy analogy is to say that analysis of SCR can serve a similar role to that of BOLD signal in fMRI in that it allows one to infer a hidden variable (e.g. sympathetic nerve firing or neural response) from observable data (some kind of timecourse be that SCR or BOLD signal).

In Experiment I reported in **Chapter 4**, I was interested in ascertaining whether a stress manipulation was effective in inducing a sympathetic response. Hence mean SCL pre and post manipulation (each recorded over 2 minute periods) were compared.

2.3 Mini-International Neuropsychiatric Interview; Chapter 3

The Mini International Neuropsychiatric Interview (MINI) (Sheehan et al., 1998) is based on the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) used by clinicians and researchers to diagnose and classify mental disorders. It is a structured interview carried out by a trained researcher which proceeds through sections, with each section asking questions about a specific mental disorder. These include; Major Depressive Disorder (MDD), Bipolar Disorder, Obsessive Compulsive Disorder, panic disorder, agoraphobia, post-traumatic stress disorder, alcohol and substance abuse/dependence, psychotic symptoms, anorexia, bulimia and generalized anxiety disorder.

Training to conduct the MINI with a participant involved 2 stages: (1) myself observing 6 separate MINIs administered by a researcher who was already trained in administering the MINI; (2) myself being observed by a trained researcher carrying out 4 MINI interviews with feedback provided after each interview.

The MINI was used in **Chapter 3** to diagnose potential MDD patients with depression and to examine other disorders depressed patients were experiencing or had experienced in the past. It was also used to ensure that control participants did not have depression or any other mental disorders either at the time of study or previously.

2.4. Mood/Personality Questionnaires

2.4.1 Beck Depression Inventory; Chapters 3-6

The Beck Depression Inventory (BDI) (Beck et al., 1961) is a 21-question self-rating scale used to provide a measure of individual's current depression symptoms. For each of the 21

items, participants select 1 statement from 4 available which are scored from 0 to 3 (example: 0 = I do not feel sad; 1 = I feel sad; 2 = I am sad all the time and I can't snap out of it; 3 = I am so sad or unhappy that I can't stand it). Participants are asked to select the statement that most concurs with how they have been feeling the past few days. Scores across all 21 items are summed to create an overall score. Scores range from 0 to 63 with the following scores (Beck et al., 1988) suggested as guidelines:

- Scores below 10 indicate normal mood.
- Scores 10-18 indicate mild depression.
- Scores 19-29 indicate moderate depression.
- Scores 30 and above indicate severe depression.

2.4.2 State Trait Anxiety Inventory; Chapters 3 and 4

The STAI-6 (Martean and Bekker, 1992) is a short version of the Spielberger State Trait Anxiety Inventory (Spielberger, et al., 1983) which has shown reliability and sensitivity to different degrees of anxiety (Martean and Bekker, 1992). Participants report their current anxiety state according to 6 statements (e.g. I am worried) on a 4-point Likert scale (1 = *not at all* to 4 = *very much*). Scores range from 6 to 24 with high scores indicating high levels of anxiety.

2.4.3 Life Orientation Test Revised; Chapters 3-6

The Life Orientation Test Revised (LOT-R) is a 10-question self-rating scale used to provide a measure of individuals' trait optimism (Scheier and Carver, 1985; Scheier et al., 1994). Participants rate how much they agree with 10 statements (for example: In

uncertain times I usually expect the best) on a scale from 1 to 5 (1 = I disagree a lot; 2 = I disagree a little; 3 = I neither agree nor disagree; 4 = I agree a little; 5 = I agree a lot). The convention is to convert these to a scale from 0 to 4 (i.e. 1 is subtracted from participants responses). Four of the items on the questionnaire are filler items and are not used for the scoring. Three of the remaining items are reversed scored (i.e. 0=4; 1=3, 2=2; 3=1; 4=0). The six items are then summed to create an overall trait optimism score. Scores can range from 0 (pessimistic) to 24 (optimistic). Reliability of the LOT-R in measuring trait optimism has been found across a number of studies (e.g. Robinson-Whelen et al., 1997; Carver et al., 2005; Sharot et al., 2007; Puri and Robinson, 2007; Rasmussen et al., 2009).

Chapter 3

Neural substrates of Unbiased Belief Updating in Depression

3.1 Introduction

Contrary to traditional psychological theories that maintain that good mental health is subserved by accurate beliefs in relation to reality, a large body of literature suggests that biases may promote adaptive functioning. In particular, positive illusions (including overly positive evaluation of the self, unrealistic optimism and an exaggerated sense of control) are argued to enhance mental health by encouraging productivity, social interaction, subjective happiness and physical health (McKay and Dennett, 2010; Taylor et al., 2003); but see (Colvin and Block, 1994; Compton, 1992). In healthy individuals, positive illusions are especially apparent under circumstances of adversity (Taylor and Armor, 1996) which may enhance resiliency to stressful life events. By contrast, moderately depressed individuals have been reported to display a less positive, but relatively unbiased, view of the self (Coyne and Gotlib, 1983), the future (Strunk and Adler, 2009) and sense of control (Alloy and Abramson, 1979) – dubbed “depressive

realism” (but see: Allan et al., 2007; Moore and Fresco, 2012). Importantly however, severely depressed individuals often show negative biases in these domains (Roiser et al., 2011) which can predict fatal outcomes (Oquendo et al., 2004).

It has recently been found that depressed patients lack the positive skewed belief updating bias seen in healthy controls when they update their expectations about the future (Korn et al., 2013). These findings raise an intriguing question as to what differs in depression in terms of underlying neural substrates that support an unbiased belief formation about the future.

Previously it has been demonstrated that, in healthy participants, biased updating in response to positive and negative news is mediated by a relatively weak correlation between brain activity and negative estimation errors, but intact coding of positive estimation errors (Sharot et al., 2011). Here, utilizing the belief-updating task in combination with functional brain imaging I ask whether depression is associated with neural responses that are likely to support a more unbiased integration of information about the future.

3.2 Materials and Methods

3.2.1 Participants

Thirty individuals aged 18 to 65 participated in the study (half unmedicated depressed patients and half healthy controls). Depressed participants were identified through Camden and Islington Foundation Trust Psychological Treatment Services or recruited by advertisement. Healthy controls were recruited from UCL psychology subject pool and matched to depressed participants for age, gender and level of education. One control

participant was subsequently excluded from the analysis due to a high score on the BDI (>10). None of the participants had taken antidepressant medication for at least six weeks prior to undertaking the study due to a variety of personal choices unrelated to the study itself. No participants had a period of substance or alcohol abuse in the six months prior to undertaking the study. See **Table 3.1** for demographic and clinical information.

	Controls	MDD Patients
N (Male)	14 (9)	15 (9)
Age¹	30.36 (8.35)	31.47 (9.16)
Level of Education*²	2.64 (0.81)	2.60 (0.88)
BDI³	2.21 (2.75)	25.80 (9.97)
N with History of Alcohol/Substance Abuse	0 (0%)	4 (26%)
N receiving psychotherapy	0 (0%)	2 (13%)
Age of Onset of First Depressive Episode	0 (0%)	18.64 (7.01)
N with at Least Two Depressive Episodes	0 (0%)	14 (93%)
N Previously Attempting Suicide	0 (0%)	2 (13%)

Table 3.1: Demographic and clinical characteristics of participants. Figures represent mean (SD) unless stated otherwise * Coded as follows: 1) High School; 2) Bachelors Degree; 3) Masters Degree; 4) PhD; ¹Independent sample t test: $t(27) = 0.34 (>0.73)$; ²Independent sample t test: $t(27) = -0.13 (>0.89)$; ³Independent sample t test: $t(27) = 8.71 (<0.05)$

Before the study, all participants were assessed for psychiatric disorders by a trained researcher using the Mini International Neuropsychiatric Inventory (MINI) (Sheehan et al., 1998). Training involved the researcher observing 6 separate MINI interviews and then being observed carrying out 4 MINI interviews with detailed feedback provided after each interview. The MINI confirmed that depressed participants had experienced depressive episodes in the past and met criteria for a major depressive episode at the time of undertaking the study. For controls, the MINI confirmed that participants had not experienced any depressive episodes during their lifetime. The MINI was also used to verify that participants in both groups had no other past or present psychiatric conditions, other than anxiety disorders in the depressed participants. Participants that did

not conform to any of the above were not invited to participate in the study further. All participants completed the BDI (controls: mean (range) BDI = 2.21 (0-9); MDD: mean (range) BDI = 25.80 (13-44)). The majority of depressed participants were mildly (n=5) or moderately (n=7) depressed, with a minority (n=3) severely depressed.

All participants gave informed consent and were paid for their participation. The study was approved by the London Queen Square Research Ethics Committee.

3.2.2 Procedure

The procedure was identical to a previous study (Sharot et al., 2011) and is outlined in **Chapter 2**. Participants went through three practice trials. The session began with a short structural scan, followed by four functional runs consisting of 40 trials each (all 80 events were presented twice). Finally, an additional longer structural scan was performed.

3.2.3 Behavioural task

The paradigm was adapted from previous studies (Chowdhury et al., 2014; Korn et al., 2013; Moutsiana et al., 2013; Sharot et al., 2011, 2012a, 2012b, 2012b) and depicted in **Experimental Methods, Chapter 2**. Participants had up to 6s to respond using a button box with four buttons in each hand. Each button corresponded to one digit. The digits 0 through 7 could be used to enter the estimated likelihoods in the ‘happen’ estimation and digits 2 through 9 in the ‘not happen’ estimation. If the participant failed to respond, then that trial was excluded from all subsequent analyses (mean trials with no response = 1.38, s.d. = 2.83). A fixation cross then appeared for 1–5 s (jittered). Next, the event description appeared again for 2 s, together with the average probability of that event to occur (or not

occur, depending on ‘happen’ or ‘not happen’ sessions). Finally, a fixation cross appeared for 1–3 s (jittered). See **Figure 3.1** for task timings.

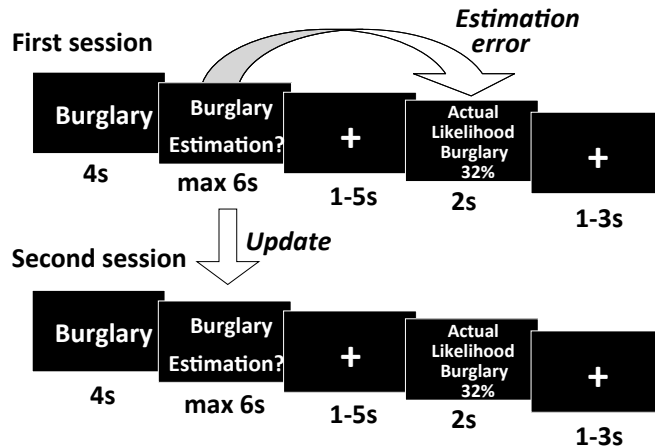


Figure 3.1: Timeline of task. On each trial, participants were presented with a short description of one of 80 adverse events (4 seconds). They were then asked to estimate how likely this event was to occur to them. Participants had up to 6 seconds to input a response. Following display of a fixation cross (1–5 seconds jittered) they were presented with the average probability of that event occurring to a person living in the same sociocultural environment as them (2 seconds). Trials ended with a fixation cross (1–3 seconds jittered).

3.2.4 Behavioural analysis

Behavioural analysis was conducted as described previously (Chowdhury et al., 2014; Korn et al., 2013; Moutsiana et al., 2013; Sharot et al., 2011, 2012a, 2012b, 2012b) and in **Experimental Methods, Chapter 2**. Trials in which the initial estimate was equal to the average presented were excluded from subsequent analyses (mean = 1.83 trials, s.d. = 1.65) as these could not be categorized into either condition. Average update scores were entered into a two (valence: good/bad) by two (group: MDD/Control) repeated-measures ANOVA.

Average memory error scores for good news and bad news were calculated for each participant and entered into a two (valence: good/bad) by two (group: MDD/Control) repeated-measures ANOVA. ANOVAs were also performed on scores of all other scales

(negativity, emotional arousal, vividness, familiarity, past experience) as well as on other task measures (initial estimates, reaction times, estimation errors, number of trials).

3.2.5 MRI scanning

Scanning was performed at the Wellcome Trust Centre for Neuroimaging at UCL using a 3T Siemens Allegra scanner with a Siemens head coil. Functional images were acquired as echo-planar (EPI) T2*-weighted images. Time of repetition (TR) = 2.73 sec, time of echo (TE) = 30 ms, flip angle (FA) = 90, matrix = 64 X 64, field of view (FOV) = 192 mm, slice thickness = 2 mm. A total of 42 axial slices (-30° tilt) were sampled for whole brain coverage, in-plane resolution = 3mm x 3mm.

3.2.6 fMRI data analysis

Statistical Parametric Mapping (SPM5, Wellcome Trust Centre for Neuroimaging, <http://www.fil.ion.ucl.ac.uk/spm/>) was used for fMRI data analysis. After discarding the first six dummy volumes, images were realigned to the seventh volume, unwarped, normalized to a standard EPI template based on the Montreal Neurological Institute (MNI) reference brain, resampled to 2 mm × 2 mm × 2 mm voxels and spatially smoothed with an isotropic 8-mm full-width at half-maximum Gaussian kernel. Low frequency artefacts were removed using a 1/128 Hz high-pass filter and temporal autocorrelation intrinsic to the fMRI time series was corrected using an AR(1) process. For each participant, a design-matrix was created with event onsets time-locked to the temporal positions of: event presentation; presentation of cue prompting response; motor response; and presentation of information (see **Figure 3.1** for timeline of task). These were modelled as durations of 4, 0, 0 and 2 seconds respectively. For all task components (except for motor

responses), regressors were subdivided into two conditions: trials of events for which participants received good news and trials of events for which they received bad news, resulting in seven regressors for each session. These events were convolved with a canonical hemodynamic response function (HRF) to create regressors of interest. Motion correction regressors estimated from the realignment procedure were entered as covariates of no interest.

To identify regions tracking estimation errors, absolute estimation errors were entered as parametric regressors modulating the events in which information was presented. For each condition (that is, for trials in which information was better than expected and trials in which information was worse) regions showing significant effects across *both* healthy and depressed participants were identified ($p < 0.05$, cluster-level corrected across the whole brain; images first thresholded at $p < 0.001$, uncorrected). Owing to the fact that past research (Sharot et al., 2011) shows that individual differences in the UBT are best predicted by region(s) inversely tracking bad news estimation errors, betas from the peak voxel in the region(s) tracking bad news estimation errors were extracted and compared between depressed participants and controls using an independent sample t test ($p < 0.05$) in SPSS. Previous research suggests that peak voxel activity can be a better predictor of electrophysiological measures of activation than average cluster (Arthurs and Boniface, 2003).

In addition, the right inferior frontal gyrus (rIFG) was selected as an a-priori ROI because previous findings (Sharot et al., 2011) showed that the degree to which BOLD signals in this region track bad news estimation errors differentiated between participants with high and low trait optimism. Thus it was examined whether a similar difference

existed between depressed and controls in the anatomically defined rIFG using small volume correction (SVC).

All activations are displayed on sections of the standard MNI reference brain. Anatomical labels were assigned using the Talairach Daemon database (University of Texas Health Science Center San Antonio; <http://www.talairach.org/>) according to peak voxels in Talairach and Tournoux coordinate space. rIFG was anatomically defined by creating an ROI mask using WFU Pickatlas (<http://fmri.wfubmc.edu/software/PickAtlas>).

3.3 Results

3.3.1 Unbiased updating in MDD, biased updating in controls

Unbiased updating in depressed individuals was found, but a valence-dependent updating bias in healthy individuals. Specifically, a group (MDD/healthy) by valence (good news/bad news) ANOVA revealed a significant interaction ($F(1,27) = 9.16, p < 0.01$, **Figure 3.2**). Replicating previous findings (Chowdhury et al., 2014; Korn et al., 2013; Moutsiana et al., 2013; Sharot et al., 2011, 2012a, 2012b), healthy participants updated their beliefs to a greater extent in response to good news relative to bad news ($t(13) = 5.00, p < 0.001$; 93% of healthy participants showed greater updating in response to good news). No such difference was observed in the MDD group ($t(14) = 1.49, p > 0.15$; 60% of depressed participants showed greater updating in response to good news, see also (Korn et al., 2013) for similar findings in hospitalized, medicated, depressed patients). The interaction was further characterised by greater updating in response to bad news ($t(27) = 2.96, p < 0.01$) in the MDD group compared to the healthy controls, with no significant difference in updating between groups in response to good news ($t(27) = -0.37, p > 0.70$).

These results suggest that depression, in contrast to good mental health, is related to a lack of discounting of bad news, resulting in unbiased updating of beliefs in response to good and bad news.

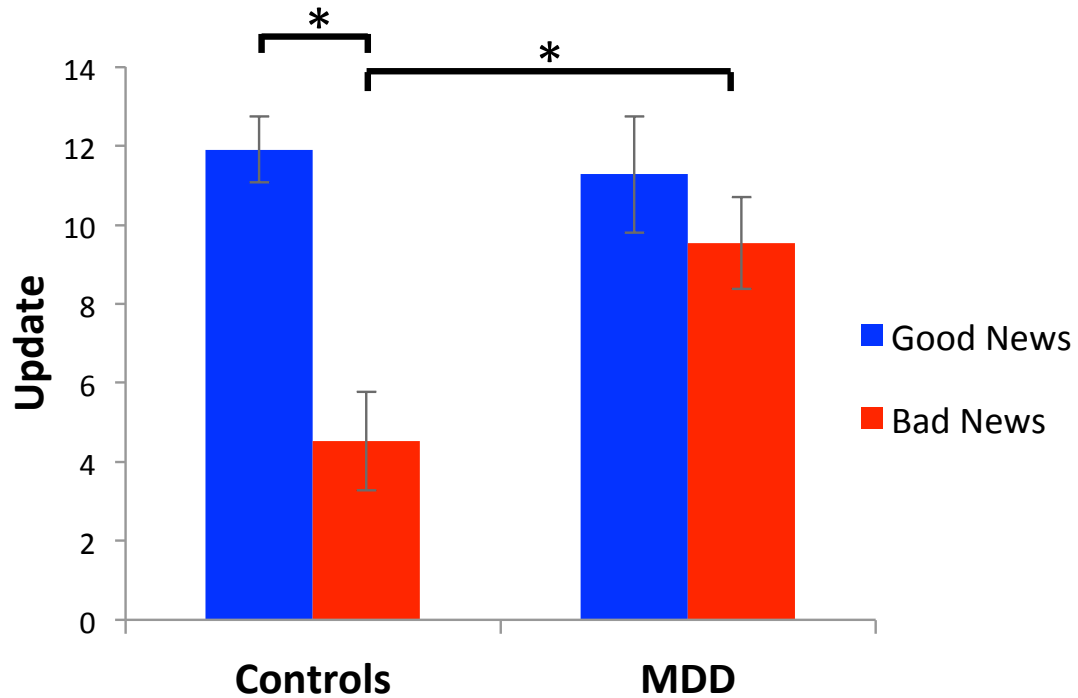


Figure 3.2: Unbiased updating in MDD but biased updating in controls. After receiving good news that presented an opportunity to adjust beliefs in a positive direction, healthy participants updated their estimations to a greater extent than after receiving bad news that called for adjustments in a negative direction. In contrast, depressed participants updated their beliefs to a similar extent after receiving good and bad news, and updated their beliefs more than healthy individuals when receiving bad news. Error bars represent s.e.m. * $P < 0.05$, two-tailed independent/paired samples t-test.

3.3.2 Control variables

To examine whether the relationship between depression and updating could be explained by any other factor, I tested for a relationship between depression and all other variables recorded.

3.3.2.1 Memory

After the scanning session, participants were asked to indicate the actual probability (as previously presented) of each event occurring on average. Memory errors were calculated as the absolute difference between the actual probability previously presented and the participants' recollection of that statistic (see **Experimental Methods, Chapter 2**). Memory errors did not differ between groups (see **Table 5.2**) and there was no interaction with valence ($F(1,27) = 1.07, p = 0.31$).

3.3.2.2 Life event ratings

Participants rated life events on five scales (past life experience with the events, familiarity with the events, ability to imagine them vividly, emotional arousal and negativity). The scores revealed that past experience and familiarity with the adverse life events, as well as the ability to vividly imagine the events and the subjective sense of emotional arousal in response to the events, *did not differ* between MDD and healthy controls and did not interact with valence (see **Table 3.2** for statistics). However, how negative the participants rated the events did reveal a group by valence interaction ($F(1,27) = 6.76, p = 0.02$). MDD patients rated life events that they received bad news for during the experiment as less aversive than events they received good news for. This was the opposite in the case of healthy controls who rated life events for which they received bad news as more aversive than life events for which they received good news. Thus I repeated the main analysis of update scores while controlling for differential scores of negativity (i.e. scores on good news trials minus scores on bad news trials). After entering these scores as covariates, the group by valence interaction on update scores remained

significant ($F(1,26) = 5.10, p < 0.05$). Thus, differential update *could not* be explained by differences in the degree of the perceived negativity of the events, by familiarity or by past experience with the events, by whether the events were imagined vividly or experienced as more or less emotionally arousing.

3.3.2.3 Task factors

There were no differences across groups in the number of missed responses, nor the number of good news and bad news trials. These factors did not differ across valence nor did valence interact with group (see **Table 3.2** for statistics). Thus, MDD participants did not miss more responses than controls and were not more likely to encounter good news trials than bad news trials. The magnitude of the estimation errors did not differ between groups (bad news: $t(27) = -0.02, p > 0.99$, good news $t(27) = 0.64, p > 0.53$). Participants' priors - their initial estimates of the probability of the events - did not differ across groups however they did correlate with BDI scores ($r = 0.38; p < 0.05$). In other words, the more depressed the individual, the more likely they were to estimate their chances of encountering aversive events as greater (also see Strunk et al., 2006). Initial estimates also correlated with updating for good news ($r=0.52, p<0.01$) and for bad news ($r = 0.41, p<0.05$). Reaction time for first estimates did not differ between groups. However, reaction time for second estimates did reveal a valence by group interaction ($F(1,27) = 6.24, p < 0.05$) with MDD participants slightly faster than controls to re-estimate their likelihood of encountering an event they previously received good news for. After entering the difference in second estimate reaction time as a covariate along with initial estimates and differential scores of negativity (see above), the group by valence interaction on update scores remained significant ($F(1,24) = 5.39, p < 0.05$).

Questionnaire and variables	MDD mean (SD)		CONTROLS mean (SD)	
	Good news	Bad news	Good news	Bad news
<i>Subjective Scales Questionnaire: All scales 1 = low to 6 = high</i>				
Familiarity ^v	3.88 (0.82)	3.62 (0.84)	4.06 (0.99)	3.67 (0.92)
Prior experience ^v	1.65 (0.31)	1.46 (0.24)	1.50 (0.68)	1.39 (0.64)
Vividness ^v	3.87 (0.84)	3.52 (0.75)	3.68 (0.89)	3.50 (0.81)
Emotional arousal	3.24 (0.79)	2.95 (0.70)	3.29 (0.85)	3.34 (0.95)
Negativity ^{v*g}	3.97 (0.71)	3.76 (0.61)	3.98 (0.68)	4.09 (0.67)
<i>Task-related variables</i>				
Memory errors	11.26 (4.37)	10.39 (3.38)	11.23 (3.29)	11.45 (1.86)
Initial estimates ^v	43.25 (7.35)	21.01 (3.2)	44.07 (5.22)	19.84 (4.47)
Reaction time first estimate (ms)	2183.01 (663.41)	2153.83 (578.97)	2298.73 (790.90)	2303.33 (716.04)
Reaction time second estimate (ms) ^{v*g}	1762.00 (533.63)	1819.19 (566.46)	1997.20 (648.74)	1891.72 (618.93)
Number of Trials	36.53 (12.44)	34.67 (8.55)	34.64 (12.30)	36.71 (7.04)

Table 3.2: Control Variables. ^v Main effect valence, $p < 0.05$; ^g Main effect group, $p < 0.05$ (NB: there was no group difference for any of the variables); ^{v*g} Interaction effect (group x valence), $p < 0.05$.

3.3.2.4 Framing

Whether participants were asked to estimate the likelihood of the events happening in the future, or never happening, did not alter the results. A group (MDD, healthy) by valence (good news/bad news) by frame (happen/not happen) ANOVA revealed the expected two-way interaction of group by valence ($F(1,27) = 4.92$, $p < 0.04$), which is driven by controls updating more on good news trials than bad and MDD showing unbiased updating. However, there were no other significant interactions with group. Note, that reaction times did not differ between frames ($t(28) = 0.96$, $p > 0.34$) nor groups ($t(28) = 1.09$, $p > 0.28$) nor was there an interaction between group and frame ($F(1,27) = 0.131$, $p > 0.7$).

3.3.3 Unbiased updating in MDD explained by adequate use and neural tracking of negative estimation errors

3.3.3.1 Learning parameters

Learning parameters for good (α_G) and bad (α_B) are calculated by quantifying the relationship, on a trial-by-trial basis for each participant, between an estimation error and the subsequent update (see **Chapter 2**). While participants learned from the information presented to them [mean learning parameters relating an individual's estimation errors to update was significantly different from zero; $t(28) = 10.79$, $p < 0.001$], their ability to do so was differentially related to depression symptoms as a function of valence. Specifically, the greater the depression symptoms of the participant, the greater the learning parameter (α_B) was for bad news. Learning parameters for good news (α_G) were not reliably associated with depression symptoms (**Figure 3.3**). This was evident in a positive correlation between BDI and learning parameters in trials when participants received bad news ($r = 0.36$, $p = 0.05$) and no correlation between BDI and learning scores when receiving good news ($r = -0.09$, $p > 0.62$). The difference between these two correlations was statistically significant ($Z = 2.17$ $p < 0.05$, Steiger's Z test).

These findings suggest a likely computational principle that mediates the observed unbiased belief formation in depression. Specifically, they point to estimation errors as providing a learning signal whose impact on update depends on an interaction between depressed mental state and whether this new information calls for an update in a positive or negative direction.

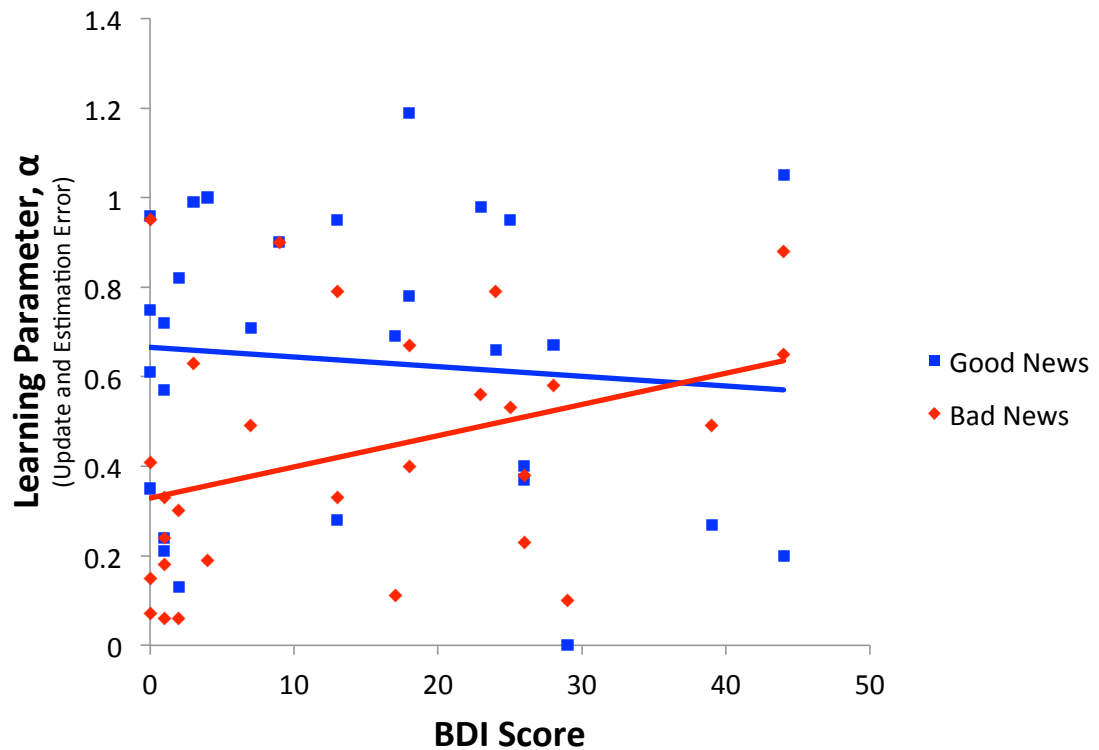


Figure 3.3: Relationship between depression (BDI) and learning parameters (α_G , α_B) from good and bad news.

3.3.3.2 fMRI data

Given the above results I next examined the fMRI data to identify how BOLD signals track estimation errors in response to information that entails a belief adjustment in either a positive or a negative direction in depressed and healthy individuals. Absolute estimation errors on each trial were entered as a parametric regressor modulating the time point at which participants were presented with information regarding the average probability of events. From this analysis, first I identified regions where BOLD signal correlated with estimation errors for either good or bad news on a trial by trial basis across *all* participants.

BOLD signal correlated positively with good news estimation errors in the left inferior frontal gyrus (left IFG: peak voxel in Talairach coordinates: $-50, 17, -4$; $k = 293$; $z = 4.31$, **Figure 3.4a**) and bilateral superior frontal gyrus (bilateral SFG: $-6, 60, 26$; $k = 174$; $z = 4.01$, **Figure 3.4b**). In addition BOLD signal correlated negatively with bad news estimation errors in the right inferior parietal lobule (right IPL: $65, -27, 36$; $k = 185$; $z = 4.30$, **Figure 3.4c**) and positively with bad news errors in Superior Temporal Gyrus ($-44, -50, 14$; $k = 1174$; $z = 4.92$) and Superior Frontal Gyrus ($-2, 50, 31$; $k = 203$; $z = 3.80$). There were no voxels in which activity correlated negatively with good news estimation errors. These 5 ROIs comprise the entire set of regions identified in the current dataset at this threshold (FWE cluster level corrected after voxel-wise thresholding at $p < 0.001$).

Next I examined whether the extent to which brain activity tracked estimation errors was related to depression. Past research (Sharot et al., 2011) has shown that brain regions in which BOLD signal *inversely* correlates with bad news estimation errors, predicts individual differences in trait optimism. I thus compared betas (relating BOLD signal to estimation errors) from the peak voxel in rIPL for participants in the MDD group and control group. Indeed, this revealed that BOLD response in the rIPL of depressed participants tracked bad news errors with greater fidelity than was the case for healthy controls ($t(27) = -2.27$, $p < 0.05$, **Figure 3.4d**; note that betas are negative, which indicate an inverse correlation – the larger the magnitude of the negative number, the stronger the relationship between BOLD signal and bad news estimation errors).

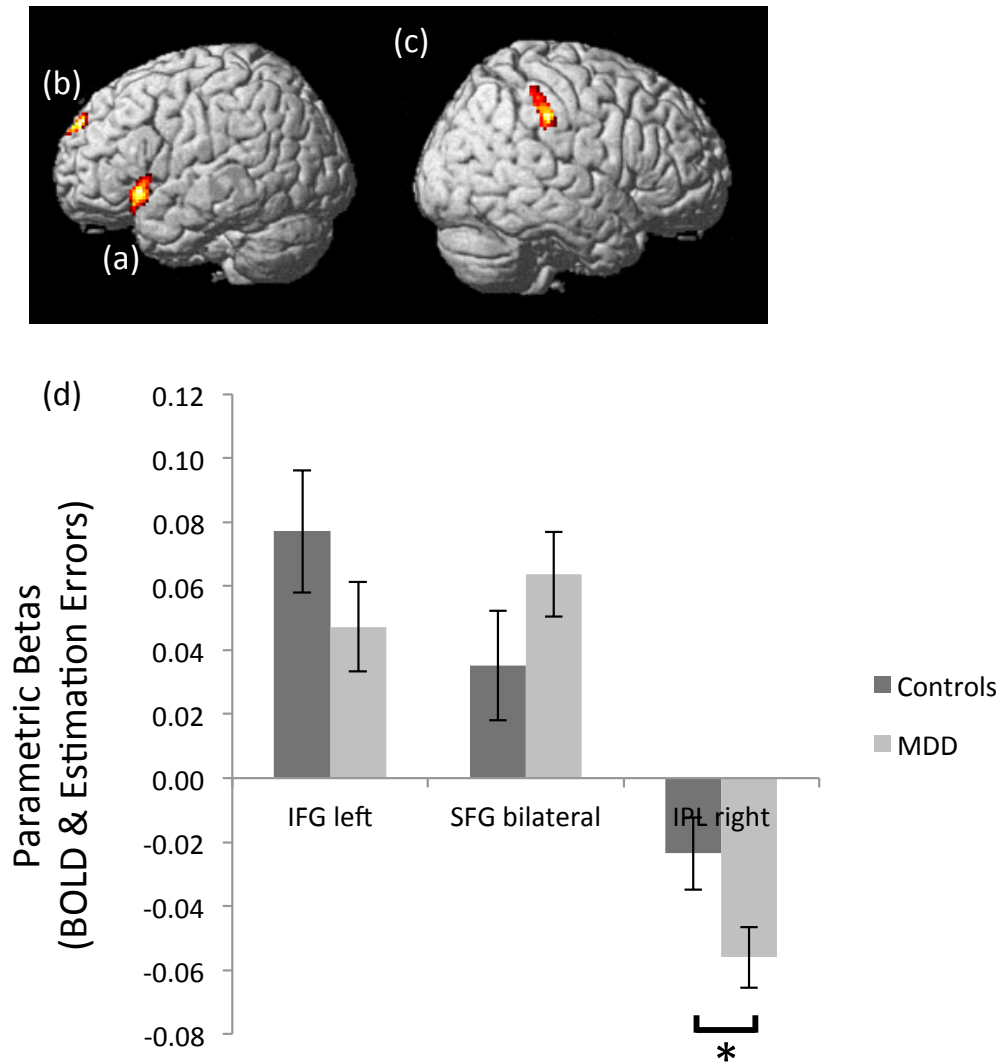


Figure 3.4: Brain activity tracking estimation errors. (a,b) Regions in which BOLD signal tracked participants' estimation errors on a trial-by-trial basis across both groups in response to good news regarding future likelihoods included the left IFG (a) and bilateral SFG (b) ($P < 0.05$, FWE cluster level corrected). (c) BOLD signal tracking participants' estimation errors in response to bad news was found in the right IPL ($P < 0.05$, FWE cluster level corrected). (d) Parameter estimates of the parametric regressors from peak voxels in each of the 3 regions tracking estimation errors. The right IPL showed a stronger inverse correlation between BOLD activity and bad news errors in depressed individuals relative to healthy individuals. Error bars represent s.e.m. * $P < 0.05$, two-tailed independent samples t-test.

In addition, I tested for differences in the anatomically defined right IFG, as it has previously been found that participants with low trait optimism were more likely to have a negative correlation between BOLD signal and bad news estimation errors in this area in

the exact same task (Sharot et al., 2011). Indeed, in this dataset there was a stronger negative correlation between BOLD activity in the right Inferior Frontal Gyrus (rIFG) and bad news estimation errors in depressed patients compared to healthy controls (46, 2, 34; $k=27$; $z=3.87$, $p < 0.05$ FWE, small volume corrected, betas plotted in **Figure 3.5**).

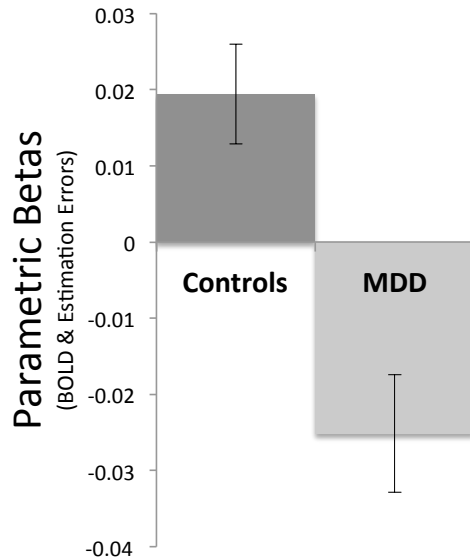


Figure 3.5: Brain activity tracking estimation errors in rIFG. Plotted are the parametric betas from the peak voxel (46, 2, 34) from rIFG cluster ($k=27$) identified by contrasting control group with MDD group betas and applying small volume correction across the anatomical rIFG (note I avoid subsequently conducting ttests as voxel selection is biased; betas are plotted out here for visualisation purposes only).

None of the above effects can be explained by the magnitude of the estimation errors – as reported above these did not differ between groups. In other words, rIFG representation of errors in response to bad news differentiated depressed individuals from healthy controls. Together, these results suggest that adequate computational exploitation, and representation of, negative estimation errors in depression underlies a relatively unbiased belief formation.

3.4 Discussion

A substantial body of research now suggests that optimal mental health is associated with unrealistic positive beliefs regarding the self (McKay and Dennett, 2010; Taylor and Armor, 1996; Taylor et al., 2003). According to Bandura (Bandura, 1989), for example, if self-efficacy beliefs were merely to mirror what people could reasonably accomplish, people would seldom fail but neither would they mount the extra effort required to go beyond ordinary performance. If indeed biased beliefs regarding the self are adaptive, there should be a mechanism that promotes formation of such skewed views, one that could be hypothesized to be *unbiased* during maladaptive mental states.

The results here show that clinically depressed participants updated their beliefs in proportion to the error made whether it called for updating in a desirable or undesirable direction, consistent with past results (Korn et al., 2013). In contrast, healthy individuals were less likely to update beliefs when information called for adjustment in a pessimistic direction. This behaviour was mediated by a diminished coding of “bad news” estimation errors in right IPL in healthy individuals, while depression was associated with close coding of negative estimation errors. These results suggest that adequate computational use, and representation of, negative estimation errors in depression underlies a relatively unbiased belief formation. The finding that mild depression may be related in some domains to an absence of a positive bias, rather than a presence of a negative bias also raises an interesting possibility for future research. Namely, testing whether the absence of positively biased belief updating could predict the onset of a depressive episode among individuals at risk for depression (see also section **7.2.1** for further discussion of this).

Interestingly, across groups activity increased for a better than expected outcome in regions tracking positive estimation errors, and dipped for a worse than expected outcome within right IPL. Depressed individuals were also more likely to show a pattern of inverse correlation between BOLD signal and bad news estimation errors in the right IFG than controls. One hypothesis is that this negative correlation in rIPL represents a “confirmation signal” such that when expectations are validated by the information provided (i.e. first estimations and base rates are closely aligned) there is a larger brain response relative to when expectations do not match up with this information. If this hypothesis about the functional role of rIPL is true, it suggests that when receiving bad news, depressed individuals make a larger distinction between instances in which this news conforms and instances in which it does not conform to pre-existing beliefs about the world. As a result, depressed individuals are potentially better able to realise the need to engage in learning in instances where a belief change is called for (i.e. in the absence of a confirmation signal). It is also of note that this pattern resembles that of dopaminergic neurons signalling prediction errors (Schultz et al., 1997). Indeed, it has previously been shown that increasing dopamine function (via administration of L-DOPA) enhances an update bias in healthy individuals by impairing updating from negative information even further (Sharot et al., 2012b). Dopamine neurons are known to project to the regions identified here (Fallon and Moore, 1978; Gerfen, 1992; Goldman-Rakic, 2000), and it is of interest that abnormal functioning of the dopaminergic system has been related to depression (Papakostas, 2006) and thus may underlie the *lack* of discounting of negative news observed in the disease.

The results in this study were not explained by how well participants subsequently

recalled the information presented to them, as memory for the data provided did not differ across groups. This renders it unlikely that the findings are driven by general differences in cognitive or mnemonic abilities. Note also that the participants performed exactly the same task on trials in which they received good news and on trials in which they received bad news, thus valence dependent differences cannot be explained by one group having specific problems with percentages, generally updating less/more or any domain general cognitive function. Neither did the results reflect specific characteristics of events including familiarity with aversive events, how negative the events were perceived to be, how emotionally arousing participants found the events, nor their past life experience with the aversive events. In other words, the depressed participants did not have more experience with these stressful life events and so this cannot explain these results.

These findings suggest that a positive state of mental health is linked to biased processing and interpretation of information in a manner that supports positively skewed views of the future, while depression is associated with a pattern of activity that supports more unbiased harvesting of information. The data, however, cannot point to causation. Depression may lead to more unbiased updating, or neural systems that supports more unbiased updating of beliefs may generate depression (see section **7.2.1** for further discussion of this). Furthermore, as depression progresses, and/or in more severe cases, a negative bias may be observed (the majority of patients in this study were moderately, clinically depressed).

It has been suggested that people create positive life affirming illusions to enable them to cope with uncertainty and anxiety regarding future dire events (Becker, 1997; Varki, 2009). These illusions are particularly apparent under aversive circumstances and

promote resilience in such situations (Taylor et al., 1984; Wood et al., 1985). A system that does not allow the creation of such perceptions may promote angst and undermine coping strategies resulting in a downward spiral of the effect of stressful life events on mental health.

Chapter 4

Updating Beliefs Under Threat

4.1 Introduction

The tendency to underweight bad news has significant consequences for individuals and society. Yet, it is unknown whether this bias is a fixed human tendency, or rather one that fluctuates in response to acute changes in the environment. Whilst studies have demonstrated that humans are more likely to integrate desirable information into self-relevant beliefs than undesirable information (Sharot et al., 2011; Eil and Rao, 2012; Mobius et al., 2012; Korn et al., 2012), all have examined belief updating in relatively safe and non-threatening environments.

It is possible that humans have evolved a stable pattern of updating beliefs. Indeed, neural function supporting the optimism bias has been related to relatively stable personality characteristics (Sharot et al., 2007, 2011). Another possibility however is that belief updating is state dependent, allowing for fluctuations in response to environmental demands. For example, in relatively safe surroundings, where potential harm is low, an

asymmetry in information integration may be prominent, leading to biased expectations (Johnson and Fowler, 2011). Yet in environments rife with threats, a physiological/psychological response may trigger changes to how information is integrated into beliefs. This could potentially lead to more balanced information integration, which may be adaptive in environments where potential costs are high. In support of the latter possibility, it has been shown that improving the habitat of non-human animals leads to greater positive biases in decision-making (Matheson et al., 2008), while housing animals in unpredictable environments (Harding et al., 2004) or subjecting them to stressors on a regular basis (Rygula et al., 2013) eliminates these biases. Studies in humans report that when people find themselves in a threatening environment, such as prior to an exam (Dewberry and Richardson, 1990) or after a natural disaster (Burger and Palmer, 1992), or terrorist attack (Klar, Zakay and Sharvit, 2002), they show a reduction in optimism.

Here, I examine whether the physiological and psychological response to a perceived threat results in alterations to how individuals integrate good and bad news into their beliefs. Participants were either exposed to a threat manipulation in the lab (Experiment I, manipulation condition), to a control manipulation (Experiment I, control), or tested in a real life environment (firefighters tested between emergency calls, Experiment II). After verifying that threat induced typical psychological (Experiment I & II) and physiological (Experiment I) changes, including changes to skin conductance, cortisol levels and self-ratings of state anxiety, participants completed the belief update task (Chowdhury et al., 2014; Garrett and Sharot, 2014; Garrett et al., 2014; Korn et al., 2013; Kuzmanovic et al., 2015; Moutsiana et al., 2013, 2015; Sharot et al., 2012a, 2012b;

also **Chapter 2**). I designed the experiment such that the threat was unrelated to the information presented in the task. This way I could test whether the effects of perceived threat on information integration were general rather than specific to the threat.

4.2 Experiment I

4.2.1 Materials and Methods

4.2.1.1 Participants

Thirty-six participants recruited via the UCL subject pool participated in the study. Participants gave informed consent and were paid for their participation. The study was approved by the Research Ethics Committee of the University College London. One participant's responses resulted in only two good news trials (out of a possible 40), which prevented a meaningful learning parameter from being calculated, thus this participant's data had to be excluded. Two participant's cortisol samples were insufficient for analysis, and samples of six participants who were suspected to have depression (BDI score greater than 10) were never sent to be analyzed. Thus analysis that includes cortisol scores is given for $n = 27$. Note, however, that either excluding those participants all together from all analysis or including them as done here generated the same results. Each participant was randomly assigned to either the threat manipulation condition (13 females, 6 males, mean age = 26.37 years, $SD = 6.58$) or the control condition (10 females, 6 males, mean age = 24.94 years, $SD = 3.82$).

4.2.1.2 Manipulation Procedure

Participants assigned to the threat manipulation were told that they would be exposed to an uncomfortable, stressful, event at the end of the study. Specifically, they were informed that at the end of the experiment they would be required to deliver a speech on a surprise topic, which would be recorded on video and judged live by a panel of staff members. They were shown an adjacent room across a double mirror window where chairs and tables were already organized for the panel. In addition, participants were presented with six difficult mathematical problems which they were asked to try and solve in 30 seconds. This manipulation is a variation of the Tier Social Stress Test (TSST; Birkett, 2011) with the main difference between the typical TSST procedure and the one used here being that participants were threatened by the possibility of a stressful social event, and completed the main task under threat, but the threat was never executed.

Participants assigned to the control condition were informed that at the end of the experiment they would be required to write a short essay on a surprise topic, which would not be judged. They were then presented with six elementary mathematical problems to solve in 30 seconds.

4.2.1.3 Manipulation Check

I examine if the manipulation resulted in the following psychological and physiological changes, which are typically observed in studies using variations of TSST (Birkett, 2011).

1. *STAI-S*. Before and after the induction procedure, participants filled out a short-form of the State scale of the Spielberger State Trait Anxiety Inventory (STAI-S) developed by Marteau and Bekker (Marteau and Bekker, 1992), see **Chapter 2**, section 2.4.2.

2. *Skin conductance level (SCL)*. SCL is an index of sympathetic tone which reflects changes in autonomic arousal. Skin conductance was recorded for 2 minutes pre and post

induction whilst participants stared at a fixation cross using disposable electrodermal gel electrodes (Biopac, EL507) attached to the distal phalanx of the pointer and middle fingers of the participants' non-dominant hand. Skin conductance responses were monitored using a MP36R system (BIOPAC Systems, Inc., Goleta, CA) and analysed with BIOPAC software *AcqKnowledge*. The difference in mean SCL in each period was taken as a change in participants' autonomic arousal levels.

3. Cortisol Level. To measure changes in participants' cortisol levels, saliva samples were collected using Salivette collection devices, (Salimetrics, UK). Four samples were taken at different time points: before the induction procedure (baseline: t0); immediately after the induction procedure but prior to undertaking the task (+10 min: t1); halfway through the task (+30min: t2); after the task and completion of post experiment questionnaires (+1hr: t3). The experiment was conducted between 2pm and 4pm, restricted to these times to control for the diurnal cycle of cortisol. Samples were stored at -80°C before being assayed. Analysis of salivary cortisol was completed by Salimetrics. Intra-assay and inter-assay coefficients of variation were all below 6.1% ($M = 1.5\%$, $SD = 1.2$). Cortisol values were measured in $\mu\text{g/dL}$. Shapiro-Wilk (SW) tests on cortisol levels at each sample period revealed that these were not normally distributed (one sample SW < .01 for all four sample intervals). As a result, cortisol values were log transformed. Change in cortisol levels for each participant was calculated as the difference between cortisol levels at time periods t1, t2 and t3 from baseline cortisol levels at t0.

4.2.1.4 Behavioural Task

The task was adopted from past studies (Chowdhury et al., 2014; Garrett and Sharot, 2014; Garrett et al., 2014; Korn et al., 2013; Kuzmanovic et al., 2015; Moutsiana et al.,

2013, 2015; Sharot et al., 2012a, 2012b) and outlined in **Chapter 2**. In this version of the task, participants were presented with one of the two lists (List A or List B, **Table 2.1**) of 40 negative life events. The lists used were counterbalanced over participants.

Events were presented for 3s, and participants were asked to estimate how likely the event was to happen to them in the future. Participants had up to 5s to respond. If the participant failed to respond, that trial was excluded from all subsequent analyses ($M = 1.31$, $SD = 1.39$). Following presentation of a fixation cross (5-10s jittered) participants were then presented with the base rate of the event in a demographically similar population for 2s followed by a fixation cross (5-10s jittered). In a second session, immediately after the first, participants were asked again to provide estimates of their likelihood of encountering the same events. Base rates were not presented again in this second session. Note that the timings for the fixation cross are different in this version of the task to those used in the other chapters. This was to enable SCR responses during the task to be analysed in an event related manner. Since the typical SCR response occurs over a longer time period than BOLD response, a longer fixation period is needed to separate out SCR responses.

4.2.1.5 Statistical analysis

For each subject, I separately calculated the weight the subject put on estimation errors in updating their beliefs due to good news (referred to as the *good news learning parameter*, α_G) using this equation:

$$Update_t = \alpha_G * |estimation\ error_t|$$

(where t indexes trial number over good news trials)

and the weight the subject put on estimation errors in updating their beliefs due to bad news (this is referred to as *bad news learning parameter, α_B*) using this equation:

$$Update_t = \alpha_B * |estimation\ error_t|$$

(where t indexes trial number over bad news trials)

Trials for which the estimation error was zero were excluded from subsequent analyses as these could not be categorized into either condition ($M = 0.89$ trials, $SD = 0.92$). Shapiro-Wilk tests were applied to check the values of learning parameters α_G and α_B were normally distributed.

To determine whether learning parameters for good (α_G) and/or for bad (α_B) news were affected by the threat manipulation, the learning parameters were submitted to a 2 by 2 ANOVA with valence (good/bad news) as a repeated-measure and group (threat manipulation/control) as a between-subjects factor.

To explore whether differences in learning parameters related to any of the specific physiological and psychological changes, I constructed a general linear model (GLM) with learning parameters (α) entered as the dependent variable and changes in SCL, STAI and cortisol as independent variables. This was done separately for good news learning parameters (α_G) and bad news learning parameters (α_B). To control for general changes in learning parameters and be able to detect valence-specific effects I entered α_G as a covariate when estimating α_B and vice versa (Moutsiana et al., 2013). In addition I controlled for priors (i.e. mean initial estimations) and mean ratings on all subjective scales including memory scores (Moutsiana et al., 2013). These scores correspond to good news trials in the GLM for α_G and bad news trials in the GLM for α_B . For α_B the formula for the regression in full therefore is as follows:

$$\alpha_B = \beta_0 + \beta_1 * \text{Change in SCL} + \beta_2 * \text{Change in STAI} + \beta_3 * \text{Change in Cortisol} + \beta_4 * \text{Mean Initial Estimates} + \beta_5 * \text{Memory} + \beta_6 * \text{Mean Vividness Rating} + \beta_7 * \text{Mean Familiarity Rating} + \beta_8 * \text{Prior Experience Rating} + \beta_9 * \text{Emotional Arousal Rating} + \beta_{10} * \text{Negativity Rating} + \beta_{11} * \alpha_G$$

4.2.2 Results

Participants assigned to the threat manipulation were told that they would be exposed to an uncomfortable, stressful, event at the end of the study. Specifically, they were informed that at the end of the experiment they would be required to deliver a speech on a surprise topic, which would be recorded on video and judged live by a panel of staff members (for full details on threat and control manipulation see methods).

Subjective reports of anxiety and physiological measures of skin conductance level (SCL) and cortisol show that the manipulation was effective. Specifically, following the manipulation self-reported anxiety (**Figure 4.1a**) and SCL (**Figure 4.1b**) showed an increase relative to before (baseline), which was greater in the threat manipulation group relative to controls (self-report anxiety: $t(33) = 4.16, p < 0.001, 95\% \text{ CI} = [1.88, 5.50]$; SCL: $t(33) = 3.32, p = 0.002, 95\% \text{ CI} = [0.56, 2.34]$) with no difference in SCL drift nor variance between time points for the threat manipulation group (drift: $t(19) = -0.62, p > 0.25$; variance: $t(19) = -0.02, p > 0.25$) nor for controls (drift: $t(15) = 1.33, p = 0.20$; variance: $t(15) = -0.68, p > 0.25$).

A repeated measures ANOVA was conducted on cortisol levels at t1, t2 and t3 relative to baseline cortisol levels (at t0). Time was entered as a 3 level repeated factor and group (threat/control) as a between subject factor. This revealed a main effect of time ($F(2,48) = 8.64, p < 0.01$) and a trend for an effect of group ($F(1,24) = 2.97, p = 0.098$). The

time*group interaction was not significant ($F(2,48)=1.27, p=0.29$). A reduction in cortisol levels over time is an effect that has previously been observed as participants become familiar with a novel experimental context (Stones et al., 1999). That there is not a greater difference in cortisol levels between groups could be due to a number of reasons. It could be a type II error (i.e. insufficient power with the sample size to detect an effect) and/or the presence of additional variables that affect cortisol levels that I do not control for (e.g. women's menstrual cycle). Across participants, these different measures were correlated with each other (Self-reported anxiety & SCL: $r = 0.38, p = 0.02$; SCL & Cortisol: $r = 0.47, p = 0.01$; trend for Cortisol & anxiety: $r = 0.33, p = 0.09$).

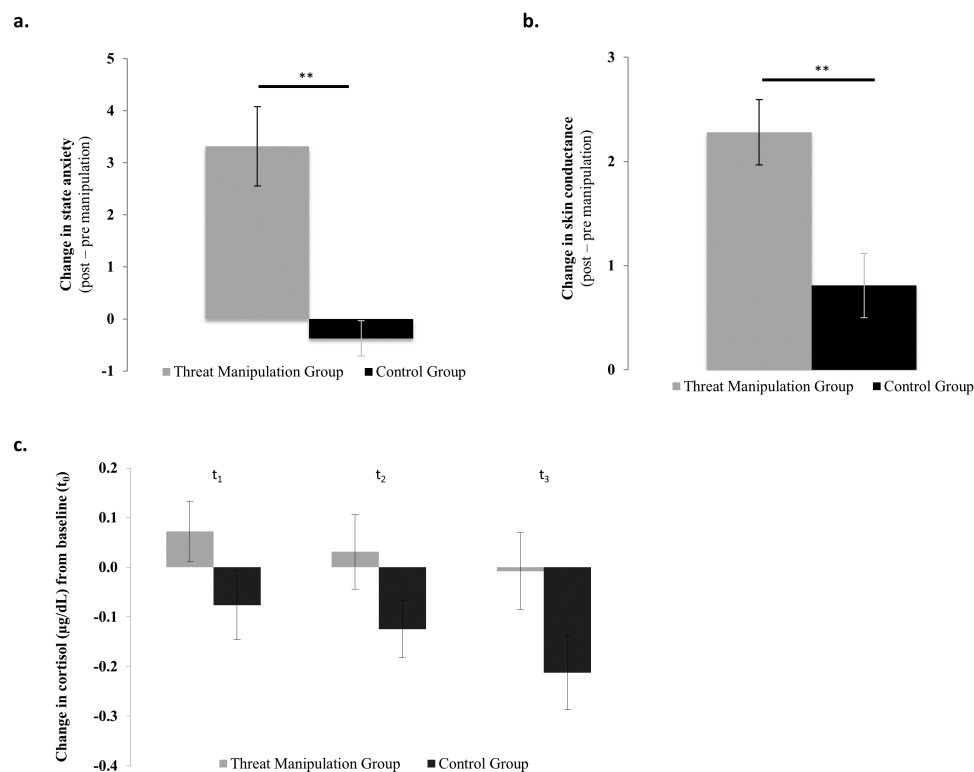


Figure 4.1: Manipulation Check. Changes (after manipulation – before manipulation) in (a) self-reported state anxiety and (b) skin conductance were greater in the threat manipulation group compared to the control group. (c) cortisol levels decreased over time but there was no difference between groups or a group*time interaction $** p < .050$; Error bars represent standard error of the mean.

Examining learning parameters from the UBT, the results revealed a significant interaction between group (control/threat) and valence (good news/bad news) on the learning parameters (α), controlling for all other variables (see below): $F(1,17) = 4.95$, $p = 0.04$, $\eta_p^2 = 0.23$. Planned follow up t-tests revealed that the threat manipulation eliminated biased information integration (Garrett et al., 2014; Moutsiana et al., 2013; Sharot et al., 2011). Participants in the threat manipulation group were more likely to effectively integrate bad news into their beliefs relative to those in the control group (significant difference in bad news learning parameter α_B : $t(33) = 2.44$, $p = 0.02$, independent sample ttest), while integration of good news (α_G) did not differ between groups (no significant difference in good news learning parameter α_G : $t(33) = 0.61$, $p > 0.25$, independent sample ttest). Moreover, whereas I observed asymmetric information integration in the control group, such that the learning parameter was larger for good news than bad ($t(15) = 3.34$, $p = 0.004$, 95% CI = [0.05, 0.25], paired sample ttest), replicating past studies (Moutsiana et al., 2013; Sharot et al., 2011), this asymmetry was absent in the threat manipulation group ($t(18) = .91$, $p > 0.25$, 95% CI = [-0.05, 0.12], paired sample ttest; **Figure 4.2**). There were no floor (one sample ttest against a test value of 0) or ceiling (one sample ttest against a test value of 1) effects for α_G and α_B in the threat manipulation or control group (all at $p < .001$).

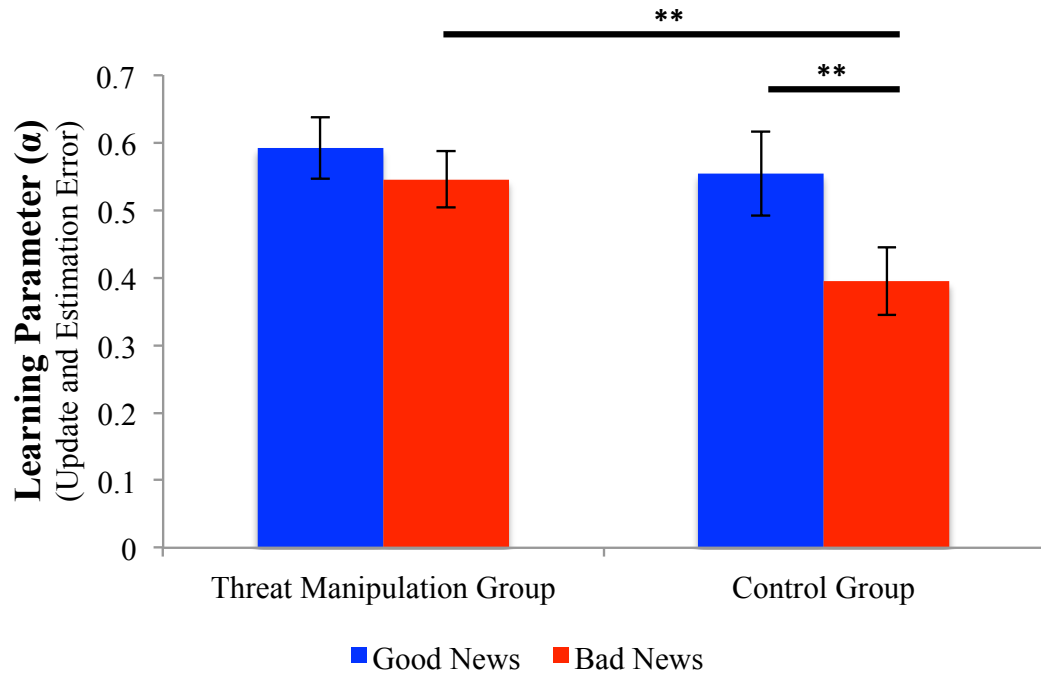


Figure 4.2: Bias in learning parameters (α_G, α_B) vanishes under threat manipulation. While the control group showed asymmetrical learning parameters (α) in response to good and bad news, this bias vanished in the threat manipulation group, due to an increase in α_B (learning parameter for bad news). The Group*Valence interaction was significant, controlling for all covariates (see Methods). ** $p < 0.05$ independent/paired sample t test as appropriate; Error bars represent standard error of the mean.

Results are not driven by priors, baseline anxiety, depression symptoms, nor by valence-dependent effects on memory, attention, emotional arousal, negative valence, vividness, familiarity, past experience, RTs. The following covariates were included in the ANOVA above: initial STAI, initial SCL, participants' priors (i.e. initial estimates of likelihoods), differential rating scores for good news and bad news trials on all scales (emotional arousal, negative valence, vividness, familiarity, past experience), differential RTs, BDI scores, standard deviation of the criterion and of the predictor of the learning parameter. In addition, to test memory for the information presented (rather than information integration), subjects were asked to provide the base rate presented to them of each event. Memory errors were calculated as the absolute difference between the base

rate and the participants' recollection of that statistic (as done before: see methods). Memory bias (i.e. the difference between memory errors for good and bad news) was added as a covariate. As all these variables were included in the ANOVA for which statistics are provided in the previous section, none of them can explain these results. Moreover, none of these factors showed a significant interaction between valence and group (see **Table 4.1** for statistics on all covariates).

Past studies show that asymmetric information integration in this task is not associated with an asymmetry in memory. However, to be certain that memory did not affect the results here either, I not only added memory scores as a covariate in all the analysis but also submitted memory scores to a group (threat manipulation/control) by valence (good news/ bad news) ANOVA, which did not reveal an interaction ($F(1,33) = 0.62, p > 0.25$). Thus, valence dependent changes in learning parameters across groups (i.e. how participants use information to update beliefs regarding personal risk) cannot be attributed to memory or encoding/attention. Note that while information can be better or worse than expected, all stimuli are negative (i.e. robbery, card fraud), thus comparison is never between positive and negative stimuli, but rather between information that is better or worse than expected.

Questionnaire and variables	Threat Group mean		Control Group mean	
	Good News	Bad News	Good News	Bad News
<i>Subjective Scales Questionnaire</i>				
Vividness*	3.37 (.75)	2.96 (.60)	3.72 (.79)	3.00 (.70)
Familiarity*	3.74 (.99)	3.43 (.75)	4.31 (.75)	3.82 (.99)
Prior experience*	1.75 (.65)	1.57 (.69)	1.69 (.44)	1.37 (.30)

Emotional arousal	3.55 (.74)	3.21 (.71)	3.56 (.98)	3.44 (.75)
Negativity	4.05 (.52)	3.85 (.62)	3.97 (.86)	4.10 (.67)
<i>Task-related variables</i>				
Number of Trials	18.16 (4.32)	19.74 (4.76)	18.00 (5.10)	19.56 (4.73)
Memory errors	9.06 (2.10)	10.29 (3.01)	10.36 (3.42)	10.57 (3.38)
Estimation errors (absolute)	18.87 (5.15)	19.69 (3.34)	18.05 (4.85)	16.94 (2.86)
Reaction time first estimate (ms)*	2742.88 (502.70)	2908.69 (521.98)	2606.70 (635.29)	2649.01 (624.72)
Reaction time second estimate (ms)*	2266.81 (545.52)	2200.15 (541.02)	2294.33 (478.73)	2174.91 (404.10)

Table 4.1: Task-related variables, subjective scales, memory and reaction times in Experiment I. * Main effect valence, $p < 0.05$

Together, these results indicate that threat in the environment selectively enhances the ability to adjust beliefs in response to bad news.

What therefore could account for the selective fluctuations in information integration of bad news? To examine which of the changes to the psychological and physiological measures (SCL, cortisol level, self-reported anxiety) could independently explain alteration in learning parameters for bad news, a linear regression was performed with α_B entered as the dependent variable and changes in STAI, SCL, and cortisol as independent variables (all entered together in one regression). To ensure that effects were valence-specific and could not be accounted for by general changes to information integration, α_G was added as a covariate. I also controlled for priors (i.e. mean initial estimations) and mean ratings on all subjective scales including memory scores. The analysis revealed that changes in self-reported anxiety indicated by STAI ($t = 2.26$, $p = 0.04$, $b_i = 0.028$, $\eta_p^2 = 0.25$) and physiological arousal indicated by SCL ($t = 2.11$, $p = 0.05$, $b_i = 0.057$, $\eta_p^2 = 0.23$) explained the variance in integration parameters for bad news,

but cortisol did not ($t = -0.56, p > 0.25, b_i = -0.04, \eta_p^2 = 0.02$; see **Table 4.2** for parameter estimates of covariates). Specifically, participants who showed the greatest increase in SCL and self-reported anxiety were most likely to change their beliefs in proportion to the difference between their prior and the bad news received (**Figure 4.3**). This pattern of results remains if BDI is added as an additional covariate in the regression. Note that the null result for cortisol may indicate either that the increase in bad news information integration is not associated specifically with cortisol level increase, or a Type II error. For completeness I repeated the analysis on α_G (including α_B and all other factors as covariates) and found no significant effects (change in STAI: $t = -0.18, p > 0.25, b_i = -0.003, \eta_p^2 = 0.00$; change in SCL: $t = 0.96, p > 0.25, b_i = 0.034, \eta_p^2 = .06$; change in cortisol: $t = 0.23, p > 0.25, b_i = .018, \eta_p^2 = 0.00$).

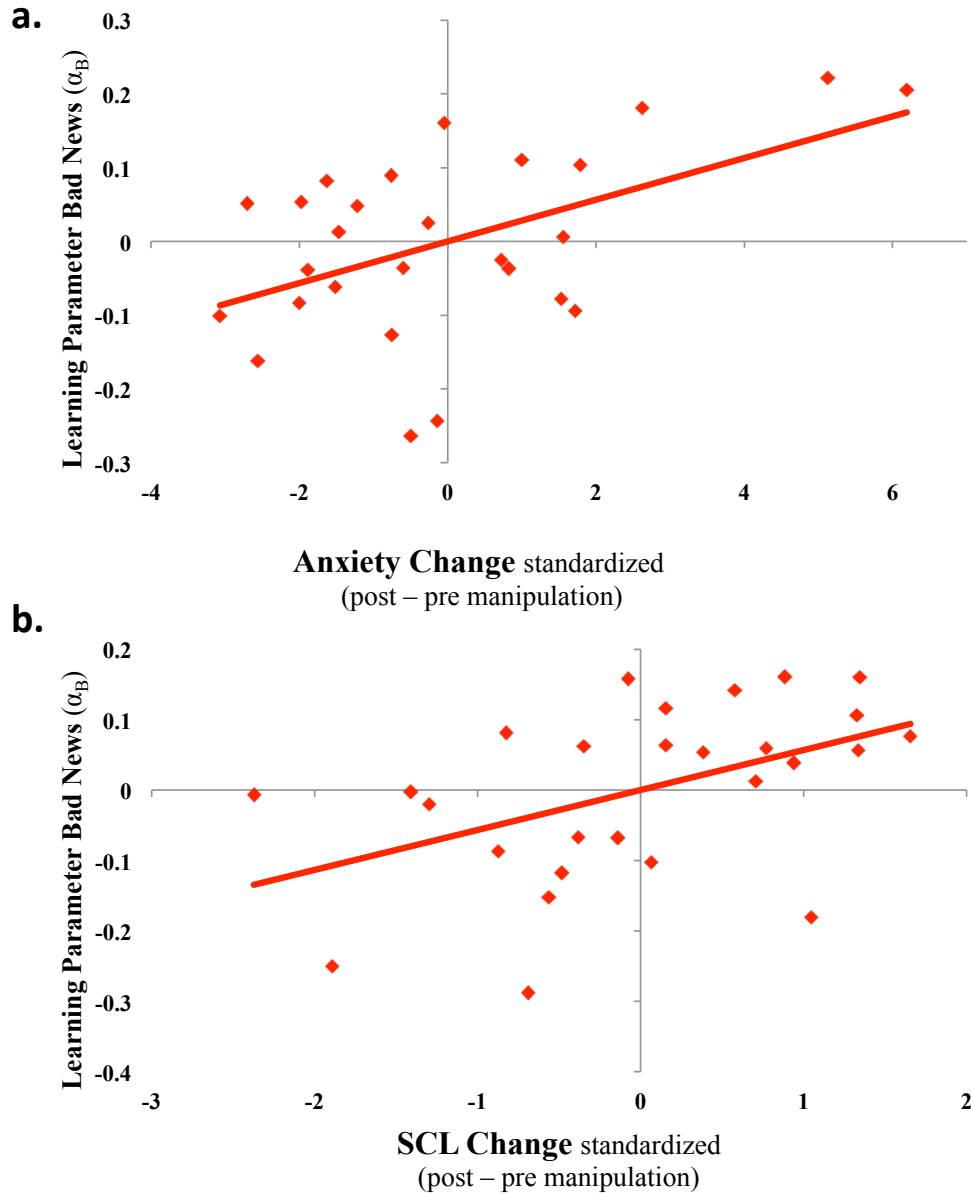


Figure 4.3: Greater integration of bad news related to state anxiety and SCL. Following the manipulation, an increase in both **a.** self-reported state anxiety ($b_i = .028$, $p = 0.039$, $\eta_p^2 = 0.25$) and **b.** skin conductance (SCL) ($b_i = .057$, $p = 0.052$, $\eta_p^2 = 0.23$) were related to larger bad news learning parameters (α_B). Plotted are the standardized learning parameters for bad news, α_B , from a linear model that also includes: learning parameters for good news (α_G), initial estimates (priors), mean ratings on all subjective scales, memory scores, changes in cortisol and changes in skin conductance (in the case of plot a) and changes in self-reported anxiety (in the case of plot b).

	b_i	Std. Error	t	p	95% Confidence Interval		η_p^2
					Lower Bound	Upper Bound	
					Mean Initial Estimations	.01	
Memory	.00	.02	-.18	.860	-.04	.03	.00
Vividness rating	-.06	.06	-.99	.340	-.19	.07	.06
Familiarity rating	.07	.05	1.62	.125	-.02	.17	.15
Prior experience rating	-.07	.07	-1.01	.330	-.21	.08	.06
Emotional arousal rating	-.13	.05	-2.52	.023	-.24	-.02	.30
Negativity rating	-.02	.08	-.29	.775	-.19	.14	.01
Learning parameter, good news (α_G)	.35	.17	2.01	.063	-.02	.72	.21

Table 4.2: Parameter estimates of covariates in Experiment I. Priors (i.e. mean initial estimations), mean ratings on all subjective scales, memory scores and α_G (learning parameters for good news) were entered as covariates to account for fluctuations in α_B (learning parameters for bad news).

4.3 Experiment II

Next I set out to extend the findings from Experiment I in a natural setting. Here, I did not manipulate threat. Rather, I measured the relationship between integration of good and bad news and perceived threat in a naturally volatile environment. Specifically, firefighters from the state of Colorado performed the UBT while on duty at their respective fire stations.

4.3.1 Materials and Methods

4.3.1.1 Participants

Thirty three operational staff stationed across seventeen fire stations within the South Metro Fire and Rescue Authority of the State of Colorado in the United States participated in the study. Five of these participants failed to complete the study leaving 28 participants

(1 female, 27 males, mean age = 42.75 years, $SD = 9.75$). A link to an online version of the experiment was sent by email to operational staff inviting them to participate in the study whilst on duty. Employees were given 18 days to attempt the experiment. They were permitted to take the experiment once in this time period and were explicitly requested to do so whilst on shift (i.e. in the station between calls). Participation in the experiment was anonymous and voluntary and participants were not paid.

4.3.1.2 Task, stimuli and control variables

An online version of the UBT used in Experiment I was designed using Qualtrics Survey Software (Qualtrics, Provo, UT). The task began by asking basic demographic questions (age, gender, marital status, level of education and number of children) and some questions pertaining to their work (including how long they had worked in the service, how many people they supervised, number of emergency they went on, what their rank in the service was) and social environment (social support at work and outside, and stress experienced at home).

After providing this information, participants read task instructions on screen at their own pace and then undertook a practice session comprising 3 practice trials. The task was the same as in Experiment I, except that some of the descriptions were adapted for language differences between American English and British English (for example the abbreviation “ER” replaced the term “A&E”). Furthermore, mindful of the firefighters’ unpredictable time constraints, memory for the information given and subjective ratings (past experience with the event and negativity) were elicited for half the stimuli and participants completed a short version of the state scale of the STAI at the beginning of

the study (Chlan et al., 2003), without providing physiological measures of autonomic arousal.

Statistical analysis: Linear regressions were performed using ordinary least squares implemented using SPSS version 22 for bad news and good news separately, with learning parameters (α_G and α_B) entered as the dependent variable and self-reported anxiety as the independent variable. To ensure that effects were valence specific and could not be accounted for by general changes to information integration, learning parameters for good news (α_G) were added as a covariate when examining learning parameters for bad news (α_B) and vice versa. For each condition I also controlled for priors, memory, mean ratings of negativity and mean ratings of past experience as done in Experiment 1. For α_B the formula for the regression in full therefore is as follows:

$$\alpha_B = \beta_0 + \beta_1 * \text{Anxiety} + \beta_2 * \text{Mean Initial Estimates} + \beta_3 * \text{Memory} + \beta_4 * \text{Prior Experience Rating} + \beta_5 * \text{Negativity Rating} + \beta_6 * \alpha_G$$

4.3.2 Results

The results of Experiment II expanded the results of Experiment I in a “real life” context. Specifically, a linear regression was carried out in which learning parameters for bad news (α_B) were regressed on self-reported anxiety, controlling for priors (i.e. mean initial estimations) and mean ratings on subjective scales including memory scores. In addition, to ensure effects were valence-specific and could not be accounted for by general changes in learning parameters, α_G was added as a covariate. This analysis revealed that self-reported anxiety significantly explained the variance in α_B ($t = 3.96$, $p = 0.001$, $\eta_p^2 = 0.43$,

$b_i = 0.054$, 95% CI = [0.026, 0.082]; **Figure 4.4** and **Table 4.3**). Thus, the higher the anxiety reported by a firefighter, the more likely the firefighter was to integrate bad news into beliefs in proportion to the difference between her/his priors and the information given. This pattern of results remains if BDI is added as an additional covariate in the regression. Next I conducted the same analysis on learning parameters for good news (α_G , with α_B and all other factors added as covariates). This revealed a *trend* in the opposite direction than for α_B ($t = -1.92$, $p = 0.07$, $\eta_p^2 = 0.117$, $b_i = -0.045$, 95% CI = [-0.094, 0.004]), such that greater anxiety was related to a trend for *less* information integration in response to good news. There were no floor or ceiling effects for α_G or α_B (all at $p < .001$).

Hence, the results indicate that anxiety is related to a selective enhancement of the ability to adjust beliefs in response to bad news. I selected to test firefighters on duty as I reasoned that these individuals may exhibit a large range of state anxiety due to the volatility of their working environment. I emphasize, however, that the results are not necessarily contingent on the environment being volatile, nor are anxiety levels necessarily a result of their work environment.

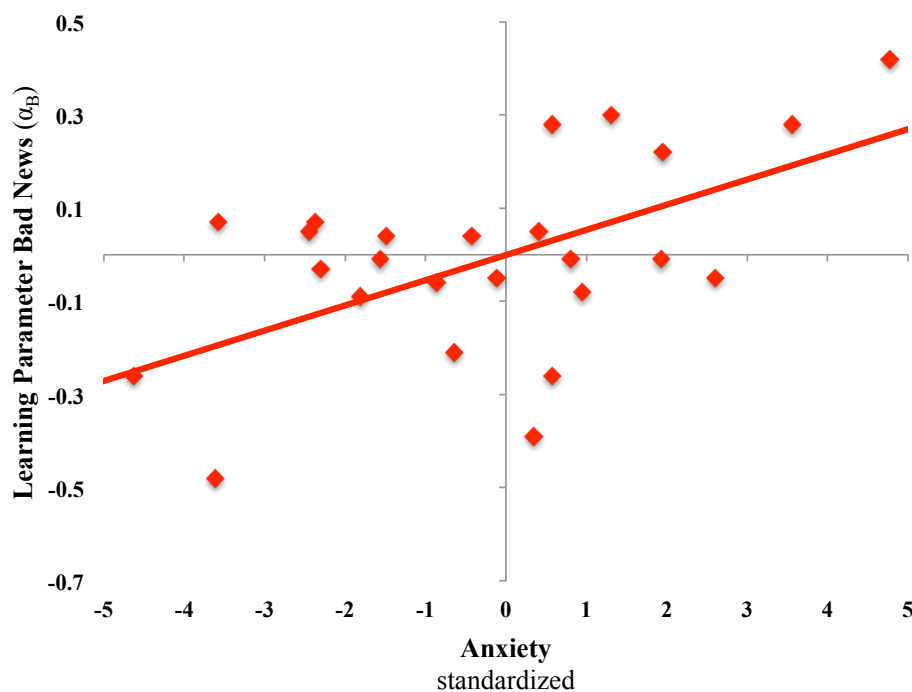


Figure 4.4: State anxiety in firefighters related to greater integration of bad news. State anxiety levels of firefighters ($b_i = 0.054, p = 0.001, \eta_p^2 = 0.43$) whilst on shift were related to larger bad news learning parameters (α_B). Plotted are the standardized α_B from a linear model that also includes initial estimates (priors), mean ratings on subjective scales, memory scores and learning parameters for good news (α_G).

	b_i	Std. Error	t	p	95% Confidence Interval		η_p^2
					Lower Bound	Upper Bound	
Mean Initial Estimations	-0.002	0.006	-0.301	0.766	-0.014	0.011	0.004
Memory	0.018	0.011	1.67	0.11	-0.004	0.039	0.117
Prior experience rating	-0.059	0.082	-0.721	0.479	-0.229	0.111	0.024
Negativity rating	0.105	0.053	1.965	0.063	-0.006	0.215	0.155
Learning parameter, good news (α_G)	0.426	0.123	3.461	0.002	0.17	0.682	0.363

Table 4.3: Parameter estimates of covariates in Experiment II. Priors (i.e. mean initial estimations), mean ratings on subjective scales, memory scores and α_G (learning parameters for good news) were entered as covariates to account for fluctuations in α_B (learning parameters for bad news).

4.4 Discussion

Together, these results provide evidence that asymmetry in belief formation is absent under perceived threat. Specifically, Experiment I shows that in a low threat environment individuals integrated information asymmetrically, faithfully incorporating good news into their existing beliefs while relatively disregarding bad news (Eil and Rao, 2011; Sharot et al., 2011). Under perceived threat, however, this asymmetry disappeared; participants showed an increased capacity to integrate bad news into prior beliefs. Increased physiological arousal and state anxiety were found to correlate with enhanced integration of unfavourable information into beliefs. The latter result was then replicated in firefighters, who often experience real life threat and on duty.

Given that enhanced information integration for bad news under threat was valence specific (i.e. observed only for bad news while controlling for any variance in information integration of good news), it could not reflect a general improvement in information integration. Furthermore, the effect on belief formation was selective and did not affect the evaluation of the stimuli presented (including emotional evaluation of the stimuli, sense of familiarity and vividness) or memory for the information presented. The latter suggests that modulation of attention is an unlikely explanation. Depression scores were also controlled for in the analysis reported thus differences in depression symptoms cannot account for the results.

It is also of interest that in the studies reported here, the source of the threat was unrelated to the information in the task. Thus, state anxiety had a valence-specific, yet general, effect on how participants used information to alter their beliefs (i.e. in response to a social threat the participants did not selectively increase their response to information

about social judgment, but to negative information in general). This may be adaptive, as threat may signify a dangerous environment that requires a general enhancement of caution. However, in situations where perceived anxiety is chronically high one would predict prolonged balanced information integration and thus an absence of optimism bias. For example, in depression, anxiety is often prolonged, and both balanced information integration (Garrett et al., 2014 and **Chapter 3**) and lack of optimism (Strunk et al., 2006) are observed. Taken together, it is possible that in individuals predisposed to affective disorders, a stressful life event can result in prolonged levels of anxiety and thus prolonged periods of increased sensitivity to negative information. This in turn will form pessimistic beliefs, which will lead to greater perceived threat.

In sum, these results provide evidence that asymmetric information integration is not set in stone, but changes acutely in response to the environment, decreasing under perceived threat. Such flexibility could be adaptive, potentially enhancing our likelihood to respond to warnings with caution in environments where future costs may be high, but enabling us to maintain positive beliefs otherwise, a strategy that has been suggested, on balance, to increase well-being (McKay and Dennett, 2010).

Chapter 5

How Robust is the Optimistic Update Bias for Estimating Self-risk and Population Base Rates?

5.1 Introduction

An open question is whether the update bias exists only when adjusting beliefs about the self or whether it is observed also when adjusting beliefs about the population at large (base rates). This is important for understanding biases in risk estimation for two reasons. First, when estimating own risk people may incorporate both base rates and diagnostic information in their calculations. For example, if someone is estimating their likelihood of cancer they may consider the known frequency in their population together with knowledge about themselves (i.e. do I smoke? do I exercise?). Thus, a bias in updating base rates may effect updating for self-risk. Second, it has been suggested that people tend to be optimistically biased when considering the self, less so when considering others (Weinstein, 1980). It is thus of interest to examine whether or not the optimistic updating bias previously found for self risk will expand to base rates.

To examine selective updating in estimating risk about oneself and the population I adjusted the UBT (see **Experimental Methods, Chapter 2**). Participants completed a revised version of the UBT where they estimated their own risk for 80 different negative events and also estimated the base rates of these events. On each trial they were then given explicit information regarding the base rates, and in a second session they estimated both again (see **Figure 5.1a, procedure**). The rationale in examining how participants update their estimates of base rates when receiving this information from the experimenter is that although a participant may recall the base rate presented accurately s/he may be uncertain of the validity of that information. For example, they may believe they have additional/more-up-to-date information regarding base rates that the experimenter does not know about. Thus when the participant incorporates the new information into his/her existing beliefs they may still do so in a biased manner.

The paradigm enabled quantification of how participants adjust their beliefs about the self and the population in response to new information in two instances: (1) when they learn that the average likelihood of encountering a negative life event is lower than their own estimates (good news, **Figure 5.1b**) and (2) when it is greater (bad news, **Figure 5.1c**). To examine the robustness of the bias I further investigated if the results differ if valence is empirically defined in two different ways: (1) by comparing the information presented to the participants' estimate of their own probability of encountering a negative event; (2) by comparing the information presented to the participants' estimate of the population base rate. By asking participants to rate the extent to which they found the information presented to them desirable or undesirable I could also examine whether these subjective ratings are driven more or less by deviations from: (1) estimations of self risk

(2) estimations of the population base rate. Finally, additional experimental factors that may influence the results (such as memory for the information provided and priors) were tested.

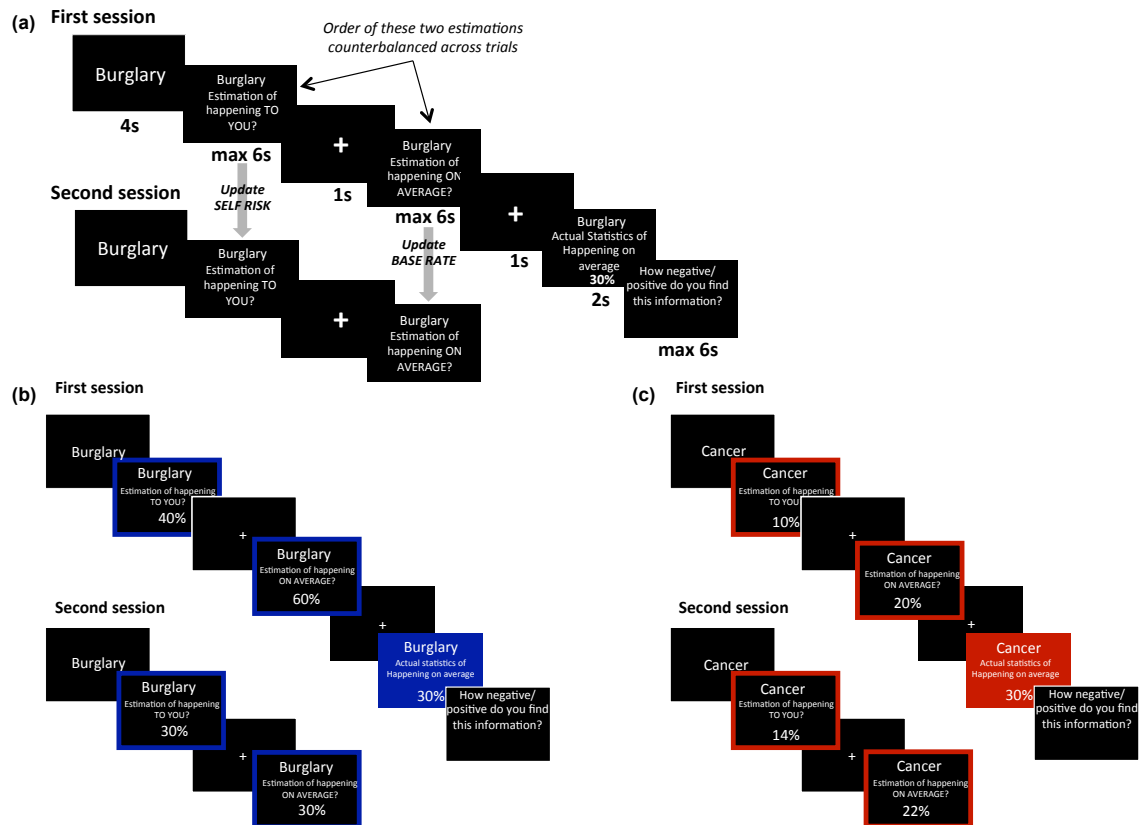


Figure 5.1: Paradigm. (a) In each trial, participants were presented with a short description of 1 of 80 adverse events and asked to estimate how likely this event was to occur to themselves in the future and how likely the event was to happen on average in the population. They were then presented with the base rate in a demographically similar population. Finally, participants were asked to rate how negative/positive they found this information. The second session was the same as the first except that the base rate was not presented and participants did not submit any ratings. Examples of trials in which the participant's estimate of the event occurring to themselves and the base rate was (b) higher or (c) lower than the provided base rate. In the specific examples shown here, under either classification scheme therefore these trials would be categorized as good news and bad news trials respectively.

5.2 Materials and Methods

5.2.1 Participants

The study was approved by the UCL Psychology Ethics Committee. Written informed consent was obtained from all participants. Thirty-two participants aged 18 to 33 participated in the study (mean age = 22.93; s.d. = 3.64). An additional six participants originally completed the task but were excluded due to BDI Scores above 11, indicating possible major depression disorder. All participants were recruited from UCL psychology subject pool. Participants were paid for their participation.

5.2.2 Task and Procedure

5.2.2.1 Stimuli

Stimuli consisted of eighty short descriptions of adverse life events (e.g. passenger in a car accident, home burglary – see **Experimental Methods, Chapter 2, Table 2.1**).

5.2.2.1 Procedure

The paradigm was modified from previous studies (Chowdhury et al., 2014; Garrett et al., 2014; Moutsiana et al., 2013, 2015; Sharot et al., 2011, 2012a, 2012b) and that described in **Experimental Methods, Chapter 2**. The version used in this experiment is depicted in **Figure 5.1**.

In this version of the task, participants were asked to separately estimate how likely each event was to happen to **themselves** in the future and how likely the event was to happen **on average** in the population. In half of the trials the order of these estimations was reversed (i.e. participants were first asked to estimate how likely the event was to

happen on average and then to estimate their own likelihood). Participants were then shown the actual frequency of the event in a demographically similar population for 2s. Finally, participants were asked to rate on a 7 point scale (1=Very negative; 7 = Very Positive) how negative/positive they found this information. Participants had up to 6s to give each estimation and rating. If the participant failed to submit a response for either estimation or rating, that trial was excluded from all consequent analyses (mean trials with missing response = 2.50, s.d. = 2.78).

In a second session, immediately after the first, participants were asked again to provide estimates of their likelihood and the average likelihood of encountering the same events (order reversed in half the trials) so that it could be assessed how they updated both estimations regarding the self and estimations regarding base rates.

5.2.3 Data Analysis

5.2.3.1 Classification, Updates and Bayesian Comparison

Trials were divided into ones in which participants received good or bad news using 2 different classification criteria:

1. Trials were classified as in previous studies (Chowdhury et al., 2014; Korn et al., 2013; Kuzmanovic et al., 2015; Moutsiana et al., 2013, 2015; Sharot et al., 2011, 2012a, 2012b) and as described in **Experimental Methods, Chapter 2**. Trials were classified as good news when the participant initially overestimated the probability of the event occurring to *themselves* relative to the provided base rate. Conversely, trials were classified as bad news when the participant initially underestimated the probability of the event occurring to themselves relative to the provided base rate. Trials in which

participants' estimates of their own likelihood were exactly equal to the provided base rate were excluded from the analysis (mean number excluded trials = 1.72; s.d. = 1.25).

2. Trials were classified as good news when the participant initially overestimated the base rate relative to the provided base rate. Trials were classified as bad news when the participant underestimated the base rate relative to the provided base rate. Trials in which participants' estimates of the base rate were exactly equal to the provided base rate were excluded from the analysis (mean number excluded trials = 1.53; s.d. = 1.02).

For both of these classifications, update terms were calculated as described in **Chapter 2** only there are now two update terms for each type of information: one for self risk and one for base rate. So for good news trials were calculated as:

$$\text{Update}_{\text{Self_Risk}} = \text{First_Estimation}_{\text{Self_Risk}} - \text{Second_Estimation}_{\text{Self_Risk}}$$

$$\text{Update}_{\text{Base_Rate}} = \text{First_Estimation}_{\text{Base_Rate}} - \text{Second_Estimation}_{\text{Base_Rate}}$$

For bad news trials, update terms were calculated as:

$$\text{Update}_{\text{Self_Risk}} = \text{Second_Estimation}_{\text{Self_Risk}} - \text{First_Estimation}_{\text{Self_Risk}}$$

$$\text{Update}_{\text{Base_Rate}} = \text{Second_Estimation}_{\text{Base_Rate}} - \text{First_Estimation}_{\text{Base_Rate}}$$

For each type of update (Self_Risk/Base_Rate) and classification scheme, update scores were entered into separate ANOVAs and follow up t-tests were conducted. In addition, a Bayesian analysis was conducted to assess the degree to which updating for Self Risk corresponded to that of a Bayesian agent under each classification method. This is described below.

Bayes theorem (Bayes and Price, 1763; Stuart, 1994) suggests that participants estimate how likely an event is to occur given the available evidence for and against the event. In instances where one examines the evidence for two competing statements (e.g. that an event will happen vs not happen), the theorem can be formalised (Kahneman and Tversky, 1973) as:

$$P(h/e) = \frac{P(e/h) \times P(h)}{P(e/h) \times P(h) + P(e/\sim h) \times P(\sim h)}$$

Using this formulation, when estimating how likely an event is, such as how likely one is to be burgled in the future, given the evidence for and against the event occurring in the future, one would estimate:

- (1) What the average burglary rate is, say in the neighbourhood, the *prior*: $P(h)$. Note that in these instances where events are binary (you are either burgled or not burgled), the prior of the event *not* happening ($P(\sim h)$) will be given by: $1-P(h)$.
- (2) The degree of evidence (e), given that the proposition (h, being burgled in this example) is true ($P(e/h)$) versus the degree of evidence (e) given that the proposition (h) is false ($P(e/\sim h)$). This can be expressed as a *likelihood ratio* (*LHR*): $P(e/h) / P(e/\sim h)$
- (3) Finally, (1) and (2) can then be integrated to arrive at an estimate of how likely the event is to happen given the evidence for and against ($P(h/e)$), using the formula above.

In the design used here, participants are asked to provide estimates of priors, $P(h)$, for each event along with estimates of their own likelihood of experiencing each event which, if it is assumed that participants arrive at through a process of Bayesian integration,

can be construed as $P(h/e)$. Hence we have 2 of the 3 pieces of the Bayesian equation, the missing one being the likelihood ratio. A recent analysis (Shah et al., 2016) has shown that an “odds ratio” formulation of the Bayes rule equation above (Kahneman and Tversky, 1973; Shah et al., 2016) can be rearranged to calculate the likelihood ratio participants are implied to have used given their priors and self estimates. The formula and analysis is originally described in Shah et al. (2016) but I restate it briefly here. The odds ratio form of Bayes rule (presented in Shah et al., 2016) is:

$$\textit{Posterior Odds} = \textit{Prior Odds} \times \textit{LHR}$$

Where LHR is the likelihood ratio. Formally this is expressed as:

$$\frac{P(h/e)}{1 - P(h/e)} = \frac{P(h)}{1 - P(h)} \times \textit{LHR}$$

This can be rearranged to find the likelihood ratio participants are implied to have used to arrive at their estimate of $P(h/e)$ given their prior $P(h)$. Specifically:

$$\textit{LHR} = \frac{P(h/e)}{1 - P(h/e)} \div \frac{P(h)}{1 - P(h)}$$

Expressing this in terms of measures in the UBT (where $P(h/e)$ is assumed to be participants’ first estimates of their own risk and $P(h)$ participants first estimates of base rates):

$$\textit{LHR} = \frac{\textit{First_Estimation}_{\textit{Self_Risk}}}{1 - \textit{First_Estimation}_{\textit{Self_Risk}}} \div \frac{\textit{First_Estimation}_{\textit{Base_Rate}}}{1 - \textit{First_Estimation}_{\textit{Base_Rate}}}$$

Following Shah et al. (2016), once this implied likelihood ratio has been calculated, it is then possible to compare participants update behavior to a rational Bayesian agent by calculating what a Bayesian posterior would be:

$$\text{Posterior Odds} = \text{Prior Odds} * \text{LHR}$$

Where Prior Odds are given by:

$$\text{Prior Odds} = \text{Base rate}/(1-\text{Base Rate})$$

Converting Posterior Odds to a percentage is then derived by dividing Posterior Odds by (1+ Posterior Odds). Hence a Bayesian 2nd estimate is given by:

$$\text{Bayesian 2}^{\text{nd}} \text{ Estimate} = \text{Posterior Odds}/(1+\text{Posterior Odds})$$

From this, Bayesian Update is calculated for desirable trials as:

$$\text{Bayesian Update} = \text{First_Estimation}_{\text{Self_Risk}} - \text{Bayesian 2}^{\text{nd}} \text{ Estimate}$$

For undesirable trials, Bayesian Update is calculated as:

$$\text{Bayesian Update} = \text{Bayesian 2}^{\text{nd}} \text{ Estimate} - \text{First_Estimation}_{\text{Self_Risk}}$$

Finally, to assess how well participants' actual updates correspond to a Bayesian update, Bayesian updates can be regressed on actual update of participants and the resulting regression coefficients examined at the group level (a similar approach is taken in Eil and Rao, 2011).

To summarise, this analysis proceeds in four steps:

- (1) Calculate *implied* likelihood ratio using estimate of own likelihood and base rate estimate.

- (2) Calculate “Bayesian” 2nd estimate and update using implied likelihood ratio and base rate provided to participant.
- (3) Compare actual update to Bayesian update by regressing Bayesian update on actual update separately for good and bad news for each participant.
- (4) Compare regression coefficients at the group level (i.e. across participants).

5.2.3.2 Desirability Ratings

On each trial, after being provided with base rates, participants were asked to rate on a 7 point scale (1=Very negative; 7 = Very Positive) how negative/positive they found this information. To test whether these desirability ratings were driven more or less by the extent to which the statistical information differed from participants’ estimations regarding the self or regarding base rates, two estimation error terms quantifying for each trial the difference between participants’ initial estimates and the information presented were calculated as:

$$\text{Estimation_Error}_{\text{Self_Risk}} = \text{First_Estimate}_{\text{Self_Risk}} - \text{Provided_Base_Rate}$$

$$\text{Estimation_Error}_{\text{Base_Rate}} = \text{First_Estimate}_{\text{Base_Rate}} - \text{Provided_Base_Rate}$$

For each participant, each set of estimation errors was correlated with their desirability ratings. To statistically test for a difference in the strength of these correlations accounting for the additional correlation between the 2 sets of estimation errors these were compared these using Steiger's Z-test.

5.3 Results

5.3.1 Do participants update in a Bayesian manner?

Ascertaining estimates of participants' own likelihood and base rate estimates allows calculation of what a Bayesian update would be on each trial. To assess the degree of correspondence between Bayesian updates and participants' actual updates for self risk, linear regressions were carried out with Bayesian Update for self risk as the independent variable and participants' actual update for self risk as the dependent variable. This was run for each participant for good news and bad news separately and under each classification scheme. Slopes indicate the degree to which participants' updates correspond to that of a Bayesian update. Under classification one and classification two, there was a tighter correspondence between Bayesian update and subjects actual update for good news compared to bad news (**Figure 5.2**), consistent with previous studies (Eil and Rao, 2011; Mobius et al., 2012).

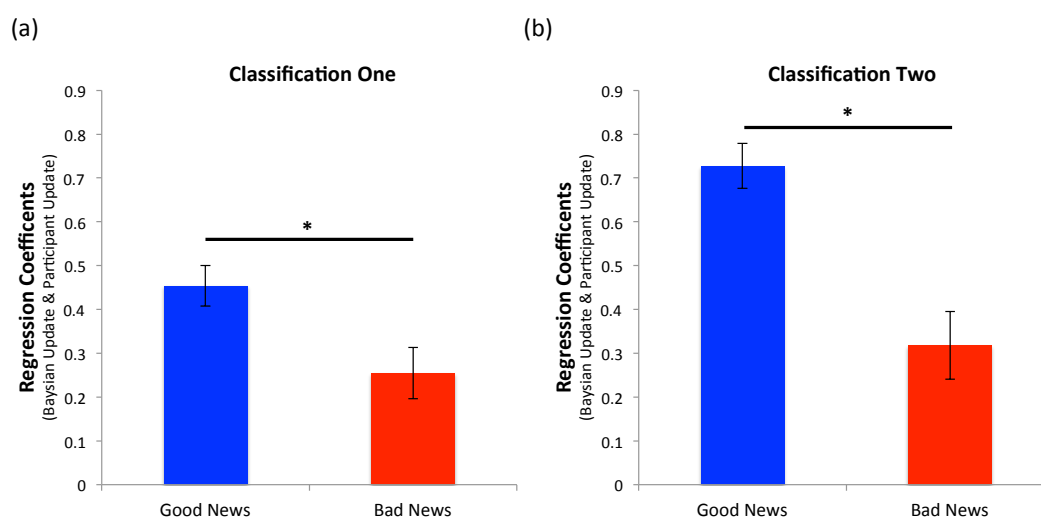


Figure 5.2: For each participant, on each trial the update a Bayesian agent would undertake was calculated. I then examined the extent to which participants' actual update corresponded to Bayesian agents. Under each classification method, participants' update showed a stronger relationship with a Bayesian agent when they received good news compared to when they received bad news. * indicates significantly different, paired sample t-test ($P < 0.05$).

5.3.2 Is updating beliefs regarding personal risk and base rates biased?

As detailed below, these results show biased updating for self risk (see **Figure 5.3**). Specifically, updating for self risk is greater in response to good news relative to bad news under both classification schemes of desirability. However, for base rates biased updating can be accounted for by priors.

5.3.2.1 Updating of personal risk

Replicating previous findings (Chowdhury et al., 2014; Garrett et al., 2014; Moutsiana et al., 2013, 2015; Sharot et al., 2011, 2012a, 2012b), classification scheme 1 revealed that participants updated their beliefs regarding self risk more when the information regarding base rates was better than their estimate of self risk compared to when it was worse ($t(31) = 6.09, P < 0.01$). This bias persists under classification 2 when classifying desirability of information as dependent on base rates (i.e. if information regarding base rates is greater or smaller than participants' estimate of these base rates). Participants were more likely to update their beliefs about their own likelihood of encountering a negative life event when the base rate was better than their initial estimate of the base rate compared to trials in which the base rate was worse ($t(31) = 4.83, P < 0.01$). These results replicate previous findings of valence dependent updating studies (Chowdhury et al., 2014; Garrett et al., 2014; Moutsiana et al., 2013, 2015; Sharot et al., 2011, 2012a, 2012b) and confirm that valence dependent updating for self risk is not contingent on classification scheme.

As detailed below, these findings hold after accounting for possible confounding factors. Specifically, as I describe in section **5.3.2.3**, examining all additional factors and ratings revealed four possible confounds: (1) under classification 1 ratings of past

experience differed for stimuli for which subjects received good and bad news; (2) under classification 2 magnitude of estimation errors differed for stimuli for which subjects received good and bad news; (3) under classification 2 there were a greater number of trials for which subjects received good news than bad news (4) initial estimates of personal risk and base rates differed (this is true for both classifications).

In these instances where confounds were identified, mean update scores for each participant were entered into a repeat measures ANOVA with valence (good/bad news) as a factor and the relevant scores added as covariates. In the case of past experience, the covariate was the mean difference between past experience for good news trials and past experience for bad news trials for each participant. For estimation errors, the covariate was the average difference in estimation errors for good news trials and bad news trial for each participant. For number of trials the covariate was the difference in number of good news and bad news trials for each participant. For initial estimates, the covariate was simply the mean initial estimate for each participant.

When controlling for these additional factors in the respective classifications a main effect of valence remained (classification 1: $F(29) = 8.02$, $P < 0.01$, classification 2: $F(29) = 10.72$, $P < 0.01$) confirming the robustness of the bias.

5.3.2.2 Updating of base rates

Participants updated their beliefs about the base rate more when the presented base rate was better than their estimate of self risk, compared to trials in which the base rate was worse (i.e. classification 1: $t(31) = 5.58$, $P < 0.01$) and also when the actual base rate was better than their initial estimate of this base rate compared to when it was worse (i.e.

classification 2: $t(31) = 2.43, P < 0.03$). As detailed below, however, the finding is abolished under classification two when accounting for priors.

Specifically, as I describe in section **5.3.2.3**, examining all additional factors and ratings revealed four possible confounds: (1) under classification 1 ratings of past experience differed for stimuli for which subjects received good and bad news; (2) under classification 1 and 2 magnitude of estimation errors differed for stimuli for which subjects received good and bad news; (3) under classification two the number of trials for which subjects initially overestimated the base rate and thus received good news was greater than the number of trials in which they underestimated it and thus received bad news; (4) initial estimates of personal risk and base rates differed (this is true for both classifications). In these instances where confounds were identified, mean update scores for each participant were entered into a repeat measures ANOVA with valence (good/bad news) as a factor and the relevant scores added as covariates.

When controlling for these additional factors in the respective classifications a main effect remained under classification 1 ($F(29) = 4.59, p < 0.05$), but not 2 ($F(29) = 0.64, p > 0.4$). Looking at the latter in detail revealed that biased updating of base rates was mostly contingent on the differences in the number of trials for which subjects received good news and bad news. With that covariate alone, biased updating for base rates was abolished ($F(29) = 2.37, p > 0.13$). Without it the effect remained ($F(29) = 4.12, p = 0.05$).

5.3.2.3 Control variables

Below I detail examination of any experimental factors (i.e. memory, familiarity, past experience, perceived negativity, estimation errors, priors, number of trials) that might differ for trials in which subjects received good and bad news. I do this under both

classification schemes and for both estimations of self risk and base rates. As I described in the results reported in sections **5.3.2.1** and **5.3.2.2**, update bias was re-examined after controlling for any differences found in these variables.

Memory: Participants remembered information presented to them equally well, irrespective of whether it was good or bad news and irrespective of whether good and bad news was classified according to method 1 ($t(31) = 0.68, P > 0.50$) or method 2 ($t(31) = -0.01, P > 0.99$).

Note that a participant may recall the base rate presented accurately but be uncertain of the validity of that information. For example, they may believe they have additional/more-up-to-date information regarding base rates that the experimenter does not know about. Thus recollection of these numbers and the participant's second estimate of the base rate may differ. Comparing these two scores (i.e. recollection of base rates presented and second estimation of base rates) revealed they were not significantly different from each other, but there was a trend towards recollection of base rates being slightly higher than second estimation of base rates ($t(31) = -1.76, p = 0.09$).

Familiarity, perceived negativity, past experience: Questionnaire scores revealed that participants did not rate events for which they received good and bad news as differing in familiarity (i.e. how familiar they are with the stimuli from friends, family TV etc.) or negativity (how negative they perceive the event to be) under either classification method (see **Table 5.1**). However, under classification one participants rated events for which they received good news as greater on past experience compared to events in which they received bad news ($t(31) = 3.02, P < 0.01$). This difference was controlled for in the results reported in **5.3.2.1** and **5.3.2.2** by calculating for each participant the difference in

mean past experience ratings for good and bad news trials. These difference scores were then added as a covariate in a repeated measures ANOVA with valence (good/bad news) as a repeated factor and mean update scores as the dependent variable.

Priors (first estimates and number of trials): In accordance with past research (e.g. see Weinstein, 1980), participants believed their own likelihood of encountering a negative event was lower than their estimate of the base rate (initial estimate of self risk was lower than estimated base rate ($t(31)=-4.30$, $P<0.01$). This was observed in 84% of the participants, suggesting they believed they would fare better than average.

Furthermore, under classification 2 they would often overestimate the base rate relative to the base rate presented to them, such that the number of trials in which they received good news regarding base rates was larger than for bad news (i.e. ratio of good news trials to all trials was larger than 0.5 $t(31) = 4.93$, $p<0.01$). There were no significant differences in the number of trials in which they received good and bad news for self risk. These differences were controlled for in the results reported in sections **5.3.2.1** and **5.3.2.2**.

Questionnaire and variables	Classification 1, mean (SD)		Classification 2, mean (SD)	
	Good News	Bad News	Good News	Bad News
<i>Subjective Scales Questionnaire: All scales 1 = low to 6 = high</i>				
Familiarity	4.05 (1.04)	3.93 (1.16)	3.99 (1.09)	3.97 (1.16)
Prior experience	1.39 (0.37)*	1.23 (0.33)*	1.28 (0.31)	1.33 (0.36)
Negativity	3.97 (0.85)	4.09 (0.92)	4.05 (0.88)	4.00 (0.91)
<i>Task-related variables</i>				
Number of Trials	36.75 (13.60)	39.03 (12.66)	45.38 (8.90)*	30.59 (8.35)*
Memory errors	14.06 (5.89)	13.46 (3.81)	13.83 (5.91)	13.82 (4.10)
Initial estimate <i>self risk</i>	44.28 (6.83)*	19.30 (5.28)*	35.79 (9.03)*	24.47 (8.05)*
Initial estimate <i>base rates</i>	41.97 (5.80)*	29.40 (5.17)*	41.94 (5.02)*	26.19 (5.02)*
Estimation Error <i>self risk</i>	20.37 (4.79)	18.90 (2.63)	18.32 (4.54)*	21.43 (4.23)*
Estimation Error <i>base rates</i>	20.47 (3.83)*	15.43 (2.00)*	18.87 (3.79)*	16.68 (2.63)*

Table 5.1: Participants' ratings of familiarity, prior experience, negativity, memory errors, initial estimates, number of trials and estimation errors. *significant difference between desirable and undesirable variable ($p < 0.05$) within same classification.

Estimation Errors: As the magnitude of the update is likely to be related to the magnitude of the initial estimation error (i.e. the difference between the participant estimate and the information provided) it is critical to examine for differences in the magnitude of the errors for good and bad news trials.

Under classification two there was a difference between good news and bad news estimation errors for self risk ($t(31) = -2.52, P < 0.02$). Under both classifications there were differences between good and bad news estimation errors for base rates (classification one: $t(31) = 6.40, P < 0.01$; classification two: $t(31) = 2.40, p < 0.03$). These differences were controlled for in the results reported in sections **5.3.2.1** and **5.3.2.2** by calculating for each participant the difference in mean absolute estimation errors for good and bad news trials. These difference scores were then added as a covariate in a repeated measures ANOVA with valence (good/bad news) as a repeated factor and mean update scores as the dependent variable.

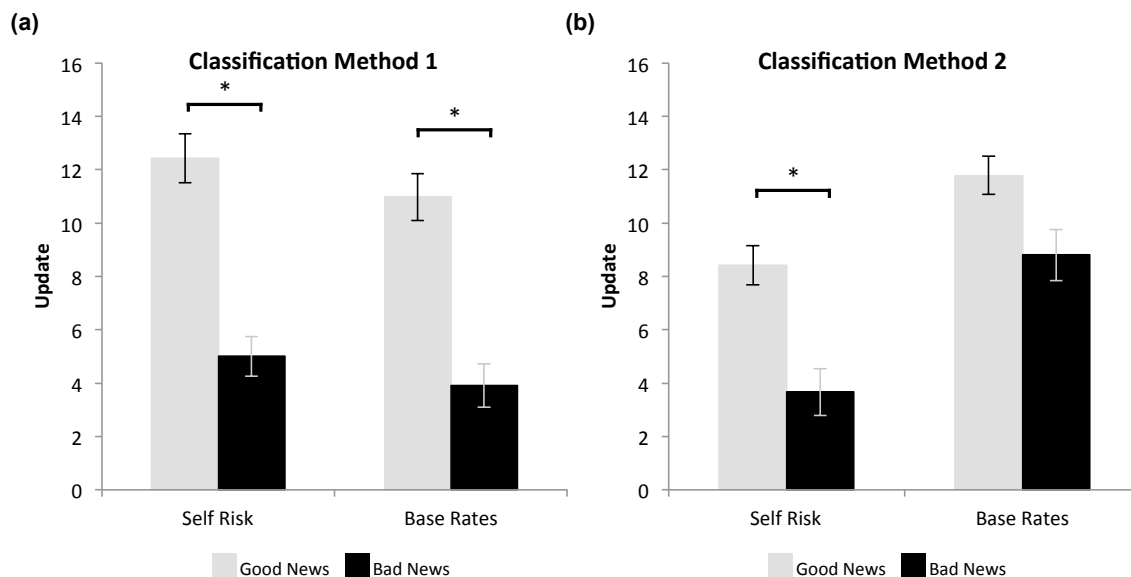


Figure 5.3: Update for self risk and base rates under different classifications of desirability. Participants update estimates of their self risk more when the information they received was good news compared to bad news. Participants update estimates of their base rates more when the information they received was good news compared to bad news under classification one but not under classification two after controlling for relevant covariates. Update calculated as first minus second estimation for good news trials and the reverse for bad news trials (positive values therefore indicate a move towards the information presented). Error bars are SEM; *indicates statistical significance at a threshold of $P < 0.05$, two tailed after controlling for all relevant covariates. Trials classified as good news when the participant overestimated the probability of the event occurring and bad news when the participant underestimated the probability of the event occurring: (a) to themselves relative to the provided base rate; (b) in the population relative to the provided base rate.

5.3.3 Do estimation errors underlie desirability and update?

5.3.3.1 Updating and estimation errors

Formal models suggest that learning from information that disconfirms one's expectations is mediated by a prediction error signal that quantifies a difference between expectation and outcome (Schultz et al., 1997; Schultz, 1998; Sutton and Barto, 1998). Previously it has been shown that an analogous mechanism underpins belief updating in this task (Sharot et al., 2011). Specifically, the difference between participants' initial estimations and the information provided (that is, estimation error) predicts subsequent updates, as

would be expected from learning models (Sutton and Barto, 1998). The strength of this association is indicative of learning. A previous study (Sharot et al., 2011) has shown that such learning is valence-dependent, being greater for information that offers an opportunity to adopt a more optimistic outlook than for information that calls for a more pessimistic outlook.

Here, I examine if updating for beliefs regarding the self and base rates are better predicted by estimation errors derived from beliefs regarding the self and base rates, and how this interacts with valence. To this end I conducted linear regressions for each participant with both Estimation Errors of Self Risk (unsigned) and Estimation Errors of Base Rate (unsigned) predicting: (1) Good News Update of self risk, (2) Bad News Update of self risk, (3) Good News Update of base rate, (4) Bad News Update of base rate, under both classifications. Regression coefficients were then tested at the group level.

As seen in **Figure 5.4**, estimation errors for self risk were better at predicting update for self risk and estimation errors for base rates were better at predicting update for base rates. In addition, the strength of the association between self-estimation errors and self-updating was valence dependent under classification one ($t(31) = 4.60, p < 0.01$), replicating previous findings (Sharot et al., 2011). Under classification 2 base rate estimation errors gain more predictive power in explaining some of the variance previously explained by self estimation errors. This resulted in neither type of estimation error alone showing a valence dependent difference in predicting self update. Rather, there was a main effect of valence such that estimation errors (of base rates and self together) were more predictive of update in response to good news than bad news information ($F(31) = 4.78, p < 0.05$).

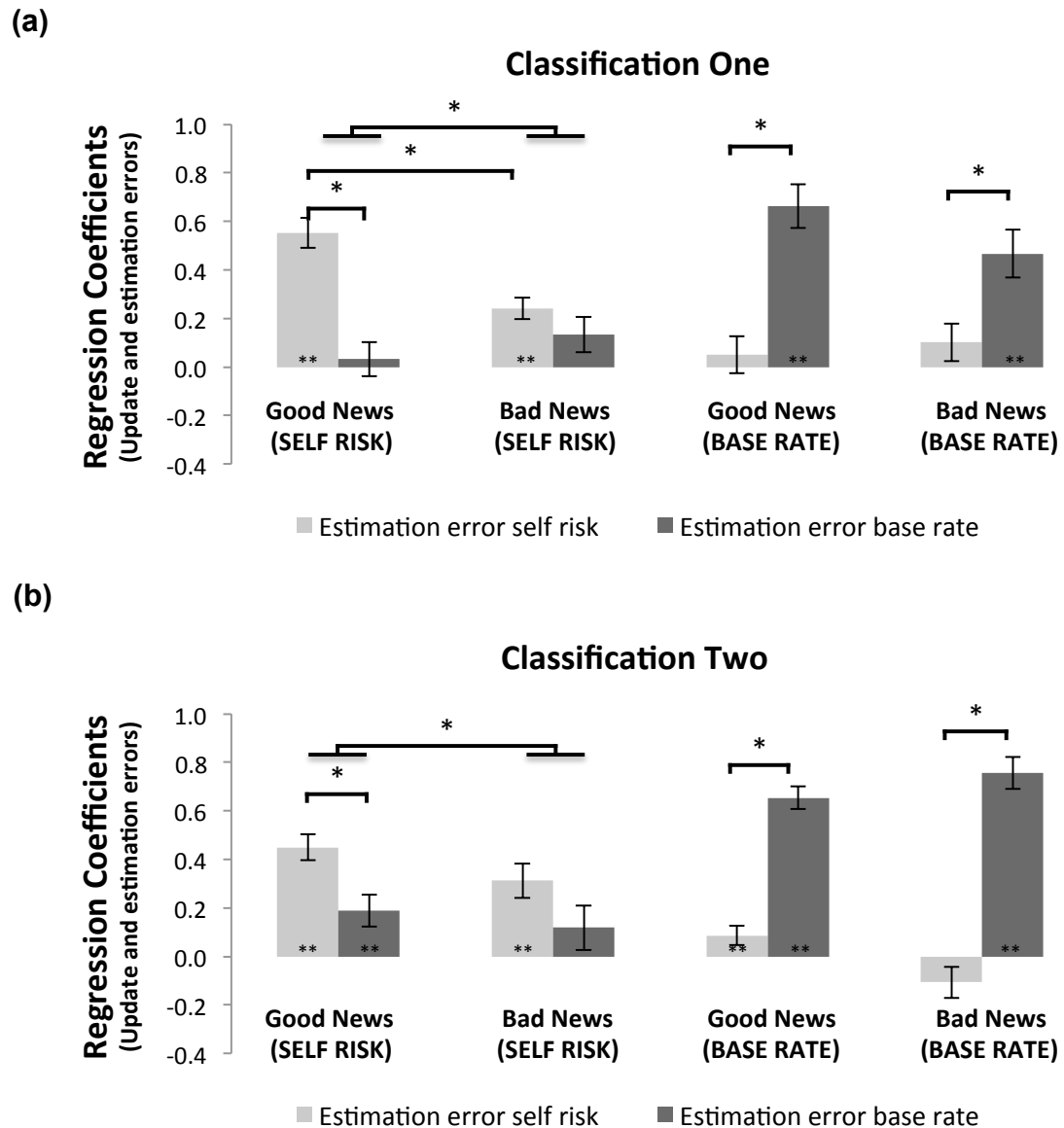


Figure 5.4: Regression coefficients predicting update from estimation errors. Estimation Error of self risk (i.e. the difference between a participants' estimate of self risk and provided base rate) significantly predicted update of self risk both for trials in which subjects received good and bad news, and under both classifications. Estimation Error of base rate (i.e. the difference between a participants' estimate of base rate and provided base rate) significantly predicted update of base rate both for trials in which participants received good and bad news, and under both classifications. Error bars are SEM; * indicates statistical significance at a threshold of $P < 0.05$, two tailed paired sample t- test after controlling for relevant factors; ** indicates significantly different to a mean of 0, one sample t-test ($P < 0.05$).

5.3.3.2 Desirability and estimation errors

Participants rated how desirable information was. I investigated whether subjective desirability ratings were driven by the extent to which the statistical information differed from participants' estimates self risk, estimates of base rates, or both.

For each participant each set of estimation errors was correlated with their desirability ratings across trials. There was a positive correlation between desirability ratings and estimation errors for self risk (mean $r = 0.50$, significantly different from zero across the population $p < 0.01$) and estimation errors for base rate (mean $r = 0.55$, significantly different from zero across the population $p < 0.01$). Specifically, participants rated information as increasingly desirable as the information provided diverged from their own estimate such that the former was a lower number. Steigers Z did not reveal a significant difference between the two sets of correlations ($Z = -0.5$, $p > 0.60$), suggesting that desirability is associated with both sets of estimation errors to a similar extent.

5.3.3.3 Effects of question order and frame

To examine whether the question order (i.e. if subject estimated their own likelihood first and then base rate or vice versa) and frame (i.e. if they were required to estimate likelihood of the event happening or not happening) influenced updating, a 3 way repeated measure ANOVA on updating scores was conducted with question order (self estimate/base rate first), frame (happen/ not happen) and valence (good/bad) entered as repeated factors under each classification. Two effects were revealed.

First, there was an interaction between valence and order for updating self risk under classification two ($F(31)=7.14$, $p < 0.02$, **Figure 5.5b**) and a trend under classification one

($F(31)=4.03$, $p=0.05$, **Figure 5.5a**) was observed. The interaction was characterized by greater valence-dependent updating when subjects estimated their own vulnerability before estimating base rate. This interesting result suggests that biased updating is reduced yet still significant when we first consider population base rates and only then our own likelihood (classification one; $F(31)=20.67$, $p<0.01$; classification two; $F(31)=6.39$, $p<0.05$). This may be because initial self-estimates tend to be more accurate when reported after estimates of base rates ($t(31) = 1.74$, $p = 0.09$) and/or because undesirable information of base rates may be more difficult to ignore under such ordering.

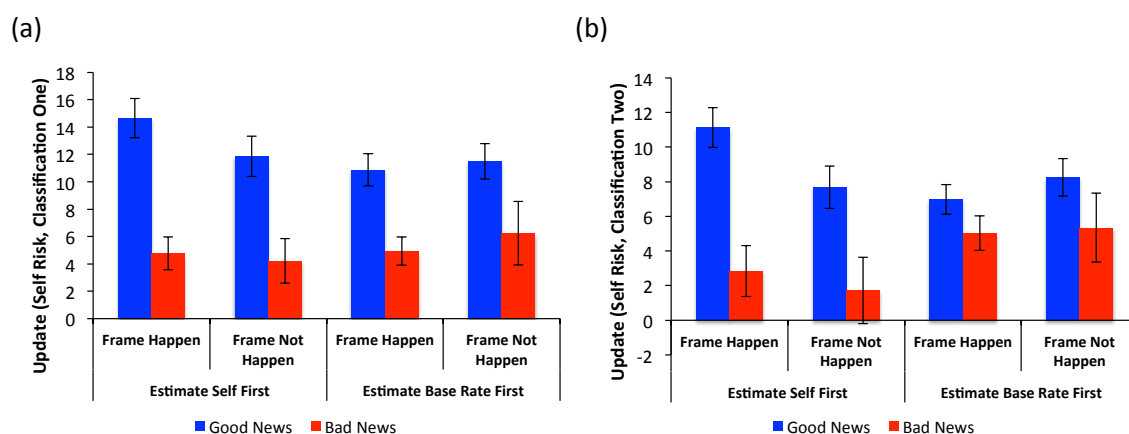


Figure 5.5: Examining order and frame effects for updates of self risk revealed (a) a trend for a valence by order interaction under classification one ($F(31)=4.03$, $p=0.05$). (b) This interaction was significant under classification two ($F(31)=7.14$, $p<0.02$). In both instances, the interaction was characterized by greater valence-dependent updating when subjects estimated their own vulnerability before estimating base rate.

Second, a main effect of frame for updating of base rates was found under classification one ($F(31)=5.40$, $p<0.05$, **Figure 5.6**) with updating for “happening” being greater than updating for “not happening”. However, the effect was not significant under classification two, nor for self risk under either classification

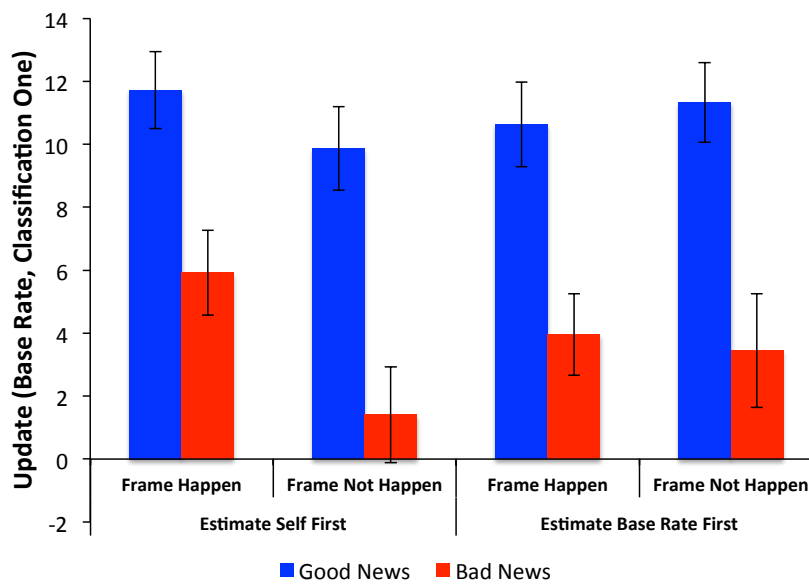


Figure 5.6: Examining order and frame effects for updates of base rates revealed a main effect of frame for updating under classification one ($F(31)=5.40$, $p<0.05$).

5.4 Discussion

These results give further evidence that individuals selectively update their beliefs when estimating their own risk; updating their estimates to a greater extent and more akin to a Bayesian agent in response to information that offers an opportunity to adjust predictions in an optimistic direction relative to information that can reduce optimism. Importantly, these results further show that valence-dependent updating of self risk reported previously (Chowdhury et al., 2014; Garrett et al., 2014; Moutsiana et al., 2013, 2015; Sharot et al., 2011, 2012a, 2012b) is not contingent on the specific method by which trials are divided into “good news” and “bad news”.

In the original task (Sharot et al., 2011) participants estimate their own probability of encountering negative events and then receive information regarding the population base rates. Trials are then labelled “good news” if the base rates provided are better than the participants’ estimate of self risk and “bad news” if the base rates provided are worse

than the participants' estimate of self risk. Such a division has proved useful when examining how the brain codes for the difference between ones' estimate of self risk and information regarding base rates. Specifically, it has been found in the past that the left inferior frontal gyrus, medial frontal lobe and cerebellum track the magnitude of the difference between a person's estimate of self risk and information regarding base rates when that information is better than the person's estimate of their own vulnerability, while the right inferior frontal gyrus codes for the magnitude of such errors when the information is worse (Sharot et al., 2011).

One can imagine a scenario where an individual holds a different estimate of their self risk and of the base rate and receives information regarding the base rate that is worse than their estimate of their self risk but better than their estimated base rate. Under this scenario whether the information is good news or not may be ambiguous. In this study by asking participants to label the information themselves as good or bad and dividing the trials in two different ways (according to the participants' estimates of their self risk and according to the participants' estimates of base rates) I show that the update bias exists under both methods of division.

It has been suggested that when estimating vulnerability (Kahneman and Tversky, 1973) a person may take into account both population base rates and diagnostic information to reach a prediction regarding personal risk. Thus, information regarding base rates will result in adjusting both estimates of personal risks and estimated base rates. Selective updating of either may lead to biased estimates for self risk. Here, I show that updating of beliefs regarding population base rates is not as robust and clear-cut as updating for self risk. Specifically, biased updating for base rates could partially be

explained by priors; subjects tended to overestimate base rates such that there were more trials in which they ended up receiving desirable information. When accounting for this difference the bias for updating base rates did not survive under classification 2.

Ample evidence suggests that people's perception of their vulnerability is biased in a positive direction (Weinstein and Klein, 1995). This study supports these past findings and demonstrates the robustness of the effect as it is observed under different empirical definitions of desirability of information provided.

Chapter 6

Optimistic Update for Positive Life Events? An Unbiased Test

6.1 Introduction

When it comes to estimating future events (rather than evaluating abilities and characteristics) most (e.g. Baker and Emery, 1993; Kuzmanovic et al., 2015, 2016; Sharot et al., 2011) but not all (Weinstein, 1980; Wiswall and Zafar, 2015), studies have examined predictions specifically regarding aversive events (such as illness and violent acts). In the literature at present (to my knowledge) there has been only one previous peer reviewed study that examined updating of beliefs regarding a future positive life event (Wiswall and Zafar, 2015). That study revealed that people update beliefs to a greater extent in response to evidence suggesting they are likely to earn more than they thought, relative to evidence suggesting they are likely to earn less. While that study suggests that optimistic updating of beliefs is indeed observed for positive life events, it is unknown whether biased updating for positive life events is greater, smaller, or equal than for

negative life events. As unrealistic optimism consists both of overestimating the likelihood of positive events and underestimating the likelihood of negative events (Sharot, 2011; Weinstein, 1980), the question of whether the same mechanism underlies both types of events equally is important for understanding optimism.

6.1.1 Obtaining statistics for positive events.

In the original belief updating task (Sharot et al., 2011), participants are asked to estimate their likelihood of experiencing 80 aversive events in their lifetime (*first estimates*). They are then presented with the likelihood of these events in their population (*information*) and subsequently asked to estimate their likelihoods again (*second estimate*). Trials are then divided into ones where participants received *good news* (they learn that an aversive event is less likely than they thought) and trials where participants received *bad news* (they learn an aversive event is more likely than they thought). *Update* is calculated as the difference between the first and second estimate. When attempting to adapt this task to study positive life events researchers face a number of potential confounds which, if ignored, will lead to invalid conclusions. A number of these including the importance of controlling for estimation errors are discussed in **Chapter 2**, however one remains.

Specifically, whilst validated statistics regarding the likelihood of encountering negative events in one's life-time are well documented (such as the likelihood of suffering different type of illness or being a victim of crime, see **Chapter 2** for full list of negative life events used in the UBT), statistics about the occurrence of positive life events are not readily available. This is problematic, as the UBT requires the use of many trials and stimuli. Yet, it is practically impossible to find even 40 positive life events accompanied

by validated statistics. One may be tempted to make up statistics for positive events to use in a study. However, the invalidity of such made up statistics will quickly become apparent to the subject, introducing a serious confound, which will make the exercise useless.

Here, to generate meaningful stimuli for both positive and negative life events I altered the UBT as follows: I asked participants to estimate their likelihood of encountering everyday life events in the *upcoming month*. I obtained the frequencies of such events by asking over 200 participants to report whether different common positive and negative life events occurred to them at least once in the past month. I then used this data to construct a list of base rates for each event (i.e. the likelihood of each life event occurring at least once in a given month in the sample). I then run the UBT on an alternate, but demographically well-matched, set of participants, asking them to estimate their likelihood of encountering these events in the next month.

6.2 Materials and Methods

6.2.1 Construction of stimuli

6.2.1.1 Participants

300 participants located in the United States completed the survey on Mechanical Turk. As in past studies of the UBT (Garrett and Sharot, 2014; Moutsiana et al., 2013, 2015), I excluded participants with a high Beck Depression Inventory (BDI) score indicating potential depression. 73 participants were excluded for having a BDI score greater than 11 (final sample = 227). Participants were all between 20 and 30 years of age (inclusive).

Completion of the survey took approximately 25 minutes and participants were compensated \$3.5 for their time.

6.2.1.2 Task

The survey began by collecting basic demographic information from participants (age, level of education, marital status, employment status, monthly income). Two training trials were then presented to familiarize participants with the task. Participants were then presented with 100 different commonly occurring life events for 3 seconds each. These were a mixture of positive events (for instance: “Discovered a new song you like”, “Laughed at a joke”) and negative events (for instance: “Had an argument with a family member”). Whilst the event was displayed on screen, participants were instructed to recall whether this event had happened to them in the past 4 weeks. They were then asked to indicate either (1) Yes: This event occurred to me at least once in the past 4 weeks; or (2) No: This event did not occur to me in the past 4 weeks. The order of these two options was counterbalanced. Participants had unlimited time to make a response (**Figure 6.1**).

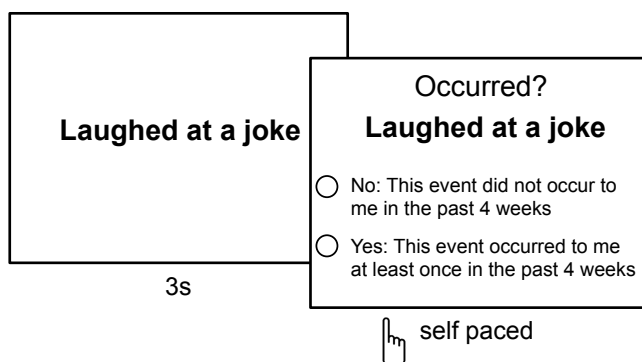


Figure 6.1: Construction of stimuli. Participants were presented with 100 commonly occurring positive and negative life events and were instructed to recall whether this event had happened to them in the past 4 weeks. Data was then used to construct a list of base rates.

After completing the survey, participants rated each event on a 5 point likert scale (1=Very Negative; 2=Negative; 3=Neutral; 4=Positive; 5=Very Positive) and then completed the BDI (Beck et al., 1961) and LOTR (Scheier et al., 1994). The survey was constructed and presented using web based survey service Qualtrics.

6.2.1.3 Analysis and event selection

For each event, the percentage of participants who indicated the event had occurred to them in the past month (out of all participants who completed the study) was calculated. A subset of the events (n=54) were selected for use as stimuli. Events were selected such that the range of each type of event (positive and negative) was normally distributed around a mean of 50% (standard deviations: positive events = 0.17, negative events = 0.21).

6.2.2 Update Bias Task

6.2.2.1 Participants

200 participants located in the United States (age range 20 and 30) completed the survey on Mechanical Turk. 56 participants were subsequently excluded for having a BDI score above 11 indicating possible depression. A further 2 participants were excluded because the range of their responses were limited, resulting in zero trials in either the “good news” bin or “bad news” bin, making comparison impossible (final n = 142). There were no differences in age, education, income, marital status or employment status between this set of participants and participants that had completed the base rate survey used to construct the base rate statistics (all $P > 0.20$). Completion of the survey took approximately 1 hour and participants were compensated \$7 for their time.

Task. The survey began with an attention check designed to filter out participants that did not read instructions prudently. Then, demographic information was collected (age, level of education marital status, employment status, monthly income) and 2 training examples were run to familiarize participants with the task.

In the first session, on each trial participants were presented with 1 of 54 life events (see **Table 6.1** for list of events) and asked to imagine the event happening to them in the month ahead. They were then asked to estimate how likely that event was to happen to them in the next 4 weeks. Participants were instructed to type in an estimate between 5% and 95%. Trials with responses outside this range were excluded from analysis. Participants were then shown the base rate statistic of the event happening in the next 4 weeks, which ranged from 15% to 85% (see **Figure 6.2**). They were told that the statistic was the average likelihood of this event happening at least once in the next four weeks to someone from the same socioeconomic environment as them. In a second session, participants were asked to re-estimate how likely each event was to happen to them in the next 4 weeks.

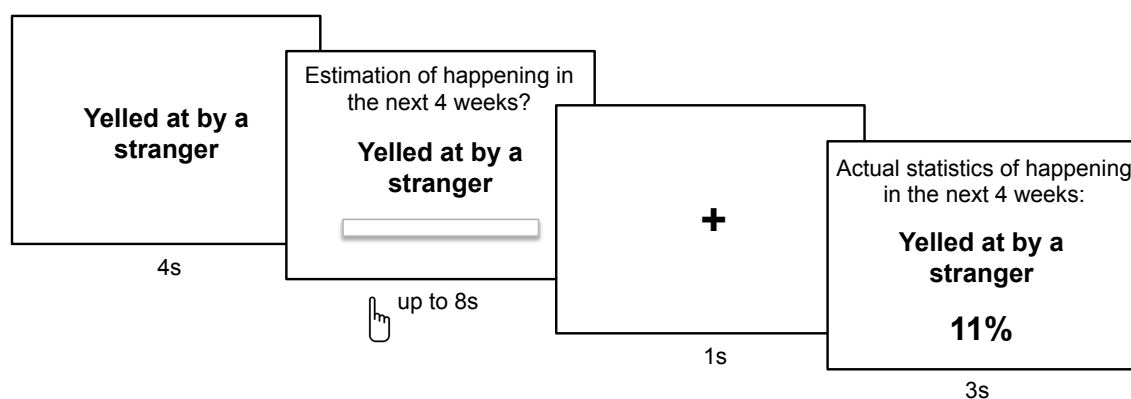


Figure 6.2: Update Bias Task. On each trial, participants were presented with a short description of one of 54 events and asked to estimate how likely this event was to occur to them. They were then presented with the average probability of that event occurring to a person like themselves (calculated from the previous task). In a second, session, participants were asked to re-estimate how likely the event was to occur to themselves. For each event

an update term was calculated as the difference between the participant's first and second estimations, such that positive numbers indicate a move toward the base rate.

Life Event	Base Rate %	Mean Desirability Rating
<i>Attend a party</i>	45%	4.12
<i>Cook dinner for friends</i>	36%	4.06
<i>Donate money to a needy person or cause</i>	37%	4.26
<i>50 hours or more sleep in a single week</i>	56%	3.65
<i>Exercise at least twice in a week</i>	70%	4.19
<i>Finish reading a book</i>	41%	3.95
<i>Fix a broken possession</i>	39%	3.96
<i>Find or receive a gift of a dollar or more</i>	56%	4.21
<i>Get a haircut</i>	45%	3.56
<i>Get invited to a party</i>	58%	4.00
<i>Get taken out for dinner</i>	61%	4.19
<i>Have a sexual encounter that you enjoy</i>	69%	4.59
<i>Have a supervisor or teacher praise your work</i>	54%	4.28
<i>Have an out of town friend visit you</i>	30%	4.15
<i>Have your photo taken</i>	75%	3.32
<i>Invite a non-family member to a meal</i>	49%	3.89
<i>Learn a new skill related to work or school</i>	48%	4.19
<i>Make a purchase in excess of \$50 for your personal enjoyment</i>	65%	3.90
<i>Meet with your supervisor</i>	56%	3.19
<i>Participate in a game of sport</i>	29%	3.71
<i>Play a board game</i>	29%	3.80
<i>Play with a pet</i>	75%	4.26
<i>Run into an old friend that you haven't seen in a long time</i>	30%	4.19
<i>Receive a pay check</i>	81%	4.48
<i>Receive a complement about how you dress</i>	54%	4.26
<i>Shop for clothes</i>	56%	3.69
<i>Successfully teach someone a new skill or concept</i>	50%	4.11
<i>Take a day or more of annual leave</i>	19%	3.48
<i>Try out a new food or dish</i>	74%	4.00
<i>Try out a new hobby, craft, or sport</i>	31%	4.01
<i>Go out of town for leisure</i>	36%	4.21
<i>Wish a friend a happy birthday</i>	67%	4.08
<i>Win a competitive game of sport</i>	22%	4.03
<i>Burn something that you are cooking</i>	41%	2.01
<i>Embarrass yourself</i>	60%	1.84

<i>Family or friend get ill</i>	56%	1.63
<i>Find out that someone you know personally has died</i>	15%	1.30
<i>Get lost</i>	26%	1.93
<i>Get rejected by someone</i>	17%	1.65
<i>Get sick or suffered a physical illness</i>	41%	1.62
<i>Have a disagreement with a friend</i>	43%	2.06
<i>Have a headache</i>	82%	1.99
<i>Hear about a natural disaster</i>	85%	1.84
<i>Hear of a terrorist attack</i>	35%	1.73
<i>Hurt someone's feelings</i>	52%	1.85
<i>Ill one day because of overdrinking</i>	21%	1.83
<i>Received a phone call from a telemarketer</i>	52%	2.07
<i>Saw a dead animal/human</i>	56%	1.67
<i>Stay up past 2 AM for school or work</i>	40%	2.44
<i>Stuck in traffic</i>	71%	2.00
<i>Teased at/made fun of</i>	35%	1.92
<i>Get lied to</i>	60%	1.67
<i>Receive a utility bill</i>	78%	2.51
<i>Clean the bathroom</i>	78%	3.12

Table 6.1: List of the stimuli used in the study, their respective base rates and mean desirability rating. Base rates were generated from an independent set of participants tasked with reporting whether each event had happened to them at least once in the previous month. Life events were classified as positive or negative separately for each participant according to their own rating. Hence some events may be classified as positive for some participants but negative for others. There was, however, a high level of agreement among participants (interclass correlation coefficient = 0.75). Life events rated as neutral (i.e. neither positive nor negative) were not included in the analysis. On average 27 events were categorized as positive, 18 as negative and 7 as neutral.

After completion of the task, I tested participants' memory for the information presented. Participants were asked to recall the information previously presented of each event. Subsequently, participants were then asked to rate all life events according to how positive or negative they found them on a likert scale (1=very negative, 2=negative, 3=neutral, 4=positive, 5=very positive). See **Table 6.1** for mean ratings of desirability across participants used to compile base rates for each event. They were also asked to rate past experience with each event ("Has this event happened to you before?" From 1 = never to 6 = very often).

Three quarters of participants (75%) also rated all events on: vividness (“How vividly could you imagine this event?” From 1 = not vivid to 6 = very vivid); familiarity (“Regardless if this event has happened to you before, how familiar do you feel it is to you from TV, friends, movies and so on?” From 1 = not at all familiar to 6 very familiar); and arousal (“When you imagine this event happening to you how emotionally arousing is the image in your mind?” From 1 = not arousing at all to 6 = very arousing). The scores of these are reported in **Table 6.2** Participants then completed the BDI and the LOTR. The survey was constructed and presented using web based survey service Qualtrics.

	Positive Life Events, mean (SD)		Negative Life Events, mean (SD)	
	Good news	Bad news	Good news	Bad news
Ratings				
<i>Subjective Scales</i> <i>Questionnaire:</i> <i>1 = low to 6 = high</i>				
Familiarity^{L, V*L}	4.68 (0.77)	5.11 (0.70)	4.75 (0.84)	4.35 (0.99)
Vividness^{L, V*L}	4.42 (0.82)	4.83 (0.65)	4.34 (0.84)	3.97 (0.99)
Emotional arousal^{V, L, V*L}	3.75 (0.91)	3.92 (1.00)	3.89 (0.83)	3.10 (1.27)
Prior experience^{V, L, V*L}	3.80 (0.81)	4.51 (0.68)	3.75 (0.84)	3.18 (0.91)

Table 6.2: Additional ratings provided by 75% of the participants. Prior Experience rated by all participants. A 2*2 repeated measures ANOVA was conducted for each subjective rating with life event type (positive/negative life event) and valence (desirable/undesirable) as factors. ^L Main effect *life event* valence (positive/negative) $p < 0.05$; ^V Main effect information *valence* (good news/bad news), $p < 0.05$; ^{V*L} Interaction (valence by life event), $p < 0.05$

6.2.2.2 Analysis

Life events were categorized as negative or positive for each participant individually according to their own evaluation. Specifically, events were classified as positive if the

participant rated the event as 4 (positive) or 5 (very positive) in the ratings section of the task, and negative if rated as a 1 (very negative) or 2 (negative). Events with a neutral rating of 3 were excluded from the analysis.

For each type of event, participants could receive either “good news” or “bad news” depending on whether the participant initially overestimated or underestimated the probability of the event relative to the base rate (see **Table 6.3**). Specifically, if their first estimate was lower than the base rate presented, the information would be categorized as “good news” if the life event was positive and “bad news” if the life event was negative (column 1, **Table 6.3**). If their first estimate was higher than the base rate presented, the information would be categorized as “bad news” if the event was rated as a positive life event and “good news” if the event was rated as a negative life event (column 2, **Table 6.3**). Trials in which the initial estimate was equal to the statistic presented were excluded from subsequent analyses as these could not be categorized into either condition (less than one negative life event trial and less than one positive life event trial on average per participant).

	Initial Estimate < Base Rate	Initial Estimate > Base Rate
Positive Life Event (e.g. Get invited to a party)	Good News	Bad News
Negative Life Event (e.g. Have a headache)	Bad News	Good News

Table 6.3: Categorization of events

Belief update was calculated for each trial and participant as the difference between first and second estimate. As done previously (Garrett et al., 2014; Moutsiana et al., 2013; Moutsiana, et al., 2015; Sharot, Kanai, et al., 2012) update was calculated such that positive scores indicate a move towards the base rate, regardless of event type and

valence categorization, and negative scores a move away from the base rate. Mean update scores for each participant were entered into a 2 (good/bad news) by 2 (positive/negative life event) repeated measures ANOVA. Controlling for (1) the difference in memory for good news trials and bad news trials, both for positive and negative stimuli, (2) the difference in number of good news trials and bad news trials, both for positive and negative stimuli (3) the difference in absolute estimation errors for good news trials and bad news trials, both for positive and negative stimuli (estimation error = | first estimate – base rate |). These were controlled for by entering the difference scores (6 in total) as covariates.

6.3 Results

There was an asymmetry in updating such that participants updated more in response to good news than bad news. This bias was not significantly different for positive and negative events. Specifically, entering update scores into a 2*2 repeated measures ANOVA with desirability of information (good/bad news) and life event type (positive/negative life event) as factors (controlling for differences in memory, differences in number of trials and differences in estimation errors) revealed a main effect of desirability of information ($F(1,135)=6.29, p<0.02$), no effect of event type ($F(1,135)=0.08, p=0.78$) and no interaction ($F(1,135)=0.31, p=0.58$). The main effect of desirability was characterized by greater updating in response to good news compared to bad news for both positive life events (mean good news update=8.71, mean bad news update=7.79) and negative life events (mean good news update=10.66, mean bad news update=6.62). **Figure 6.3.**

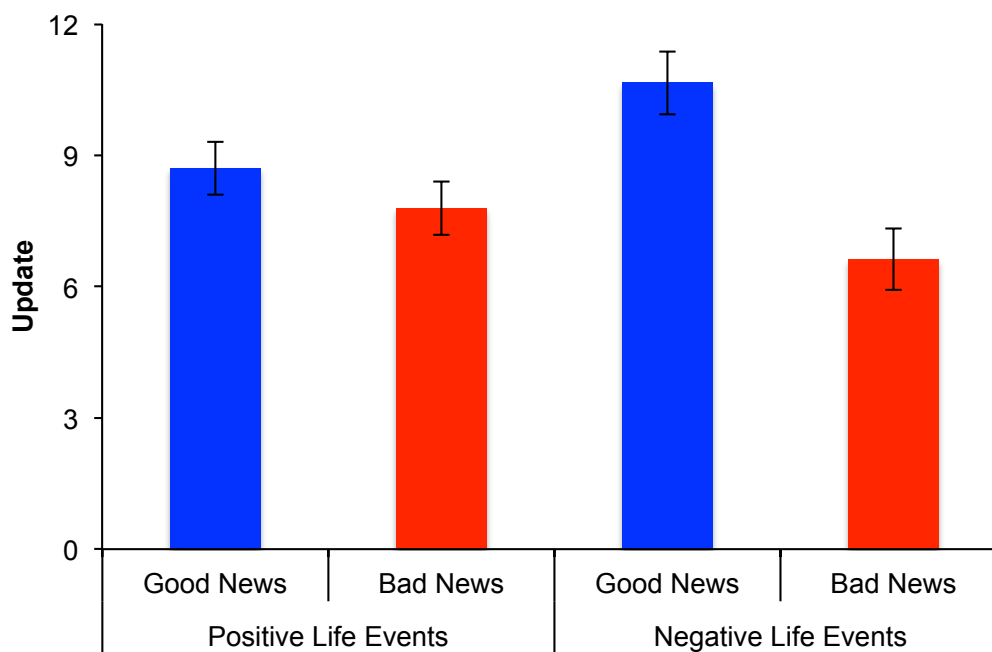


Figure 6.3: Biased Updating. Updating is greater for good news compared to bad news (main effect of desirability of information: $F(1,135)=6.29$, $p<0.02$). Error bars represent SEM.

6.4 Discussion

In this study, because valid base rates regarding the likelihood of positive events occurring during a person's lifetime are difficult to come by, I altered the UBT. Specifically, I elicited real statistics for both positive and negative life events by sampling a large group of participants prior to conducting the study, asking them to indicate which of 100 different events did and did not happen to them in *the last four weeks*. This provided a set of base rates, a subset of which were used to test a second group of participants, matched to the first, on the UBT. The second set of participants were also asked to estimate the likelihood of events happening in the next four weeks, rather than in a lifetime.

The current results show a valence dependent asymmetry in how participants update their beliefs, consistent with other findings (Eil and Rao, 2011; Korn et al., 2012; Kuzmanovic et al., 2015, 2016; Lefebvre et al., 2016; Mobius et al., 2012; Sharot, 2011; Sharot and Garrett, 2016; Sharot et al., 2011, 2012b). Here, I observe this asymmetry when people update their beliefs regarding their likelihood of experiencing every day events in an upcoming month. In particular, participants updated their beliefs to a greater extent when receiving good news regarding the likelihood of experiencing future life events in the next four weeks relative to bad news. Such an asymmetry in belief updating has been suggested as a mechanism supporting optimism (Sharot and Garrett, 2016).

Whilst unrealistic optimism has been previously reported for both positive (e.g. winning an award) and negative (e.g. divorce) life events (Weinstein, 1980), the extent of asymmetric updating for positive and negative life events has never been compared. The aim of this study was to compare optimistic updating (i.e. updating more when receiving desirable compared to undesirable information) for future *positive* life events with that for future *negative* life events. These results show no statistical differences in the pattern of updating for positive and negative life events - in both cases participants (n= 142) updated their beliefs to a greater extent when receiving desirable information compared to undesirable information, regardless of whether the information was regarding a positive life event (such as: "Receive a complement about how you dress") or a negative life event (such as: "Hurt someone's feelings").

In sum, this study (1) extends the finding of an asymmetry in belief updating to everyday life events and (2) reveals a similar pattern of asymmetric updating for positive and negative life events.

Chapter 7

General Discussion

The studies that comprise this thesis (Chapters 3-6) comprise two sections. The first section - Chapters 3 and 4 – set out to explore individual variation in biased information integration, investigating first how it is altered in patients with clinical depression (**Chapter 3**) and when individuals encounter a threatening environment (**Chapter 4**). The second section - Chapters 5 and 6 - set out to test the robustness of biased information integration. Specifically, these studies examine whether biased information integration exists under different definitions of good and bad news (**Chapter 5**) and whether biased information integration occurs for positive as well as negative life events (**Chapter 6**).

This discussion will provide a brief summary of the aims, findings and limitations of each chapter. Finally it will provide some suggestions for future research.

7.1 Summary and limitations of experimental investigations

7.1.1 Chapter 3: Neural Substrates of Unbiased Belief Updating in Depression

7.1.1.1 Summary

In this study, I investigate whether biased information integration is altered in patients with MDD. I use fMRI in conjunction with the UBT administered to clinically depressed patients and healthy controls to characterize brain activity that supports belief updating and differences in belief updating between clinically depressed patients and healthy controls. The behavioural results reveal an interaction between valence (whether a piece of news is good or bad) and group (MDD versus Control) for updating. This is characterized by greater updating in response to bad news by the MDD group relative to the control group. As a result, the MDD group do not show the bias in updating observed in the control group.

The fMRI results in this study suggest that unbiased belief updating in depression is mediated by strong neural coding of estimation errors in response to both good news (in left inferior frontal gyrus and bilateral superior frontal gyrus) and bad news (in right inferior parietal lobule and right inferior frontal gyrus) regarding the future. In contrast, intact mental health was linked to a relatively attenuated neural coding of bad news about the future. These findings identify a potential neural substrate which mediates the breakdown of biased updating in Major Depression Disorder, which may be important for mental health.

7.1.1.2 Limitations

One limitation of the study reported is that the depressed patients recruited were moderately rather than severely depressed. Another study that also used the UBT to examine differences between MDD and healthy controls (Korn et al., 2013) used a severer sample of depressed patients (mean BDI of 32.6 compared to 25.8 in the patient group used in **Chapter 3**). This study also found that MDD patients were unbiased in their updating. However, this was the result of reduced updating of good news rather than enhanced updating of bad news as I report here.

Another limitation is that as a result of using a between subject design it was not possible to directly show that asymmetric updating varies with depression within an individual. A within subject design testing MDD patients both in and out of a depressive episode at different time points would be needed to test this.

7.1.2 Chapter 4: Updating Beliefs Under Threat

7.1.2.1 Summary

In this study I investigate if an asymmetry in information integration is altered in response to stress. I provide evidence that biased information integration is state dependent and evaporates under stress. This pattern of results was observed in a controlled laboratory setting (Experiment I) and in firefighters on duty (Experiment II). Such flexibility in how individuals integrate information may enhance the likelihood of responding to warnings with caution in threatening environments, whilst maintaining a positivity bias otherwise, a strategy that may increase well-being.

7.1.2.2 Limitations

A limitation to this study is that in Experiment II, whilst it was possible to obtain self report measures of stress from participants (firefighters), it was not possible to get physiological measures (SCR, cortisol) as it was for Experiment I, owing to participants conducting the study online and whilst on call, rather than in the laboratory. Another drawback with Experiment II is that it was not possible to tie self report measures of stress to a specific external event. Whilst likely that some stress reported is due to the nature of the job, it is unknown whether a particular aspect of the job is the stressor (such as how many call outs, how stressful the last call out was) and/or whether factors unrelated to the job are also contributing.

A limitation of both Experiments I and II is that as a result of using a between subject design it was not possible to directly show that asymmetric updating varies with stress within an individual. A within subject design testing participants twice, once when under stress and once when not under stress would be required to test this.

7.1.3 Chapter 5: How Robust is the Optimistic Update Bias for Estimating Self-risk and Population Base Rates?

7.1.3.1 Summary

The aim of the study reported in **Chapter 5** was to examine the robustness of the update bias by examining whether it exists when a different classification scheme is used to classify events as good and bad news. I ran a version of the UBT in which participants were asked to provide estimates of the base rate of each event as well as estimate their own likelihood. Events were categorised as good news or bad news in two different ways: (1) According to whether estimate for *own likelihood* was above or below the base rate

presented, as done in previous studies using the UBT (Chowdhury et al., 2014; Korn et al., 2013; Kuzmanovic et al., 2015, 2016; Moutsiana et al., 2013, 2015; Sharot et al., 2011, 2012a, 2012b); (2) According to whether estimate for the *base rate* was above or below the base rate presented (also see Kuzmanovic et al., 2015, 2016).

Participants update beliefs regarding risk in the population in an asymmetric manner regardless of the different empirical definitions of desirable information (i.e. under both classifications, biased updating is present). In contrast, valence-dependent updating of base rates were accounted for by priors. The presence of a bias in updating for the self and absence of a bias in updating for base rates is consistent with a separate study, conducted by a different research group using a different version of the UBT (Kuzmanovic et al., 2015, 2016).

I analysed the degree to which participants updating patterns correspond to that of a Bayesian agent. I did this separately for receipt of good news and bad news, again under both classifications of good/bad news. The results from this analysis revealed that there was a tighter correspondence between updating by a Bayesian agent and by human participants when in receipt of good news compared to when in receipt of bad news. This was the case under both classification schemes. This is consistent with other studies that have examined updating beliefs about personality traits such as attractiveness (Eil and Rao, 2011) and intelligence (Eil and Rao, 2011; Mobius et al., 2012) using a Bayesian approach.

This study expands our knowledge of information integration and how it is guided by valence. It shows that individuals update information about their own likelihoods to a greater extent, and in a more Bayesian manner, when receiving good news compared to

bad news. This is not something selective to a specific method of classifying good and bad news.

7.1.3.2 Limitations

A limitation of this study is that the design does not capture a full distribution of participants priors. This is also the case in other studies reported in this thesis. However, it is a particular limitation of this study since it prevents a full characterisation of how a Bayesian agent would update given the distribution of priors. If a full prior distribution was elicited from participants (for example, by having them provide repeat estimates, draw out a probability distribution or provide confidence intervals), as done in another study (Eil and Rao, 2011), this would enable a more accurate prediction of how a Bayesian agent should update following presentation of each base rate, particularly in instances where prior beliefs are skewed.

7.1.4 Chapter 6: Optimistic Update for Positive Life Events? An Unbiased Test

7.1.4.1 Summary

The majority of studies that have examined belief updating about the likelihood of negative life events occurring in the future have used negative aversive life events exclusively (Chowdhury et al., 2014; Garrett and Sharot, 2014; Korn et al., 2013; Kuzmanovic et al., 2015, 2016; Moutsiana et al., 2013, 2015; Sharot et al., 2011, 2012a). In the study reported in **Chapter 6**, I examined belief updating for positive *and* negative life events. Owing to an absence of statistically meaningful statistics for positive life events that match the negative life events that are typically used, I examined this using a

new approach. Specifically, I ran an online version of the UBT using everyday life events as stimuli. To get meaningful base rates for these events I asked a large group of online participants to report whether each life event had occurred to them at least once in the previous month. This data set was used to calculate the average likelihood of each event occurring in this group in a month. The UBT was then run on a separate group of participants (well matched to the first group for demographics) who were asked to estimate the likelihood of each event to occur in the following month, presented with base rates ascertained from the first group and subsequently asked to re-estimate their own likelihood of each event to occur the following month.

The results from this study revealed that an overall update bias exists when both positive and negative life events are used (main effect of valence, no main effect of life event or interaction between valence and life event). As well as extending the remit of biased updating to positive and negative life events, this study also expands previous findings in other ways. It shows that biased updating occurs for everyday life events and that it exists in a new different demographic sample - MTurk workers living in the US – which has not been tested before (Garrett and Sharot, 2014; Kuzmanovic et al., 2015, 2016; Sharot et al., 2011, 2012a; Moutsiana et al., 2013).

7.1.4.1 Limitations

A limitation is that whilst participants rated how positive/negative they considered each event, they were not asked to rate how important or significant each event would be to them if it occurred and this might be just as important a factor as valence. This is discussed in greater detail in the next section.

7.2 Key questions for future research

7.2.1 The interplay between anxiety, biased updating and depression

The results reported in **Chapter 3** revealed an association between depression symptomology and the degree of positive bias in information integration. Separate to this, **Chapter 4** revealed an association between anxiety and the degree of positive bias in information integration. Characterising the way in which anxiety, depression and biased belief updating relate to and impact one another (along with their co-morbidity) are key questions for future research.

A recent cognitive neuropsychological model of depression (Clark et al., 2009; Roiser et al., 2011) provides a framework for understanding the factors that lead to alterations in information processing and at what stages these factors could fit into both the onset and development of depression. Simplified, the model proposes that genetic factors and stressful life events cause alterations in monoamine transmission which disrupt the functioning of low level brain circuits. These result in negative affective processing biases (biasing individuals towards negative stimuli and/or away from positive stimuli) in a range of domains such as emotional perception (Gilboa-Schechtman et al., 2002) and emotional memory recall (Bradley et al., 1995). Importantly, these negative affective processing biases play a causal role in the initial development of depression rather than being a symptom of it. Over time, these negative affective processing biases cause negative schemata's to develop (Beck, 1967) which, once established, can lead to the development of subsequent top down negative affective processing biases such as deficits in affective cognitive control (Clark et al., 2009) which can exacerbate the state of depression and/or contribute to its persistence.

One hypothesis about the directionality of the different factors, consistent with both the cognitive neuropsychological model of depression (Roiser et al., 2011) and current findings in relation to the UBT, is that stress and/or genetic factors cause individuals to be more sensitive to negative information, possibly as a result of changes in dopamine function (dopamine has previously been shown to alter belief updating in the UBT, see Sharot et al., 2012b). This in turn leads these individuals to form pessimistic beliefs (Strunk et al., 2006) which contribute to the onset of depression. Under this hypothesis, alterations in belief updating occur in part as the result of stress and occur prior to the initial onset of depression.

An alternative hypothesis is that initial onset of depression occurs as the result of alterations to low level information processing biases specifically such as emotional attention. It does not occur as the result of changes to information integration in relation to beliefs; since these can be construed as more abstract and high level forms of information processing, alterations occur at a later point in time. Depression is often comorbid with anxiety (Baldwin et al., 2002; Cameron, 2005) and findings in **Chapter 4** showed a relationship between anxiety and individuals sensitivity to negative information. Under this hypothesis, alterations in belief updating also occur after the initial onset of depression and either mediates or is mediated by anxiety.

Both of these hypothesis are consistent with the fact that throughout the life span anxiety (Deeming, 2013), depression (Blanchflower and Oswald, 2008) and balanced integration of information (Chowdhury et al., 2013; Moutsiana et al., 2013) follow an inverse U shape; being low for teenagers and the elderly and high in middle-aged. Note

that the two hypothesis are not mutually exclusive and increased sensitivity to negative information could be a factor in both the onset of depression and its perseverance.

Longitudinal designs could be used to help crystallise how information processing, depression and anxiety interact with one another. One such study (along the lines of Chan et al., 2007) could examine belief updating in individuals who have not been diagnosed with depression but are reported to have high levels of anxiety such as individuals with high levels of neuroticism. If it can be established that highly neurotic individuals lack a positive bias in belief updating and subsequently observed that they develop depression, this would lend support to the view that unbiased belief updating can occur pre depression onset and may be a cause (rather than just a symptom) of it. An alternate longitudinal study could examine individuals who have not been diagnosed with depression and do not have a history of anxiety but are at risk of developing depression in their lifetime (e.g., individuals with first degree relatives that suffer from depression). If it can be established that those individuals that proceed to develop depression exhibit a positive bias in belief updating pre depression onset and observed that this bias subsequently reduces, this would suggest that alterations in the bias are a symptom of depression once it has formed and set in. Note that in this latter scenario, anxiety may still play a key role in altering belief updating. One possibility for instance is that once an episode of depression has formed, individuals are less able to regulate attention away from aversive stimuli and this can be anxiety inducing (which, in line with the results of **Chapter 4**, could cause individuals to be more responsive to bad news in the UBT). Section 7.2.9 discusses in more detail the possible relationship between attention, biases in belief updating and depression and how this might be tested.

Using the UBT longitudinally in this way can therefore help identify the stage or stages at which biases in belief updating fit into the causes of depression. This can have important implications for treatment in terms of determining whether pharmacological or behavioural approaches (such as cognitive behavioural therapy) are most likely to be effective (Rosier et al., 2011).

7.2.2 Neurological mechanism by which stress alters information integration

The physiological measures taken in **Chapter 4** do not enable identification as to the neurological mechanism by which threat increases integration of bad news. One possibility is that the stress response to threat interferes with top down control mechanisms that normally inhibit integration of bad news. It has been shown that stress impairs frontal lobe function reducing cognitive control and modulation of subcortical regions (see review in Yu, 2016). Interfering with left IFG activity, one of the frontal regions identified as tracking information in **Chapter 3**, using TMS has been shown to abolish asymmetric information integration in the UBT task by increasing information integration from unwanted information (Sharot et al., 2012a) which suggests that ordinarily the IFG may inhibit information integration from negative information and that interfering with its activity using TMS releases this inhibition. Furthermore, individuals with strong white matter connectivity between the left IFG and subcortical regions involved in valuation, emotion and learning show greater valence dependent asymmetry in information integration and reduced integration from bad news, possibly because of increased ability for frontal-subcortical modulation (Moustiana et al., 2015). Taken

together, stress may mimic the effects found in individuals following TMS, which diminishes activity in cognitive control related regions, releasing the tendency to dismiss negative information. A second possibility is that the stress reaction to threat directly boosts integration of negative information. For example, it has been shown that negative, but not positive, prediction errors in the striatum are amplified under stress (Robinson et al., 2013). It may be that the representation of negative, but not positive, estimation errors are also amplified under stress. Future studies are needed to reveal the neural mechanism by which stress abolishes asymmetric information integration.

7.2.3 Understanding the process of likelihood ratio formation

Chapter 5 examined the degree to which updating correspond to that of a Bayesian agent. However the analysis reported relies on a number of assumptions. A strong one is that participants first estimates of their own likelihood has been calculated accurately through a process of Bayesian integration; this needs to be assumed in order to calculate the likelihood ratio participants have, implicitly, used to arrive at their estimates. However, it is not known empirically whether participants do actually arrive at such estimates through Bayesian integration and - if so – how well they implement this calculation. In reality, participants may have noisy priors and/or likelihood ratios leading to noisy estimates when these are integrated to form posteriors. In addition, participants may have difficulty with the process of integration itself which will also lead to noisy estimates. Note that past studies in belief updating (and the results in this study) suggest that participants are often not perfectly matched by Bayesian agents. Mobius et al. (2012) for instance show that participants update conservatively relative to Bayesian agents when expressing beliefs

about their intelligence and this is the case both when they receive good and bad news (although they are more conservative in the case of bad news).

Another assumption is that likelihood ratios remain constant between participants' first and second estimates. This permits that the likelihood ratio implied from participants' first estimates be used to calculate their "Bayesian" 2nd estimate. Likelihood ratios may not remain fixed in this way, however. Intuitively, the process of ascertaining evidence pertaining to whether an event will happen is a complex one. Taking burglary as an example again, an individual is likely to have a multitude of potential pieces of evidence they could bring to mind, some of which are more likely to be present if the event happens (e.g. live on ground floor, have no security system, etc.) and some of which are more likely to be present if the event does not happen (e.g. have vicious guard dog, member of neighbourhood watch scheme, etc.). What are the factors that determine which pieces of evidence are called to mind and how these are weighed up against each other to determine a likelihood ratio? The complexity inherent here makes it possible that likelihood ratios vary each time one is asked to make an estimate rather than remain fixed. In addition it is also possible that estimates of likelihood ratios may themselves depend in part on valence. Consider if one gets bad news. A potential mechanism to prevent full integration of this information could be to subsequently recall more instances of, or give greater weight to, pieces of evidence that would be present if the event does not happen (and less instances of evidence that would be present if the event were to happen). Highly relevant to this idea is the theme of motivated reasoning (Kunda, 1990) which proposes that individuals need evidence to support a judgement but the quantity of cognitive processing (Kruglanski and Ajzen, 1983; Pyszczynski and Greenberg, 1987; Ditto and Lopez, 1992) and the range of

cognitive strategies employed (Kunda, 1990) to generate this evidence varies depending on the conclusion individuals would prefer to arrive at. In one study (Sanitioso et al., 1990), participants were asked to generate autobiographical memories of instances in which they had behaved in an introverted and extraverted way. Beforehand, participants were either led to believe that introversion or that extroversion was a desirable character trait. Participants led to believe that introversion was a desirable trait generated a greater number of memories for introversion and vice versa for participants led to believe that extroversion was a desirable character trait. This suggests that memory search was biased towards reaching evidence consistent with possessing a desirable character trait. Similar biases in memory search have been found in other studies (Ross et al., 1981; Markus and Kunda, 1986).

Applied to the UBT, a possibility is that part of the mechanism that gives rise to asymmetric updating is a biased search for and selection of evidence towards a belief one wants to hold. Such a bias will influence the extent to which presented base rates are integrated to form new beliefs. Identifying the factors involved in generating likelihood ratios and understanding how these factors are influenced by valence is an important area for future research and key to building comprehensive picture of belief updating in a Bayesian framework.

7.2.4 The role of motivation in asymmetric updating

A suggestion outlined in **Chapter 1** is that one of the constraints of the bias is that there needs to be motivation on the part of the agent to disregard undesirable information. In the case of own likelihoods, agents derive positive utility from maintaining a positive picture

of their future and negative utility by shifting to worse views of their future (Loewenstein, 2006; Brunnermeier and Parker, 2004).

An interesting question that comes out of this is whether biased updating varies as the relevance of the information to the participant is manipulated. For example, are participants also relatively unbiased in updating when the focus of the update is not themselves specifically but someone close to them (e.g. a spouse, close friend, work colleague) who they might be motivated to maintain positive views of? One might expect to also see a bias in these instances and that as the importance of the individual the information concerns to the participant decreases, biased updating also decreases (as individuals are only motivated to maintain positive views of themselves and those close to them). And how might biased integration of information vary if people actively dislike the person under consideration (e.g. their worst enemy)? One possibility is that biased updating flips in the other direction such that people update more when they get “bad news” about that persons likelihood of encountering a disease and offers reluctance to update when receiving good news⁴.

The question of motive is also raised by the study presented in **Chapter 6**. In this study, participants are asked to rate how positive or negative they find each life event. This is a conceptually different question, however, to ascertaining how much participants want each event to happen to them that month. For instance, most people would rate “Finding £20” as a positive life event. But the importance of this event happening in the month ahead may well vary from person to person (depending on individual income levels

⁴ Here the terms “good news” and “bad news” become somewhat ambiguous as what is good news for the target of the information (e.g. the participants nemesis) may constitute bad news for the participant if they prefer that they have a higher likelihood of negative life events happening to them and a lower likelihood of positive life events happening to them. Here I refer to good and bad news from the perspective of the target of the information (not the participant).

for instance). If motivation is important, one would expect the bias to covary with the importance of the event to the individual as well as with how positive/negative the event is.

7.2.5 Environmental volatility

Chapter 6 revealed that asymmetric updating is influenced by the environment individuals are operating in. An important feature of this study though was that the source of risk in the environment was unrelated to the UBT, yet participants information integration in response to bad news in the UBT was enhanced. An interesting question that comes out of this is whether biased information integration might also vary according to volatility in instances where the source of the threat is related to the structure of the task participants are engaged in.

It has been shown in reinforcement learning contexts that the magnitude of learning rates depends in part on the uncertainty of the action value being learnt. Specifically, learning rates have been shown to adapt to volatility inherent in the environment in which the learning takes place. For instance, in a two armed bandit task, when stimulus-outcome associations are learned in stable environments whereby action-outcome contingencies remain fixed, learning rates are lower compared to when learning occurs in a volatile environment in which stimulus-outcome contingencies sporadically change (Behrens et al., 2007). High learning rates necessitate that an individual places a greater weight on the most recent outcomes which is the best strategy when outcomes are unpredictable as unexpected outcomes might signal a change in learned contingencies. But in stable environments, a better strategy is to place less weight on unexpected outcomes as

these are more likely to be the result of noise in the relationship between actions and outcomes rather than a change in the underlying structure of the task. More recently it has been shown that individuals high in trait anxiety are worse at adapting their learning rates according to the volatility inherent in the environment (Browning et al., 2014).

As discussed in **Chapter 1**, the nature of learning in the UBT is conceptually different to that which takes place in reinforcement learning. A key difference is that individuals in the UBT see a new event (stimulus) on each trial rather than face the same stimuli repeatedly with the opportunity to learn about the properties of this over a succession of trials. Nonetheless, the UBT could be adapted to explore whether learning from information adapts to the volatility of this information and explore whether valence interacts with this adaptation.

7.2.6 Bias modification

The findings presented in **Chapter 4** suggest that asymmetric information integration is not something that is fixed but can potentially be altered under certain conditions. There may be instances in which it would be beneficial to try to reduce asymmetric information integration such as when it prevents individuals from taking prudent precautionary actions (such as eating healthily, depleting retirement savings prudently). Below I briefly outline two methods that could be tried as a means of attempting to alter expectations.

(1) Use of imagination: It has been shown that individuals determine the likelihood of an event based in part on how easy it is to picture mentally (Kahneman and Tversky, 1982; LeDoux, 1992). Asking participants to imagine life events and providing them with

cues to enhance how vividly the events are imagined could therefore be used as an effective tool for increasing estimates of likelihoods of events occurring in the future.

(2) Use of reasoning: Studies have shown that contemplating and generating reasons for why an outcome might occur increases the perceived likelihood that the event will occur (Gregory et al., 1982; Hoch, 1984). Asking participants to select specific reasons for why different life events in the future may occur might increase estimates of their likelihood of encountering these events.

7.2.7 The role of genetics in belief updating

An interesting question for future research is whether genetic factors account for individual differences in asymmetric learning? There is good reason to hypothesize that genes related to dopaminergic function will predict individual differences in asymmetric belief formation. First, enhancing dopaminergic function via the administration of L-DOPA (L-3,4-dihydroxyphenylalanine) reduces learning from undesirable information (Sharot et al., 2012b). Second, within reinforcement-learning tasks, individual differences in learning from positive and negative prediction errors has been respectively related to genetic polymorphisms of the DARPP32 (dopamine and cAMP-regulated phosphoprotein 32, also known as PPP1RB) and DRD2 (dopamine receptor D2) genes (Frank, 2004; Frank and Hutchison, 2009). DARPP32 encodes an intracellular protein that is concentrated in the striatum. This protein, phosphatase 1, is inhibited when phosphorylated by dopamine receptor D1 stimulation, thus enabling cortico-striatal synaptic plasticity. DRD2 alleles determine dopamine receptor D2 affinity; this is primarily expressed by striatopallidal medium spiny neurons that are sensitive to dips of

dopamine below baseline. In addition, a polymorphism of the COMT (catechol-O-methyltransferase) gene, which is associated with individual differences in prefrontal dopamine function, has been shown to be related to the in/ability to learn from new information that does not fit with a subject's prior beliefs (Doll et al., 2011). COMT encodes an enzyme that breaks down extracellular dopamine (Meyer-Lindenberg et al., 2005, 2007), modulating dopamine levels and D1 receptor availability in prefrontal cortex (Gogos et al., 1998; Matsumoto et al., 2003; Slifstein et al., 2008), thus possibly influencing striatal activity indirectly by affecting prefrontal neurons that project to striatum (Krugel et al., 2009). An open question is whether the same genetic polymorphisms also predict learning from positive and negative estimation errors for motivated beliefs.

7.2.9 The role of attention in belief updating, depression and rumination

Dysfunction of executive control is thought to play a key role in depression (Roiser et al., 2011; Ottowitz et al., 2002) and rumination, a trait closely associated with depression. It is proposed that top down executive control plays an important regulatory function, enabling individuals to shift their attention away from negative emotional stimuli (Ottowitz et al., 2002; Joorman, 2010). When this is impaired therefore, it results in increased attention towards negative stimuli and leads to the persistence of negative thoughts, both of which are observed in rumination and depression (Nolen-Hoeksema et al., 2008; Armstrong and Olatunji 2012; Ellenbogen et al., 2002). Consistent with this, responses in Dorsolateral Prefrontal Cortex (DLPFC), a key brain region for cognitive control (MacDonald, et al., 2000; Miller & Cohen, 2001) have been shown to be impaired in patients with depression

and rumination (Mayberg et al., 1999; Wagner et al., 2006; Kross et al., 2009; Koster et al., 2011).

Given these findings, a possible mechanism by which healthy individuals are able to be less responsive to bad news in the UBT is via attentional processes whereby cognitive control is exerted to bias attention away from aversive information (i.e. bad news) with the consequence that it is integrated less than good news. Depressed patients in contrast due to impaired cognitive control function in prefrontal regions (Mayberg et al., 1999; Wagner et al., 2006; Kross et al., 2009; Koster et al., 2011) may have less capacity to shift attention away from aversive information and integrate it to a greater degree as a result. Consistent with this hypothesis, applying TMS to frontal brain regions has been shown to increase participants ability to integrate bad news (Sharot et al., 2012a). Strength of connectivity between frontal and subcortical brain regions (including limbic structures involved in emotional processing such as the amygdala) has also been shown to correlate with the extent of update bias with greater connectivity corresponding to less updating from bad news (Moutsiana et al., 2015).⁵

There are a number of ways future studies could aim to examine whether attention modulates integration of information in the UBT. One means of testing this would be to examine if increasing executive control in depressed patients, decreases their ability to integrate bad news. Previous studies (Siegle et al., 2007; Siegle et al, 2014) have shown that getting depressed individuals to complete attention training exercises over a period of time (2 weeks) can result in reductions in rumination (measured pre and post training

⁵ Note however that inconsistent with this hypothesis is the lack of differences in emotional ratings found for bad news life events between depressed patients and healthy controls in Chapter 3. Although inconsistent, this does not rule the hypothesis out however since ratings were taken after the UBT had been completed rather than during presentation of the aversive information itself.

using a self report questionnaire) possibly as the result of enhanced prefrontal function. One approach therefore (along the lines of Siegel et al., 2014) would be to examine whether integration of bad news is reduced in depressed patients after they have undertaken a period of attention training exercises. Physiological indicators of attention such as pupil dilation (e.g. during presentation of negative information) could be used to examine if any differences in information integration pre and post attention training can be related to markers of DLPFC cognitive control. Pupil dilation has previously been shown to vary with DLPFC activity in a cognitive task (Siegel et al., 2011).

Another approach would be to examine if decreasing executive control in healthy individuals increases their ability to integrate bad news. This could be tested by introducing a cognitive load manipulation into the UBT (for instance requiring participants to remember numbers before and after information presentation). Since cognitive load is believed to interfere with cognitive control processes mediated by frontal brain regions, if biased updating is the result of top down control, cognitive load ought to disrupt the bias.

7.3 Conclusion

This thesis has shown the pervasiveness of asymmetric updating by examining its existence when we consider everyday positive and negative life events (**Chapter 6**) and use different means of classifying information as good or bad (**Chapter 5**). Humans are shown to integrate information into beliefs based on the desirability of beliefs at hand, updating beliefs to a greater degree (**Chapters 3-6**) and more akin to a Bayesian agent (**Chapter 5**) when information presents good news relative to bad news. Errors in

prediction are coded differently by the brain and are used to differing degrees to update subsequent beliefs (**Chapter 3**). In clinical depression, this neural coding is enhanced for bad news and correspondingly, depressed patients exhibit unbiased updating as a result of greater updating from undesirable information. Just as the bias is shown to dissipate in depression (**Chapter 3**), it also reduced when individuals are faced with a threatening environment (**Chapter 4**). This suggests that the bias is not a fixed means of integrating information but one that imposes itself in safe environments with positive effects on our affective state (Loewenstein, 2006; Brunnermeier and Parker, 2004; Bracha and Brown, 2012), health (Taylor et al., 2000), and motivation (Varki, 2009; Bénabou and Tirole, 2002).

List of Abbreviations

Abbreviation	Details
ANOVA	Analysis of Variance
BDI	Beck Depression Inventory
BOLD	Blood Oxygenated Hemodynamic
CBT	Cognitive Behavioural Therapy
DLPFC	Dorsolateral Prefrontal Cortex
DSM-5	Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition
DTI	Diffusion Tensor Imaging
fMRI	Functional magnetic resonance Imaging
IFG	Inferior frontal gyrus
LHR	Likelihood Ratio
LOTR	Life Orientation Test Revised
MRI	Magnetic Resonance Imaging
MDD	Major Depressive Disorder
MINI	Mini International Neuropsychiatric Inventory
rIPL	Right Inferior Parietal Lobule
RT	Reaction Time
STAI	State Trait Anxiety Inventory
TSST	Tier Social Stress Test
TMS	Transcranial Magnetic Stimulation
UBT	Update Bias Task
SEM	Standard Error of the Mean
s.d.	Standard Deviation
SPM	Statistical Parametric Mapping
SCL	Skin Conductance Level
SCR	Skin Conductance Response
SVC	Small Volume Correction
TR	Repetition Time

Bibliography

Allan, L.G., Siegel, S., and Hannah, S. (2007). The sad truth about depressive realism. *Q. J. Exp. Psychol.* 60, 482–495.

Allison, P.J. (2003). Dispositional Optimism Predicts Survival Status 1 Year After Diagnosis in Head and Neck Cancer Patients. *J. Clin. Oncol.* 21, 543–548.

Alloy, L.B., and Abramson, L.Y. (1979). Judgment of contingency in depressed and nondepressed students: Sadder but wiser? *J. Exp. Psychol. Gen.* 108, 441.

Andersson, J., Hutton, C., Ashburner, J., Turner, R., and Friston, K. (2001). Modeling geometric deformations in EPI time series. *NeuroImage*, 13, 903–919.

Armor, D.A., and Taylor, S.E. (2002). "When predictions fail: the dilemma of unrealistic optimism", in *Heuristics and Biases: The Psychology of Intuitive Judgment*, eds T. D. Gilovich, D. W. Griffin, and D. Kahneman (New York, NY: Cambridge University Press), 334–347.

Armstrong, T., and Olatunji, B.O. (2012). Eye tracking of attention in the affective disorders: A meta-analytic review and synthesis. *Clinical Psychology Review*, 32, 704–723.

Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., and Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature neuroscience*, 6(2), 115-116.

Aron, A. R., Robbins, T. W., and Poldrack, R. A. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences*, 8(4), 170-177.

- Arthurs, O.J., and Boniface, S.J. (2003). What aspect of the fMRI BOLD signal best reflects the underlying electrophysiology in human somatosensory cortex? *Clin. Neurophysiol.* 114, 1203–1209.
- Baker, L.A., and Emery, R.E. (1993). When every relationship is above average. *Law Hum. Behav.* 17, 439–450.
- Baldwin DS, Evans DL, Hirschfeld RM, Kasper S. (2002) Can we distinguish anxiety from depression? *Psychopharmacol Bull*, 36:158-165.
- Bandura, A. (1989). Human agency in social cognitive theory. *Am. Psychol.* 44, 1175.
- Barber, B.M., and Odean, T. (1999). The courage of misguided convictions. *Financ. Anal. J.* 55, 41–55.
- Baxter, L., Schwartz, J.M., Phelps, M.E., Mazziotta, J.C., Guze, B.H., Selin, C.E., and Sumida, R.M. (1989). Reduction of prefrontal glucose metabolism common to three types of depression. *Archives of General Psychiatry*, 46, 243–250.
- Baxter, M.G., and Murray, E.A. (2002). The amygdala and reward. *Nat. Rev. Neurosci.* 3, 563–573.
- Bayes, M., and Price, M. (1763). *An Essay towards Solving a Problem in the Doctrine of Chances.* By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S. *Philos. Trans.* 53, 370–418.
- Beck, A.T., Ward, C.H., Mendelson, M., Mock, J., and Erbaugh, J.K. (1961). An inventory for measuring depression. *Arch. Gen. Psychiatry* 4, 561.
- Beck A.T. (1967). *Depression: Clinical, Experimental and Theoretical Aspects.* Harper & Row: New York.
- Beck, A.T., Steer, R.A., and Carbin, M.G. (1988). Psychometric properties of the Beck Depression Inventory: Twenty-five years of evaluation. *Clin. Psychol. Rev.* 8, 77–100.
- Becker, E. (1997). *The denial of death* (Simon and Schuster).

- Behrens, T.E., Woolrich, M.W., Walton, M.E., and Rushworth, M.F. (2007). Learning the value of information in an uncertain world. *Nature neuroscience*, 10(9), 1214-1221.
- Bénabou, R., and Tirole, J. (2002) Self-confidence and personal motivation. *Q. J. Econ.* 117, 871–915
- Birkett, M.A. (2011). The Trier Social Stress Test Protocol for Inducing Psychological Stress. *J. Vis. Exp.*
- Bishop, C.M. (2006). *Pattern Recognition and Machine Learning* (Springer Verlag).
- Blanchflower, D.G., and Oswald, A.J. (2008). Is well-being U-shaped over the life cycle?. *Social science & medicine*, 66(8), 1733-1749.
- Bower, G.H. (1981). Mood and memory. *American psychologist*, 36(2), 129.
- Bracha, A., and Brown, D.J. (2012) Affective decision making: a theory of optimism bias. *Games Econ. Behav.* 75, 67–80
- Bradley, B.P., Mogg, K., Millar, N., and White, J. (1995) Selective processing of negative information: effects of clinical anxiety, concurrent depression, and awareness. *J Abnorm Psychol* 104: 532–536.
- Brown, J.D. (1986). Evaluations of Self and Others: Self-Enhancement Biases in Social Judgments. *Soc. Cogn.* 4, 353–376.
- Browning, M., Behrens, T.E., Jocham, G., O'Reilly, J.X., and Bishop, S.J. (2015). Anxious individuals have difficulty learning the causal statistics of aversive environments. *Nature neuroscience*, 18(4), 590-596.
- Brunnermeier, M.K., and Parker, J.A. (2004). *Optimal Expectations* (No. w10707). National Bureau of Economic Research.
- Burger, J.M., and Palmer, M.L. (1992). Changes in and Generalization of Unrealistic Optimism Following Experiences with Stressful Events: Reactions to the 1989 California Earthquake. *Pers. Soc. Psychol. Bull.* 18, 39–43.

- Cahill, D.P. (2015) *Wishful Thinking, Fast and Slow*, Doctoral Dissertation, Harvard University Graduate School of Arts and Sciences.
- Cameron, O. G. (2005). Anxious-depressive comorbidity: effects on HPA axis and CNS noradrenergic functions. *Essential psychopharmacology*, 7(1), 24-34.
- Carver, C.S., and Scheier, M.F. (2014). Dispositional optimism. *Trends Cogn. Sci.* 18, 293–299.
- Carver, C.S., Smith, R.G., Antoni, M.H., Petronis, V.M., Weiss, S., and Derhagopian, R.P. (2005). Optimistic personality and psychosocial well-being during treatment predict psychosocial well-being among long-term survivors of breast cancer. *Health Psychology*, 24(5), 508.
- Chan, S.W., Goodwin, G.M., and Harmer, C.J. (2007). Highly neurotic never-depressed students have negative biases in information processing. *Psychol Med* 37: 1281–1291.
- Chang, E.C.-H. (2001). *Optimism & pessimism: Implications for theory, research, and practice.* (American Psychological Association).
- Chase, H.W., Frank, M.J., Michael, A., Bullmore, E.T., Sahakian, B.J., and Robbins, T.W. (2009). Approach and avoidance learning in patients with major depression and healthy controls: relation to anhedonia. *Psychol. Med.* 40, 433.
- Chlan, L., Savik, K., and Weinert, C. (2003). Development of a Shortened State Anxiety Scale From the Spielberger State-Trait Anxiety Inventory (STAI) for Patients Receiving Mechanical Ventilatory Support. *J. Nurs. Meas.* 11, 283–293.
- Chowdhury, R., Sharot, T., Wolfe, T., Duzel, E., and Dolan, R.J. (2014). Optimistic update bias increases in older age. *Psychol. Med.* 44, 2003–2012.
- Christakou, A., Gershman, S.J., Niv, Y., Simmons, A., Brammer, M., and Rubia, K. (2013). Neural and Psychological Maturation of Decision-making in Adolescence and Young Adulthood. *J. Cogn. Neurosci.* 25, 1807–1823.

Christopoulos, G. I., Tobler, P. N., Bossaerts, P., Dolan, R. J., and Schultz, W. (2009). Neural correlates of value, risk, and risk aversion contributing to decision making under risk. *The Journal of Neuroscience*, 29(40), 12574-12583.

Clark, L., Chamberlain, S.R. and Sahakian, B.J. (2009). Neurocognitive mechanisms in depression: implications for treatment. *Annu Rev Neurosci* 32: 57–74.

Collins, A.G.E., and Frank, M.J., (2014). Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychol. Rev.* 121, 337–366.

Colvin, C.R., and Block, J. (1994). Do positive illusions foster mental health? An examination of the Taylor and Brown formulation.

Compton, W.C. (1992). Are positive illusions necessary for self-esteem: A research note. *Personal. Individ. Differ.* 13, 1343–1344.

Coyne, J.C., and Gotlib, I.H. (1983). The role of cognition in depression: A critical appraisal. *Psychol. Bull.* 94, 472.

d'Acremont, M., Lu, Z. L., Li, X., Van der Linden, M., and Bechara, A. (2009). Neural correlates of risk prediction error during reinforcement learning in humans. *Neuroimage*, 47(4), 1929-1939.

d'Acremont, M., Schultz, W., and Bossaerts, P. (2013). The human brain encodes event frequencies while forming subjective beliefs. *The Journal of Neuroscience*, 33(26), 10887-10897.

Deeming, C. (2013) Addressing the Social Determinants of Subjective Wellbeing: The Latest Challenge for Social Policy. *Journal of Social Policy*, 42, pp 541-565

Dewberry, C., and Richardson, S. (1990). Effect of Anxiety on Optimism. *J. Soc. Psychol.* 130, 731–738.

Dickerson, S.S., and Kemeny, M.E. (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological bulletin*,

130(3), 355.

Ditto, P. H., and Lopez, D.F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, 63(4), 568.

Doll, B.B., Hutchison, K.E., and Frank, M.J. (2011). Dopaminergic Genes Predict Individual Differences in Susceptibility to Confirmation Bias. *J. Neurosci.* 31, 6188–6198.

Douglas, C., Bateson, M., Walsh, C., Bédoué, A., and Edwards, S.A. (2012). Environmental enrichment induces optimistic cognitive biases in pigs. *Appl. Anim. Behav. Sci.* 139, 65–73.

Dunning, D., Heath, C., and Suls, J.M. (2004). Flawed Self-Assessment: Implications for Health, Education, and the Workplace. *Psychol. Sci. Public Interest J. Am. Psychol. Soc.* 5, 69–106.

Eil, D., and Rao, J.M. (2011). The Good News-Bad News Effect: Asymmetric Processing of Objective Information about Yourself. *Am. Econ. J. Microecon.* 3, 114–138.

Ellenbogen, M.A., Schwartzman, A.E., Stewart, J., and Walker, C. (2002). Stress and selective attention: The interplay of mood, cortisol levels, and emotional information processing. *Psychophysiology*, 39, 723–732.

Fallon, J.H., and Moore, R.Y. (1978). Catecholamine innervation of the basal forebrain IV. Topography of the dopamine projection to the basal forebrain and neostriatum. *J. Comp. Neurol.* 180, 545–579.

Fjell, A.M., Sneve, M.H., Grydeland, H., Storsve, A.B., and Walhovd, K.B. (2016). The Disconnected Brain and Executive Function Decline in Aging. *Cerebral Cortex*, bhw082.

Figner, B., and Murphy, R.O. (2011). Using skin conductance in judgment and decision making research. *A handbook of process tracing methods for decision research*, 163-184.

Forgas, J.P., Bower, G.H., and Krantz, S.E. (1984). The influence of mood on perceptions of social interactions. *Journal of Experimental Social Psychology*, 20(6), 497-513.

- Frank, M.J. (2004). By Carrot or by Stick: Cognitive Reinforcement Learning in Parkinsonism. *Science*, 306, 1940–1943.
- Frank, M.J., and Hutchison, K. (2009). Genetic contributions to avoidance-based decisions: striatal D2 receptor polymorphisms. *NSC* 164, 131–140.
- Frank, M.J., Moustafa, A.A., Haughey, H.M., Curran, T., and Hutchison, K.E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc. Natl. Acad. Sci.* 104, 16311–16316.
- Galasso, A., and Simcoe, T.S. (2011). CEO Overconfidence and Innovation. *Manag. Sci.* 57, 1469–1484.
- Garrett, N., and Sharot, T. (2014). How robust is the optimistic update bias for estimating self-risk and population base rates? *PloS One* 9, e98848.
- Garrett, N., Sharot, T., Faulkner, P., Korn, C.W., Roiser, J.P., and Dolan, R.J. (2014). Losing the rose tinted glasses: neural substrates of unbiased belief updating in depression. *Front. Hum. Neurosci.* 8.
- Geers, A. L., Wellman, J.A., and Lassiter, G.D. (2009). Dispositional optimism and engagement: the moderating influence of goal prioritization. *Journal of personality and social psychology*, 96(4), 913.
- Gerfen, C.R. (1992). The neostriatal mosaic: multiple levels of compartmental organization. 43–59.
- Gershman, S.J. (2015). Do learning rates adapt to the distribution of rewards? *Psychon. Bull. Rev.* 22: 1320-1327.
- Gilboa-Schechtman, E., Erhard-Weiss, D., and Jeczemien, P. (2002). Interpersonal deficits meet cognitive biases: memory for facial expressions in depressed and anxious men and women. *Psychiatry Res* 113: 279–293.
- Giltay, E.J., Geleijnse, J.M., Zitman, F.G., Buijsse, B., and Kromhout, D. (2007). Lifestyle and dietary correlates of dispositional optimism in men: The Zutphen Elderly Study.

Journal of psychosomatic research, 63(5), 483-490.

Glaser, D. (2004). Variance components. In R. Frackowiak, ed., *Human Brain Function*, Academic Press.

Greening, S. G., Finger, E. C., and Mitchell, D. G. (2011). Parsing decision making processes in prefrontal cortex: response inhibition, overcoming learned avoidance, and reversal learning. *Neuroimage*, 54(2), 1432-1441.

Goense, J.B., and Logothetis, N.K. (2008). Neurophysiology of the BOLD fMRI signal in awake monkeys. *Current Biology*, 18(9), 631-640.

Gogos, J.A., Morgan, M., Luine, V., Santha, M., Ogawa, S., Pfaff, D., and Karayiorgou, M. (1998). Catechol-O-methyltransferase-deficient mice exhibit sexually dimorphic changes in catecholamine levels and behavior. *Proc. Natl. Acad. Sci.* 95, 9991–9996.

Goldman-Rakic, P. (2000). Localization of function all over again. *NeuroImage* 11, 451–457.

Gregory, W.L., Cialdini, R.B., and Carpenter, K.M. (1982). Self-relevant scenarios as mediators of likelihood estimates and compliance: Does imagining make it so? *J. Pers. Soc. Psychol.* 43, 89.

Harding, E.J., Paul, E.S., and Mendl, M. (2004). Cognitive bias and affective state. *Nature* 427, 312.

Hoch, S.J. (1984). Availability and interference in predictive judgment. *J. Exp. Psychol. Learn. Mem. Cogn.* 10, 649–662.

Hoorens, V. (1993). Self-enhancement and Superiority Biases in Social Comparison. *Eur. Rev. Soc. Psychol.* 4, 113–139.

Jobin, J., Wrosch, C., and Scheier, M.F. (2014). Associations between dispositional optimism and diurnal cortisol in a community sample: When stress is perceived as higher than normal. *Health psychology*, 33(4), 382.

Johnson, D. (2009). *Overconfidence and war*. Harvard University Press.

- Johnson, D.D.P., and Fowler, J.H. (2011). The evolution of overconfidence. *Nature* 477, 317–320.
- Johnson, E.J., and Tversky, A. (1983). Affect, generalization, and the perception of risk. *Journal of personality and social psychology*, 45(1), 20.
- Joormann, J. (2010). Inhibition and emotion regulation in depression. *Current Directions in Psychological Science*, 19, 161–166.
- Joormann, J., and Gotlib, I. H. (2010). Emotion regulation in depression: Relation to cognitive inhibition. *Cognition and Emotion*, 24(2), 281–298.
- Kahneman, D., and Tversky, A. (1973). On the psychology of prediction. *Psychol. Rev.* 80, 237-251.
- Kahneman, D., and Tversky, A. (1982). The simulation heuristic. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 201-208). New York: Cambridge University Press.
- Kahneman, D. and Tversky, A. (1984). Choices, Values, and Frames. *American Psychologist*. 39 (4): 341–350.
- Kahneman, D. and Tversky, A. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*. 5 (4): 297–323.
- Kendler, K.S., Kessler, R.C., Neale, M.C., Heath, A.C., and Eaves, L.J. (1993). The prediction of major depression in women: toward an integrated etiologic model. *Am J Psychiatry* 150: 1139–1148.
- Kim, E.S., Park, N., and Peterson, C. (2011). Dispositional Optimism Protects Older Adults From Stroke The Health and Retirement Study. *Stroke*, 42(10), 2855-2859.
- Kirschbaum, C., Pirke, K.M., and Hellhammer, D.H. (1993). The ‘Trier Social Stress Test’—a tool for investigating psychobiological stress responses in a laboratory setting. *Neuropsychobiology*, 28(1-2), 76-81.
- Kirschbaum, C., and Hellhammer, D.H. (1994). Salivary cortisol in psychoneuroendocrine

research: recent developments and applications. *Psychoneuroendocrinology*, 19(4), 313-333.

de Kloet, E.R., Oitzl, M.S., and Joëls, M. (2011). Stress and cognition: are corticosteroids good or bad guys? *Trends Neurosci.* 22, 422–426.

Konishi, S., Nakajima, K., Uchida, I., Kameyama, M., Nakahara, K., Sekihara, K., and Miyashita, Y. (1998). Transient activation of inferior prefrontal cortex during cognitive set shifting. *Nature neuroscience*, 1(1), 80-84.

Konishi, S., Nakajima, K., Uchida, I., Sekihara, K., and Miyashita, Y. (1998). No-go dominant brain activity in human inferior prefrontal cortex revealed by functional magnetic resonance imaging. *European Journal of Neuroscience*, 10(3), 1209-1213.

Körding, K.P., and Wolpert, D.M. (2006). Bayesian decision theory in sensorimotor control. *Trends Cogn. Sci.* 10, 319–326.

Korn, C.W., Prehn, K., Park, S.Q., Walter, H., and Heekeren, H.R. (2012). Positively Biased Processing of Self-Relevant Social Feedback. *J. Neurosci.* 32, 16832–16844.

Korn, C.W., Sharot, T., Walter, H., Heekeren, H.R., and Dolan, R.J. (2013). Depression is related to an absence of optimistically biased belief updating about future life events. *Psychol. Med.* 44, 579–592.

Koster, E.H.W., De Lissnyder, E., Derakshan, N., and De Raedt, R. (2011). Understanding depressive rumination from a cognitive science perspective: The impaired disengagement hypothesis. *Clinical Psychology Review*, 31, 138–145.

Kross, E., Davidson, M., Weber, J., and Ochsner, K. (2009). Coping with emotions past: The neural bases of regulating affect associated with negative autobiographical memories. *Biological Psychiatry*, 65, 361–366.

Krugel, L.K., Biele, G., Mohr, P.N.C., Li, S.-C., and Heekeren, H.R. (2009). Genetic variation in dopaminergic neuromodulation influences the ability to rapidly and flexibly adapt decisions. *Proc. Natl. Acad. Sci. U. S. A.* 106, 17951–17956.

- Kruger, J., and Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J. Pers. Soc. Psychol.* 77, 1121–1134.
- Kruglanski, A.W., and Ajzen, I. (1983). Bias and error in human judgment. *European Journal of Social Psychology*, 13, 1-44.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498.
- Kuzmanovic, B., Jefferson, A., and Vogeley, K. (2015). Self-specific Optimism Bias in Belief Updating Is Associated with High Trait Optimism. *J. Behav. Decis. Mak.* 28, 281–293.
- Kuzmanovic, B., Jefferson, A., and Vogeley, K. (2016). The role of the neural reward circuitry in self-referential optimistic belief updates. *NeuroImage* 133, 151–162.
- LeDoux, J.E. (1992). Brain mechanisms of emotion and emotional learning. *Curr Opin Neurobiol.* 2(2): 191-7.
- LeDoux, J.E. (2000). Emotion circuits in the brain. *Annu. Rev. Neurosci.* 23, 155–184.
- Lefebvre, G., Lebreton, M., Meyniel, F., Bourgeois-Gironde, S., and Palminteri, S. (2016). Asymmetric reinforcement learning: computational and neural bases of positive life orientation. *bioRxiv* 038778.
- Loewenstein, G. (2006). Social science. The pleasures and pains of information. *Science* 312, 704–706.
- Logothetis, N.K., Pauls, J., Augath, M., Trinath, T., and Oeltermann, A. (2001). Neurophysiological investigation of the basis of the fMRI signal. *Nature* 412, 150–157.
- Lovallo, D., and Kahneman, D. (2003). Delusions of success. *Harv. Bus. Rev.* 81, 56–63.
- Lupien, S.J., and McEwen, B.S. (1997). The acute effects of corticosteroids on cognition: integration of animal and human model studies. *Brain Res. Rev.* 24, 1–27.

MacDonald, A.W., Cohen, J.D., Stenger, V.A., and Carter, C.S. (2000) Dissociating the role of the dorsolateral prefrontal and anterior cingulate cortex in cognitive control. *Science* 288, 1835–1838.

Markus, H., and Kunda, Z. (1986). Stability and malleability of the self-concept. *Journal of Personality and Social Psychology*, 51, 858-866.

Marteau, T.M., and Bekker, H. (1992). The development of a six-item short-form of the state scale of the Spielberger State-Trait Anxiety Inventory (STAI). *Br. J. Clin. Psychol. Br. Psychol. Soc.* 31 (Pt 3), 301–306.

Maslow, A.H. (1950). Self-actualizing people: a study of psychological health. *Personality Symposium* 1, 11–34.

Matheson, S.M., Asher, L., and Bateson, M. (2008). Larger, enriched cages are associated with “optimistic” response biases in captive European starlings (*Sturnus vulgaris*). *Appl. Anim. Behav. Sci.* 109, 374–383.

Matthews, G., and Wells, A. (2004). Rumination, depression and metacognition: the S-REF model. In C. Papageorgiou & A. Wells (Eds.), *Depressive rumination: Nature, theory and treatment* (pp. 125-151). Chichester, UK: Wiley.

Matsumoto, M., Weickert, C.S., Akil, M., Lipska, B.K., Hyde, T.M., Herman, M.M., Kleinman, J.E., and Weinberger, D.R. (2003). Catechol O-methyltransferase mRNA expression in human and rat brain: evidence for a role in cortical neuronal function. *NSC* 116, 127–137.

Mayberg, H.S., Liotti, M., Brannan, S.K., McGinnis, B.S., Mahurin, R.K., Jerabek, P.A., and Fox, P.T. (1999). Reciprocal limbic cortical function and negative mood: Converging PET findings in depression and normal sadness. *American Journal of Psychiatry*, 156, 675–682.

McClure, S. M., Daw, N. D., and Read Montague, P. (2003). A computational substrate for incentive salience. *Trends in Neurosciences*, 26, 423–428.

- McKay, R.T., and Dennett, D.C. (2010). The evolution of misbelief. *Behav. Brain Sci.* 32, 493.
- Meyer, H.H. (1975). The pay-for-performance dilemma. *Organ. Dyn.* 3, 39–50.
- Meyer-Lindenberg, A., Kohn, P.D., Kolachana, B., Kippenhan, S., McInerney-Leo, A., Nussbaum, R., Weinberger, D.R., and Berman, K.F. (2005). Midbrain dopamine and prefrontal function in humans: interaction and modulation by COMT genotype. *Nat. Neurosci.* 8, 594–596.
- Meyer-Lindenberg, A., Straub, R.E., Lipska, B.K., Verchinski, B.A., Goldberg, T., Callicott, J.H., Egan, M.F., Huffaker, S.S., Mattay, V.S., and Kolachana, B. (2007). Genetic evidence implicating DARPP-32 in human frontostriatal structure, function, and cognition. *J. Clin. Invest.* 117, 672–682.
- Miller, E.K., and Cohen, J.D. (2001). An Integrative Theory of Prefrontal Cortex Function. *Annu. Rev. Neurosci.* 24, 167–202.
- Mobius, M.M., Niehaus, P., Niederle, M., and Rosenblat, T.S. (2012). *Managing Self-Confidence: Theory and Experimental Evidence.*
- Moore, M.T., and Fresco, D.M. (2012). Depressive realism: A meta-analytic review. *Clin. Psychol. Rev.* 32, 496–509.
- Moutsiana, C., Garrett, N., Clarke, R.C., Lotto, R.B., Blakemore, S.-J., and Sharot, T. (2013). Human development of the ability to learn from bad news. *Proc. Natl. Acad. Sci.* 110, 16396–16401.
- Moutsiana, C., Charpentier, C.J., Garrett, N., Cohen, M.X., and Sharot, T. (2015). Human Frontal–Subcortical Circuit and Asymmetric Belief Updating. *J. Neurosci.* 35, 14077–14085.
- Niv, Y., Edlund, J.A., Dayan, P., and O’Doherty, J.P. (2012). Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. *J. Neurosci.* 32, 551–562.

- Nolen-Hoeksema, S., Wisco, B.E., and Luybomirsky, S. (2008). Rethinking rumination. *Perspectives on Psychological Science*, 3, 400–424.
- Novotny, P., Colligan, R.C., Szydlo, D.W., Clark, M.M., Rausch, S., Wampfler, J., Sloan, J.A., and Yang, P. (2010). A pessimistic explanatory style is prognostic for poor lung cancer survival. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer* 5, 326–332.
- O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., and Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature neuroscience*, 4(1), 95-102.
- Ogawa, S., Tank, D.W., Menon, R., Ellermann, J.M., Kim, S.G., Merkle, H., and Ugurbil, K. (1992). Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. *Proceedings of the National Academy of Sciences of the United States of America*, 89, 5951.
- Oquendo, M.A., Galfalvy, H., Russo, S., Ellis, S.P., Grunebaum, M.F., Burke, A., and Mann, J.J. (2004). Prospective study of clinical predictors of suicidal acts after a major depressive episode in patients with major depressive disorder or bipolar disorder. *Am. J. Psychiatry* 161, 1433–1441.
- Ottowitz, W.E., Dougherty, D.D., and Savage, C.R. (2002). The neural network basis for abnormalities of attention and executive function in major depressive disorder: Implications for application of the medical disease model to psychiatric disorders. *Harvard Review of Psychiatry*, 10, 86–99.
- Paling, J. (2003). Strategies to help patients understand risks. *BMJ* 327, 745–748.
- Papakostas, G.I. (2006). Dopaminergic-based pharmacotherapies for depression. *Eur. Neuropsychopharmacol.* 16, 391–402.
- Paton, D. (2003). Disaster preparedness: a social-cognitive perspective. *Disaster Prev. Manag.* 12, 210–216.
- 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,

1042–1045.

Phelps, E.A. (2006). Emotion and Cognition: Insights from Studies of the Human Amygdala. *Annu. Rev. Psychol.* 57, 27–53.

Post, R. M. (1992). Transduction of psychosocial stress into the neurobiology. *Am J Psychiatry*, 149, 999-1010.

Puri, M., and Robinson, D. (2007). Optimism and economic choice, Vol. 86.

Robinson, O.J., Overstreet, C., Charney, D.R., Vytal, K., and Grillon, C. (2013) Stress increases aversive prediction error signal in the ventral striatum. *Proc. Natl. Acad. Sci.* 110, 4129–4133.

Pyszczynski, T., and Greenberg, J. (1987). Toward an integration of cognitive and motivational perspectives on social inference: A biased hypothesis-testing model. (In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 20, pp. 297–340). New York: Academic Press.)

Ross, M., McFarland, C., and Fletcher, G.J.O. (1981). The effect of attitude on recall of past histories. *Journal of Personality and Social Psychology*, 10, 627-634.

Rasmussen, H.N., Scheier, M.F., and Greenhouse, J.B. (2009). Optimism and physical health: A meta-analytic review. *Annals of behavioral medicine*, 37(3), 239-256.

Robinson-Whelen, S., Kim, C., MacCallum, R.C., and Kiecolt-Glaser, J.K. (1997). Distinguishing optimism from pessimism in older adults: Is it more important to be optimistic or not to be pessimistic? *Journal of personality and social psychology*, 73(6), 1345.

Roiser, J.P., Elliott, R., and Sahakian, B.J. (2011). Cognitive Mechanisms of Treatment in Depression. *Neuropsychopharmacology* 37, 117–136.

Russell, S., and Norvig, P. (1995). *Artificial Intelligence: A Modern Approach*. Prentice Hall Englewood Cliffs NJ.

- Rygula, R., Papciak, J., and Popik, P. (2013). Trait Pessimism Predicts Vulnerability to Stress-Induced Anhedonia in Rats. *Neuropsychopharmacology* 38, 2188–2196.
- Salmeto, A.L., Hymel, K.A., Carpenter, E.C., Brilot, B.O., Bateson, M., and Sufka, K.J. (2011). Cognitive bias in the chick anxiety–depression model. *Brain Res.* 1373, 124–130.
- Scheier, M.F., and Carver, C.S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology*, 4, 219-247.
- Scheier, M.F., Carver, C.S., and Bridges, M.W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. *J. Pers. Soc. Psychol.* 67, 1063–1078.
- Scheier, M.F., Matthews, K.A., Owens, J.F., Schulz, R., Bridges, M.W., Magovern, G.J., Sr, and Carver, C.S. (1999). Optimism and rehospitalization after coronary artery bypass graft surgery. *Arch. Intern. Med.* 159, 829.
- Schultz, W., Dayan, P., and Montague, P.R. (1997). A Neural Substrate of Prediction and Reward. *Science* 275, 1593–1599.
- Schultz, W. (1998). Predictive reward signal of dopamine neurons. *J. Neurophysiol.* 80, 1–27.
- Shah, P., Harris, A. J. L., Hahn, U., Catmur, C., and Bird, G. (2016). A pessimistic view of optimistic belief updating. *Cognitive Psychology*. 90, 71-127.
- Sharot, T. (2011). The optimism bias. *Curr. Biol.* 21, R941–R945.
- Sharot, T., and Garrett, N. (2016). Forming Beliefs: Why Valence Matters. *Trends Cogn. Sci.* 20, 25–33.
- Sharot, T., Riccardi, A.M., Raio, C.M., and Phelps, E.A. (2007). Neural mechanisms mediating optimism bias. *Nature*, 450, 102–105.
- Sharot, T., Korn, C.W., and Dolan, R.J. (2011). How unrealistic optimism is maintained in the face of reality. *Nature Neuroscience*, 14(11), 1475–1479.

- Sharot, T., Kanai, R., Marston, D., Korn, C.W., Rees, G., and Dolan, R.J. (2012a). Selectively altering belief formation in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17058–17062.
- Sharot, T., Guitart-Masip, M., Korn, C.W., Chowdhury, R., and Dolan, R.J. (2012b). How dopamine enhances an optimism bias in humans. *Curr. Biol.* CB 22, 1477–1481.
- Sheehan, D.V., Lecrubier, Y., Sheehan, K.H., Amorim, P., Janavs, J., Weiller, E., Hergueta, T., Baker, R., and Dunbar, G.C. (1998). The Mini-International Neuropsychiatric Interview (MINI): the development and validation of a structured diagnostic psychiatric interview for DSM-IV and ICD-10. *J. Clin. Psychiatry* 59, 22–33.
- Shefrin, H. (2009). *How Psychological Pitfalls Generated the Global Financial Crisis* (Rochester, NY: Social Science Research Network).
- Sherman, S.J. (1980). On the self-erasing nature of errors of prediction. *J. Pers. Soc. Psychol.* 39, 211–221.
- Siegle, G.J., Price, R.B., Jones, N.P., Ghinassi, F., Painter, T., and Thase, M.E. (2014). You gotta work at it pupillary indices of task focus are prognostic for response to a neurocognitive intervention for rumination in depression. *Clinical Psychological Science*, 2(4), 455-471.
- Siegle, G.J., Steinhauer, S.R., Friedman, E.S., Thompson, W.S., and Thase, M.E. (2011). Remission prognosis for cognitive therapy for recurrent depression using the pupil: Utility and neural correlates. *Biological Psychiatry* 69, 726–733.
- Siegle, G.J., Ghinassi, F., and Thase, M.E. (2007). Neurobehavioral therapies in the 21st century: Summary of an emerging field and an extended example of cognitive control training for depression. *Cognitive Therapy & Research*, 31, 235–262.
- Slagter, H.A., Georgopoulou, K., and Frank, M.J. (2015). Spontaneous eye blink rate predicts learning from negative, but not positive, outcomes. *Neuropsychologia*, 71, 126–132.

- Slifstein, M., Kolachana, B., Simpson, E.H., Tabares, P., Cheng, B., Duvall, M., Gordon Frankle, W., Weinberger, D.R., Laruelle, M., and Abi-Dargham, A. (2008). COMT genotype predicts cortical-limbic D1 receptor availability measured with [11C]NNC112 and PET. *Mol. Psychiatry* 13, 821–827.
- Solberg Nes, L., Evans, D.R., and Segerstrom, S.C. (2009). Optimism and college retention: mediation by motivation, performance, and adjustment¹. *Journal of Applied Social Psychology*, 39(8), 1887-1912.
- Spielberger, C.D., Gorsuch, R.L., Lushene, R., Vagg, P.R., and Jacobs, G.A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Stones, A., Groome, D., Perry, D., Hucklebridge, F., and Evans, P. (1999). The effect of stress on salivary cortisol in panic disorder patients. *J. Affect. Disord.* 52, 197–201.
- Strunk, D.R., and Adler, A.D. (2009). Cognitive biases in three prediction tasks: A test of the cognitive model of depression. *Behav. Res. Ther.* 47, 34–40.
- Strunk, D.R., Lopez, H., and DeRubeis, R.J. (2006). Depressive symptoms are associated with unrealistic negative predictions of future life events. *Behav. Res. Ther.* 44, 861–882.
- Stuart (1994). *Kendalls Advanced Theory of Statistics: v. 1* (Chichester: Wiley-Blackwell).
- Sutton, R.S., and Barto, A.G. (1998). *Reinforcement learning: An introduction*. MIT Press.
- Svenson, O. (1981). Are we all less risky and more skillful than our fellow drivers? *Acta Psychol. (Amst.)* 47, 143–148.
- Taylor, S.E., and Armor, D.A. (1996). Positive illusions and coping with adversity. *J. Pers.* 64, 873–898.
- Taylor, S.E., and Brown, J.D. (1988). Illusion and well-being: a social psychological perspective on mental health. *Psychol. Bull.* 103, 193.

- Taylor, S.E., Lichtman, R.R., and Wood, J.V. (1984). Attributions, beliefs about control, and adjustment to breast cancer. *J. Pers. Soc. Psychol.* 46, 489.
- Taylor, S.E., Kemeny, M.E., Reed, G.M., Bower, J.E., and Gruenewald, T.L. (2000). Psychological resources, positive illusions, and health. *Am. Psychol.* 55, 99–109.
- Taylor, S.E., Lerner, J.S., Sherman, D.K., Sage, R.M., and McDowell, N.K. (2003). Portrait of the self-enhancer: well adjusted and well liked or maladjusted and friendless? *J. Pers. Soc. Psychol.* 84, 165.
- Tiger, L. (1979). *Optimism: The Biology of Hope* (Kodansha International).
- Tindle, H.A., Chang, Y.F., Kuller, L.H., Manson, J.E., Robinson, J.G., Rosal, M. C., and Matthews, K.A. (2009). Optimism, cynical hostility, and incident coronary heart disease and mortality in the Women's Health Initiative. *20 Circulation*, 1 (8), 656-662.
- Tindle, H., Belnap, B.H., Houck, P.R., Mazumdar, S., Scheier, M.F., Matthews, K. A., and Rollman, B.L. (2012). Optimism, response to treatment of depression, and rehospitalization after coronary artery bypass graft surgery. *Psychosomatic medicine*, 74(2), 200.
- Varki, A. (2009). Human uniqueness and the denial of death. *Nature* 460, 684–684.
- Volkow, N.D., Wang, G.J., Fowler, J.S., Logan, J., Gatley, S.J., MacGregor, R.R., et al. (1996). Measuring age-related changes in DA D2 receptors with [C]raclopride and with [F]N-methylspiroperidol. *Psychiatry Research*, 67, 11–16.
- Von Neumann, J., and Morgenstern, O. (1953). *Theory of games and economic behavior*. Princeton University Press.
- Wagner, G., Sinsel, E., Sobanski, T., Kohler, S., Marinou, V., Mentzel, H. J., and Schlosser, R.G. (2006). Cortical inefficiency in patients with unipolar depression: An event-related FMRI study with the Stroop task. *Biological Psychiatry*, 59, 958–965.
- Weinstein, N. (1980). Unrealistic Optimism About Future Life Events. *J. Pers. Soc. Psychol.* 39, 806.

Weinstein, N.D., and Klein, W.M. (1995). Resistance of personal risk perceptions to debiasing interventions. *Health Psychol.* 14, 132.

Wiecki, T.V., and Frank, M.J., (2010). Neurocomputational models of motor and cognitive deficits in Parkinson's disease. *Prog. Brain Res.* 183, 275–297 .

Wiswall, M., and Zafar, B. (2015). How Do College Students Respond to Public Information about Earnings? *J. Hum. Cap.* 9, 117–169.

Wood, J.V., Taylor, S.E., and Lichtman, R.R. (1985). Social comparison in adjustment to breast cancer. *J. Pers. Soc. Psychol.* 49, 1169.

Yu, R. (2016). Stress potentiates decision biases: A stress induced deliberation-to-intuition (SIDI) model. *Neurobiol. Stress* 3, 83–95.