Electrophysiological Evidence for a Universal Reward Prediction Error Encoder in Reinforcement Learning

Harry Stewardson

A thesis submitted to the School of Psychology at the University of East Anglia in fulfilment for the degree of Doctor of Philosophy.

July 2022

Abstract

In human decision-making, choices are made with the ultimate aim of maximising favourable outcomes while minimising unfavourable ones. This is not determined merely by an assessment of outcomes being good or bad, but by them being better or worse than anticipated before their occurrence. The difference between expected and attained outcomes can be quantified in the form of reward prediction errors, whereby positive values represent rewards and negative values instead represent punishments. Research in Neuroeconomics demonstrates a convincing argument that the human brain generates signals representative of such reward prediction errors for the purposes of reinforcement learning. Isolating such signals would stand to tell a great deal about not only typical human decision-making, but also prove beneficial for furthering the understanding of non-normative behaviour such as psychopathological gambling and addiction. The present thesis develops this research by focussing on the feedback-related negativity, an electrophysiological component largely attributed to the reflection of a neural encoding of reward prediction errors in reinforcement learning tasks. While the feedback-related negativity is the most widely researched component regarding this topic, methodological differences in its study and the intricacies of component overlap has led to substantial debate regarding the precise type of reward prediction error encoding it is said to represent. In this thesis I have broken down these types of encoding and accurately identified that most represented in the time interval of the feedback-related negativity; counter to many claims, I have demonstrated this encoding to not only be present using experimental designs that have previously led to its misidentification, but also present within the pre-existing literature following thorough meta-analytical approaches utilizing Bayesian methods. In line with these findings, I propose the feedback-related negativity as a signal representative of universal, or general-purpose, reward prediction error encoding in all manner of human reinforcement learning scenarios.

Contents

1. General Introduction ---- 10

1.1 Reward Prediction Errors ---- 10

1.1.1 RPE valence ---- 10

1.1.2 RPE size ---- 11

1.1.3 A neural RPE ---- 12

1.2 The FRN ---- 12

1.2.1 Operationalising the FRN ---- 13

1.2.2 Encoders ---- 14

1.2.2.1 The RPE sign encoder ---- 14

1.2.2.2 The RPE-utility encoder ---- 14

1.2.2.3 RPE size encoders ---- 15

1.3 Causes of Competing Claims ---- 15

1.3.1 Methodological differences between studies ---- 16

1.3.2 Component overlap ---- 17

1.4 General-purpose RPE Encoding ---- 17

1.4.1 General-purpose RPE assumptions ---- 18

1.4.2 Thesis summary ---- 18

1.5 Glossary ---- 20

2. Reward, Salience and Agency in ERPs for Appetitive and Aversive Contexts ---- 22

2.1 Chapter Abstract ---- 22

2.2 Introduction ---- 22

2.2.1 The FRN: RPE-utility claims ---- 23

2.2.2 The FRN: Motivational salience claims ---- 24

2.2.3 The role of agency ---- 24

2.2.4 Predictions and rationale ---- 25

2.3 Method ---- 26

2.3.1 Participants ---- 26

2.3.2 Experimental Design ---- 27

2.3.3 Procedure ---- 27

2.3.4 EEG Recording ---- 28

2.3.5 EEG Analysis ---- 29

2.3.6 Statistical Analyses ---- 29

2.3.7 Computational Modelling of RPE utility ---- 30

2.4 Results 31
2.4.1 Behavioural Results 31
2.4.2 FRN 31
2.4.3 RPE-utility Encoding 33
2.5 Discussion 34
2.5.1 Results summary 34
2.5.2 RPE-utility encoding 34
2.5.3 Passive RPE encoding 35
2.5.4 UPE encoding 36
2.5.5 Chapter conclusions 36
3. Reward Prediction Error in the ERP Following Unconditioned Aversive Stimuli 37
3.1 Chapter Abstract 37
3.2 Introduction 37
3.2.1 The FRN and unconditioned stimuli 37
3.2.2 Methodological challenges 39
3.2.3 Predictions and rationale 40
3.3 Method 41
3.3.1 Participants 41
3.3.2 Shock Apparatus 41
3.3.3 Shock Calibration 41
3.3.4 Experimental Design 42
3.3.5 Procedure 43
3.3.6 EEG Recording 45
3.3.7 EEG Analysis 45
3.3.8 Statistical Analysis 45
3.4 Results 48
3.4.1 Behavioural Results 48
3.4.2 RPE Encoding 48
3.4.3 Startle Response 50
3.5 Discussion 51
3.5.1 Results summary 51
3.5.2 Timing differences 52
3.5.3 Single-domain limitations 53
3.5.4 Affect in RPEs 54
3.5.5 Chapter conclusions 54

4. Bayesian Contrasts Analysis of the Reward Prediction Error Literature 55
4.1 Chapter Abstract 55
4.2 The Confounding of RPE-sign, RPE-utility and Motivational Salience 55
4.2.1 The three encoders 55
4.2.2 Component overlap 56
4.3 The Importance of Context 56
4.3.1 Context-free and context-dependent encoding 56
4.3.2 The context dependency of RPE-sign, RPE-utility and motivational salience 57
4.4 Disambiguating the Waveform 57
4.4.1 RPE valence 58
4.4.2 RPE size 58
4.4.3 Context dependency 59
4.4.3.1 Experiment Form 59
4.4.3.2 The meta-analysis approach 59
4.4.4 Predicting the three encoders 60
4.5 Summary 61
4.6 Method 61
4.6.1 Inclusion and exclusion criteria 61
4.6.2 Search strategies 62
4.6.3 Coding procedures 64
4.6.4 Primary analysis 64
4.6.4.1 The limitation of single effects as a means of identifying encoders 64
4.6.4.2 Exhaustive Bayesian contrasts 65
4.6.4.3 Resolving component overlap: PCA 66
4.6.4.4 Resolving component overlap: Frontal – parietal Bayesian contrasts 66
4.6.5 Supporting analyses 67
4.6.5.1 Bayesian repeated-measures ANOVAs 67
4.6.5.2 Standard operationalisations 68
4.7 Results 69
4.7.1 Primary analysis 69
4.7.1.1 Exhaustive Bayesian contrasts 69
4.7.1.2 PCA 70
4.7.1.3 Frontal – parietal Bayesian contrasts 73
4.7.2 Supporting analyses 74
4.7.2.1 Bayesian repeated-measures ANOVAs 74

4.7.2.2 Standard operationalisations ---- 75
4.8 Additional Results ---- 78
4.8.1 Explaining the reduced aversive FRN ---- 79
4.8.1.1 Overlapping motivational salience ---- 79
4.8.1.2 Aversive neglect ---- 81
4.8.2 Testing cases for component overlap and aversive neglect ---- 81
4.8.2.1 Effects of blocking type ---- 81
4.8.2.2 Effects of cue type ---- 82
4.8.3 Resolving the reduced aversive FRN ---- 84
4.9 Discussion ---- 85
4.9.1 Results summary ---- 85
4.9.2 Evidence for motivational salience encoding ---- 86
4.9.2.1 Parietal encoding at 360 ms ---- 86
4.9.2.2 Frontal encoding at 280 ms ---- 86
4.9.3 Evidence for RPE-sign encoding ---- 87
4.9.4 Chapter conclusions ---- 87
5. Evidence for Parietal Reward Prediction Errors using Great Grand Average Meta-analysis ---- 89
5.1 Chapter Abstract ---- 89
5.2 Introduction ---- 89
5.2.1 The FRN ---- 90
5.2.2 The P3 ---- 90
5.2.3 Summary ---- 91
5.3 Method ---- 91
5.3.1 Inclusion and Exclusion Criteria ---- 91
5.3.2 Search Strategies ---- 92
5.3.3 Coding Procedures ---- 93
5.3.4 Statistical Methods ---- 94
5.4 Results ---- 94
5.4.1 Effect of RPE Valence ---- 94
5.4.2 Effect of Valence*Size ---- 97
5.4.3 Independent Components vs. Information Relay ---- 100
5.5 Discussion ---- 101
5.5.1 Results summary ---- 101
5.5.2 Limitations of difference waves ---- 102
5.5.3 Evidence of parietal RPE encoding ---- 102

5.5.4 Multiple RPE encoders 103
5.5.5 Chapter conclusions 104
6. General Discussion 105
6.1. Evidence of General-purpose RPE Encoding 105
6.1.1 RPE or salience 105
6.1.1.1 Insights from agency 105
6.1.1.2 Insights from primary/secondary reinforcement 106
6.1.2 Insensitivity to domains 107
6.1.3 Insensitivity to RPE size modulation 108
6.2 Additional Encoding 109
6.2.1 Evidence of motivational salience 109
6.2.2 Evidence of RPE-utility 110
6.3 Thesis conclusions 111
7. References 113

Figures

Figure 1. FRNs from studies manipulating outcome domain*participant agency (Chapter 2) 26
Figure 2. One trial of experiment (Chapter 2) 28
Figure 3. RPE encoding in the feedback-locked waveform (Chapter 2) 33
Figure 4. One trial of the shock calibration procedure (Chapter 3) 42
Figure 5. One trial of experiment (Chapter 3) 44
Figure 6. Simple waveforms, FRN, and scalp topographies thresholded FRN and Bayesian contrasts (Chapter 3) 49
Figure 7. Scalp topography of physical salience encoding and Bayesian contrasts (Chapter 3) 50
Figure 8. Grand average startle response (Chapter 3) 51
Figure 9. Predicted amplitudes for each encoder (Chapter 4) 60
Figure 10. Relative probabilities for the presence of each encoder, including a null encoder, for frontal waves across all time points (Chapter 4) 69
Figure 11. Factors 1 – 3 extracted from voltage waveforms using PCA (Chapter 4) 70
Figure 12. Factor 1 extracted from voltage waveforms using PCA (Chapter 4) 71
Figure 13. Factor 2 extracted from voltage waveforms using PCA (Chapter 4) 72
Figure 14. Factor 3 extracted from voltage waveforms using PCA (Chapter 4) 73
Figure 15. Relative probabilities for the presence of each encoder, including a null encoder, for frontal-parietal difference waves across all time points (Chapter 4) 74
Figure 16. Bayes factors for each main effect and interaction (Chapter 4) 75

Figure 17. Effect of domain by experiment form (Chapter 4) ---------- 76
Figure 18. Effect of size by experiment form (Chapter 4) ---------- 77
Figure 19. Effect of domain*size by experiment form (Chapter 4) ---------- 78
Figure 20. The FRN for domain form studies (Chapter 4) ---------- 79
Figure 21. Hypothetical example of component overlap (Chapter 4) ---------- 80
Figure 22. Reduced aversive FRN (Chapter 4) ---------- 82
Figure 23. The effect of cue type (Chapter 4) ---------- 84
Figure 24. The FRN for studies with abstract cues (Chapter 4) ---------- 85
Figure 25. Effect of RPE valence at frontal and parietal sites (Chapter 5) ---------- 95
Figure 26. Effect of RPE valence broken down by whether likelihood or magnitude is used to modulate RPE size (Chapter 5) ---------- 95
Figure 27. Funnel plot for the parietal effect of RPE valence (Chapter 5) ---------- 96
Figure 28. Effect of RPE size broken down by whether likelihood or magnitude is used as its modulator (Chapter 5) ---------- 96
Figure 29. Simple ERPs for RPE size, for rewarding outcomes only (Chapter 5) ---------- 97
Figure 30. Simple ERPs for RPE valence, for large outcomes only (Chapter 5) ---------- 97
Figure 31. Effect of valence*size at frontal and parietal sites (Chapter 5) ---------- 98
Figure 32. Effect of valence*size broken down by whether likelihood or magnitude is used to modulate RPE size (Chapter 5) ---------- 98
Figure 33. Funnel plot for the parietal effect of valence*size (Chapter 5) ---------- 99
Figure 34. Effect of valence*size broken down by moderator level (Chapter 5) ---------- 99
Figure 35. Heat map depicting superiority of parietal-frontal correlations over parietal-parietal correlations (Chapter 5) ---------- 101

Tables

Table 1. Predictive contrasts for each encoder, including a null encoder (Chapter 3) ---------- 47
Table 2. Studies used for the meta-analysis (Chapter 4) ---------- 63
Table 3. Predictive contrasts for each encoder, including a null encoder, across experiment forms (Chapter 4) ---------- 65
Table 4. Main effects and interactions for each encoder (Chapter 4) ---------- 68
Table 5. PCA factor summaries (Chapter 4) ---------- 71
Table 6. Studies used for the meta-analysis (Chapter 5) ---------- 93

Declaration and Acknowledgements

The work completed and submitted for this research degree at the University of East Anglia has not been previously included as part of any other degree or qualification, either at the University of East Anglia or any other institution. As part of this programme of study, relevant scientific seminars and conferences were attended, with the below work contributing towards presentations and publications.

I would like to thank my supervisor, Dr Thomas Sambrook, for his immeasurable patience and support throughout my studies and for his vast contributions towards my own development, and that of this thesis.

Publications:

Stewardson, H. J., & Sambrook, T. D. (2020). Evidence for parietal reward prediction errors using great grand average meta-analysis. International Journal of Psychophysiology, 152, 81-86.

Stewardson, H. J., & Sambrook, T. D. (2021a). Reward, Salience, and Agency in Event-Related Potentials for Appetitive and Aversive Contexts. Cerebral Cortex, 31(11), 5006–5014. https://doi.org/10.1093/cercor/bhab137

Stewardson, H. J., & Sambrook, T. D. (2021b). Reward prediction error in the ERP following unconditioned aversive stimuli. Scientific reports, 11(1), 1-10.

Stewardson, H. J., & Sambrook, T. D. (2022). Valence precedes value in neural encoding of prediction error. Psychophysiology. **(Submitted)**

Conferences:

EPS Online – 01/2021

EPS Decision-making in Voluntary Action Workshop – 04/2021

EEGLAB Virtual Workshop – 06/2021

Word count: 41,575

Signed:  **Date: 08/07/2022**

1. General Introduction

1.1 Reward Prediction Errors

Decision-making is characterised by adaptive control (Thorndike, 1898). An agent learns the best actions to take in order to maximise beneficial situations and minimise detrimental ones, switching to new behaviours when the outcome of a given situation is worse than expected. A key aspect of this adaptive control is that learning is not simply driven by the understanding of outcomes being good or bad, but by them being better or worse than was expected prior to their occurrence. For example, consider purchasing a packet of cookies. For this particular packet of cookies, it might be expected that there are five cookies inside, as our prior knowledge about them (as indicated by the label, say) tells us that this should be the case. However, upon opening the packet you find that there are only four cookies inside. Disappointed, you're now less likely to purchase those cookies again in the knowledge that you might receive less cookies than expected.

The above scenario is an example of a reward prediction error (RPE), which is a term representing the difference between the value of an obtained outcome and its prior expected value (EV). RPEs include two terms; their valence, defined by whether the outcome was better (positive) or worse (negative) than the EV, and their size, defined by how *much* better or worse the outcome was compared to the EV. Both of these terms are critical to the behavioural adjustment an agent makes following the outcome of an event, dictating whether behaviour should be adjusted or not, and to what extent it should be adjusted in order to optimise the outcomes of future events. The following subsections cover the terms of RPE valence and size in greater detail, before introducing the basis of this thesis; the neural processing of RPEs for human decision-making.

1.1.1 RPE valence

The valence of an outcome defines whether it is good or bad (for instance, the receipt of money is good, but the receipt of an electric shock is bad). While the valence of an outcome itself is no doubt an important consideration in reinforcement learning, the valence of an RPE is different and perhaps more informative. Throughout the reinforcement learning literature, valence has been used to describe both the inherent goodness or badness of outcomes as well as to describe RPEs, and so to avoid confusion throughout this thesis, we describe outcomes not as being of good or bad valence, but of *appetitive or aversive outcome domain*. Thus, the term valence here is used exclusively to describe the good or bad basis of RPEs.

Since RPE's consist of the difference between the value of an outcome and the EV, RPE valence determines whether the outcome of an event was better or worse than expected prior to its occurrence. Typically, positive RPEs (+RPEs) are a response to either the occurrence of rewards, such as a packet of cookies containing one more cookie than expected, or the omission of punishments, such as a packet of cookies containing the usual number of cookies when you expected there to be one less. On the contrary,

negative RPEs (-RPEs) are a response to either the omission of rewards or the occurrence of punishments. Note here the key distinction between outcome domain and RPE valence; The receipt of cookies is inherently a good outcome, or an *appetitive domain* outcome, as for the most part receiving cookies (like receiving all food stuffs) is considered to be a good thing. However, if the cookies received were of less quantity than expected, this is a bad outcome in RPE valence terms (a -RPE). Considering this, RPE valence is important for optimising choices regardless of whether they result in appetitive or aversive domain outcomes; less of a bad thing is good in a situation where only bad things can happen, with the opposite being true for good things.

1.1.2 RPE size

While RPE valence is at the crux of adaptive decision-making, the size of RPEs is also important. RPEs promote positive or negative reinforcement, with changes in the size of the RPE determining the degree of reinforcement. For instance, if you were to receive £5 for a task where £2 was expected, it would be a +RPE and thus promote positive reinforcement. If a different task (also with an expected £2 reward) resulted in the receipt of £10 however, then the positive reinforcement would be even greater (an overall RPE of +£8 rather than +£3). Given the choice of the two tasks again in the future, you would optimise your earnings by choosing the second task, as this behaviour produced a greater +RPE in comparison to the first task. In this manner, fully adaptive decision-making is not just defined by seeking out reward and avoiding punishment, but also by maximising the size of rewards (+RPEs) and minimising the size of punishments (-RPEs).

An additional point to consider is that RPE size can be altered not only by manipulating the magnitude of an outcome (such as the aforementioned difference between getting £5 or £10 for a particular task), but also by manipulating the likelihood of an outcome. For example, consider a task where you press a button that rewards you with either 10p or nothing at all. If the chances of that button rewarding the 10p is 50%, then the EV of the button press is +5p (10/2 = 5), thus winning the 10p would result in a +RPE of 5p. However, if the chances of winning the 10p is only 25%, then the EV is instead reduced to only +2.5p (10/4 = 2.5), thus winning the 10p would instead result in a +RPE of 7.5p. In this manner, increasingly unlikely outcomes are of increasingly greater RPE size.

1.1.3 A neural RPE

With RPEs representing a quantifiable, signed term for behavioural adjustment as detailed above, neuroeconomics considers them as central to human reinforcement learning, and it is highly likely that the brain indeed processes RPEs for this purpose (Sutton & Barto, 1998). Computationally, reinforcement learning takes shape as a trial-and-error process; an agent's goal is to piece together a mapping of world states (the events surrounding a given decision) and actions (the options available to the agent for that decision) over repeated trials in order to maximise rewarding outcomes, or minimise punishing ones. This process defines reinforcement learning theory (RL-theory), which directly ties this quantitative behavioural adjustment to the phasic increases and decreases of activity in the midbrain dopamine system (Schultz, 2002).

Specifically, RL-theory dictates that the brain computes RPEs as a result of an actor-critic architecture. In this model, brain regions such as the amygdala or prefrontal cortex (the "actors") send motor signals after making actions to the anterior cingulate cortex (ACC, the "critic"), which essentially tells the actor how appropriate that action was. This ultimately generates an RPE (Houk et al., 1995, p. 216; Schultz et al., 1997), before the critic then decides on how the actor should adjust to improve the chances or quality of future reward. This process of evaluating events is powerfully non-specific to outcome domain and thus applies to all manner of events; RPE's should not only describe utilitarian behavioural adjustment from rewards and punishments like with the cookies example detailed above, but should also describe performance-level behavioural adjustment from correctness or error feedback (Nieuwenhuis et al., 2004). However, this precise functionality of RPE encoding is questioned frequently within the literature, with some researchers disagreeing such a generic mechanism for reinforcement learning exists (Gehring & Willoughby, 2002). This disagreement ultimately stems from research failing to find common ground in where, when and how, computationally, RPE encoding takes place in the brain, and it is this debate on the neural foundation of RPEs that encompasses this thesis.

The next section of this chapter starts to unravel the causes for debate by first discussing some of the different types of RPE encoding that have been identified, before moving on to the methodological inconsistencies that may have led to the differing findings across the pool of reinforcement learning research. To do this, we focus the next section of this thesis largely on research investigating the *feedback-related negativity* (FRN), which is arguably the most prominent physiological component suggested to be tied to RPE encoding.

1.2 The FRN

The primary neurophysiological method used in the RPE literature has been electroencephalography (EEG) which, thanks to its temporal precision, has allowed for observable responses in decision-making tasks to be traced to the nearest millisecond. Of the EEG research conducted to date, the perhaps most

pivotal paper in the search for a neural RPE comes from Holroyd and Coles (2002). In their paper, the authors used EEG to observe two event-related potentials (ERPs) following outcomes in decision-making tasks, and made the claim that these signals represented an underlying neural mechanism for reinforcement learning. These ERP components, the error-related negativity (ERN) and the aforementioned FRN, were suggested to be linked in that both reflected a prediction error produced at the ACC; in line with the actor-critic architecture as described by RL-theory. While the ERN is observed as the participant becomes aware of their making an error, the FRN response is somewhat different, and has become the central component of interest in the reinforcement learning literature.

The FRN, sometimes known as the reward positivity, or RewP (Proudfit, 2015), is a frontocentral scalp signal that occurs around 240 – 340 ms after feedback following a choice. Specifically, it arises when a participant is unaware of whether a choice will result in success or failure at the time of making that choice, with a distinctive negative-going deflection in the waveform occurring when the participant is given feedback detailing the outcome. Holroyd and Coles suggested that this deflection is more negative for -RPEs than for +RPEs, and that the size of the deflection is related to the size of the RPE. For example, consider a simple task where a participant is told to guess whether a flipped 10p coin will land heads-up or tails-up. If the participant correctly guesses the result, then they win the coin, else they win nothing. As such, the size of the EV for deciding either heads *or* tails is equal to 5p; the success or failure of the participant's choice only comes to light once the coin lands. If the participant is incorrect in their guess, then they do not win the 10p coin and the RPE generated is equal to the forfeited EV, -5p. If they are correct, then the 10p coin is won and the generated RPE is instead equal to +5p. Crucially, the RPE is generated at the moment the heads or tails nature of the coin is revealed, with a more negative wave for the -5p RPE then for the +5p RPE.

We must detail that the proposed increased positivity of +RPEs compared to -RPEs is in fact arbitrary; there is no reason that +RPEs must be more positive in amplitude compared to -RPEs. This direction is chosen only due to negative-to-positive shifts in RPE valence being *typically* encoded by increases in voltage positivity in the literature. In line with this, we therefore quantify amplitudes in terms of positivity throughout this thesis.

1.2.1 Operationalising the FRN

As detailed at the start of this chapter, RPEs are constituted of both the valence and size of events in respect to prior expectations, and there are multiple ways in which these terms may be neurally encoded for the purposes of reinforcement learning. In this sense, while Holroyd and Coles' research demonstrated electrophysiological evidence of RPE processing in line with RL-theory, it has proven difficult to discern whether the FRN is in fact representative of a single, highly generic RPE encoder responsive to both the valence and size of events, representative of multiple less-generic RPE encoders that are responsive to some events but not others (Gehring & Willoughby, 2002), or indeed not representative of RPE encoding at all (Talmi et al., 2013). In short, the FRN signal and, for the purposes of this thesis, the wider reinforcement

learning waveform, may represent one or more types of encoding. The following subsections detail the most prominent types of encoding that have been identified in the literature, starting with the encoding most commonly attributed to the FRN; *RPE-sign encoding*.

1.2.2 Encoders

1.2.2.1 The RPE sign encoder

RPE-sign encoding is defined by a sensitivity to RPE valence, but not size. This effect of RPE valence is typically operationalised as a difference wave calculated as -RPEs – +RPEs, revealing a characteristic negative-going difference wave due to +RPEs typically generating more positive amplitudes in the waveform compared to those of -RPEs. This response is critical to the adaptive learning we described earlier with the cookies example, and is demonstrative of an agent's ability to identify an event as being better or worse than expected. We detailed in the previous section that the waveform is typically more positive for +RPEs compared to -RPEs, and specifically for RPE-sign encoding this positivity should be of the same amplitude regardless of the size of the RPE. For example, consider again the coin flip task from earlier. In this task, The FRN's amplitude following a correct choice should be more positive than it is following an incorrect choice. Crucially, if the coin was swapped to a 20p (giving a larger RPE size of +10p rather than +5p), the amplitude of the FRN following a correct choice should be the same as it was beforehand when winning the 10p. In summary, an RPE-sign encoder cares about whether an outcome was better or worse than expected, but not *how much* better or worse than expected it was.

1.2.2.2 The RPE-utility encoder

The RPE-sign account of the FRN is not the only possible encoding that has been suggested in the literature, with some research having alternative claims. The first of these alternative claims is that of *RPE-utility encoding*. RPE-utility is distinct from RPE-sign in that rather than simply distinguishing between the valence of events dichotomously (an event being better or worse than expected), it distinguishes between them on a scale of utility (an event being *this much* better or worse than expected). Such an encoder fits the axiomatic criteria established by (Caplin & Dean, 2008), whereby it is argued that the FRN should demonstrate an interactive response of both RPE valence and RPE size. This interaction is necessary to sufficiently explain adaptive reinforcement learning in line with Holroyd and Coles' theory, and thus represents a neural response even more powerful than that of RPE-sign. Returning to the coin flip task example above, if the FRN were to represent an RPE-utility response then its amplitude following successfully winning the 20p coin will be of greater positivity than that of winning the 10p coin.

Furthermore, not winning the 20p coin (an EV of -10p) represents a larger -RPE than not winning the 10p coin (an EV of -5p), and this should be reflected in the FRN having a less positive amplitude.

1.2.2.3 RPE size encoders

Another alternative claim regarding the function of the FRN is that it is not representative of a sensitivity to RPE valence at all, but to RPE size. In nature, larger prediction errors are typically expected to drive faster behavioural adaptation; the worse an outcome is to an agents' expectation, the more that agent will adjust its behaviour in order to minimise or avoid that outcome in the future. Furthermore, the encoding of an effect of RPE size is useful in the case of there being separate +RPE and -RPE encoders. For example, an encoder that is only responsive to better-than-expected events may be represented in the waveform by a greater positive deflection to an RPE of +10p compared to an RPE of +5p. Crucially, effects of RPE size can be encoded in two distinct ways; as an *unsigned prediction error* (UPE), or as *motivational salience*. A UPE encoder is responsive to RPE size in an absolute sense, whereby the amplitude of the waveform is determined by the size of the RPE regardless of its valence, and regardless of the outcome's inherent size or salience. Motivational salience encoding is instead responsive not to the size of the prediction error per se, but to the salience of a motivational good in respect to prior expectation.

To more clearly exemplify the difference between the two encoders, consider a slightly different coin flip task. In one trial of this task, a correct guess of heads-up or tails-up results in the participant winning a 20p reward, with the incorrect guess instead giving a 10p reward. In the second trial of the task, a correct guess results in a 10p reward, with an incorrect guess resulting in no reward at all (0p). In both trials, the outcomes all result in RPEs equal to 5p in size, as participants will either win, or not win, a monetary value that is 5p more or less than the EV (with EV equal to 15p in trial one, and 5p in trial two). As such, an unsigned prediction error encoder would be expected to respond to all events with the same amplitude; all RPEs are 5p in size, and the encoder does not distinguish if the 5p was lost or gained. However, a motivational salience encoder would respond with a greater amplitude for the +RPEs than the -RPEs in this scenario, as the motivational good (money) is more salient than expected (20p is more salient than 15p, and 10p is more salient than 5p). Similarly, increasing the salience of a motivational good that is aversive in domain, such as electric shocks, would also increase amplitude; a more intense electric shock than expected is more salient than a less intense electric shock than expected.

1.3 Causes of Competing Claims

The types of encoding detailed above comprise the most common arguments for the underlying encoding described by the FRN in the literature. They are distinct in their responses to RPE valence and size, and so distinguishing these types of encoding, and identifying which encoding the FRN signal more likely represents, should be easily possible through manipulations of both RPE valence and size in decision-making

tasks. In spite of this, researchers continue to disagree. The following subsections detail the two primary causes for these disagreements that are later addressed throughout the studies conducted in this thesis; *methodological differences between studies* and *component overlap*.

1.3.1 Methodological differences between studies

RPE valence and size are often established through tasks manipulating the variables of outcome domain and outcome size (such as a task including trials resulting in monetary wins and monetary losses of varying magnitude). However, these are far from the only variables that can be expected to have an effect on the amplitude of the FRN. There is a great deal of heterogeneity in the methodological approaches used across the reinforcement learning literature, and while these differences may all have a profound effect on the observed waveform, controlling or manipulating them stands as an obvious next step for gaining a deeper understanding of underlying encoders. For example, participant agency (defined by participants either passively observing outcomes, or actively making decisions to trigger them) may have an effect on the waveform when considering that RPEs are tied strongly to instrumental learning. As is detailed later on in Chapter 3 of this thesis, studies employing passive observation tasks (Hird et al., 2018; Soder et al., 2020; Soder & Potts, 2017; Talmi et al., 2013) typically fail to find RPE encoding in the time interval associated with the FRN. Another methodological cause for inconclusive results may be a lack of goal direction (San Martín, 2012). For example, studies employing designs where participants are told to predict upcoming monetary wins and losses, rather than being explicitly told to seek rewards and avoid losses, may result in the perception of successfully predicted losses as positive feedback.

The requirements and demands of participant behaviour within tasks are not the only additional factors to be considered as having an effect on the waveform. Amplitudes may also differ based on the type of stimulus or cue used (money, electric shocks, food, etc.), with perceived outcome size or intensity varying widely across participants and modality. Cookies may hold a much greater inherent value (be more salient) for one person compared to another, for example, and as we detailed above differences in salience are represented by some encoder types (such as for a motivational salience encoder) but not others (such as a UPE encoder). Furthermore, RPEs can be modulated through manipulations to both outcome magnitude and likelihood, and while the details of how an RPE's size is modulated is not strictly necessary for reinforcement learning models, research has found the FRN to be insensitive to one or the other in the past (Yeung & Sanfey, 2004). It is thus possible that unique neural systems may be involved in the processing of magnitude-based and likelihood-based RPEs separately.

Another key discrepancy in the literature is that of chosen intervals for statistical reporting. The functionality of RPEs, defined by their responses to valence and size, are typically ascertained through the differencing of ERPs and as detailed earlier in this introduction, such analyses are largely attributed to the FRN component. While the choice to operationalise the FRN as a representative of RPE encoding has firm grounds, a meta-analysis of FRN research has shown the interval of investigation across studies to vary

(Sambrook & Goslin, 2015). As this thesis shall address in chapters 4 and 5, such a lack of a specific target interval for the FRN is likely to lead to differing findings across studies as to its function; as mentioned above, there is a likelihood of the FRN representing multiple RPE encoders taking shape near-simultaneously, and so even slight discrepancies in the investigated interval of interest could lead to ERPs favouring particular conclusions regarding the nature of RPE demonstrated by the FRN over others. It also leads to future research being left with an increasingly vague interval in which to apply focussed analysis; an important issue to consider if one is attempting to resolve the encoding underlying a single component amongst the maze of many encoders that can theoretically reside within a given interval.

1.3.2 Component overlap

Following on from the previous section, despite the FRN being the primary component of interest in the reinforcement learning literature, it is likely not the only component that occupies the post-feedback waveform. These overlapping components, all of which representing a response of one or more encoders, lead to the summing and cancelling of ERPs that can make the disambiguation of individual encoders extremely difficult. To exemplify component overlap, consider there is an RPE-utility encoder at around 250 ms post-feedback in a decision-making task. This encoder produces an amplitude of decreased positivity for increasingly worse-than-expected events, and increased positivity for increasingly better-than-expected events. Now consider that in the same interval a motivational salience encoder is present, giving a response of increased positivity for both increasingly better *and* worse-than-expected events. The increased positivity for increasingly better-than-expected events generated by both encoders sum together, resulting in a larger positive-going deflection at that time point. However, the decreased positivity for increasingly worse-than-expected events generated by the RPE-utility encoder is cancelled out by the increased positivity for these events generated by the motivational salience encoder. This means the resultant waveform appears to be an entirely different response; an RPE-utility encoder that is only sensitive to better-than-expected events. Because of this, the waveform itself may represent a type of encoder that is not present at all, or indeed a single encoding may be disregarded as a possibility due to incorrect assumptions of component overlap being present.

1.4 General-purpose RPE Encoding

In light of the disambiguation difficulties detailed above, one important question remains unanswered in the literature: Is there a single encoder, present for all kinds of tasks and events across human decision-making, that is responsible for reinforcement learning? Alternatively, is it instead the case that there are multiple separate RPE encoders, restricted to RPEs constituted of particular traits (such as the aforementioned possibility of separate magnitude-based and likelihood-based encoders)? This thesis ultimately aims to resolve this debate, demonstrating evidence in favour of an ever-present RPE encoder;

which we term *general-purpose RPE encoding*. Before we summarise the scope of this thesis and detail why identifying such an encoder is of interest to the literature, the next subsection covers the assumptions that can be made regarding the specific behaviour of general-purpose RPE encoding.

1.4.1 General-purpose RPE assumptions

Firstly, and quite obviously, a general-purpose RPE encoder should demonstrate *RPE* encoding. As such, motivational salience encoding for example, due to its lack of sensitivity to RPEs, is not powerful enough to fulfil the requirements of adaptive reinforcement learning that is suggested by Holroyd and Coles, and the wider RL-theory. Following this notion, a general-purpose RPE encoder must be responsive to RPE valence. A response to utility however, with increasingly large +RPEs resulting in more positive amplitudes, and increasingly large -RPEs resulting in less positive amplitudes, is not strictly necessary. As such, a general-purpose RPE encoder could demonstrate RPE-sign or RPE-utility encoding.

Secondly, considering that one of the key aspects of an RPE encoder is that it should allow for the processing of not outcome domain, but RPE valence, then a general-purpose RPE encoder should be responsive consistently across both appetitive and aversive domains. The occurrence of better-than-expected events and the omission of worse-than-expected events should result in an increased positivity in the waveform, with the opposite being true for the omission of better-than-expected events and the occurrence of worse-than-expected events. In line with this assumption, a general-purpose RPE should be functional in all manner of situations, and thus should be responsive to all manner of stimuli, be them primary reinforcers (such as food or pain) or secondary reinforcers (such as abstract shapes or fractals). While the amplitudes of the waveform for such stimuli may differ based on their modality, there should always be a similar response in that the learning system cares not about the inherent value of an outcome, but the difference between what was expected and what was obtained.

Lastly, a general-purpose RPE encoder should be insensitive to how RPE size is modulated. As detailed earlier, this information should not be necessary for reinforcement learning models, what matters is the difference between expected and received outcomes. For example, a general-purpose RPE encoder should be responsive not only during tasks employing changes in monetary value, but also in tasks where monetary value is kept constant, but expectancy of wins or losses is instead manipulated.

1.4.2 Thesis summary

Finding a general-purpose RPE encoder is of great interest, regardless of if it is established in the interval of the FRN, or indeed elsewhere in the waveform. There is a strong likelihood of RPEs being at the core of reinforcement learning and behavioural adjustment, and so a single encoder responsive to RPEs in all kinds of situations would sufficiently explain this adaptability as well as adequately supporting RL-theory.

This thesis addresses the inconsistencies present in the current literature while resolving the issues of methodological heterogeneity and component overlap, ultimately identifying the presence of a general-purpose RPE encoder. First, chapters 2 and 3 address methodological inconsistencies across studies, whereby extraneous variables including participant agency and the use of primary rather than secondary stimuli can be manipulated to resolve the competing claims about whether the FRN demonstrates the presence of an RPE encoder, or rather a motivational salience encoder. Following from this, Chapter 4 uses a recent method of meta-analysis (Sambrook & Goslin, 2015) alongside the inclusion of a key variable, experiment form, in order to better disambiguate encoders as evidenced across the literature, applying Bayesian analyses to ascertain the relative evidence of encoders being present at each given point in the waveform. Chapter 5 also applies such a meta-analysis to provide support for the evidence of general-purpose RPE encoding, while addressing the possibility of there being additional, parietal RPE encoding. This is also of interest, as while the search for general-purpose RPE encoding lies predominantly in the FRN, there are likely additional RPE encoders elsewhere as part of separate neural processes distinct in purpose. Throughout, we establish evidence in favour of the RPE-sign account of the FRN while demonstrating its presence across all manner of tasks and events, in line with the predicted response of a general-purpose RPE encoder.

1.5 Glossary

The terms used to define RPE encoders and the effects they are responsive to are notably inconsistent throughout the literature, and so this glossary sets out the use of terminology for this thesis specifically.

Expected value (EV)

The sum of all available outcomes weighted by their likelihood. As an example, the EV of a trial with the possible outcomes of either +25p or +75p is +50p.

Outcome value

The face value of an outcome. As an example, the receipt of +25p has an outcome value of exactly +25p.

Physical salience

The physical intensity of a stimulus. This is not always present; for example, while monetary deductions are negative in outcome value, the absence of the physical good (in this case, money) means there is no physical salience.

Reward prediction error (RPE)

The signed, quantitative difference between outcome value and EV. As an example, the receipt of +75p following a trial with an EV of +50p yields an RPE of +25p.

Unsigned prediction error (UPE)

The absolute difference between outcome value and EV. As an example, the receipt of +25p following a trial with an EV of 0p yields a UPE of 25p. Furthermore, the deduction of -25p following a trial with an EV of 0p also yields a UPE of 25p. Similar to an RPE, a UPE is *not* affected by the physical salience of an outcome; the receipt of +25p following a trial with an EV of 0p yields a UPE of 25p, and the receipt of +50p following a trial with an EV of +25p also yields a UPE of 25p, despite the latter outcome being of greater physical salience.

Motivational salience

Similar to UPE, but with the key difference in that motivational salience *is* affected by an outcome's physical salience. While the receipt of +25p following a trial with an EV of 0p yields a prediction error of 25p, and the receipt of +50p following a trial with an EV of +25p also yields prediction error of 25p, the latter outcome is of greater motivational salience. This is because +50p is of a greater physical salience than +25p.

RPE valence

A categorical variable describing whether an outcome is better or worse than expected, and thus refers to RPEs only. The levels of prediction error valence are *positive* and *negative*, and are typically worded throughout this thesis as *positive RPEs* (also *+RPEs*) and *negative RPEs* (also *-RPEs*). As an example, the receipt of 25p from a trial with an EV of 0p is a +RPE, whereas the receipt of 25p from a trial with an EV of 50p is a -RPE. Put simply, RPE valence refers to only the sign of an RPE, not its size.

Outcome domain

A categorical variable describing whether an outcome is inherently good or bad, and thus does not refer to RPEs. The levels of outcome domain are *appetitive* and *aversive*. As an example, the receipt of 25p is always an appetitive outcome, regardless of whether it was a positive RPE or a negative RPE. Conversely, the deduction of 25p is always an aversive outcome.

RPE size

Essentially the opposite to the above; this variable describes the quantitative size of an RPE, and thus refers to RPE's only. As an example, the receipt of 25p from a trial with an EV of 0p has an RPE size of 25p.

Outcome size

A variable describing the absolute size of an outcome. Outcome size is typically manipulated such that it has two categorical levels, *large* and *small*. Outcome size is constituted by magnitude, such as the receipt of 75p (large) compared to an alternative of 25p (small).

2. Reward, Salience and Agency in ERPs for Appetitive and Aversive Contexts

This chapter is based on a published paper:

Stewardson, H. J., & Sambrook, T. D. (2021a). Reward, Salience, and Agency in Event-Related Potentials for Appetitive and Aversive Contexts. Cerebral Cortex, 31(11), 5006–5014.

2.1 Chapter Abstract

Cognitive architectures tasked with swiftly and adaptively processing biologically important events are likely to classify these on two central axes: *motivational salience*, in other words, those events' importance and unexpectedness, and *RPE-sign*. Because of its temporal precision, EEG provides an opportunity to resolve processes associated with these two axes. As discussed in the previous chapter, a focus of attention for the last two decades has been the FRN, a frontocentral component occurring 240 – 340 ms after valenced events that are not fully predicted. Both motivational salience and value (in the form of RPE-sign encoding) are present in such events and competing claims have been made for which of these is encoded in the interval of the FRN. The present study suggests that RPE-sign encoding is the primary determinant of the FRN in active contexts, while in both passive and active contexts, a weaker and earlier overlapping motivational salience encoder may be present.

2.2 Introduction

Adaptive behaviour in an intelligent organism requires neural processes capable of extracting the full range of information associated with motivationally relevant events in the environment. Some of these processes are associated with online monitoring of unfolding events, for example whether immediate goals have been achieved, awareness of whether the motivating context is one of threat or opportunity, and orientation to unexpected events. Other processes are concerned with adjusting longer term expectations about the future profitability of the environment, and of particular actions pursued within it. A challenge for neuroscience is to isolate the neural signatures associated with these functionally distinct processes.

Foremost of concerns for any motivated agent is valence. As stated in chapter 1, valence may be determined by the biologically specified appetitive or aversive properties of primary reinforcers (food is good, and pain is bad) or, with sufficient exposure, by secondary reinforcers that precede these primary reinforcers. However, valence can also consist in whether an event is better or worse than expected from prior experience with the situation at hand, that is, it can consist in positive or negative RPEs. This is the key distinction between biologically specified valence, or outcome domain, and RPE valence, which represents a powerful computational means for learning appropriate actions in response to motivational cues. To reiterate, the sign of an RPE (positive or negative) respectively strengthens or weakens the propensity for future

selection of an action, and its size determines the degree of strengthening or weakening. By these means, RPEs can produce very effective machine learning (Sutton & Barto, 1998), and they represent a computationally efficient mechanism for learning. They can be recruited for any class of motivating event and are unaffected by whether these events are appetitive or aversive, since the RPE merely reflects the difference in obtained and expected value, and not the value of the event itself. Their importance for learning in humans and other animals is demonstrated by blocking: the finding that learners fail to learn stimulus-response contingencies when the following outcomes do not produce prediction errors (Kamin, 1967). Furthermore, midbrain dopamine neurons have been shown to encode RPEs (Schultz et al., 1997) directly linking their computation to that of RL-theory.

The ubiquity of RPE computation, and the generalisability of a single computational process over the full range of motivational events, raises the prospect that the brain might host one or more general-purpose RPE encoders recruited in any motivated learning situation. As is the primary aim of this thesis, demonstrating the presence (or indeed absence) of such encoding would improve our understanding of the computational architecture underlying motivated behaviour. In identifying RPE encoders, temporal precision is at a premium, since even a fully-modular encoder would likely be swiftly relaying computational input and output to non-RPE systems. For this reason, EEG represents an appropriate technique for the isolation of RPEs.

2.2.1 The FRN: RPE-utility claims

One candidate for such an encoding lies in the interval associated with the FRN (Holroyd & Coles, 2002; Miltner et al., 1997). This electrophysiological component is elicited by valenced feedback, occurs at a latency of 240 to 340 ms post-feedback and shows a frontocentral scalp distribution. Since it is typically observed in designs with dichotomous good vs. bad feedback it is most often operationalised as the voltage difference between a +RPE and a -RPE. The waveform for +RPEs typically shows a relative voltage positivity and this, by convention, is subtracted from the waveform for -RPEs to produce a negative-going difference wave, the FRN. The amplitude of the FRN is greater when RPEs are large (Sambrook & Goslin, 2015), suggesting it may represent the encoding of RPE size in addition to RPE valence, and thus may represent full encoding of utility in the manner of an axiomatic RPE-utility encoder (Caplin & Dean, 2008). This is supported by studies manipulating utility as a continuous variable (Cavanagh, 2015; Y. Gu et al., 2021; Sambrook & Goslin, 2014, 2016). When the generation of prediction errors (and consequent learning) is prevented in a blocking procedure, the FRN is reduced, suggesting it may play an obligatory role in the transmission of RPEs (Luque et al., 2012).

2.2.2 The FRN: Motivational salience claims

Other studies have claimed that activity in the temporospatial interval of the FRN is better interpreted as a response to motivational salience rather than RPE-sign encoding. Motivational salience consists in an outcome's ability to elicit attention due to its motivational relevance (Schultz, 2017) and is high for *both* appetitive and aversive events, and low for neutral events (Berridge & Robinson, 1998; Bromberg-Martin et al., 2010). In a counterpart to demonstrations of RPE-encoding dopamine neurons, Matsumoto and Hikosaka (2009) showed other dopamine neurons appeared to code for motivational salience, insofar as they were excited by the delivery of intrinsically motivating stimuli (e.g., juice or air puff) regardless of whether these were appetitive or aversive. These neurons were also responsive to cues predicting the imminent onset of these stimuli. Furthermore, while RPE-encoding neurons showed a reduction in firing when appetitive events were unexpectedly omitted (thus carrying the -RPE portion of the signal in the appetitive domain), this second class of dopamine neuron was unresponsive to omission, further suggesting a motivational salience encoding that treated omission as a neutral outcome, even when this was unexpected. A large number of fMRI studies have attempted to identify structures on the mesolimbic and mesocortical pathways that might receive either the RPE-sign (Fujiwara et al., 2009; O'Doherty et al., 2004) or motivational salience signal (Jensen et al., 2007; Metereau & Dreher, 2013).

A motivational salience encoder would appear to be highly adaptive in terms of orientation and attention, and we might expect its effects to be seen at some point in the feedback-locked ERP waveform. Four studies (Hird et al., 2018; Soder et al., 2020; Soder & Potts, 2017; Talmi et al., 2013) have made the case that the FRN represents a motivational salience encoding, and have done so by means of a particularly strong methodology: the use of a primary aversive reinforcer (e.g., shock, noise burst or bitter taste). This contrasts with the widespread use of monetary loss as a reinforcer. While monetary loss serves to represent -RPEs, and negative feedback generally, it is a problematic choice for specifically eliciting a motivational salience encoder. Money, through a process of robust association, does appear to be an intrinsically motivating stimulus for humans (Zaghloul et al., 2009), but it can only be an appetitive, not aversive stimulus. Furthermore, as noted above, motivational salience consists in the occurrence of motivating stimuli, not their omission, and the extent to which losses can be classed as occurrences is unclear.

2.2.3 The role of agency

The studies cited above manipulated RPEs independently in appetitive and aversive domains and reversed the polarity of the FRN in the aversive domain. This reversal occurred because the voltage positivity associated with rewards in the appetitive domain was, in these studies, shown for punishments in the aversive domain, and so the differencing of bad and good outcomes produced a positive, rather than negative-going FRN. This is parsimoniously explained by the FRN simply showing a motivational salience

encoding, in other words, the delivery rather than omission of a motivating, valenced stimulus, but with no encoding of RPE valence.

In stark contrast however, two other studies using pain as a primary aversive reinforcer (Heydari & Holroyd, 2016; Mulligan & Hajcak, 2017) found outcome domain to have no effect on the polarity of the FRN, supporting the RPE-sign account. Notably, all four studies supporting the motivational salience account employed tasks in which participants merely passively experienced rewards and punishments, while the two studies supporting RPE-sign encoding employed active, instrumental designs. Since RPEs are understood to be computational terms in instrumental learning, we should be unsurprised to find a much-reduced RPE encoding, or indeed its absence, in passive tasks. This is a ubiquitous finding, at least in the appetitive domain (Hassall et al., 2019; Philiastides et al., 2010; Sambrook & Goslin, 2015; Walsh & Anderson, 2012). Additionally, the four passive studies reported somewhat earlier and smaller FRN peaks in both appetitive and aversive domains, raising the possibility that the component observed was revealed by the reduction or removal of a larger, partially overlapping RPE encoder, present in active tasks due to its obligatory role in instrumental learning. A motivational salience encoder would fit this profile and, serving a general role in orientation and attention rather than instrumental learning, we would expect it to be present in both passive and active contexts.

2.2.4 Predictions and rationale

The interval occupied by the FRN lies from 240 to 340 ms post feedback according to a meta-analysis of fifty-five studies (Sambrook & Goslin, 2015). If this interval is occupied by an RPE-sign encoder, most strongly elicited in active tasks, but the earlier part of this interval is also occupied by a motivational salience encoder, equally active in both passive and active tasks, a specific set of effects on the waveform is predicted. During active tasks in the appetitive domain, RPE-sign encoding and motivational salience encoding should sum in the early portion of the interval, while in the aversive domain they will cancel. This should manifest in a greater latency for the FRN in the aversive domain than in the appetitive, with the true latency of the underlying neural generator assumed to lie in between. In passive tasks, where there is only one encoder, appetitive and aversive FRNs, now reflecting only motivational salience, should show the same latency, should be earlier than both active task FRNs and should be reversed for the aversive FRN. This is indeed the picture presented by a simple condition-wise averaging of waveforms in the six studies cited earlier, as shown in Figure 1.

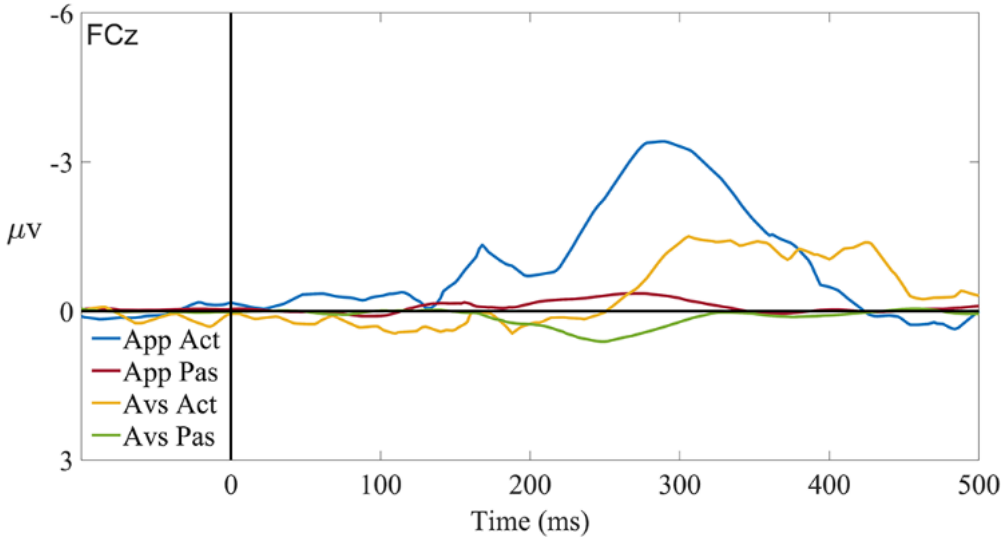

*Figure 1. Average FRNs from six studies manipulating outcome domain*participant agency. Weighted averages and standard deviations are shown, based on experiments' sample size. In designs manipulating RPE size as well as valence, the largest available RPE sizes were used. See Sambrook and Goslin (2015) for a description of the averaging technique.*

Because both the latency and amplitude of ERP components is subject to a host of incidental factors, peculiar to procedures that vary over experiments and laboratories, our aim in this study was to manipulate participant agency (active vs. passive) and outcome domain (appetitive vs. aversive) within a single experiment to establish whether an active context instated an encoding of RPE-sign in the FRN interval.

Our primary aim was thus a resolution of motivational salience vs. RPE-sign accounts by manipulation of participant agency. As a secondary aim, we wished to further characterise the active-context FRN, or any other observed active-context RPE signal in terms of its capacity to carry a continuous measure of RPE utility, in a fashion consistent with an axiomatic RPE-utility encoder (Caplin & Dean, 2008) and with that frequently used in reinforcement learning algorithms (Sutton & Barto, 1998).

2.3 Method

2.3.1 Participants

Sixty-four students from the University of East Anglia participated for course credit and an opportunity to win money. All participants were under 29 years, had no history of neural damage or other significant health problems, and were not on medication at the time of the experiment. Eight participants were excluded for excessive EEG artefacts (see EEG analysis section below) and 11 for failure to meet the learning criterion: a significantly greater selection of the optimal key in active conditions under a Chi squared goodness of fit test ($\alpha = .05$). This left a final sample of forty-five participants (31 female). The

study was approved by the ethics committee of the School of Psychology at the University of East Anglia and the experiment was undertaken with the understanding and written consent of each participant.

2.3.2 Experimental Design

Three factors were orthogonally manipulated. *Outcome domain* consisted in whether appetitive (money) or aversive (noise) outcomes were at stake, *RPE valence* consisted in +RPEs (money delivery, noise omission) vs. -RPEs (noise delivery, money omission) and *participant agency* consisted in individuals' active vs. passive role in the task.

The FRN was operationalised by a difference wave of voltages for -RPE and +RPE outcomes, thereby reducing the design to a manipulation of outcome domain and participant agency. RPE-utility encoding was expressed in terms of Pearson's *r*, calculated by a correlation of voltage with RPE utility over trials. RPE utility, a signed value between -1 and +1, was derived from a computational model of participants' choices (see below). This value was correlated, across trials, with the observed voltage at each electrode/sample to produce plots representing the strength of RPE-utility encoding.

Separate correlations were performed for +RPEs and -RPEs to avoid incorporating the effect of RPE valence, that is, the dichotomous discrimination of a +RPE from a -RPE. Since RPE utility was inferred from participant choice, only active conditions were used in this analysis, producing a 2 x 2 design, comprised of outcome domain and RPE valence. Interpretation of the results is based on both the significance and sign of *r*. If this is same-signed and significant for both +RPEs and -RPEs, a bivalent RPE-utility encoding is indicated, capable of ordering RPEs from much worse than expected to much better than expected on a single bivalent scale. If *r* is significant for either +RPEs *or* -RPEs, a univalent RPE-utility encoder is indicated, discriminating the RPE utility for one valence of prediction error but not the other. If *r* is opposite-signed and significant for both +RPEs and -RPEs, a (continuous) encoding of motivational salience is indicated. In all cases, outcome domain should have no effect.

2.3.3 Procedure

To standardise the aversiveness of white noise, participants were fitted with headphones, and listened to eleven 700 ms bursts of white noise of ascending volumes between 50dB and 70dB. On a second presentation, they rated the aversiveness of each of these on a visual analogue scale, and for the remainder of the experiment the volume used was that corresponding to the highest volume which the participant described as unpleasant but tolerable for the experiment.

In the main experiment, participants undertook five consecutive blocks of sixty trials each in four conditions: active-appetitive, active-aversive, passive-appetitive, and passive-aversive. Condition ordering was counterbalanced over participants. On each trial, participants observed the presentation of a simple geometric symbol that denoted, with 100% accuracy, the delivery or omission of a reward or punishment. In

aversive conditions this symbol was followed by either a 700 ms burst of noise or silence, in appetitive conditions by a 700 ms audio clip of a cash-till (indicating a £0.02 win each time this was incurred) or silence. All successful trials in the appetitive conditions incurred payment, following the recommendation of Schmidt et al. (2019).

In active conditions only, participants began each trial by choosing between two keys, one of which was predetermined to give the better symbol (e.g., denoting money delivery or noise omission) 60% of the time, while the other gave it 40% of the time. In passive blocks, the better symbol was presented 55% of the time, a figure selected in an attempt to produce comparable ratios of good to bad symbols in active vs. passive conditions, an outcome which was approximately achieved (53.3% in active). Figure 2 depicts one trial.

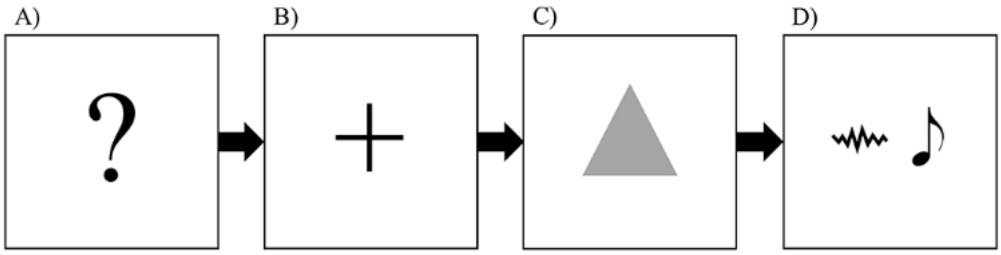

Figure 2. One trial of the experiment. A) 1000 ms in passive condition or until keypress (max 1000 ms) in active conditions B) 600–700 ms fixation C) 1000 ms truth cue D) 700 ms noise, cash-till sound or silence.

Before the first block of each condition, participants were shown a pair of symbols and instructed which of the two indicated delivery and omission of the reward or punishment at hand for that condition. The mapping of symbols to delivery or omission remained constant over all five blocks, with separate pairs of symbols used for each condition in counterbalanced form. In active conditions, participants were told that one key would deliver the better symbol more often, that they could infer this key by trial and error and that it would be randomly reset at the start of each block. Participants were told to attend to symbols in both active and passive conditions. In order to encourage attention to symbols, particularly in passive blocks, after 5% of trials, participants were required to indicate which symbol they had been shown and were fined £.10 for failing to correctly respond within 5s.

2.3.4 EEG Recording

EEG data were collected from 61 Ag/AgCl active electrodes (actiCAP, Brain Products, Gilching, Germany) mounted on an elastic cap and arranged in a standard International 10 – 20 montage referenced to the left mastoid. Vertical eye movement was monitored by a right suborbital electrode, and horizontal eye movement was monitored using an electrode on the right external canthus. Electrode impedances were kept

below 20 kΩ. EEGs were amplified using a BrainAmp amplifier (Brain Products), continuously sampled at 1000 Hz.

2.3.5 EEG Analysis

EEG data were down-sampled to 500 Hz, filtered with notch filters at 60 Hz and 50 Hz, followed by a .1 Hz high pass filter and 30 Hz low pass filter. Segments were time-locked to 300 ms before the onset of the feedback symbol to 700 ms afterwards, and were baseline-corrected using the interval -200 – 0 ms. Eye movement artefacts were removed using a criterion of a voltage change exceeding 75 μv per 200 ms in eye electrodes in the interval -300 – 700 ms. Other non-specific artefacts in the interval -200 – 700 ms were removed using a criterion of any electrode showing either a voltage change exceeding 50 μv per ms, a voltage value exceeding 100 μv relative to baseline, or activity across the epoch below 12 μv. Segments were re-referenced to the average of left and right mastoid activity and baselined once again. After exclusion of participants showing fewer than 30 trials in any of the four conditions, 73.47% of trials were retained and the average number of trials per condition was: active-appetitive, 95.86 (SD = 29.07); active-aversive, 98.71 (SD = 25.76); passive-appetitive, 93.77 (SD = 28.06); passive-aversive, 93.61 (SD = 25.53). Electrodes which malfunctioned in the course of an experiment were substituted using topographic interpolation (Perrin et al., 1989).

2.3.6 Statistical Analyses

Since the FRN's latency was predicted to vary over conditions, and because components overlapping the FRN were of interest, a predetermined window of analysis was not used, and instead the full waveform was analysed. Control for multiple comparisons was achieved with the method of Maris and Oostenveld (2007), implemented in a custom script written in the proprietary code of Brain Vision Analyzer. One sample t-tests (test value = 0) were performed in each condition on subject average voltages at each sample of the waveform and, where these were significant, agglomerated over temporally or spatially adjacent samples to produce a cluster-t statistic. The significance of this was established by comparison to a distribution of cluster-t values under the null hypothesis obtained by switching data (subject average) and test value (zero) at each sample in the cluster with a 50% probability. Twenty thousand iterations performed on data down-sampled to 100 Hz were used. Significance was indicated by the proportion of this null distribution with values higher than the observed cluster-t value. To remove transient activity unlikely to reflect genuine components, only clusters covering a minimum of 25 samples were assessed, and clusters were discarded if their cluster-t value failed to meet an alpha threshold of 0.025, Bonferonni-corrected by the number of clusters found in the initial agglomeration process. Having identified clusters of activity, condition effects were established by performing analysis of variance at the site where the FRN was maximal (Fz), on mean voltage over the interval determined as significant by cluster randomisation above. The same process was

used to establish clusters of RPE-utility encoding activity, with *r* rather than difference wave voltage serving as the test data.

2.3.7 Computational Modelling of RPE utility

While some studies of the FRN build a continuous measure of RPE utility directly into their design (Pedroni et al., 2011), the great majority present trials which, at the moment of feedback, provide one of a simple pair of categorical outcomes. This was true of the six key studies cited earlier, so retaining this design feature was paramount for a valid comparison. While simple dichotomous feedback cannot reveal information about the encoding of RPE size, it can reveal an encoding of RPE valence, which is sufficient to address the motivational salience vs. RPE-sign debate, the primary aim of this experiment. The secondary aim of the experiment, identifying neural correlates of continuous RPE utility, did require a manipulation of RPE size, and this was addressed by incorporating a second step: the estimate of trial-by-trial RPEs based on a computational model of reinforcement learning. Q learning (Sutton & Barto, 1998) was selected on the basis of its parsimony and wide usage. While this is one of a class of model-free learning algorithms with somewhat different assumptions, in a direct comparison of parameter estimates derived from three such models, Q learning, SARSA and Actor-Critic, Walsh and Anderson (2011) showed parameter estimates to be very similar over models, and consequently there to be no significant effect of which model was used on the amplitude of FRNs built from RPE estimates generated by the model. The purpose of fitting was not to best characterise participants' learning but simply to generate regressors for an analysis of RPE-utility encoding that were unavailable from the dichotomous feedback design chosen.

In the model, participants choose one of two actions, a_1 or a_2, at the opening state *s* (the prompt to press a key) on each trial, *t*. The value of each action is denoted by *Q*. The value for *Q* is updated at the end of each trial based on the prediction error, δ, adjusted by the learning rate, α.

$$Q(s_t, a_{i,t})^{new} = Q(s_t, a_{i,t}) + \alpha.\delta$$

The prediction error is given by the difference of *Q* and *r*, the reward obtained.

$$\delta_t = r_t - Q(s_t, a_{i,t})$$

Q could take values between 0 and 1 (initialised at 0.5), *r* could take values 0 or 1 and δ could take values between -1 and 1. The learning rate, α, was fitted on a participant-wise basis using maximum likelihood estimation from observed choices. This necessitated the incorporation of a choice rule. A standard softmax rule was used, incorporating the inverse temperature parameter β (also fitted) to derive the probability, P, of each action.

$$P(a_{i,t}) = \exp[\beta.Q(s_t, a_{i,t})] / \Sigma a' \exp[\beta.Q(s_t, a_{i,t})]$$

Each participant was fitted individually using the L-BFGS-B method (Byrd et al., 1995) of the *optim* function in R (R Core Team, 2020). In order to increase the stability of the neural regressors, and following Daw (Delgado et al., 2011, p. 6), α and β were set as a constant over participants based on their fitted median, and the model was re-run to establish trial by trial RPEs.

Fits were compared to those obtained from a model in which participants held a fixed preference for one of the two keys, modelled as a single free parameter, and a null model with no free parameters, in which participants chose randomly. Fitting code for Q learning was based on that described in Sambrook et al. (2018), documented in full as part of the of the catlearn package (Walsh & Anderson, 2012; Wills et al., 2018) and is available at: https://github.com/thomasdsambrook/Q.

2.4 Results

2.4.1 Behavioural Results

In the active task, participants chose the correct key 68% of the time, with all participants showing a significant preference for this key under individual Chi squared goodness of fit tests against a conventional $\alpha = .05$. At the group level, preference for the correct key was unaffected by outcome domain (paired $t(44) = .99$, $p = .33$). Participants' responsiveness to feedback was further established by the model fit being superior to that of the alternative models. Median raw log likelihoods with interquartile ranges were: Q learning -282.74 (-312.25 to -253.85); fixed side preference -392.22 (-414.85 to -384.91); null -415.89 (no range). The fit remained superior after converting each model's raw log likelihoods to a Bayesian Information Criterion in order to compensate for the two free parameters in the Q learning model (aggregate level Bayes Factors: Q learning vs. key preference 9,566, Q learning vs. null 10,173). On an individual basis, the Bayes Factor for the Q learning model was superior in all but four cases when compared to the fixed preference model, and in all cases when compared to the null. The learning rate α was fitted with a median value of 0.50 and an interquartile range of 0.26 to 0.80, the inverse temperature β was fitted with a median value of 3.79 with an interquartile range of 2.41 to 5.95.

2.4.2 FRN

Figure 3A shows simple waves for the eight conditions, all of which display the P2-N2-P3 complex that is typical in the feedback-locked ERP. In Figure 3B, waveforms for bad and good outcomes are differenced to produce FRNs by condition. Monte Carlo cluster randomisation of the FRN revealed a significant frontocentral cluster of activity for the active appetitive condition between 210 and 270 ms

(Monte Carlo p = .00195), for the active aversive condition between 270 and 320 ms (Monte Carlo p = .00095) and the passive appetitive condition between 220 – 260 ms (Monte Carlo p = .00140). Importantly, and as predicted by our primary hypothesis, the FRN for the active aversive condition was negative going and same-signed as the active appetitive condition.

Mean voltages at Fz in the Monte Carlo determined intervals of significance (and for passive aversive from 220 – 260 ms) were entered into a 2 x 2 outcome domain x participant agency analysis of variance. This revealed no significant effect of outcome domain, but an effect of participant agency ($F = 4.66$, $p = .036$, $\sigma^2 = .096$) and a significant interaction ($F = 5.47$, $p = .024$, $\sigma^2 = .11$). The effect of participant agency was present for the aversive domain ($t(44) = 3.21$, $p = .003$) but not appetitive ($t(44) < 1$). This is the expected pattern if an encoding of RPE-sign, present only in active conditions, is removed in passive conditions to reveal a weaker motivational salience encoding, oppositely signed in the aversive domain but same-signed in the appetitive. The effect was not sufficiently strong to produce an FRN with a significant (one-sample t-test vs. baseline) reversed polarity in the passive aversive condition. Nevertheless, a simple effects contrast of passive appetitive and passive aversive FRNs at Fz in the interval 220 – 260 ms revealed a significant difference ($t(44) = 2.58$, $p = .013$) as predicted by a motivational salience encoding.

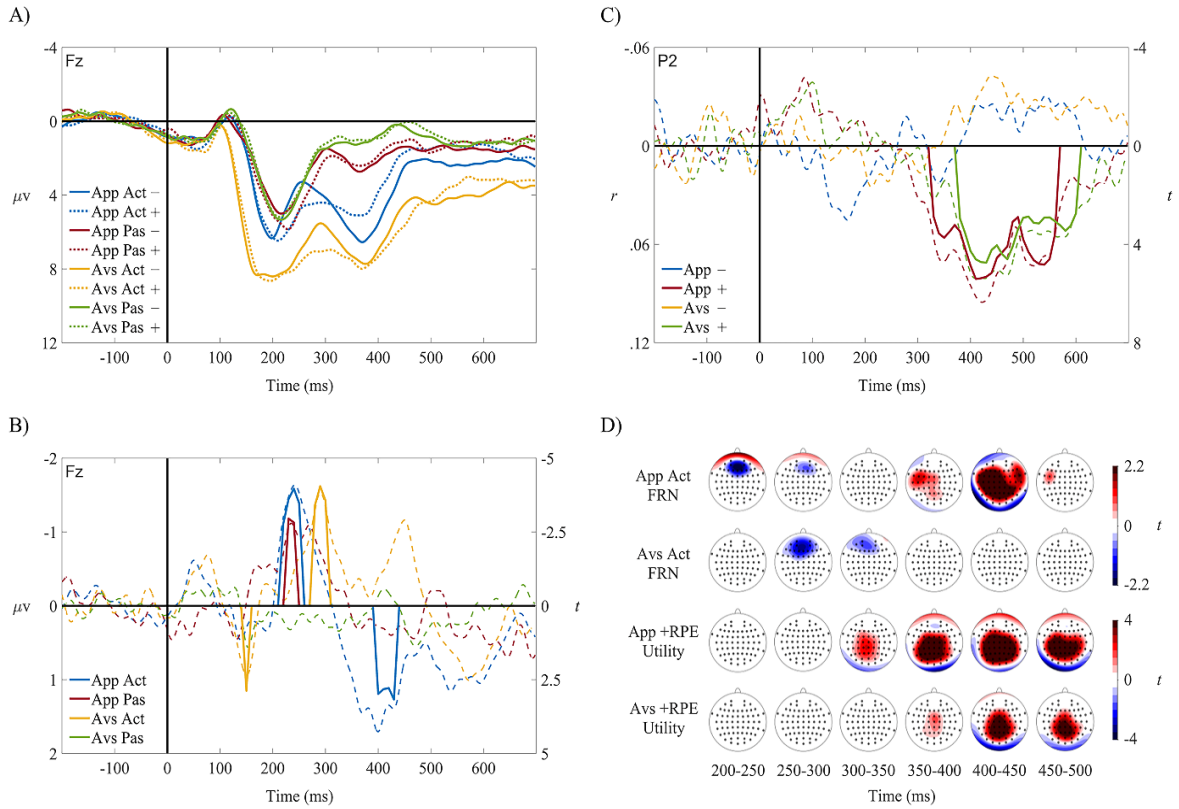

Figure 3. RPE encoding in the feedback-locked waveform. (A) shows simple waveforms by outcome domain, participant agency and RPE sign. (B) shows FRNs (dashed lines) by outcome domain and participant agency, obtained by differencing +RPE and –RPE simple waveforms in (A). Bold lines indicate significance under a t-test (see right hand axis) in intervals retained after cluster randomisation. (C) shows the correlation, r, of voltage and utility (dashed lines), and significant intervals after cluster randomisation (bold lines). Scalp maps in (D) show the time course of the principal effects: FRNs in active conditions and RPE-utility encoding for +RPEs.

2.4.3 RPE-utility Encoding

Monte Carlo cluster randomisation of Pearson *r* values, derived from the correlation of RPE utility and voltage, revealed a single cluster of activity for each condition. Large, overlapping centroparietal clusters encoded appetitive +RPE utility (300 – 690 ms, peak P3, Monte Carlo p < .00005) and aversive +RPE utility (350 – 620 ms, peak P2, Monte Carlo p = .00018), with increasing RPE utility associated with voltage positivity. Smaller, spatiotemporally distinct clusters encoded appetitive -RPE utility (130 – 190 ms, peak FC1, Monte Carlo p = .00015, with decreasing RPE utility associated with positive voltage) and aversive -RPE utility (400 – 510 ms, peak C1, Monte Carlo p = .00080, with decreasing utility associated with positive voltage). As such, no encoder of motivational salience active in both outcome domains could be shown,

while an RPE-utility encoder, restricted to +RPEs, was indicated by the conjunction of the +RPE clusters, in an interval extending from 370 – 570 ms, maximal at P2.

2.5 Discussion

2.5.1 Results summary

There has been ongoing debate as to whether activity in the FRN interval represents an encoding of RPE-sign or motivational salience. This study found that, in active conditions, the FRN was same-signed for both appetitive and aversive stimuli, suggesting the FRN demonstrates RPE-sign encoding during instrumental learning. Previous studies showing a reversal of FRN polarity in the aversive domain may have failed to elicit this encoder as a consequence of employing a passive design, and so revealed an earlier overlapping motivational salience encoder instead. Consistent with our hypotheses, the active aversive FRN was delayed, and the negativity of the passive aversive FRN was greater reduced than that of the passive appetitive FRN. Nevertheless, no significant reverse-signed passive aversive FRN was seen. This may be attributable to cancelling due to the presence of a weak RPE-encoding component elicited even in passive tasks, something that has been frequently observed in the literature (Walsh & Anderson, 2012). This would act to cancel the motivational salience encoding in the aversive domain and augment it in the appetitive domain, the pattern observed in the data. Such an effect should of course be present in the four other passive designs featuring in Figure 1 and yet these studies do report aversive reverse-signed and significant FRNs. As such, this aspect of our account merits further investigation, possibly with designs manipulating the availability of prediction errors (e.g., using blocking), or task instructions asking participants to attend to either motivational salience or RPE valence (Gu et al., 2021).

Overall, the common polarity of the active-context FRN over appetitive and aversive domains is supportive of the possibility of a single, instrumental learning signal that is blind to the actual value of the stimuli at hand (noise, money, etc). The FRN was spatially similar for the two outcome domains and the latency differences were potentially attributable to component overlap as described earlier. Future studies applying source localisation should further assess the evidence for such a common encoding of RPEs over appetitive and aversive domains.

2.5.2 RPE-utility encoding

As a secondary aim, the study explored encoding of RPE-utility. It found evidence that, in a later interval, RPE-utility was indeed encoded rather than the simply dichotomous RPE sign, as shown by a positive correlation between +RPE size and voltage, regardless of outcome domain. Notably, the size of +RPEs in the aversive domain were strongly represented, despite these “outcomes” constituting an omission.

Since orientation and avoidance responses would be inappropriate to omitted events, a learning signal is strongly implicated. The presence of +RPE encoding but not -RPE encoding (also found in Sambrook & Goslin, 2016 for both monetary gains and losses) is notable, and echoes claims made elsewhere that feedback locked ERPs constitute a response to better than, rather than worse than expected outcomes (Foti et al., 2011; Gu et al., 2021; Holroyd et al., 2008) where -RPE-utility is encoded. Notably, encoding of RPE-utility did not occur in the interval associated with the FRN however, suggesting the possibility of separate RPE-sign and RPE-utility encoding, evidence for which has been presented elsewhere (Fouragnan et al., 2015; Philiastides et al., 2010).

2.5.3 Passive RPE encoding

This study was designed in order to assess the *relative* evidence for two competing hypotheses of activity in the FRN interval: whether this represents RPE-sign or motivational salience. While we argue that RPE-sign underlies the dominant response in active contexts, we would not claim that encoding in this interval is restricted simply to this property. Clearly, the existence of an overlapping motivational salience encoder is central to our interpretation of the data. An encoder responding to passive RPEs is also implicated by this study and such has prior plausibility. The learning of stimulus-outcome associations in passive contexts is adaptive for learning about environmental contingencies that may serve future action. Animals will approach stimuli that have previously undergone positive Pavlovian reinforcement, for example (Brown & Jenkins, 1968). Representations of stimulus value would likely be maintained by RPEs just as would action values in an instrumental context, and this is suggested by single cell studies, many of which employ passive learning contexts (Niv & Schoenbaum, 2008).

Whether scalp activity associated with active and passive RPEs is based on the same or different encoders is yet to be resolved. The greater scalp amplitude following RPEs in active tasks may simply reflect general arousal effects (Yeung et al., 2005). Alternatively, it may indicate the activation of distinct encoders, differentially conducted through to the scalp. In Holroyd and Coles' 2002 model, the FRN is generated specifically when an RPE coincides with activation of motor controllers in the anterior cingulate cortex, in other words, in active tasks. This model is supported by the observation that the FRN is largely eliminated when a delay between action and outcome is introduced (Weinberg et al., 2012). If the FRN is dependent on such an eligibility trace, this would imply a highly modular signal in an actor-critic architecture assigned to learning action values only. fMRI studies have shown the dorsal striatum to activate specifically in active contexts, while the ventral striatum responds to both active and passive (Balleine et al., 2007; O'Doherty et al., 2004). While source analysis of the FRN has produced conflicting results, the dorsal striatum has been cited as a source of the FRN using joint EEG/fMRI (Carlson et al., 2011) and PCA informed source localization (Cohen et al., 2011). If so, an alternative source or sources will need to be found for passive RPEs and this must await source localization applied specifically to passive tasks.

2.5.4 UPE encoding

As we alluded to in chapter 1, another encoding claimed to lie in the feedback-locked waveform is UPE encoding (Hauser et al., 2014). Sometimes referred to as "simple surprise", this constitutes another form of salience that is distinct from motivational salience due to it having a V-shaped relationship with utility within each outcome domain and regardless of inherent outcome size, treating both omissions and deliveries as equally salient. Gu et al. (2021) were able to reverse the polarity of the FRN for monetary losses by task instructions that stressed predictive accuracy (UPE), and then reinstate the normal polarity when instructing participants to focus on reward (RPE). A further property that may be encoded is outcome domain. While we have made the case for RPE-sign encoding in the FRN being at last partially outcome domain-independent, a component encoding for outcome domain is likely, since the generation of adaptive behaviour requires that +RPEs in aversive environments not be confused as desirable events in any absolute sense (Boureau & Dayan, 2011).

2.5.5 Chapter conclusions

In short, even very ambitious multifactorial designs will struggle to unconfound all the computational terms present in feedback processing. The inevitability of component overlap worsens this problem when using the ERP method, and there are likely to be a host of non-learning processes involved with orientation and cognitive control that affect the latency and amplitude of observed ERP components, and which will be variably elicited across studies, depending on the stimuli and design employed. As stated earlier, there are strong grounds for supposing that both RPE-sign and motivational salience are encoded in the brain. It is largely an empirical question when and where these occur on the scalp, and our aim in the current study has not been to afford either a greater theoretical weight, but to resolve their contribution to the feedback-locked ERP waveform, and in particular, the FRN, a much-studied component. The present study shows that RPE valence, outcome domain and participant agency are critical determinants of the FRN, and that when they are experimentally controlled, an encoding of RPE-sign is revealed.

3. Reward Prediction Error in the ERP Following Unconditioned Aversive Stimuli

This chapter is based on a published paper:

Stewardson, H. J., & Sambrook, T. D. (2021b). Reward prediction error in the ERP following unconditioned aversive stimuli. Scientific reports, 11(1), 1-10.

3.1 Chapter Abstract

Temporal difference (TD) methods of reinforcement learning generate RPEs at the earliest time at which a revision in reward or punishment likelihood is signalled, for example by a conditioned stimulus. Midbrain dopamine neurons, believed to compute RPEs, generate this signal in response to both conditioned and unconditioned stimuli, as predicted by TD learning. Electroencephalographic recordings of human participants have suggested that the FRN is generated when this signal is carried to the cortex. If this is so, the FRN should be expected to reflect a response that is equivalent for conditioned and unconditioned stimuli. However, very few studies have attempted to measure the FRN for unconditioned stimuli. The present study attempted to elicit the FRN in response to a primary aversive stimulus (electric shock) using a design that varied RPE while holding physical intensity constant. The FRN was strongly elicited, but earlier and more transiently than typically seen in the literature, suggesting that it may incorporate other processes than the midbrain dopamine system.

3.2 Introduction

A challenge facing neuroscientific studies of RPEs is linking research at the microscopic and mesoscopic scales. Single cell recordings have demonstrated RPE encoding in midbrain dopamine neurons, occurring with a latency of 50 – 110 ms and peaking at 200 ms (Hollerman et al., 1998; Mirenowicz & Schultz, 1996; Schultz et al., 1997). Meanwhile, the ERP technique has demonstrated how the FRN may represent the encoding of an RPE (Miltner et al., 1997; Sambrook & Goslin, 2014) . The FRN is localised most frequently to the cingulate cortex, though sometimes to the striatum (Walsh & Anderson, 2012), both targets of midbrain dopamine neurons. A pertinent question is whether the FRN reflects the transmission of a midbrain RPE signal to corticolimbic structures, as Holroyd and Coles' (2002) original theory claims.

3.2.1 The FRN and unconditioned stimuli

According to RL-theory, midbrain dopamine cells respond to both appetitive and aversive stimuli, coding +RPEs (e.g., delivered juice or omitted air puff) with increased phasic firing and negative RPEs (e.g.,

omitted juice or delivered air puff) with phasic dips (Hollerman et al., 1998; Ungless et al., 2004). This implies that diverse motivational events may be encoded on a single scale of RPE utility. Another notable characteristic of midbrain dopamine cells is that they respond both to unconditioned stimuli (e.g., air puff or juice), and conditioned stimuli that indicate imminent delivery of these primary reinforcers (Matsumoto & Hikosaka, 2009). This accords with a TD learning model in which RPE signals that are initially incurred at the receipt of a primary reinforcer propagate back in time to the presentation of a conditioned stimulus predicting that reinforcer's delivery (Houk et al., 1995; Sutton & Barto, 1998). The similarity of neural response to conditioned and unconditioned stimuli suggests that properties of the actual stimulus are not encoded by these cells, merely its motivational *value*, and so a process of simple model-free learning reinforcement learning is at work (Schultz, 2013).

If the FRN arises from the transmission of this signal to the basal ganglia and cortex, then the same equivalent response to unconditioned or conditioned stimuli should be seen in this component. However, very little is known about the FRN's response to unconditioned stimuli. While some FRN studies have used primary reinforcers such as food or shock rather than the ubiquitous secondary reinforcer of money, the FRN is near universally measured not in response to these reinforcers but to a cue denoting their subsequent delivery. This cue may be a fully abstract "truth cue" (e.g., a geometric shape), may carry a conventional outcome valence denotation (e.g., a cross or tick) or, less frequently, may visually reference the reinforcer itself (e.g., a picture of a snack). Over the course of an experiment, this cue should be expected to acquire the status of a conditioned stimulus, and so should be expected to engage the midbrain dopamine cells. Initially however, the correspondence between the cue and the reinforcer it denotes must be consciously apprehended by participants as a result of instruction and held in working memory. While it is possible that such a mapping is routed via the midbrain dopamine system from the outset in order to create an RPE, this is not necessary, and the FRN in this case might not reflect an RPE in the strict model-free sense, but rather a more generic performance error. The idea of the FRN reflecting quantities other than RPEs is supported on a number of grounds. For example the component can be elicited, albeit in a reduced form, when feedback denotes a correct response, but one to which no reward is attached (Van den Berg et al., 2012; Weinberg et al., 2014). It is also shown by participants observing rewards or punishments incurred by others (Bellebaum, Polezzi, et al., 2010; Marco-Pallarés et al., 2010). Furthermore, there is evidence that it incorporates counterfactual knowledge about unchosen options, something difficult to reconcile with model-free reinforcement learning (Gehring & Willoughby, 2002). Thus, the FRN, at least as it is traditionally operationalised, may represent the combination of overlapping, same-polarity deflections associated with functionally and anatomically distinct processes broadly relating to error and success. Alternatively, it may reflect a single unitary response that integrates such inputs, including both RPE and performance monitoring aspects. In either case however, if at least some portion of the FRN is derived from activity of the midbrain dopamine neurons, as has been claimed, it should be expected to be sensitive to the direct presentation of unconditioned stimuli in the manner shown by these cells.

3.2.2 Methodological challenges

Measuring an ERP to a primary reinforcer, rather than to a cue, presents methodological challenges. The use of cues is beneficial as it makes it easy to study RPEs while controlling out differences in the sensory properties of unconditioned stimuli. This includes both their modality (electric shocks might produce different event related potentials from money) and intensity (large shocks might produce larger potentials than their small equivalents, even when the RPE they carry is equal). Locking the waveform to arbitrary truth cues, counterbalanced across the reinforcers they denote, deals with the possibility of sensory confounds which can otherwise never be ruled out. Additionally, it may reduce attentional, orienting responses that are specific to the onset of primary reinforcers, and which have nothing to do with RPEs or learning. This should produce better signal to noise ratio and, indeed, the FRN as locked to cues is a robustly elicited component.

There are two principal challenges when measuring RPEs in response to unconditioned stimuli. One of these is to present the stimulus at a sufficiently well specified time in order to align epochs, helping to resolve components in the averaged waveform. Food or drink ingestion is difficult to time-lock in this way, as are naturally occurring social rewards such as smiling or laughter. Studies that have attempted to use affective images to generate the FRN have fared poorly (Brown & Cavanagh, 2018; Olofsson et al., 2008), perhaps also due to weak operationalisation of those images via participant ratings. In the aversive domain, which is less often used in studies of human reinforcement learning, two temporally precise reinforcers are available however; electric shocks and bursts of white noise.

The second problem with the use of unconditioned stimuli, alluded to earlier, is that RPEs are typically overlapped with *physical salience*, that is the intensity of the stimulus (Schultz, 2016). As the physical salience of a motivationally relevant stimulus increases (e.g., the sweetness of juice or the strength of shock), its utility, in terms of RPE, changes too, with the direction depending on whether the stimulus is appetitive or aversive. It therefore becomes unclear whether prediction error or physical salience underlies the observed potential. The FRN is typically operationalised as a negative-going bad – good difference wave, indicating a positive relationship between RPE utility and voltage. In the case of an aversive stimulus such as shock, lower than expected shocks should produce a relative positivity of voltage compared to that produced by greater than expected shocks, producing a negative going high shock – low shock difference wave. However, when the FRN is locked to the actual shock, rather than a cue, such an effect could equally as well be cited as evidence for a purely sensory component responding to physical salience, in this case via voltage negativity for increasing salience. Since RPEs are rarely the object of study in experiments using electric shock, this confound is typically not addressed. However, one study removed the confound by delivering shocks of equal intensity in which one shock was higher than expected (-RPE) and the other lower (+RPE). In this study, lower than expected electric shocks produced a waveform showing a relative positivity from 60 ms onwards, although a significant effect was only reported in the context of a late P2 component, at approximately 350 ms, allowed to vary over site and latency across the fourteen participants (Hird et al., 2018). Using a similar design, though with laser evoked potentials, and only six participants, another study

found encoding of physical salience, but inspection of the tabulated data suggests no RPE-sign effect (Brown et al., 2008). Importantly, both studies cited above delivered stimuli to participants who were passive in the process, but as we demonstrated in the previous chapter, the FRN is known to be weak or absent in such tasks. The aim of the present study was to assess the evidence that the FRN is elicited in response to RPEs carried by an unconditioned stimulus in an active learning task.

3.2.3 Predictions and rationale

Satisfying the above claim would require a demonstration of neural activity consistent with the FRN's polarity, topography and latency. For polarity, +RPEs would need to show a positive voltage relative to -RPEs, thus producing the negative going bad – good difference wave by which the FRN is conventionally operationalised. Topographically, the component should be frontocentral. Temporally, it should have a similar latency to the FRN as typically reported. This latency is dependent on the terms by which the FRN is defined, however. While, as noted above, the standard operationalisation is a bad – good valence contrast, because of interest in the FRN's role in demonstrating the encoding of RPE-utility rather than merely a dichotomous comparison of RPE valence, the FRN is also sometimes operationalised in terms of the difference of this difference wave for small and large RPEs (Holroyd et al., 2003, 2009; Nieuwenhuis et al., 2002). A meta-analysis of the FRN showed that while both operationalisations peaked at around 280 ms, they showed nested intervals, with the FRN lying between 150 and 400 ms, and the difference of difference waves lying between 240 and 340 ms (Sambrook & Goslin, 2015). Since the present design did not manipulate RPE size, the wider interval was selected as the interval of interest.

In addition to investigating the effect of primary reinforcers on the FRN, the present study's use of electric shock supports an investigation of affective aspects of RPE encoding. RPEs consist in a simple differencing of expected and received value. The encoding of value implies a motivational and therefore affective element, however the functional links between RPE generation and wider affect are not fully understood. Early accounts of the FRN varied in the degree to which they stressed its affective (Gehring & Willoughby, 2002) versus computational qualities (Holroyd & Coles, 2002). While self-reported affect does not correlate well with FRN amplitudes (Li et al., 2011; Sambrook et al., 2012; Yang et al., 2013), physiological measures provide a plausible alternative, provided reinforcers are sufficiently strong to produce a measurable response, as might be expected for shock. The present study used startle in response to unexpected bursts of white noise to index the affective state of participants waiting in anticipation of shock. The startle response was chosen partly because it is used across multiple mammal species, but also because it has been shown in humans to serve as a *bivalent* measure of affect, potentiated when participants face the prospect of an aversive outcome such as shock, and reduced when they are shortly to receive an appetitive outcome such as sweet food (Andreatta & Pauli, 2015). This bivalent nature makes it ideal for capturing variance in the FRN which may be modulated by either positive or negative RPEs.

3.3 Method

3.3.1 Participants

Forty-nine students (20 males) of the University of East Anglia participated for course credit. All participants were under 29 years, had no history of neural damage or other significant health problems, and were not on medication at the time of the experiment. Four participants failed to complete the experiment because they found the procedure too painful, and one because of equipment failure. The study was approved by the ethics committee of the School of Psychology at the University of East Anglia. The experiment was undertaken with the understanding and written consent of each participant, with participants receiving course credit for their participation.

3.3.2 Shock Apparatus

Shocks were delivered by electrocutaneous stimulation using a linear isolated stimulator (Model: STMISOLA; 45 BIOPAC Systems, Inc.) charged by a stabilized current, and electrodes (Model: EL350; BIOPAC Systems, Inc.) attached to the wrist of the participant's non-dominant hand. Shocks were controlled from E-Prime code running on the presentation computer, communicating with the stimulator by means of a Measurement Computing Corporation (MCC) data acquisition device (USB-1208HS-4AO). Shock was carried by a 25 Hz sine wave lasting 160 ms and was delivered at three voltages on the basis of the calibration procedure described below.

3.3.3 Shock Calibration

In order to calibrate the aversiveness of shock on an individual basis, participants were administered a range of shock intensities and rated each on a ten-point visual analogue scale (see Figure 4). Shocks began at 30 volts and were incremented in three-volt steps until the participant rated the pain at "8", or 70 volts was reached. The shock was decremented two volts, and then in steps of three volts until a rating of "3" was reported. Shock was then incremented by a single volt, and then in steps of three volts, until the participant rated the pain at "8", or 70 volts was reached. This procedure resulted in ratings at one-volt steps. Small, medium and large shocks for the subsequent experimental session were calculated by taking the mean voltage given a rating of "3" ("very slightly unpleasant"), "5" ("moderately unpleasant") and "7" ("highly unpleasant but still tolerable for the forthcoming experiment").

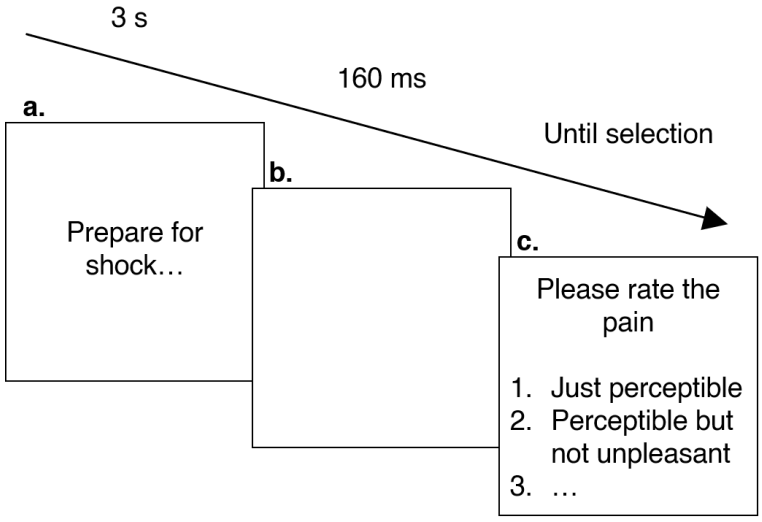

Figure 4. One trial of the shock calibration procedure. a. Warning for the approaching shock. b. Shock is delivered. c. Participant rates the pain: "1" ("just perceptible"), "2" ("perceptible but not unpleasant"), "3" ("very slightly unpleasant"), "4" ("slightly unpleasant"), "5" ("moderately unpleasant"), "6" ("unpleasant"), "7" ("highly unpleasant but still tolerable for the forthcoming experiment"), "8" ("intolerable for the forthcoming experiment"), "9" ("very painful indeed and intolerable for the forthcoming experiment"), or "10" ("excruciating and intolerable for the forthcoming experiment").

3.3.4 Experimental Design

Although three shock sizes were used, these were implemented in a 2 x 2 trial stakes x RPE valence design. In low stakes trials, participants could receive either the small or medium shock, while in high stakes trials they could receive either the medium or large shock. Once participants had been cued of the two permissible shocks, the arrival of one therefore represented a +RPE and the arrival of the other a – RPE. Cells in the design are thus referred to as low stakes positive (LP), low stakes negative (LN), high stakes positive (HP) and high stakes negative (HN). In terms of physical salience (a variable that was not present in the design, and which we wished to control out), LP shocks were small, LN and HP shocks were medium, and HN shocks were large. The critical contrast for the experiment was between LN and HP since this varies RPE valence while holding physical salience constant.

A power analysis was performed based on a meta-analysis of the cue-locked FRN (Sambrook & Goslin, 2015) in which the FRN was defined by the mean amplitude of the good – bad difference wave in the interval 228 – 334 ms. This meta-analysis used original subject-level data in 27 studies that were supplied by authors. The standardised effect sizes ranged from 0.19 – 2.03, with all FRNs correctly signed. The average standardised effect size was 1.00 (0.93 when this average was weighted by study sample size) and the average simple effect size in microvolts was -2.16. Effect sizes of *treatment effects* on the FRN were of

course much smaller but the present study was concerned merely with demonstrating the presence of an FRN.

Given an effect size of 1.00, an alpha of 0.05 and significance assessed with a one-sample t-test, a manual calculation of power (Howell, 1999) reveals that a sample size of 8 is required to achieve 0.8 power and 18 to achieve 0.99 power (10 and 21 participants if effect size is set to 0.93). However, the estimated effect size of 1.00 was based on studies that used a larger number of trials than were used in the present study. For ethical reasons the number of large shocks was limited to fifty per participant, resulting in two hundred trials per participant and one hundred trials in the critical comparison of LN and HP. The effect that varying trials has on effect size has been studied using the lateralised readiness potential (Boudewyn et al., 2018) , a component with a comparable but somewhat lower simple effect size than the FRN. This study showed that when N > 15, increasing trials beyond 30 had no effects on internal consistency of the component's amplitude. This finding tallies with a study specifically of the FRN which showed that with a sample size of 25, 20 trials were sufficient to measure FRN amplitude (Marco-Pallares et al., 2011). Because the effect size of the FRN in response to unconditioned stimuli was unknown, a sample size of 45 was selected to provide a comfortable margin of power.

3.3.5 Procedure

Participants were informed that they would take part in a learning task in which they had to choose between two fractals and attempt to minimise their exposure to painful electric shocks that followed their choice. They were told to monitor the shock received and use that to guide their subsequent choices. The meaning of the stakes cues was explained to them and therefore which two levels of shock were possible on a given trial. They were told they would experience bursts of white noise, which they should ignore since these were unrelated to the task. The sequence of events on a trial is shown in Figure 5.

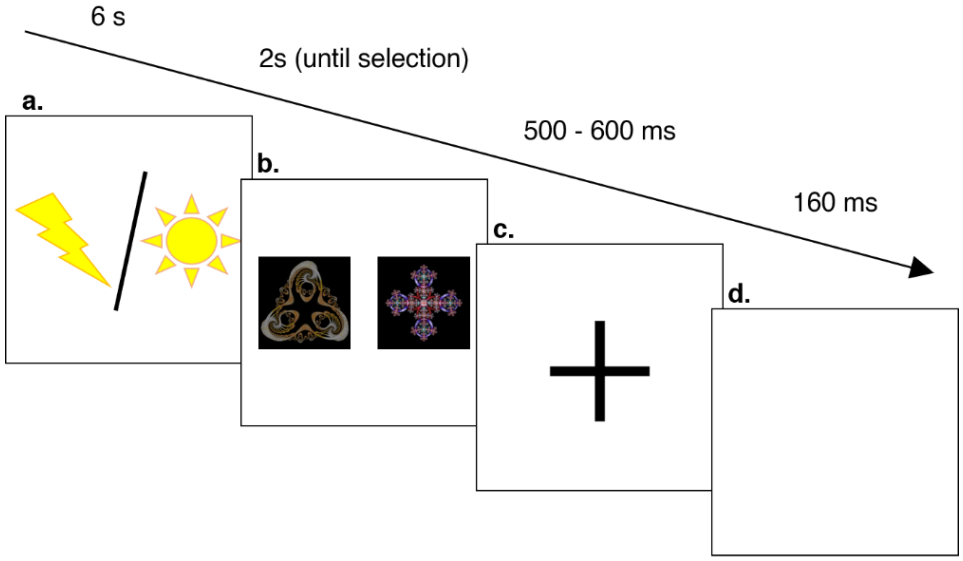

Figure 5. a. Icon indicates if trial is high or low stakes, with noise administered on 20% trials at a random point between 2 and 5 s. b. Participant selects from a pair of fractals. c. Cross indicates shock is coming. d. Shock is delivered.

Participants were first presented with a stakes cue indicating whether the trial would be high or low stakes. This cue was visible for six seconds and on 20% of occasions participants were exposed to white noise during its presentation. On these occasions a 105 dB burst of white noise was delivered via headphones for one second during a randomly determined interval from two to five seconds. This was followed by the presentation of a pair of fractals (unique to each level of stakes, and each block of the experiment), which remained on screen for 2 s or until participants selected one of the fractals with a keypress. This keypress was the sole choice made by participants in a trial. This was followed by an image of a cross lasting 500 – 600 ms and then the delivery of a shock. Unbeknownst to participants, shock outcomes were predetermined such that half of the high stakes trials resulted in a large shock and half in a medium shock, presented in a pseudorandom order. Half of the low stakes trials resulted in a small shock and half in a medium shock. Keypresses therefore had no relationship with the size of the shock that was administered. Despite the task being unsolvable, participants were instructed to try and work out which key would minimise shock.

Participants undertook five blocks of forty trials each, with each block containing twenty high, and twenty low stakes trials, presented in a random order. The symbols denoting whether a trial was high or low stakes were held constant for a given participant but counterbalanced across participants. At the conclusion of each block, participants were asked which key they considered to have been most favourable. Since one key would generally be associated with lower shock by chance association, this served as an index of participants' learning effort in the absence of a genuinely learnable task.

3.3.6 EEG Recording

EEG data were collected from 61 Ag/AgCl active electrodes (actiCAP, Brain Products, Gilching, Germany) mounted on an elastic cap and arranged in a standard International 10 – 20 montage referenced to the left mastoid. Vertical eye movement was monitored by a right suborbital electrode, and horizontal eye movement was monitored using an electrode on the right external canthus. Electrode impedances were kept below 20 kΩ. EEGs were amplified using a BrainAmp amplifier (Brain Products), continuously sampled at 500 Hz.

3.3.7 EEG Analysis

Data were filtered with notch filters at 60 Hz and 50 Hz, followed by a .1 Hz high pass filter and 30 Hz low pass filter. Segments were time-locked to 300 ms before the onset of shock to 700 ms afterwards and were baseline-corrected using the interval -200 – 0 ms. Eye movement artefacts were removed using the semi-automatic ICA-based ocular artefact rejection function within BrainVision Analyzer software. Other non-specific artefacts in the interval -200 – 700 ms were removed using a criterion of 50 μV change per ms, a voltage value exceeding 100 μV relative to baseline, or activity across the epoch below 12 μV. Segments were re-referenced to the average of left and right mastoid. Electrodes which malfunctioned in the course of an experiment (an average of 0.56 per participant) were substituted using topographic interpolation (Perrin et al., 1989). Data were further down-sampled to 100 Hz for the cluster-randomisation procedure described below. Mean trials per condition were as follows: LP, 48.59 (SD = 1.99), LN, 48.02 (SD = 3.22), HP, 47.73 (SD = 3.89), HN, 47.23 (SD =4.69).

3.3.8 Statistical Analysis

Two analysis strategies were employed to reveal RPE encoding. The first strategy was to assess the evidence for a significant LN – HP difference wave. This represents a comparison of -RPEs and +RPEs, thus conforming to a conventional analysis of the FRN, and controls for physical salience. The full waveform was analysed for evidence of such a difference wave, and cluster randomisation using the method of Maris and Oostenveld (2007) was used to compensate for multiple comparisons (See Fieldtrip (Oostenveld et al., 2010) version 20200607). In this process, one sample t-tests, performed over participants, were used to establish significance of the difference wave at each sample (electrode x timepoint combination), and clusters of spatially and temporally adjacent significant samples were created (minimum 25). Each cluster was assigned a Monte Carlo p value based on positioning its sum of t-values in a null distribution of clusters created from 400,000 permutations of the data under Monte Carlo simulation. The Monte Carlo p value returned by this procedure was Bonferroni-corrected by the number of clusters found at the initial stage. All t-tests reported in this study underwent this correction for multiple comparisons. On an exploratory basis, the same

procedure was used to identify the timing and location of activity encoding physical salience. The contrast used for this purpose was high vs low stakes trials since this varies the aggregate level of physical salience while ensuring equal frequency of positive and negative prediction errors. Finally, it should be noted that no comparison was made between FRNs of high and low stakes (LN-LP vs. HN-HP) since these difference waves confound prediction error with stimulus intensity, the effects of which cannot be assumed to increase linearly over the range used in the experiment.

The second strategy employed a more fine-grained analysis of the evidence for RPE encoding in the context of what were expected to be strong effects of physical salience and the possibility that RPE encoding would behave differently under high and low stakes conditions. To separate out effects of RPE and physical salience encoding, we specified, for each of the six pairwise contrasts available from the 2 x 2 design, the expected direction of effect that would be produced by five possible encoders. These encoders were: RPE encoded by either negative or positive voltage shifts (with positive shifts corresponding to the conventional FRN), physical salience encoded by either negative or positive voltage shifts, and a null encoder in which neither property was encoded. Evidence for these encoders' activity at a given point in the waveform thus consisted in the degree to which the observed contrasts met the predictions. These predictions are given in Table 1. In a number of contrasts the predicted effect is that of no effect and, since this is as diagnostically meaningful as an effect, a Bayesian approach was taken to place testing of the two possibilities on an equal footing. One-sample Bayesian t-tests were used to test each contrast. These were run at each sample using the ttestBF function of the BayesFactor package in R (Morey & Rouder, 2015), and the resulting Bayes' factors were expressed as posterior probabilities in support of each of the predictions shown in the cells of Table 1. Evidence for each encoder was calculated by taking the product of the posterior probabilities of the six predicted effects specified in that encoder's column. This evidence was normalised across encoders by dividing each encoder's compound probability by the sum of the compound probabilities for all five encoders. The resulting probabilities, which summed to one, thus represented the *relative* evidence for each encoder compared to the others, and in particular the null encoder. Clusters of activity associated with each encoder were determined by thresholding these normalised compound probabilities at .975, making their significance comparable with the standard t-tests used elsewhere in the paper.

Table 1. Rows designate the six pairwise contrasts available from the design. Columns 4 – 8 show the predicted effects that five different encoders will show in each contrast. The quantity compared in these contrasts is positivity of voltage. The emboldened cell identifies the single critical contrast used for the first analysis strategy.

			Encoding of				
			RPE		**Physical salience**		**Nothing**
Trial contrast	**Salience contrast**	**RPE contrast**	**Voltage +**	**Voltage -**	**Voltage +**	**Voltage -**	
HN/HP	high/med	bad/good	HN < HP	HN > HP	HN > HP	HN > HP	HN = HP
HN/LN	high/med	bad/bad	HN = LN	HN = LN	HN > LN	HN > LN	HN = LN
HN/LP	high/low	bad/good	HN < LP	HN > LP	HN > LP	HN > LP	HN = LP
HP/LN	med/med	good/bad	**HP > LN**	HP < LN	HP = LN	HP = LN	HP = LN
HP/LP	med/low	good/good	HP = LP	HP = LP	HP > LP	HP > LP	HP = LP
LN/LP	med/low	bad/good	LN < LP	LN > LP	LN > LP	LN > LP	LN = LP

For startle response analysis, EEG data were filtered with a 28 Hz high pass filter and notch filters at 60 Hz and 50 Hz, and rectified to isolate high frequency electromyographic effects (Blumenthal et al., 2005). Startle response was measured at the right suborbital electrode. Activity at this site was averaged in a rolling 50 ms window and baseline corrected from -50 – 0 ms. Segments were removed as artefactual if activity exceeded ± 5 μV in the baseline (4.89% of trials) and removed as non-startle if no activity exceeding ± 5 μV was found in the interval 0 – 200 ms (49.45% of remaining trials). Remaining startle trials (mean per participant 17.73, range 2 – 36) were averaged for each participant to provide participant average startle response waveforms. Because two participants showed no startle responses in one of the two stakes conditions, they were excluded from the analysis of the effects of stakes on startle. Because of the great variation in the number of startle trials per participant, t-tests and correlations seeking to establish group level effects were weighted by the number of blink trials provided by each participant. Significance was measured at $\alpha = .05$ (two tailed) and no outliers were found for any of the analyses when assessed against a criterion of $z > 3.29$ (Tabachnick & Fidell, 2013).

3.4 Results

3.4.1 Behavioural Results

Because shock outcomes were predetermined, the reinforcement learning task as presented to participants was intrinsically unlearnable. Nevertheless, evidence that participants engaged in reinforcement learning could be assessed by comparing the frequency with which participants took actions congruent with the prior outcome (win-stay, lose-shift) compared to actions that were incongruent (win-shift, lose-stay). There was strong evidence of such outcome sensitivity across the 200 trials: $\overline{X}_D$= 20.73, s_D = 23.04, 95% CI [13.72, 27.73], paired t = 5.97, df = 43, p < .001, d = 0.90) Additionally, when queried at each block end as to which fractal in a pair was associated with the smaller amount of shock (ten queries over the experiment, corresponding to five blocks x two stakes), participants were successful for more pairs than not: $\overline{X}_D$= 2.59, s_D = 3.07, 95% CI [1.66, 3.52], paired t = 5.60, df = 43 p < .001, d = .84).

3.4.2 RPE Encoding

Simple grand average waveforms for the four conditions, and the LN – HP difference wave are shown in Figures 6a and 6b, both plotted at C2 where the difference wave was maximal. Figure 6c shows the scalp topography of this difference wave at points where it was significant after cluster randomisation and reveals two significant clusters for the effect.

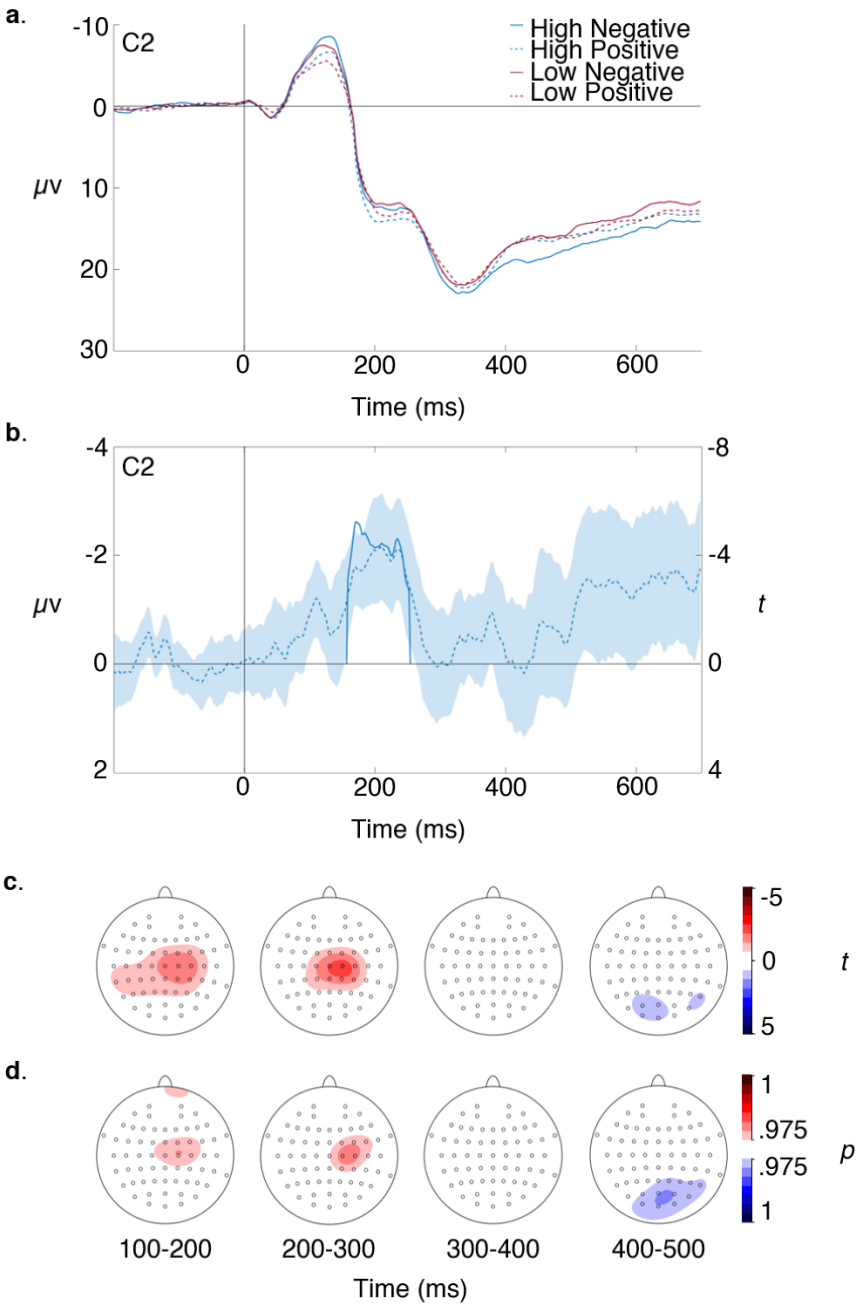

Figure 6. a. Simple waveforms. b. FRN (i.e., LN-HP difference wave) expressed as voltage (dotted line) and thresholded t-statistic over participants (bold line), with shading showing two standard errors of the mean. c. Scalp topography of the thresholded FRN t-statistic showing areas where increased RPE value is encoded with positive voltage (red) and negative voltage (blue). d. Scalp topography of thresholded Bayesian contrasts showing same effects. All plots were created in MATLAB version 2020a (https://uk.mathworks.com/products/matlab.html).

The first occurs in the interval 150 – 260 ms, maximal at C2 at 210 ms (Monte Carlo p = 0.000035), with +RPEs associated with positive voltage, thus corresponding with the conventional FRN. The second,

weaker effect, occurs in the interval 390 – 480 ms, maximal at POz at 418 ms (Monte Carlo p = 0.00053), with +RPEs associated with negative voltage. Figure 6d shows the topography of RPE encoding under Bayesian t-tests, thresholded at .975. This reveals a very similar pattern, with an effect in the interval 150 – 250 ms, maximal at C2, with +RPEs associated with positive voltage, and an effect in the interval 390 – 480, maximal at P8, with +RPEs associated with negative voltage. Figure 7 presents analogous scalp plots for the effect of physical salience.

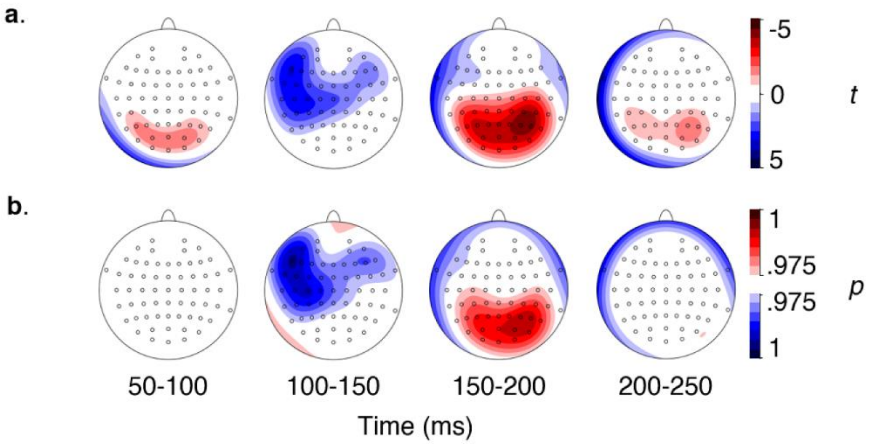

Figure 7. a. Scalp topography of physical salience encoding expressed as the thresholded t-statistic for the difference wave (LP + LN) – (HP + HN): blue indicates encoding of increasing physical salience via voltage negativity, red indicates encoding via voltage positivity. b. Scalp topography of thresholded Bayesian contrasts showing physical salience encoding via the same colour mapping.

3.4.3 Startle Response

Startle responses occurred with similar frequency in high stakes ($\overline{X}$ = 9.14, s = 5.25) and low stakes conditions ($\overline{X}$ = 9.00, s = 4.65), as shown below in Figure 8. The startle response was maximal 120 ms after noise onset. Supplementary Figure 3 shows the time course of the grand average startle response separately for high and low stakes conditions, weighted by the number of startle response trials shown by each participant.

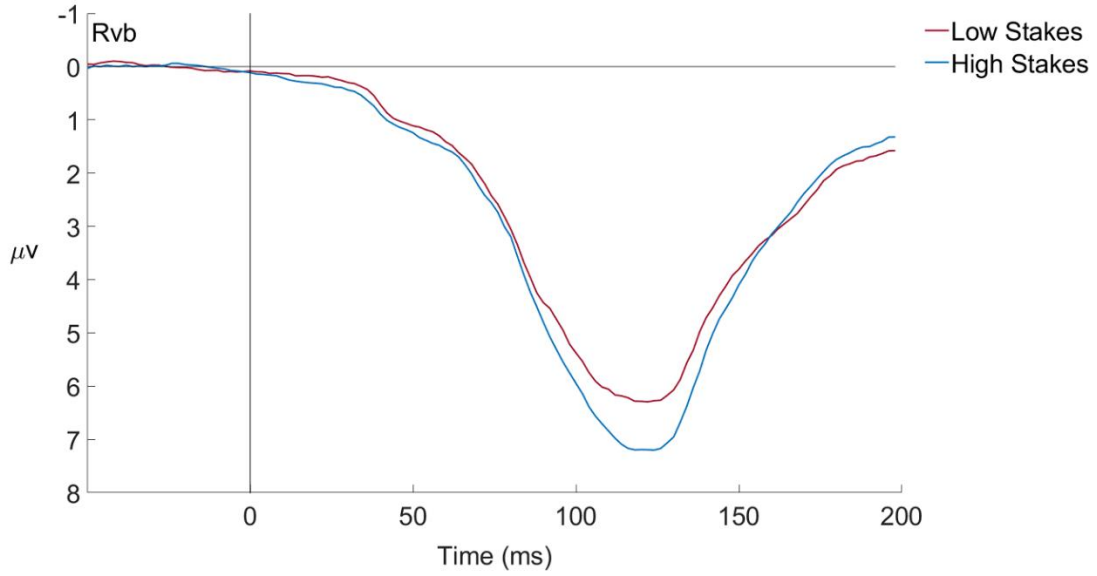

Figure 8. Grand average startle response weighted by the number of startle response trials shown by each participant.

The startle response was of significantly greater positive amplitude for high stakes compared to low: $\overline{X}_D$= 0.91 μV, s_D = 2.89, 95% CI [.01, 1.81], paired t = 2.04, d.f. = 41, p = .048, *d* = 0.31. A correlation of startle response amplitude at 120 ms (collapsing out stakes conditions) and average FRN amplitude, as measured in the interval 150 – 260 ms at C2, showed greater startle response to be significantly associated with greater FRN amplitude (Pearson's *r* = -.37, p = .013, Spearman's rho = -.32, p = .028, N= 44). Very similar values were found when the 120 ms point sample of startle response amplitude was replaced with mean startle response activity in an interval 100–140 ms following noise onset.

3.5 Discussion

3.5.1 Results summary

The present study addressed a dearth of research on event related potentials generated by RPEs associated with unconditioned stimuli. It provided evidence for an RPE-sign component that responded to an unconditioned stimulus with the same polarity and topography as the conventional cue-locked FRN, and which occupied the earlier portion of that component's typical interval. It thus provides provisional evidence that the FRN encodes the value of primary reinforcers and is not elicited simply by arbitrary "truth cues".

Timing differences between the component elicited here and the typical cue-locked FRN were apparent. The component's onset of 150 ms corresponded to that shown by the cue-locked FRN, but it peaked and ended some 70 ms earlier than is typically observed. This is not expected from TD learning, which predicts no difference in the neural response to RPEs carried by unconditioned stimuli and the

conditioned stimuli that cues should develop into over the course of learning. The shortened nature of the component raises the possibility that it shares some, but not all of the neural activity underlying the cue-locked FRN. The principal task difference in the current study was that at no point in the experiment were participants required to maintain in working memory an association between a cue and a primary reinforcer. This contrasts with cue-based experiments. While the cue-reinforcer mapping may be discarded in time as associative learning confers value directly onto the cue (making it a conditioned stimulus), waveforms averaged over the experiment will incorporate a number of trials where this mapping was in effect.

3.5.2 Timing differences

One explanation for the timing difference observed here is that in cue-based experiments the RPE signal is delayed while associative mappings held in working memory are passed to the midbrain dopamine neurons responsible for RPE generation. The fact that the component observed here is topographically similar to the typical cue-locked FRN is consistent with a common process. This is supported by an fMRI study showing that when truth cues indicating success change every trial, preventing any possibility of their acquiring value via associative learning, the midbrain still appears to activate, presumably from cortical input carrying these mappings (McDougle et al., 2020). Such an explanation speaks to the interesting possibility that all RPE computation is obligatorily dependent on a single anatomically defined pathway into which diverse inputs feed. The actual point of confluence of these inputs would remain an open question; while this might be the midbrain, mappings might instead converge in the striatum, which has been proposed to serve a generalised gating role for prediction errors (Den Ouden et al., 2012). This question would be difficult to resolve by the event related potential method which cannot detect midbrain activity and is limited in its access to subcortical structures such as the striatum.

The account above suggests that the delayed peak in the typical cue-locked FRN, relative to that shown in the present experiment, reflects the interposition of a preliminary step prior to RPE-sign encoding. An alternative explanation is that the delayed cue-locked FRN reflects the *addition* of later, functionally separate processes that happens to show the same polarity at frontocentral sites and whose spatial characteristics are unresolvable by the coarse spatial resolution of EEG. The midbrain RPE signal is typically regarded as an example of model-free reinforcement learning, but cue-locked FRN experiments entail a degree of instruction-based learning, in the form of truth cues, which can be regarded as a prime example of goal-directed, or model-based learning (Hammerstrom et al., 2021). Both such signals may contribute to the cue-locked FRN, with one study using principal components analysis to decompose the FRN into separate model-based and model-free RPEs (Sambrook et al., 2018a).

3.5.3 Single-domain limitations

Some limitations should be noted. In the present study, RPEs were studied only in the aversive domain, owing to the difficulty of finding an appetitive unconditioned stimulus that could be delivered with temporal precision. It is thus possible that the shortened latency is specific to the aversive domain. An influential single cell study found that, while neurons responding to juice delivery were the same as those responding to the conditioned stimuli predicting juice delivery, separate cell populations responded for air puff delivery compared to the conditioned stimuli predicting air puff (Matsumoto & Hikosaka, 2009). Furthermore, the -RPE signal carried by phasic reduction in cell firing occurred much earlier in response to air puff compared to conditioned stimuli for air puff (51 ms vs 179 ms).

A further consequence of restricting the design to the aversive domain is that it cannot formally distinguish between coding of RPE and motivational salience. The motivational salience of a reinforcer lies in its ability to elicit attention due to its motivational relevance without regard to its actual valence (Schultz, 2016). Appetitive and aversive events thus produce the same response in a motivational salience encoder to the extent that these events are equally motivationally relevant. The LN-HP difference wave, which controls out effects of physical salience, has been interpreted in this study as indicating RPE-sign encoding. However, the LN outcome constitutes greater motivational salience than the HP outcome since, while both outcomes are equally unexpected ($p = .5$), the LN outcome constitutes expectancy violation via an excess rather than deficit with respect to the expected amount of a motivational good (shock). The significant LN-HP difference wave observed might thus indicate a component coding for motivational salience via voltage negativity rather than RPE-sign via voltage positivity. Distinguishing between the two interpretations requires a design that varies the size of prediction errors for both appetitive and aversive goods. If the negative voltage shift we have observed as an aversive stimulus becomes more intense was also found when an appetitive stimulus becomes more intense, our findings would be interpreted as showing motivational salience encoding. If increasing appetitive stimulus intensity produced a positive shift then the RPE-sign explanation would be retained. In fact, cue-locked experiments that have included this manipulation have provided good evidence that motivational salience *is* encoded in the earlier portion of the interval associated with the FRN (Hird et al., 2018; Soder & Potts, 2017; Stewardson & Sambrook, 2021; Talmi et al., 2013). However, this encoding is carried by a voltage positivity and for this reason we can be confident that this is not the component elicited here. Thus, we can assume the observed component is the same, or related, RPE-sign encoding seen in the previous chapter and other FRN studies, rather than a new salience component of opposite polarity to that usually seen. Nevertheless, it remains the case that this possibility cannot be ruled out.

Lastly, because the number of experiments using shock to generate prediction error is so few, it is not yet clear whether shock has specific effects on the waveform, once its sensory effects are controlled out to isolate the underlying prediction error. Three cue-locked shock studies all reported later FRNs than that shown here, but were themselves highly variable (250 ms (Talmi et al., 2013), 380 ms (Mulligan & Hajcak,

2017) , 420 ms (Heydari and Holroyd, 2015). Task effects are likely to have contributed to this variability; in particular whether participants were active or passive, like we have previously shown in chapter 1.

3.5.4 Affect in RPEs

Using startle response in the context of a conditioned Pavlovian stimulus (the stakes cue), this study found modest evidence of a relationship between an affective physiological response and the expected strength of a forthcoming shock. This accords with other studies showing correlations between expected value (here, negative value) and physiological measures such as startle response and pupil dilation (Pietrock et al., 2019; Tzovara et al., 2018). Additionally, we found that, measured across participants, the average strength of the startle response incurred to noise bursts was correlated with the average amplitude of the FRN produced by instrumentally unrelated electric shocks. This effect was modest, especially when taken in the context of the low power present in between-subjects comparisons and should be viewed with caution. However, it provides provisional evidence, albeit circumstantially, for an affective element in RPE generation. Such a relationship is implied in the phenomenon of Pavlovian-instrumental transfer, in which the vigour of simple instrumental responses is increased and decreased by appetitive and aversive Pavlovian cues unrelated to the stimulus-response relationship. In humans, fMRI studies have suggested the striatum to be an important site of Pavlovian-instrumental transfer (Geurts et al., 2013; Talmi et al., 2008). An intact nucleus accumbens is necessary for startle attenuation (Koch et al., 1996), and this structure is also a major projection site for dopamine neurons implicated in RPE transmission. It appears to serve the critic role in an actor-critic architecture shared with the dorsal striatum and is regarded as a "limbic–motor interface (Cardinal et al., 2002; Mogenson et al., 1980; Mogenson & Yang, 1991), allowing Pavlovian incentive value to modulate instrumental behaviour. RPEs serve learning rather than behaviour however, raising the question of why affect might modulate them. One possibility is that this indicates an architecture that biases learning based on the profitability of the environment (Cazé & van der Meer, 2013), with this profitability carrying an affective character. Indeed, FRN amplitudes appear to be reduced for high anxiety people (Aarts & Pourtois, 2012; Andreatta et al., 2017) suggesting a reluctance to positively adjust expectations in the face of transient positive outcomes, or a seizing of negative outcomes as evidence of poor chances of future success.

3.5.5 Chapter conclusions

To end, the present study elicited a component resembling the FRN in response to an unconditioned stimulus. The finding that an FRN is generated in response to unconditioned as well as conditioned stimuli strengthens the claim that it reflects the transmission of an RPE-sign encoder from midbrain dopamine cells. However, the relatively short latency for the FRN to the unconditioned stimulus used here suggests that the typically observed cue-locked FRN may incorporate neural processes other than those carried by midbrain dopamine neurons.

4. Bayesian Contrasts Analysis of the Reward Prediction Error Literature

This chapter is based on a published paper:

Stewardson, H. J., & Sambrook, T. D. (2022). Valence precedes value in neural encoding of prediction error. Psychophysiology. **(Submitted)**

4.1 Chapter Abstract

Following the previous chapters of this thesis, we have already assessed the question of whether the FRN represents the encoding of a general-purpose RPE as opposed to alternative claims that it instead demonstrates motivational salience (Hird et al., 2018; Soder & Potts, 2017; Talmi et al., 2013). However, there remains a key issue in that the findings we have gathered thus far in support of the RPE-sign account of the FRN, as identified through the response to RPE valence and lack of response to RPE size, are confounded by the potential overlapping of two encoders similarly responsive, or indeed not responsive, to these variables. These encoders, which we have termed RPE-utility and motivational salience, also have strong ties to reinforcement learning and have both been claimed to define the FRN at some point in the literature. Consequently, in this chapter we aim to disambiguate these three encoders more rigorously by identifying the particular effects that each encoder is responsive to and applying a variety of analyses to gather evidence towards their presence throughout the feedback-locked waveform.

4.2 The Confounding of RPE-sign, RPE-utility and Motivational Salience

4.2.1 The three encoders

Before discussing the difficulty with disambiguating the three types of encoding, it is important to reiterate their functions. As stated above, all three encoders have distinct responses, or indeed lack of responses, to RPE valence and RPE size.

First is RPE-sign encoding, which is represented by a response to RPE valence but not to RPE size, such that the receipt of +RPEs generates ERPs more positive in amplitude to those upon the receipt of -RPEs. Crucially, for RPE-sign encoding the amplitude of ERPs should be unaffected by outcome size. For example, the amplitude of the ERP generated after receiving 75p when 25p was the alternative outcome should not differ from the ERP generated after receiving £7.50 when 25p was the alternative outcome, even though the latter two possible outcomes are separated by a much larger monetary value.

Second is motivational salience. This encoder is represented by a response to RPE size but not to RPE valence, being of greater amplitude for both +RPEs and -RPEs, and of smaller amplitude to events that

are neutral. Importantly, the amplitude of ERPs constituting +RPEs, and those constituting -RPEs, should not differ in polarity. For example, if +RPEs of increasing size result in relative positivities in the waveform, then -RPEs of increasing size should also result in such positivities.

Lastly is RPE-utility, which is represented by the fulfilment of the axiomatic model of RPEs (Caplin & Dean, 2008). This means that an RPE-utility encoder not only discriminates RPE valence, but also RPE size, such that ERP amplitudes increase in one direction for increasingly better +RPEs and in the other direction for increasingly worse -RPEs. For example, if the increase in the amount of a monetary reward in an instrumental learning task increases the positivity of the resultant ERP, then increasing the amount of a monetary loss should decrease the positivity of the ERP. In summary, RPE-utility is represented by an interaction of RPE valence and RPE size, increasingly responsive as utility increases.

4.2.2 Component overlap

As detailed above, the three key encoders are represented by their responses, or lack thereof, to effects of RPE valence and RPE size. We have noted previously how in the likely event of overlapping components, the waveform is comprised of the summing and cancelling of these differing responses, revealing a single confounded waveform in place of a clearer picture of the underlying encoders present. For example, the combined presence of an RPE-utility encoder and a motivational salience encoder is expected to adversely affect the resultant waveform observed. The positive amplitudes for increasingly positive outcomes, a predicted effect for both encoders, sum together. Conversely, the negative amplitudes for increasingly negative outcomes, as predicted by RPE-utility, are cancelled out by the positive amplitudes for these same outcomes as predicted by motivational salience. The final output of this summing and cancelling is an apparent RPE-utility encoder that is instead responsive to only +RPEs. While steps can be taken to minimise component overlap in the waveform, such as through the use of principal components analysis (PCA) (Sambrook & Goslin, 2016), there is another feature of the encoders, in addition to their responses to RPE valence and size, that must be considered when attempting to disambiguate them: their *context dependency*.

4.3 The Importance of Context

4.3.1 Context-free and context-dependent encoding

Aside from the confounding of the three encoders with each other, there is also an effect of experimental context that can lead to particular outcomes eliciting different responses for a single encoder. To be specific, experimental context here refers to the manner in which particular outcomes are evaluated by an encoder, with encoders being either *context-free* or *context-dependent*. Context-free encoding is

represented by the valuation of an outcome based on all other available outcomes in an experiment, whereas context-dependent encoding is represented by the valuation of an outcome instead being relative to the range of outcomes provided as alternatives as part of a given trial. Consider an experiment manipulating outcome domain (appetitive or aversive) and outcome size (large or small), whereby there are 4 total outcomes across the course of the experiment: +75p, +25p, -25p, -75p. For a context-free encoder, the expected value (EV) of a trial, that is, the sum of outcomes weighted by their likelihood, is calculated across all outcomes in the experiment, thus always equalling 0. This means that, compared to the EV, the appetitive domain outcomes of +75p and +25p are always +RPEs, and the aversive domain outcomes of -75p and -25p are always -RPEs, regardless of how the outcomes may be paired together across the trials of the experiment. For a context-dependent encoder however, EV is calculated across outcomes of a given trial only. This means that if +75p and +25p are paired together, the +25p outcome constitutes a -RPE as it is worse than the EV of +50p, and the +75p outcome thus constitutes a +RPE. Alternatively, if +25p and -25p are paired together, then the EV now equals 0, and the +25p reward constitutes a +RPE as it is now of greater value than the EV.

4.3.2 The context dependency of RPE-sign, RPE-utility and motivational salience

In nature, having both context-free and context-dependent encoding is advantageous; an individual's ability to quickly process whether one outcome is better than another, while also processing if yet better outcomes are possible in an alternative scenario, helps to optimise decision-making. Considering this notion in respect to the three target encoders of this chapter, RPE-sign most represents context-dependent encoding, with RPE-utility and motivational salience instead matching more closely to context-free encoding. In the literature RPE-sign is typically operationalised as a difference wave of worse outcomes – better outcomes *for each trial*, with RPE valence crucially defined by this trial-by-trial basis. Thus, a context-free RPE-sign response is implausible; what constitutes the worse outcome of one trial in an experiment may not constitute the worse outcome of a consequent trial. Conversely, RPE-utility and motivational salience encoders are operationalised by difference waves constituted of outcomes *across trials*. For example, motivational salience is typically calculated as the difference of the sum of all smaller outcomes – the sum of all larger outcomes, and RPE-utility as a difference of difference waves; the difference of all larger outcomes – the difference of all smaller outcomes. These operationalisations are thus implicative of both encoders being context-free, generating a response considerate of all outcomes in a given experiment. Following this, in the forthcoming sections of this chapter we treat RPE-sign encoding as context-dependent, while treating RPE-utility and motivational salience encoding as context-free.

4.4 Disambiguating the Waveform

Having established that the three target encoders of this thesis are confounded not only by their responses to RPE valence and size, but also by their context-dependencies, the next step is to consider the variables that can be manipulated to allow for their disambiguation based on these effects.

4.4.1 RPE valence

RPE valence is defined categorically by two distinct levels: outcomes that are better than the EV of a trial, and outcomes that are worse. Manipulating valence itself as a study variable however can be problematic; as we discussed earlier the valence of an outcome for a context-dependent encoder is not constant, but rather determined by the alternative outcomes available for each trial. Considering the present study aims to disambiguate such an encoder, RPE-sign, from the context-free encoders of RPE-utility and motivational salience, it is thus implausible to consider manipulating RPE valence directly. As such we instead manipulate *outcome domain*, which is a fully objective variable that is represented by outcomes being appetitive or aversive. For example, while a study using the delivery of money is rewarding (appetitive domain), a study using monetary loss is instead punishing (aversive domain).

4.4.2 RPE size

RPE size is defined continuously, essentially representing the amount of "error" of an RPE. While continuous, for the purposes of the forthcoming sections of this chapter we separate RPE size binarily, manipulated using two levels of *outcome size*: large or small. Outcome size can be constituted by the magnitude of outcomes, such as receiving 75p (large) or 25p (small), or constituted by their delivery or omission, such as receiving a monetary reward (large) or not (small).

4.4.3 Context dependency

4.4.3.1 Experiment Form

Considering that RPE-sign is a context-dependent encoder, with its neural response dependent on what constitutes the worse or better outcomes in each trial of an experiment, it gives us an additional variable with which to more easily disambiguate it from the other two encoders: *experiment form*. Referring again to the experiment example provided in the earlier section "Context-free and context-dependent encoding", the task can be separated into two distinct experiment forms. In one version of the task, the appetitive outcomes may be paired together as the two possible outcomes within trials, such that a participant is aware that they will be rewarded, but do not know if they are going to be rewarded 75p or 25p. Likewise, the aversive outcomes are also paired together, with participants not knowing if they will lose 75p or 25p. We term this the *domain form*; the trials are separated by outcome domain, and thus RPE valence is dependent on the outcome's size. In a different version of the study however, it may be that the outcomes of equal size are paired together within trials, and so participants instead make a mixed gamble: A choice to determine whether that money is won or lost. We term this the *size form*; trials here are instead separated by outcome size, with RPE valence dependent on the outcome's domain. As such, in a domain form study the reward of 25p generates a -RPE, as it is less rewarding than the alternative 75p reward it is compared against in a given trial. In a size form study, the 25p reward instead generates a +RPE, being rewarding rather than punishing like the 25p loss it is compared against.

4.4.3.2 The meta-analysis approach

While experiment form provides an improved means of encoder disambiguation, it has never been considered as a variable in the literature. Following that no studies manipulating experiment form currently exist, it must be noted that its manipulation is only possible using a meta-analytic approach. Despite this limitation, the need to use a meta-analytic approach is far from problematic. Using meta-analyses to investigate the feedback-locked waveform has its own advantages in that past data can be put to further use than originally intended, and that it provides a solution to the inconsistencies of the latency at which encoders have been studied across the literature (Sambrook & Goslin, 2015). In short, experiment form is a necessary variable if we are to better disambiguate RPE-sign, RPE-utility and motivational salience encoding, and can only be manipulated through the use of a meta-analytic study.

4.4.4 Predicting the three encoders

The three variables described above comprise a 2 (outcome domain) x 2 (outcome size) x 2 (experiment form) mixed design, with experiment form as a between-subjects variable used to better disambiguate RPE-sign encoding from that of RPE-utility and motivational salience. To detail the usefulness of experiment form, we provide plots in Figure 9 below to indicate the predicted neural responses of each encoder to each effect of outcome domain, outcome size and experiment form.

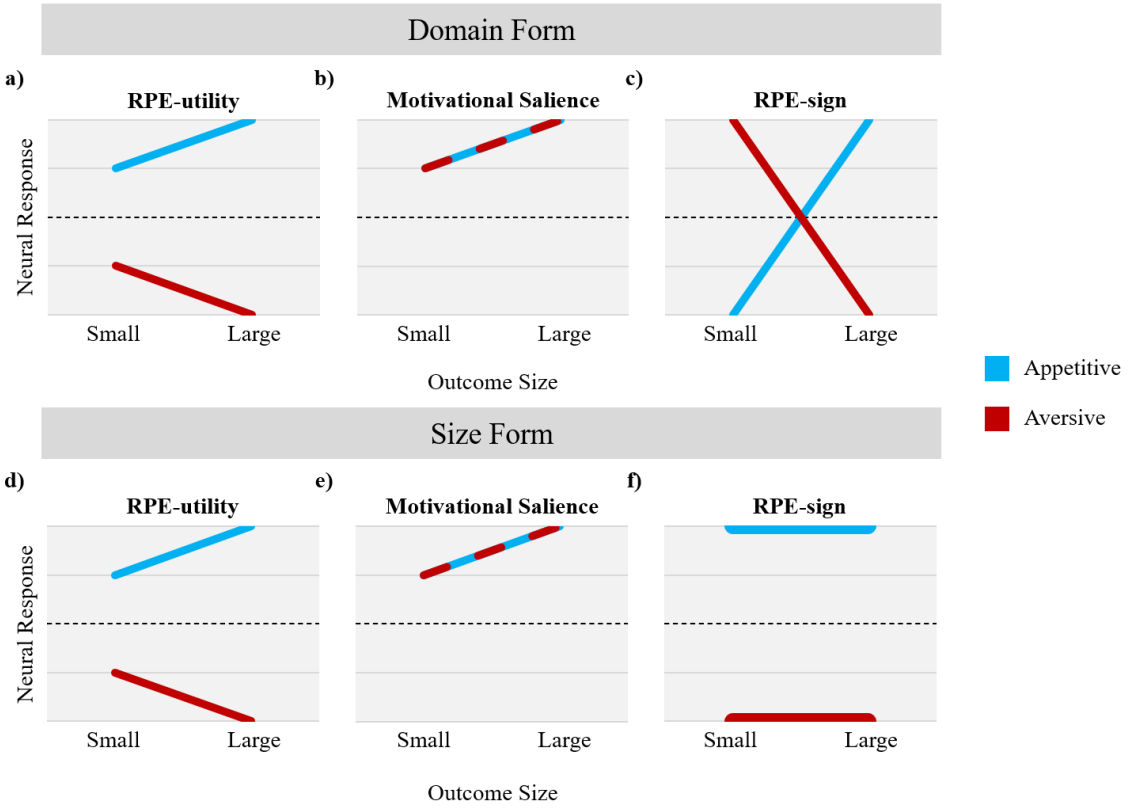

Figure 9. Predicted amplitudes (y-axes) for each encoder (columns) as a consequence of outcome domain (separate lines), outcome size (x-axes) and experiment form (rows).

As seen in Figure 9, an RPE-utility encoder is uniquely responsive to both RPE valence and RPE size. As shown in panes a) and d), the distinct RPE valence*size interaction is evident in the relative increase in positivity for increasingly large appetitive outcomes, and the relative decrease of positivity for increasingly large aversive outcomes. For the motivational salience encoder, shown in panes b) and e), the clear overlaying of the amplitudes for appetitive and aversive outcomes is indicative of its unique lack of response to RPE valence, with the change in amplitude between small and large outcomes indicative of its response to RPE size. Conversely, the identical amplitudes predicted in response to both small and large outcomes for an RPE-sign encoder, shown in panes c) and f), is indicative of its unique lack of response to

RPE size, and instead demonstrates a response to RPE valence evidenced by the oppositely-signed amplitudes for appetitive and aversive outcomes.

There is also an effect of outcome domain for RPE-sign encoding. In domain form studies an RPE-sign encoder is expected to respond with identical positive amplitudes for appetitive large and aversive small outcomes, and with identical negative amplitudes for aversive large and appetitive small outcomes. In size form studies however, RPE-sign encoding is instead expected to show identical positive amplitudes for both appetitive outcomes, and identical negative amplitudes for both aversive outcomes. The RPE-sign response is thus distinct from RPE-utility and motivational salience in different ways for each form, and, when taken together, provides a greater means of disambiguating the encoders than through the manipulation of outcome domain and outcome size only. This is most easily done by comparing panes c) and f) to the remaining four panes.

4.5 Summary

Throughout the literature, the search for a general-purpose RPE is confounded by the overlapping of multiple encoders that are similarly responsive, or indeed unresponsive, to effects of RPE valence, RPE size and their context dependencies. By manipulating outcome domain, outcome size and experiment form in a meta-analytic approach, the present study aims to better disambiguate the presence of RPE-sign, RPE-utility and motivational salience. Using a multitude of analyses, we aim to gather both relative and actual evidence towards the presence of each encoder across each time point in the feedback-locked waveform, collating the results across a pool of studies to gain a comprehensive view of the encoders that comprise decision-making.

4.6 Method

4.6.1 Inclusion and exclusion criteria

Included studies needed to manipulate outcome domain (appetitive or aversive) and outcome size (large or small) using either a domain form or size form experiment, ultimately fitting a 2 x 2 x 2 factorial design. As detailed in the introduction, outcome size could be manipulated through changes in magnitude or deliveries vs omissions. Where more than 2 levels of outcome size were possible, the most extreme levels were chosen in order to maximise the contrast between large and small outcomes. For example, if a study uses monetary rewards and losses of 25p, 50p, 75p and £1, then the 4 chosen outcomes for our analysis would be the rewards and losses of 25p (small) and £1 (large). Additionally, for studies manipulating outcome likelihood, dichotomous outcomes were used. To elaborate, consider a study where, in the appetitive domain, the participant can either win 25p or £1 on each trial. In half of these trials, there is a 25% chance of winning £1 and a 75% chance of winning 25p, with the opposite odds being true for the other half

of the trials. For this study, winning £1 when the chance was 25% would be chosen as the appetitive large outcome as it is the maximal +RPE, thus winning 25p when the chance was 75%, which is the alternative outcome in that same trial, would be chosen as the appetitive small outcome. The maximal -RPE would be losing £1 with a 25% chance and thus the minimal -RPE would be losing 25p with a 75% chance. This then gives 4 chosen outcomes out of the possible 8 from the study. The use of dichotomous outcomes in this manner is necessary for calculating RPE-sign, which as we have previously stated is responsive only to RPE valence on each given trial.

Studies manipulating participant agency were included, with active trial outcomes chosen over passive trial outcomes where possible. Regarding participant samples, studies were not included if participants consisted of an experimental population (such as clinical groups or those deriving from a screening process), unless they included healthy adults as a control group. All studies had to include simple voltage*time waveforms for each cell in the 2 x 2 x 2 design as required, however no statistics for these waves were required. The waveforms needed to be expressed at FCz, Fz, or Cz or, if unspecified, from a pool of frontocentral electrodes. Studies also providing waveforms expressed at Pz, CPz or POz or, if unspecified, from a pool of centroparietal electrodes, had these waveforms included for the purposes of an additional analysis (see the "Resolving component overlap: Frontal – parietal Bayesian contrasts" subsection below). Lastly, the feedback-locked epoch needed to run, at minimum, from -100 to 500 ms.

4.6.2 Search strategies

English language journals and books were searched using the following databases: PsycInfo, PsycArticles, ERIC, PubMed and Web of Science. The search was conducted in Endnote and the results were compiled in Zotero. The search string used was "feedback-related negativity" OR "feedback negativity" OR "reward positivity" OR "FRN" OR "feedback error-related negativity" OR "feedback error related negativity" OR "feedback correct-related negativity" OR "feedback correct related negativity", and covered journal titles, abstracts and keywords. This search string mimicked that of the search string in a previous analysis (Sambrook & Goslin, 2015), following the basis that the majority of RPE research is focussed on the FRN, and yielded 1244 studies. Studies were then checked for their fit with the inclusion and exclusion criteria, with the final sample consisting of 36 studies. Of these 36 studies, 22 had parietal waveforms. A breakdown of the studies included is shown on the next page in Table 2.

Table 2. Studies used for the meta-analysis.

Experiment	*N*	Site (Frontal)	Site (Parietal)	Figure No.
Aarts & Pourtois (2012)	60	FCz	_	3h, l
Bell et al. (2016)	36	Pool	_	5
Broyd et al. (2012)	38	Pool	_	3
Chen et al. (2017)				
Friend condition	14	Fz	Pz	2a
Stranger condition	14	Fz	Pz	2a
Chen et al. (2018)	20	FCz	Pz	2
Ernst & Steinhauser (2018)	34	FCz	Pz	5a
Forder & Dyson (2016)	33	Pool	_	2a
Glienke et al. (2015)	20	Fz	_	4
HajiHosseini et al. (2012)	26	Fz	_	2a, c
Heydari & Holroyd (2016)	30	FCz	_	4a, b
Hird et al. (2018)	20	Pool	_	2
Holroyd et al. (2006)	23	Fz	_	2
Kamarajan et al. (2009)	50	FCz	Pz	4
Kujawa et al. (2013)	22	Pool	Pz	1
Long et al. (2018)	26	FCz	Cz	3a
Luo & Qu (2013)				
Low magnitude	18	FCz	Pz	3a
High magnitude	18	FCz	Pz	3b
Mei et al. (2018a)	32	Pool	Pool	5
Mei et al. (2018b)	57	FCz	CPz	1
Parvaz et al. (2015)	25	FCz	_	2a
Peterburs et al. (2013)	28	FCz	Pz	2
Pfabigan et al. (2011)	20	FCz	Pz	1
Pfabigan et al. (2015)	31	Fz	Pz	2
Sambrook et al. (2012)	66	FCz	Pz	5
Santesso et al. (2012)	29	Fz	_	2
Sato et al. (2005)	18	Fz	Pz	1
Soder & Potts (2018)	50	Pool	_	2
Talmi et al. (2013)	20	Pool	_	4
Watts & Bernat (2018)	132	Pool	_	4
Wei et al. (2018)	22	FCz	Pz	4
Wu & Zhou (2009)				
Low expectancy	16	FCz	Pz	1
High expectancy	16	FCz	Pz	1
Yu & Zhou (2006a)	20	Fz	Pz	1
Yu & Zhou (2006b)	20	Fz	Pz	1
Zheng et al. (2017)	36	FCz	Pz	6

4.6.3 Coding procedures

Electronic copies of experiments were accessed, and screenshots were taken of grand average waveforms. Where multiple frontocentral waveforms were available they were taken in preference order FCz, Fz, Cz; where multiple parietal waveforms were available, they were taken in preference order Pz, CPz, POz. Screenshots were cropped and enlarged, and the waveforms digitised with Plot Digitizer (http://sourceforge.net/projects/plotdigitizer) by using a mouse to manually lay points along the waveform at approximately 5 ms intervals. A custom Excel macro (available as a supplementary file in Sambrook and Goslin, 2015) linearly interpolated coordinates at 2 ms intervals between the existing manually assigned ones. The digitizing process was performed twice for each waveform and a mean taken in order to provide greater accuracy. Waveforms were then replotted to make sure that they corresponded visually with the original. Further details can be found in Sambrook and Goslin (2015).

4.6.4 Primary analysis

4.6.4.1 The limitation of single effects as a means of identifying encoders

Within the RPE literature, studies typically use standard operationalisations to identify effects that are representative of particular types of encoding. For example, the operationalisation of RPE-utility, usually calculated as a difference of difference waves (the difference of appetitive and aversive small outcomes subtracted from the difference of appetitive and aversive large outcomes), is used to represent the necessary interaction of RPE valence and size required in an RPE-utility response. However, such operationalisations only show a single effect, thus ignoring any effects which may be more supportive of alternative encoders. To exemplify, consider the neural responses predicted by an RPE-utility encoder and an RPE-sign encoder for domain form studies, shown in Figure 9 panes a) and c) respectively. The domain*size effect, typically indicative of RPE-utility encoding, is also expected for the RPE-sign encoder, and so looking at this effect alone is not enough to disambiguate them.

To improve the disambiguation of encoders, we instead consider how the data at each time point fits with each contrast across the 2 x 2 x 2 design shown in Figure 9. While encoders share some contrasts, with effects then being expected for multiple encoders such as the aforementioned domain x size effect, the overall pattern across *all* contrasts for each encoder is unique. Thus, in the present study we exhaustively test across the 6 possible contrasts for each of the encoders, with evidence for each encoder at a given time point defined by the degree to which the waveform conforms to the 6 predicted contrasts. The 6 contrasts for each encoder (including a null encoder) are shown below in Table 3, which shows *identical* information to the data points in Figure 9 and has been provided to more clearly specify the predicted pair-wise contrasts for each encoder. Note that voltage has been quantified as positivity rather than negativity. For example, the top left cell predicts that the voltage for appetitive outcomes *will* be more positive than that for aversive

outcomes for the RPE-utility encoder, represented by a tick for the "App > Avs" effect. This quantification is arbitrary but was chosen pragmatically due to increases in RPE-utility, increases in motivational salience and negative-to-positive shifts in RPE-sign being typically coded by positivity in the literature.

Table 3. Predictive contrasts for each encoder, including a null encoder (columns), across experiment forms (rows).

Experiment Form	Effect	Encoding			
		RPE-utility	Motivational Salience	RPE-sign	Null
Domain Form	App > Avs	✓	✗	✗	✗
	Large > Small	✗	✓	✗	✗
	Interaction	✓	✗	✓	✗
Size Form	App > Avs	✓	✗	✓	✗
	Large > Small	✗	✓	✗	✗
	Interaction	✓	✗	✗	✗

4.6.4.2 Exhaustive Bayesian contrasts

Our analysis first involved gathering grand averages for each cell in the 2 x 2 x 2 design (appetitive large, appetitive small, aversive large, aversive small, for both domain and size form studies). All averages covered the interval of -100 – 500 ms common throughout the literature. Bayesian one-sample *t* tests were then conducted using the BayesFactor package in R (Morey & Rouder, 2015) to create Bayes factors representing the likelihood of each predicted effect relative to the two alternative effects shown in Table 3. For example, a predicted contrast of appetitive > aversive is compared to the alternative contrasts of appetitive = aversive and appetitive < aversive. The Bayes factors are converted from ratios into probabilities using the equation 1/(1+(1/BF)), where BF is the provided Bayes factor for that contrast. These probabilities comprise the evidence for the predicted effect being present. Importantly, the reason we applied Bayesian tests over non-Bayes equivalents is so that the presence of null effects, which are sometimes predicted by a pair-wise contrast, can also be considered when generating the probabilities for each encoder. For example, no discrimination between appetitive and aversive outcomes is predicted for RPE-sign encoding in domain form studies. The product of the probabilities for each of the 6 effects predicted by a given encoder is then calculated, producing a single compound probability for the presence of that encoder. The compounding of so many terms results in very small probabilities, and so we scale them relative to each other; dividing the compound probability of each encoder by the sum of the compound probabilities for all 4 encoders, leaving

us with a single *relative* probability for each encoder. This process is performed at each time point on the waveform, allowing for the evidence for each encoder across the waveform to be plotted. By plotting the relative probabilities for each encoder in this manner, we are then able to address which encoder is most likely to be present in the waveform at each time point.

4.6.4.3 Resolving component overlap: PCA

Component overlap is likely present throughout the waveform, with the relative probabilities calculated in the exhaustive Bayesian contrasts detailed above potentially constituted from the summing and cancelling of effects from multiple encoders (Sambrook & Goslin, 2014). As we previously exemplified, the presence of an RPE-utility encoder could be impacted by its summing with an overlain motivational salience encoder such that the waveform wrongly suggests it to be responsive to only +RPEs. This is just one of many such confusing aggregates of components that are possible, thus we attempted to remove the effects of component overlap by conducting temporal PCA.

PCA was performed on the grand average difference waves using the ERP PCA Toolkit Version 2.95 (Dien, 2010a) using identical means to those used by Sambrook and Goslin (2016) and following published guidelines (Dien, 2010a; Dien et al., 2005; Dien et al., 2007). Three difference waves, corresponding to the domain effect, size effect, and domain*size effect, were entered for each study. This yielded 108 observations (36 studies x 3 difference waves). Factors were retained if they explained more variance than a factor extracted from a null dataset, i.e., they passed a parallel test (Horn, 1965), and these were subjected to Promax rotation. Factors explaining very little variance in the original data, or showing no clear time course, were discarded. Remaining factors were then subjected to condition wise analysis. Factor scores for each study x difference wave combination were used as the dependent variable instead of voltages, but otherwise the exhaustive Bayesian analysis was conducted in the same way as before. For the purposes of visualisation, factors were reconstructed into voltage denominated waveforms using the product of the factor pattern matrix and the standard deviations (Dien 2010a). These could then be visually interpreted in the same manner as the original grand average difference waves.

4.6.4.4 Resolving component overlap: Frontal – parietal Bayesian contrasts

Based on the literature, encoding of motivational salience in particular is commonplace in the parietal waveform (Bellebaum, Polezzi, et al., 2010; R. Gu et al., 2011; Kreussel et al., 2012; Toyomaki & Murohashi, 2005; Yeung & Sanfey, 2004), and it is possible that this activity may be traceable in the frontal waveform. Thus, as a further test for component overlap specific to this possibility, we conducted a Bayesian contrasts analysis on a waveform comprised of frontal voltage – parietal voltage. By differencing the scalp sites in this manner, we hoped to subtract out any motivational salience encoding associated parietally from

mediofrontal encoding associated with RPE processing. This process is considerably coarser than the PCA but addresses the specific known overlap of parietal motivational salience encoding.

4.6.5 Supporting analyses

4.6.5.1 Bayesian repeated-measures ANOVAs

As detailed earlier, our use of relative probabilities produced from the exhaustive Bayesian contrasts provides more robust evidence for the presence of each of the three encoders across the waveform compared to the typical operationalisation of encoders through main effects and interactions seen more commonly throughout the literature. However, the relative evidence for a given encoder does not itself provide sufficient evidence that any of the three encoders is actually present, and so we conducted additional tests to assess the *actual* evidence for the predicted effects with which to compare against the relative evidence provided by the exhaustive Bayesian contrasts. As such we conducted a Bayesian 2 x 2 x 2 repeated-measures ANOVA at a series of time points in the waveform in order to check that specific effects predicted by encoders are present at the time points associated most with that encoding in the exhaustive Bayesian contrasts. Since no implementation of Bayesian ANOVA is currently available in R, running values across the waveform could not be easily calculated. Instead, the JASP spreadsheet-based statistical software (JASP Team, 2020) was used to conduct Bayesian repeated-measures ANOVAs at intervals of 40 ms, starting from 120 ms post-feedback. Time points prior to 120 ms were not analysed as the interval of interest for RPE processing typically begins at around 200 ms in the literature, and this reduced the number of ANOVAs that were needed. Table 4 below has been provided to show these predicted effects for each encoder.

Table 4. Main effects and interactions (rows) for each encoder (columns). For each cell, a tick represents the corresponding effect being predicted for the corresponding encoder, and a cross represents the null being predicted.

Effect	Encoding			
	RPE-utility	Motivational Salience	RPE-sign	Null
Domain	✓	✗	✓	✗
Size	✗	✓	✗	✗
Form	✗	✗	✗	✗
Domain*Size	✓	✗	✓	✗
Domain*Form	✗	✗	✗	✗
Size*Form	✗	✗	✗	✗
Domain*Size*Form	✗	✗	✓	✗

Note that while the inclusion of the Bayesian repeated-measures ANOVAs allows us a means to sanity-check against the relative evidence of the exhaustive Bayesian contrasts, inconsistencies are to be expected. Specifically, actual evidence for an effect may be present in the ANOVAs but not sufficient to provide relative evidence for an encoder in its own right, as the remaining contrasts predicted for that encoder may not be fulfilled. For example, a strong effect of size present at a given time point does not guarantee that the exhaustive Bayesian contrasts will show motivational salience to be the most dominant encoding at that same time point if the evidence for an RPE-sign encoder is greater because of an even stronger effect of domain. In summary, the exhaustive Bayesian contrasts make use of *all* contrasts, including those that are null, to more accurately predict the presence of encoders. These thus comprise the primary evidence, with the ANOVAs being supportive.

4.6.5.2 Standard operationalisations

We also present actual evidence for each encoder in the form of standard operationalisations used in the literature. While they are not as informative as the exhaustive Bayesian contrasts or the Bayesian repeated-measures ANOVAs, the benefit of including standard operationalisations is that they can be directly compared to the vast range of similar operationalisations calculated throughout the literature. As such, our study not only has additional means with which to assess the accuracy of the relative probabilities provided by the exhaustive Bayesian contrasts, but also to assess its accuracy against the effects in the literature from which it is comprised. The evidence for RPE-utility was based on the domain*size interaction calculated by the difference wave ((aversive large – appetitive large) – (aversive small – appetitive small)), corresponding

to the "RPE-FRN" (Sambrook and Goslin, 2015). Evidence for motivational salience was based on the size effect, calculated by the difference wave ((aversive small + appetitive small) – (aversive large + appetitive large)). Lastly, the RPE-sign effect was calculated by the difference wave of the better outcomes vs the worse outcomes, on each trial.

4.7 Results

4.7.1 Primary analysis

4.7.1.1 Exhaustive Bayesian contrasts

The relative probabilities (*rp*), following Bayesian one-sample *t* tests, for each encoder are shown in Figure 10. Four time points of interest can be seen, the first suggesting strong evidence for motivational salience encoding at 160 ms with an *rp* = .93. Secondly, there is modest evidence for RPE-sign encoding at 200 ms with an *rp* = .75. Thirdly, there is strong evidence for RPE-utility encoding at 320 ms with an *rp* = .93. Finally, there is very strong evidence for motivational salience encoding once again from 360 ms onwards, with *rp* > .975 at all samples.

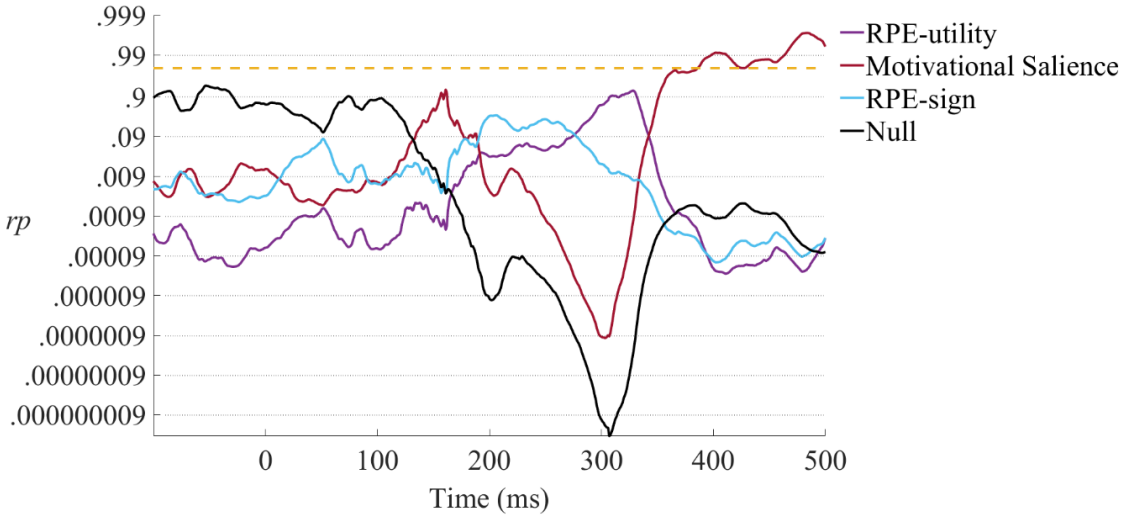

Figure 10. Relative probabilities for the presence of each encoder, including a null encoder, for frontal waves across all time points. A logarithmic y-axis is used. The yellow dotted line shows rp = .975.

4.7.1.2 PCA

Temporal components were subjected to Bayesian one-sample *t* tests using the same method as above, but using factor scores (per study for each of the cells in the 2 x 2 x 2 design) instead of voltage at each time point. Following a scree test, 5 factors were identified by the PCA. Factors 4 and 5 were weakly defined, with only ~3% of variance in the waveform explained by each component. Furthermore, they showed an oscillatory characteristic suggesting a lack of complete decomposition, thus only factors 1 – 3 are analysed. Figure 11 below shows these three factors and their sum, including the grand average ERP of the original data across all conditions. The three factors constitute 91% of the original data, as seen by the fit of the summed factors to the grand average ERP. Additionally, Table 5 shows the summaries for factors 1 – 3 collapsed across conditions.

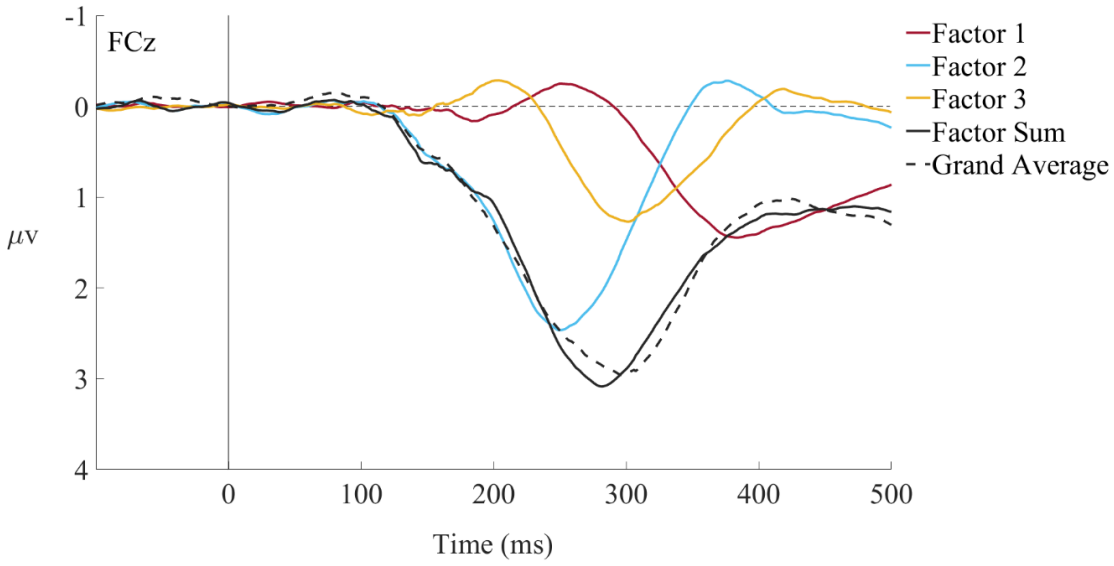

Figure 11. Factors 1 – 3 extracted from voltage waveforms using PCA, overlain with their sum and the grand average ERP across all conditions.

Table 5. Factor summaries for the 3 factors identified by the PCA.

Factor	Temporal peak (ms)	Variance explained (%)	Relative probability (*rp*)
TF1	384	.40	Motivational Salience (.99)
TF2	250	.40	RPE-sign (.74)
TF3	302	.11	RPE-utility (.73)

Factor 1 is most representative of the motivational salience encoder shown by the Bayesian contrasts in the non-PCA data after 360 ms (*rp* = .99). This can also be seen in Figure 12, where there is a clear effect of size, but not an effect of domain or domain*size. This combination of effects are to be expected of a motivational salience encoder based on the predictive amplitudes shown in Figure 9 and the predictive contrasts shown in Table 3.

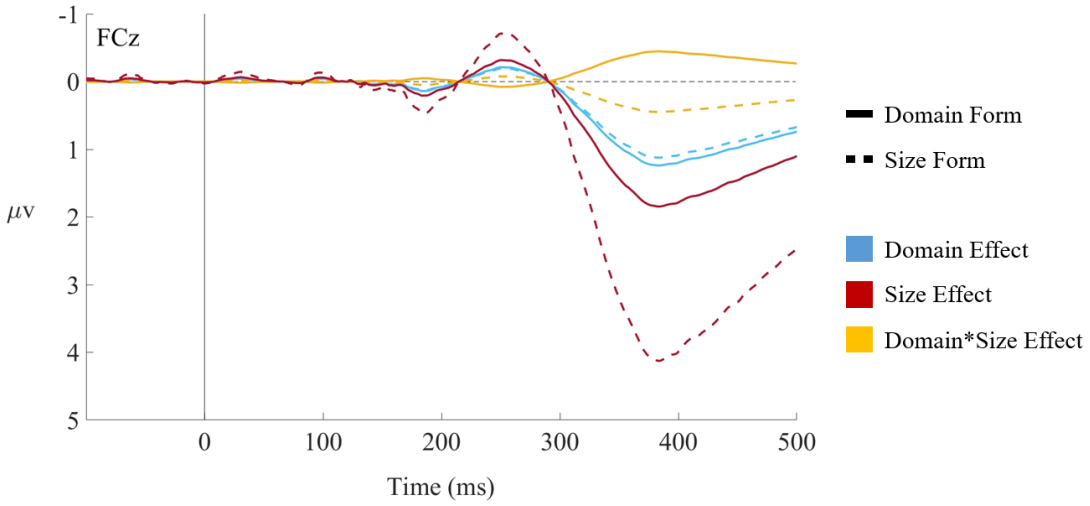

*Figure 12. Factor 1 extracted from voltage waveforms using PCA , separated by effects of domain, size, domain*size, and experiment form.*

Factor 2 is most representative of the RPE-sign encoder shown by the Bayesian contrasts in the non-PCA data at 200 ms (*rp* = .74). This can also be seen in Figure 13, whereby there is an effect of domain for size form studies and not for domain form studies. Additionally, there is an effect of domain*size for domain form studies and not for size form studies. This combination of effects is expected of an RPE-sign encoder as shown predictive amplitudes shown in Figure 9 and the predictive contrasts shown in Table 3. However, there is also a strong effect of size for domain form studies, which is not expected for RPE-sign encoding.

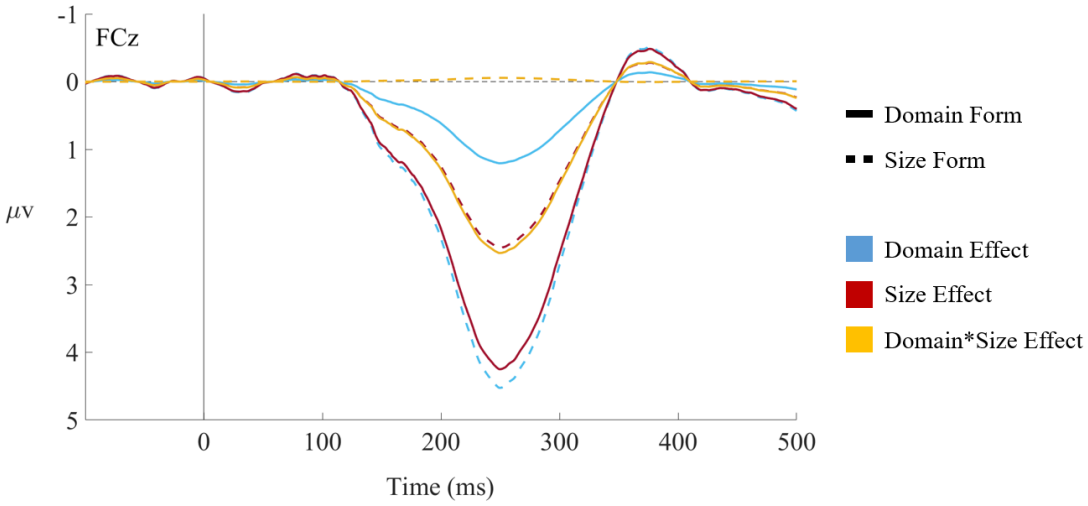

*Figure 13. Factor 2 extracted from voltage waveforms using PCA, separated by effects of domain, size, domain*size, and experiment form.*

Factor 3 is most representative of the RPE-utility encoder shown by the Bayesian contrasts in the non-PCA data at 320 ms (*rp* = .73). This can also be seen in Figure 14, where there is a distinctive effect of domain*size for both domain form and size form studies. Additionally, there is an effect of domain for size form studies. Such effects are expected of an RPE-utility encoder as shown predictive amplitudes shown in Figure 9 and the predictive contrasts shown in Table 3. However, there are also effects present that are not expected of RPE-utility encoding; an effect of size is present for both domain form and size form studies, and the domain effect for domain form studies is of a notably reduced positivity to that of size form studies.

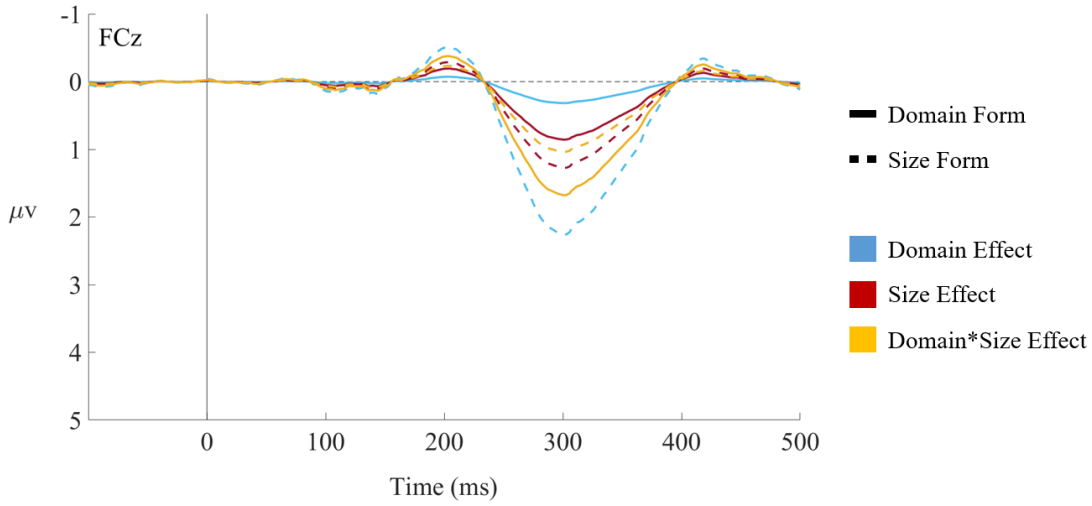

*Figure 14. Factor 3 extracted from voltage waveforms using PCA, separated by effects of domain, size, domain*size, and experiment form.*

4.7.1.3 Frontal – parietal Bayesian contrasts

As we detailed earlier, encoding of motivational salience in particular is commonplace in the parietal feedback-locked waveform, and it is possible that this activity may confound RPE encoding in the frontal waveform. Thus, we also aimed to remove possible component overlap by conducting Bayesian contrasts on a waveform comprised of frontal voltage minus parietal voltage, isolating the frontal effects. By differencing the waves in this manner, we hoped to subtract out motivational salience encoders arising parietally from frontocentral encoders associated with RPE processing. Figure 15 shows the relative probabilities of each encoder in this difference wave, following Bayesian one-sample *t* tests across all time points.

Results show that after 360 ms, the evidence for motivational salience encoding that was originally present in the exhaustive Bayesian contrasts is now greatly diminished ($rp < .1$ at all samples), indicating that this encoding is better considered as parietal. However, throughout the waveform the frontal-parietal Bayesian contrasts are inconclusive. Overall, we cannot conclude with any certainty that the evidence shown in the exhaustive Bayesian contrasts in Figure 10 are complicated by parietal component overlap or not.

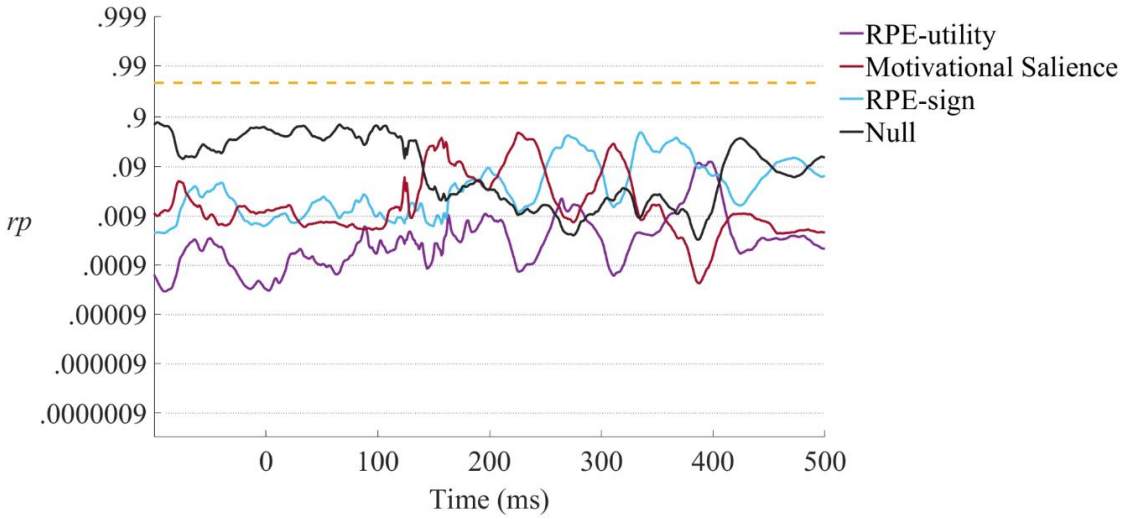

Figure 15. Relative probabilities for the presence of each encoder, including a null encoder, for frontal-parietal difference waves across all time points. A logarithmic y-axis is used. The yellow dotted line shows rp = .975.

4.7.2 Supporting analyses

4.7.2.1 Bayesian repeated-measures ANOVAs

Figure 16 shows the Bayes factors for each main effect and interaction at each 40 ms time interval. Firstly, we expected to see an effect of size at around 160 ms post-feedback, consistent with the motivational salience encoding suggested to be present at this interval by the exhaustive Bayesian contrasts in Figure 10. A size effect was predicted for this encoding based on the expected effects predictive of motivational salience detailed in the second column of Table 4. Following the Bayesian ANOVA at 160 ms, this expectation was fulfilled, with a main effect of size (BF = 104.62) found at 160 ms.

Returning to the expected effects of encoders in Table 4, we expected to see an effect of domain, an interaction of domain*size, and an interaction of domain*size*form at around 280 ms post-feedback, consistent with the RPE-sign encoding suggested at this interval by the exhaustive Bayesian contrasts. The Bayesian ANOVA at 280 ms revealed a main effect of domain (BF = 6305.26) and an interaction of domain*size (BF = 14.76), but no interaction of domain*size*form (BF = 3.10), only partly supporting expectations. Additionally, a main effect of size was found (BF = 3159.75), consistent instead with motivational salience encoding. Overall, the ANOVA generally supports the exhaustive Bayesian contrasts, with the observed effects representing RPE-sign more so than any other encoding at this interval.

Lastly, we expected to see an effect of size at around 360 ms, consistent with the motivational salience encoding suggested to be present at this interval by the exhaustive Bayesian contrasts. Once again, this effect was predicted based on the expected effects predictive of motivational salience encoding detailed

in the second column of Table 4. A Bayesian ANOVA at 360 ms revealed a main effect of size (BF = 163319.88), supporting this expectation.

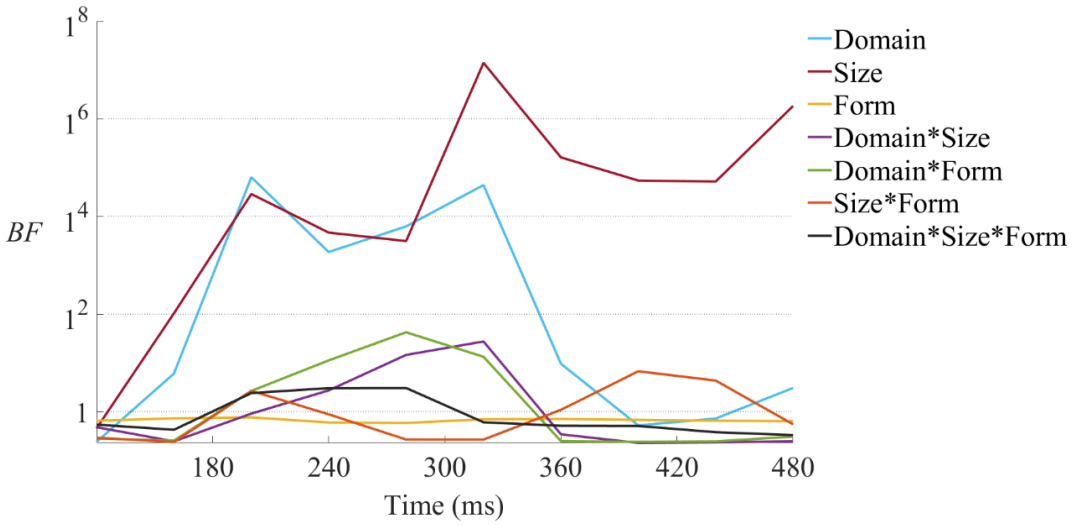

Figure 16. Bayes factors for each main effect and interaction across selected time points.

4.7.2.2 Standard operationalisations

Additional to the Bayesian repeated-measures ANOVAs, we also calculated standard difference wave operationalisations as typically seen in the literature. These effectively constitute piecemeal effects taken from the ANOVAs and tested using conventional null hypothesis significance testing. In order to be consistent with the exhaustive Bayesian contrasts, we first expected to see an effect of domain around 280 ms for size form studies (Figure 9f), consistent with the RPE-sign encoding suggested to be present at this interval.

As seen in Figure 17, a domain effect, constituted of the sum of aversive outcomes – the sum of appetitive outcomes, was found around 280 ms for size form studies. One-sample *t* tests conducted across the whole waveform revealed this effect to be significant from 153 – 369 ms, with a peak significance at 307 ms ($t(19) = -7.08$, $p < .001$). For domain form studies the effect was not significant at this time point. Thus, the standard operationalisation of the domain effect is consistent with the relative evidence for RPE-sign as shown by the exhaustive Bayesian contrasts.

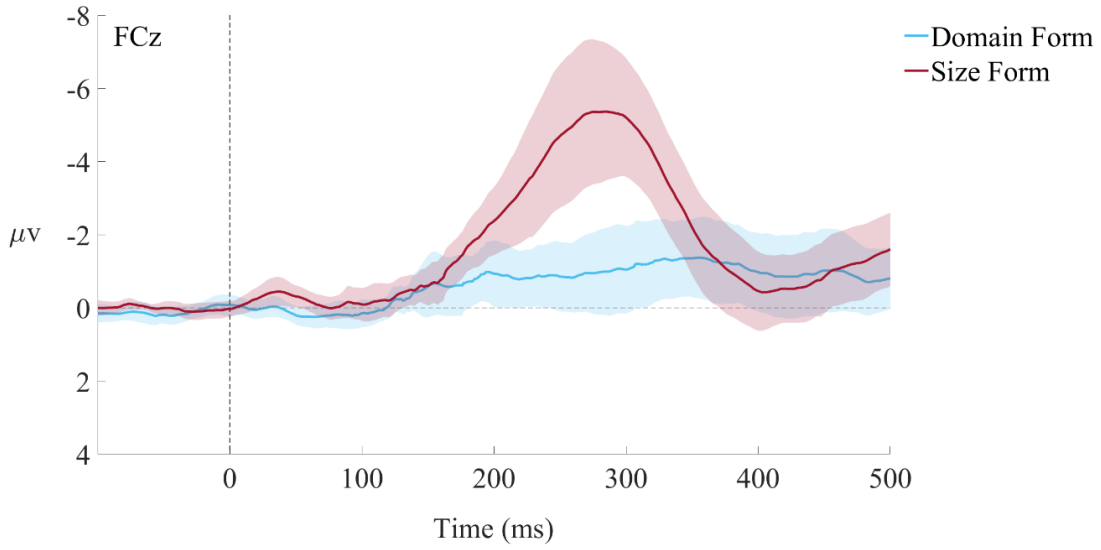

Figure 17. Effect of domain by experiment form, expressed as a great grand average voltage (sum of aversive outcomes – sum of appetitive outcomes) with shading showing two standard errors of the mean.

We next expected to see effects of size for both domain form and size form studies at around 160 ms and from 360 ms onwards (Figure 9b and 9e), consistent with the relative evidence for motivational salience encoding at these intervals suggested by the exhaustive Bayesian contrasts. Furthermore, we expected no difference between the effect of size for domain form studies and that of size form studies, as motivational salience encoding is context-free and thus is unaffected by experiment form.

As seen in Figure 18, an effect of size, constituted of the sum of small outcomes – the sum of large outcomes, was found around 360 ms for size form studies. One-sample *t* tests conducted across the whole waveform revealed this effect to be significant from 129 – 500 ms, with a peak significance at 401 ms ($t(19) = -4.41$, $p < .001$). This effect was also found for domain form studies, significant from 170 – 367 ms and peaking at 195 ms ($t(15) = -5.03$, $p < .001$). While unexpected differences for the size effect over experiment form are present, thus demonstrating potential component overlap, the size effects around 360 ms are strongly consistent with the relative evidence for motivational salience as shown by the exhaustive Bayesian contrasts.

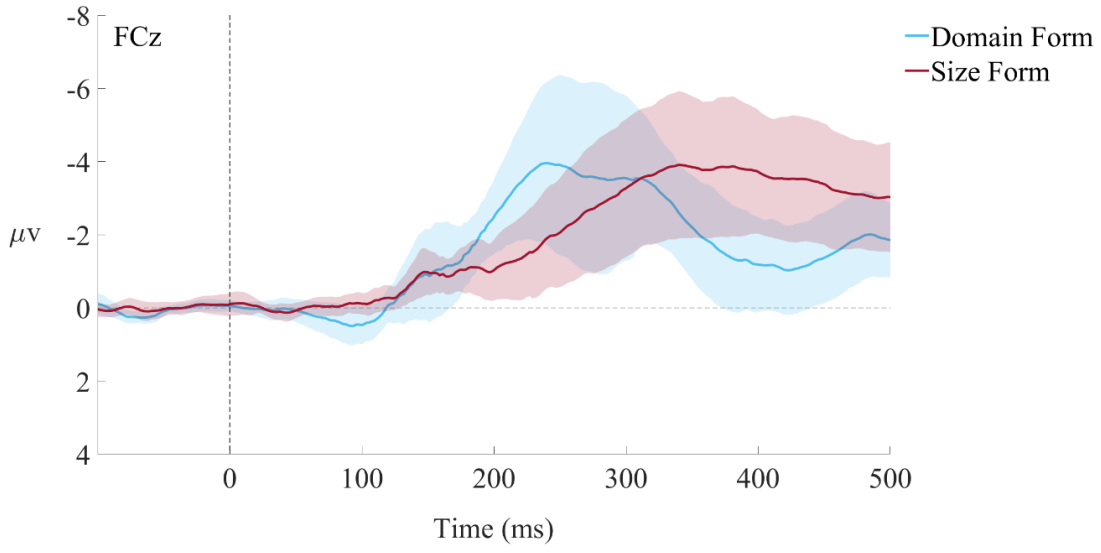

Figure 18. Effect of size by experiment form, expressed as a great grand average voltage (sum of small outcomes – sum of large outcomes) with shading showing two standard errors of the mean.

We expected to see an effect of domain*size for both domain form and size form studies at around 320 ms (Figure 9a and 9d), in line with where the relative evidence for RPE-utility encoding is strongest in the exhaustive Bayesian contrasts (*rp* = .93). We additionally expected to see this effect earlier for domain form studies at around 280 ms (Figure 9c), consistent with the relative evidence for RPE-sign encoding at this interval.

As seen in Figure 19, a domain*size effect, constituted of the difference between large outcomes – the difference between small outcomes, was found for domain form studies around 320 ms. One-sample *t* tests conducted across the whole waveform revealed the effect to be significant in the range 186 – 335 ms, peaking at 307 ms ($t(15) = -4.93$, $p < .001$). This effect was also found for size form studies, significant from 283 – 375 ms and with a peak significance at 312 ms ($t(19) = -3.01$, $p = .0072$). The significant and temporally similar domain*size effects for both domain form and size form studies seen at around 320 ms is suggestive of RPE-utility encoding, consistent with the interval in which RPE-utility evidence is strongest in the exhaustive Bayesian contrasts in Figure 10 and where the domain*size effect is strongest in the Bayesian repeated-measures ANOVAs in Figure 16. Such an encoder has been found at 320 ms using PCA (Sambrook & Goslin, 2016), occurring at a slightly greater latency than RPE-sign encoding.

Nevertheless, the domain*size effect for size form studies was found to be significantly smaller than that of domain form studies in the interval 283 – 315 ms, with the difference having a peak significance at 283 ms ($t = -3.82$, $p = .001$), which is not predicted for an RPE-utility encoder. Instead, this pattern is more indicative of RPE-sign encoding, thus supporting the evidence for RPE-sign encoding around this interval in the exhaustive Bayesian contrasts. We must note however that the smaller domain*size effect for size form

studies was still significant, which is not predicted for RPE-sign encoding, and so this suggests the possible presence of an overlapping RPE-utility encoder.

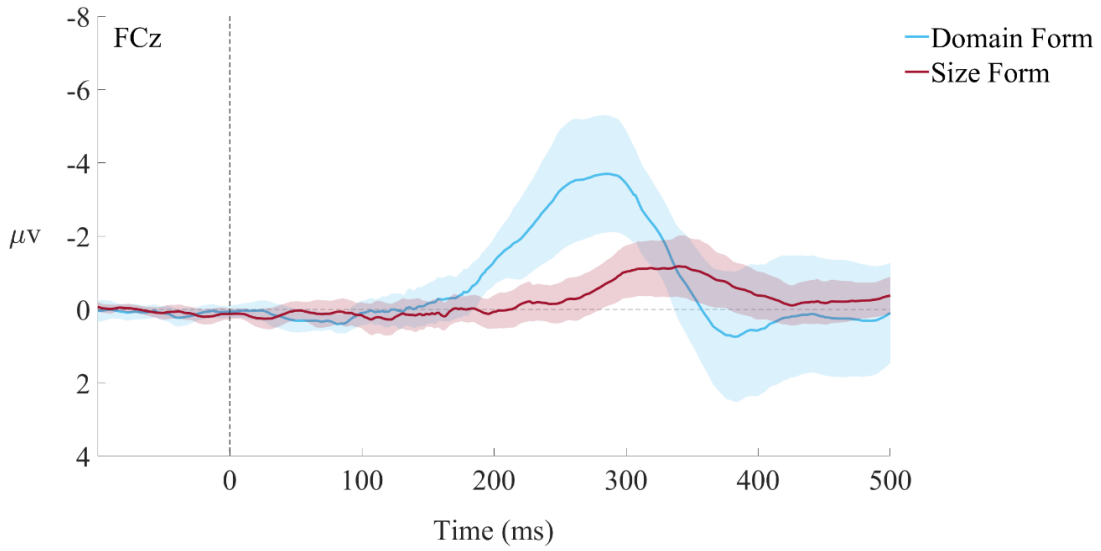

*Figure 19. Effect of domain*size by experiment form, expressed as a great grand average voltage (Large outcomes difference wave – small outcomes difference wave) with shading showing two standard errors of the mean.*

4.8 Additional Results

The results provided above demonstrate the possibility of an RPE-sign encoder at around 280 ms post-feedback. In line with the RPE literature, this encoder should be represented by an FRN, whereby worse outcomes generate a relatively more negative neural response compared to better outcomes regardless of outcome domain. As such, the RPE-sign response we observed in the previous analyses should be represented by an equal FRN for both appetitive and aversive domains. Figure 20 shows this comparison of the appetitive and aversive FRNs. Note that this test is only possible for studies where trials are separated by outcome domain, and as such only domain form studies are included.

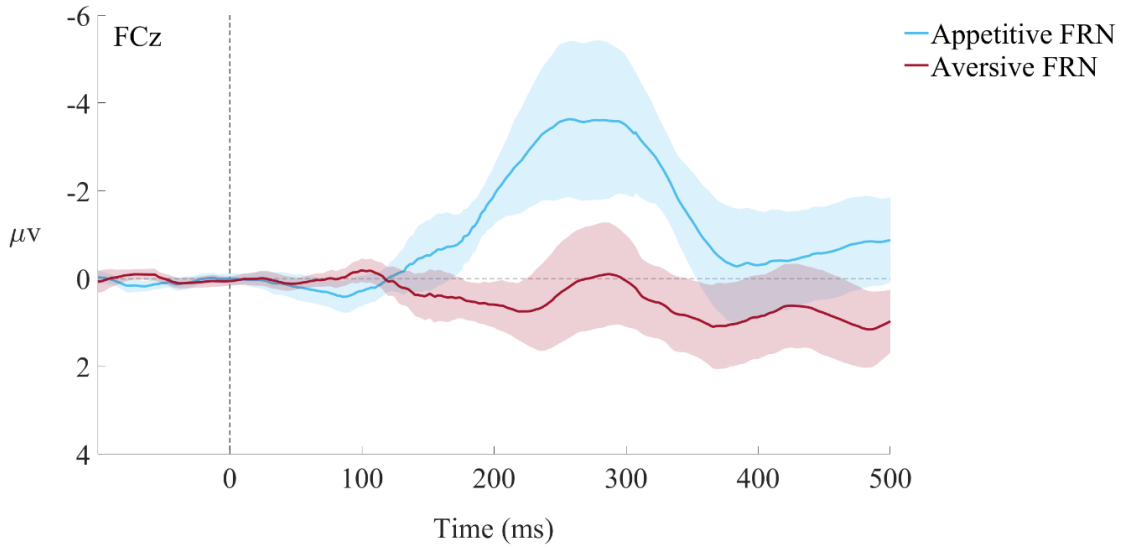

Figure 20. The FRN (worse outcomes – better outcomes on each trial) for domain form studies, separated by outcome domain. Shading shows two standard errors of the mean.

It can be seen that counter to expectation, the aversive domain FRN is greatly diminished compared to the appetitive domain FRN. At 280 ms, a repeated measures *t* test reveals a strong significant difference between the appetitive and aversive FRNs (t(15) = -2.70, p = .016), supporting the case that the aversive FRN is in fact reduced at the time interval associated with the RPE-sign encoding suggested by our exhaustive Bayesian contrasts in the previous section.

The presence of this reduced aversive FRN is problematic as it counters the possibility of the RPE-sign encoding being general-purpose. As such, the forthcoming subsections investigate two possible explanations for the effect; overlapping motivational salience encoding, and the possibility of the RPE-sign encoder being neglectful of aversive domains. We note that as there is no further need to manipulate experiment form (we have already disambiguated the RPE-sign response from RPE-utility and motivational salience), the following analyses manipulate RPE valence directly.

4.8.1 Explaining the reduced aversive FRN

4.8.1.1 Overlapping motivational salience

Firstly, the reduced aversive FRN may be due to a temporally overlapping motivational salience encoder; a highly likely case given the effect of size at 280 ms in the Bayesian repeated-measures ANOVAs, the separate motivational salience component found following PCA, and from the findings of our past studies in Chapters 2 and 3. The hypothetical plot in Figure 21 demonstrates this effect of motivational salience overlap.

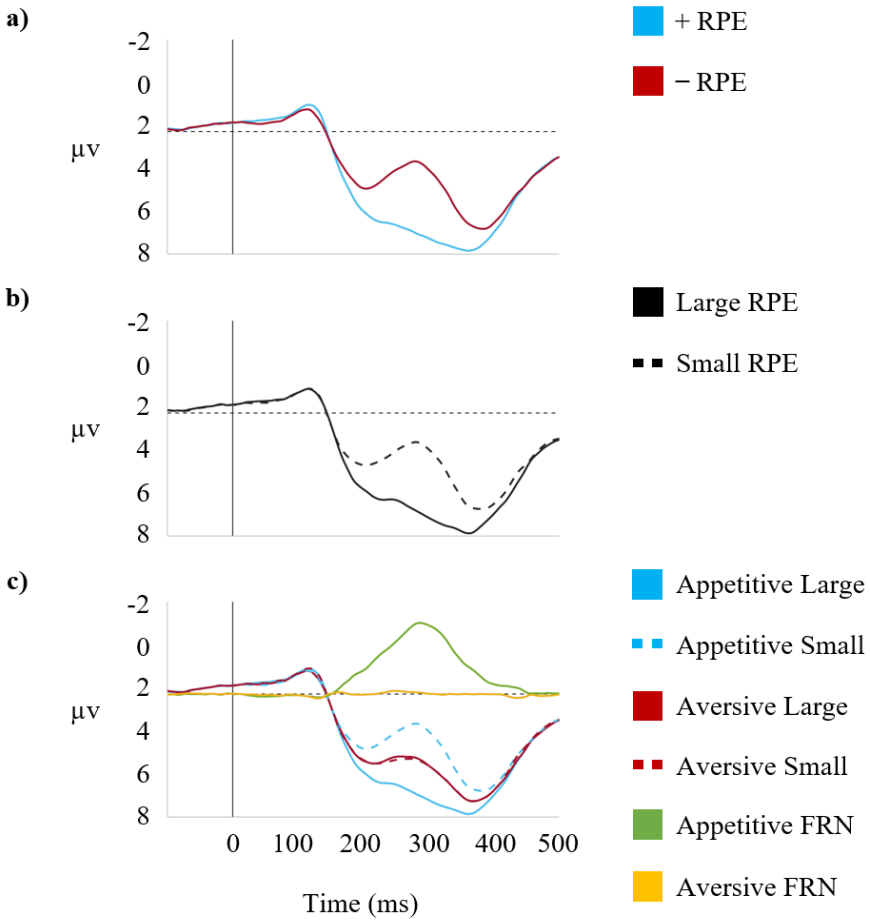

Figure 21. Hypothetical example of component overlap. a) An example RPE-sign encoder, responsive such that -RPEs produce a relative negativity in the waveform compared to +RPEs. b) An example motivational salience encoder, responsive such that small RPEs produce a relative negativity in the waveform compared to large RPEs. c) The observed ERPs following the overlap of the example RPE-sign and motivational salience encoders. Large ERPs are shifted positively, and small ERPs negatively, by the motivational salience encoding component. Appetitive ERPs are shifted further apart because of this, while aversive ERPs are shifted together, ultimately removing the aversive FRN.

Specifically, the greater positivity of large appetitive outcomes for RPE-sign encoding is enhanced by the effect of size expected for the overlapping motivational salience encoder, increasing the amplitude of the appetitive FRN difference wave. Conversely, the greater positivity of small aversive outcomes for RPE-sign encoding is reduced by the effect of size, and so the aversive FRN difference wave is reduced in amplitude. The resulting plot of small and large outcomes separated by outcome domain, as shown in Figure 21, then appears to be that of an RPE-sign encoder that neglects the aversive domain, when in actuality it is the combination of two separate encoders.

4.8.1.2 Aversive neglect

There is also the possibility that the RPE-sign encoding present at 280 ms is not that of a *complete* RPE-sign encoder, but instead of one sensitive only to outcomes of an appetitive domain. Such an encoder is identical to a standard RPE-sign encoder as detailed throughout this chapter, with the sole difference that it is entirely unresponsive to aversive outcomes. This neglect of aversive outcomes results in a change in the neural response for RPE-sign encoding in that an effect of size is predicted for domain form studies.

The possibility that the reduced aversive FRN seen in our data is due to aversive neglect is more problematic for this thesis than the claim there is an overlapping motivational salience encoder as it directly counters the possibility of the RPE-sign encoder present at 280 ms being a general-purpose RPE encoder, whereby the neural response should be indifferent to the domain of an outcome. In order to resolve whether the reduced aversive FRN is due to an overlapping motivational salience encoder or instead due to aversive neglect in the RPE-sign encoder, below we introduce two variables capable of discriminating between these cases.

4.8.2 Testing cases for component overlap and aversive neglect

4.8.2.1 Effects of blocking type

To check for aversive neglect, we analysed the waveforms of domain form studies separated out by their blocking type; this refers to whether outcome domain is manipulated across separate blocks within a study (blocked), or instead switched back and forth on a trial-by-trial basis (unblocked). The reason for this is that domain form studies that employ a blocked design are expected to have outcomes being encoded as the better or worse of two on a given trial. More explicitly, as participants repeat trials that are consistently of one domain, we can expect their attention to become focused solely on whether they got the best or worst of the outcomes available, with the effect of domain being lost as they 'forget' about the manipulation of outcome domain to a greater or lesser extent. If there is indeed aversive neglect, then we can expect the reduced aversive FRN to be largely restricted to unblocked designs, where outcome domain is more highly "present" for participants. In blocked designs the effect should be attenuated. The reduced aversive FRN, operationalised as the appetitive FRN – the aversive FRN, was compared for blocked and unblocked designs as shown in Figure 22 below.

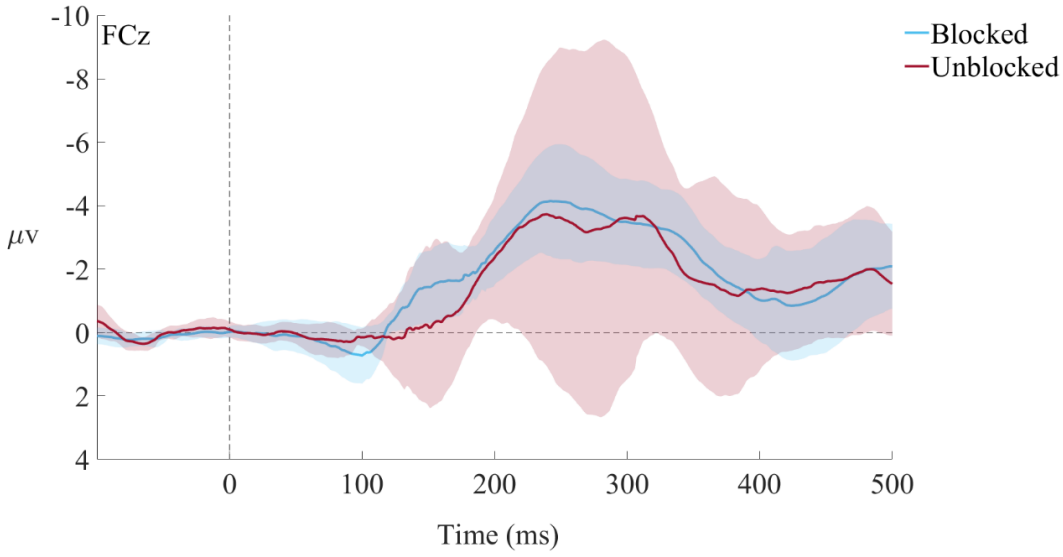

Figure 22. Reduced aversive FRN (appetitive FRN – aversive FRN) for domain form studies, separated by blocking type. Shading shows two standard errors of the mean.

Independent samples *t* tests across the whole waveform showed no significant effect of blocking type at any point in the waveform (at 280 ms, t(14) = -.18, p = .86), which can be seen in the similarity of the reduced aversive FRN effect for both blocked and unblocked designs in Figure 22. These null findings do not therefore support the presence of an RPE-sign encoder that is neglectful of aversive domains. As such, the best remaining explanation for the presence of the reduced aversive FRN is that a complete RPE-sign encoder is overlapped by a motivational salience encoder, producing a misleading impression of aversive neglect.

4.8.2.2 Effects of cue type

The previous analysis made the case for an overlapping motivational salience encoder by excluding the possibility of the RPE-sign encoder showing aversive neglect. The current analysis aims to more directly uncover evidence for an overlapping motivational salience encoder by investigating a new variable: *cue type*, referring to the representational status of cues with respect to the reinforcers they denote. *Concrete* cues are similar to the reward or punishment ultimately incurred by the subject, such as displaying "+20p" on a screen to denote the delivery of a 20p reward or displaying an image of a lightning bolt to denote the delivery of an electric shock. *Abstract* cues are arbitrarily associated with rewards and punishments. Such cues include displaying an image of a fractal to denote monetary reward, or a shape such as a blue circle to denote the delivery of an electric shock. We would expect concrete cues to be more motivationally relevant than abstract ones, thus engaging a motivational salience encoder more strongly.

Two means of manipulating motivational salience are thus available: Outcome size and cue type. The first is an unfortunate confound in studies of the FRN, whereas the latter is an optional means of investigating this confound by manipulating motivational salience without affecting the RPE. Earlier we made the case that, in the aversive domain, the motivational salience of large outcomes relative to small outcomes produces a positivity that cancels out the relative negativity produced by an RPE-sign encoder when encoding a -RPE. If this cancelling is indeed occurring, it should be possible to limit it by manipulating cue type. Specifically, because motivational salience is greater for concrete cues relative to abstract cues, an overlapping motivational salience encoder should generate a greater positivity for large outcomes when they are represented by concrete cues compared to abstract cues. Figure 23a below shows the observed effect of cue type on each of the four possible outcomes, and 15b the overall effect of cue type.

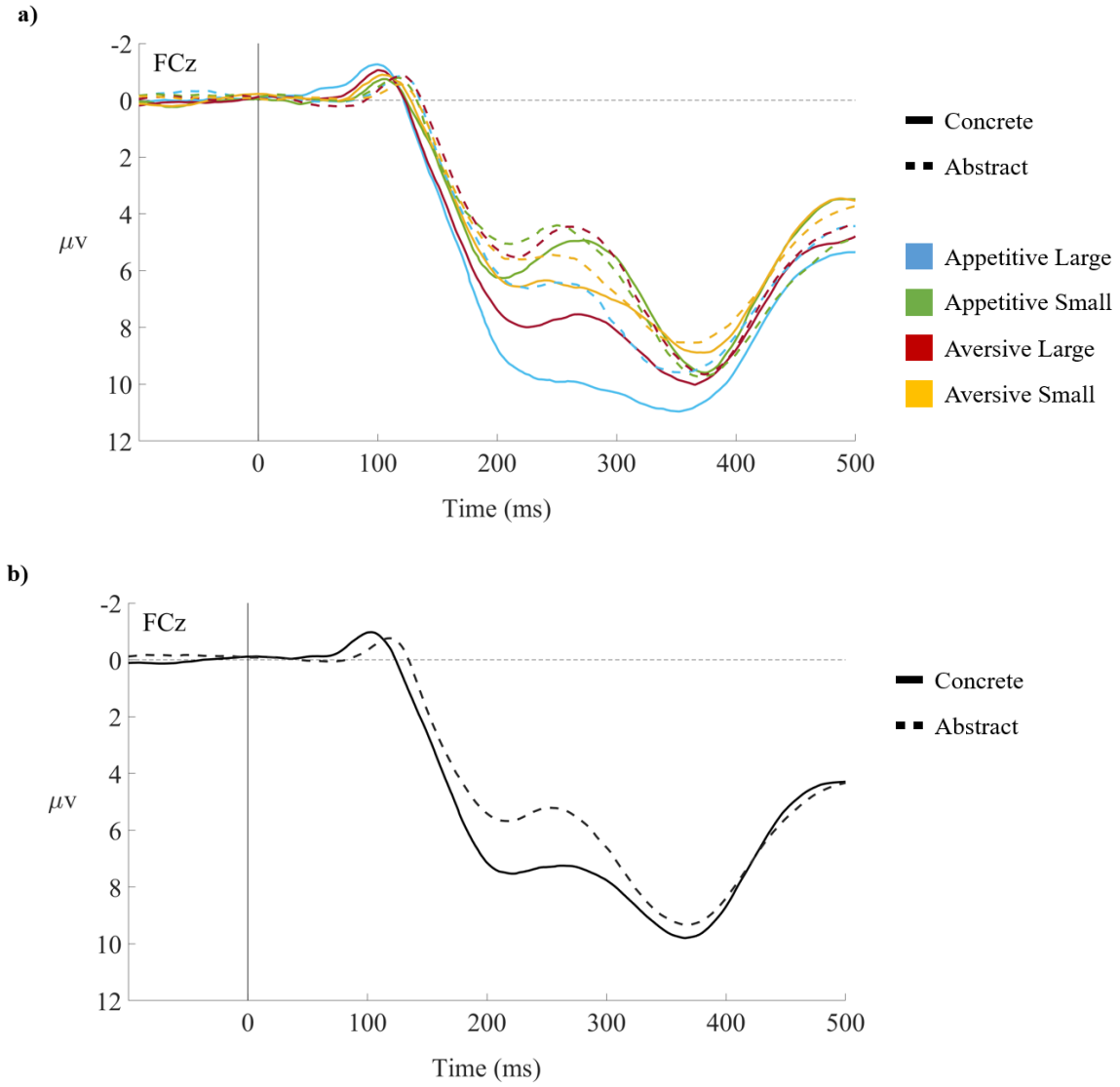

Figure 23. The effect of cue type for a) each of the four possible outcomes and b) across all outcomes.

A Repeated-measures ANOVA conducted at 280 ms revealed a significant main effect of size ($F(14) = 7.85$, $p = .014$), as well as a significant interaction of size*cue type ($F(14) = 6.23$, $p = .026$), fulfilling our expectations that large outcomes generate a greater positivity when represented by concrete cues compared to abstract cues.

4.8.3 Resolving the reduced aversive FRN

Having identified an effect of cue type on the reduced aversive FRN, we then attempted to remove the overlapping motivational salience response, thus enabling us to uncover any underlying general-purpose RPE encoder. This was done by once again plotting the appetitive and aversive FRNs, in the same manner as in Figure 20, but only including studies that use abstract cues. After a repeated-measures *t* test between the

appetitive and aversive FRNs of studies using abstract cues (Figure 24), no significant difference was revealed (at 280 ms, t(5) = -.22, p = .83), demonstrating no significant diminishing of the aversive domain FRN. As can also be seen in Figure 24, the aversive FRN is slightly delayed, mimicking the results seen in the previous chapter.

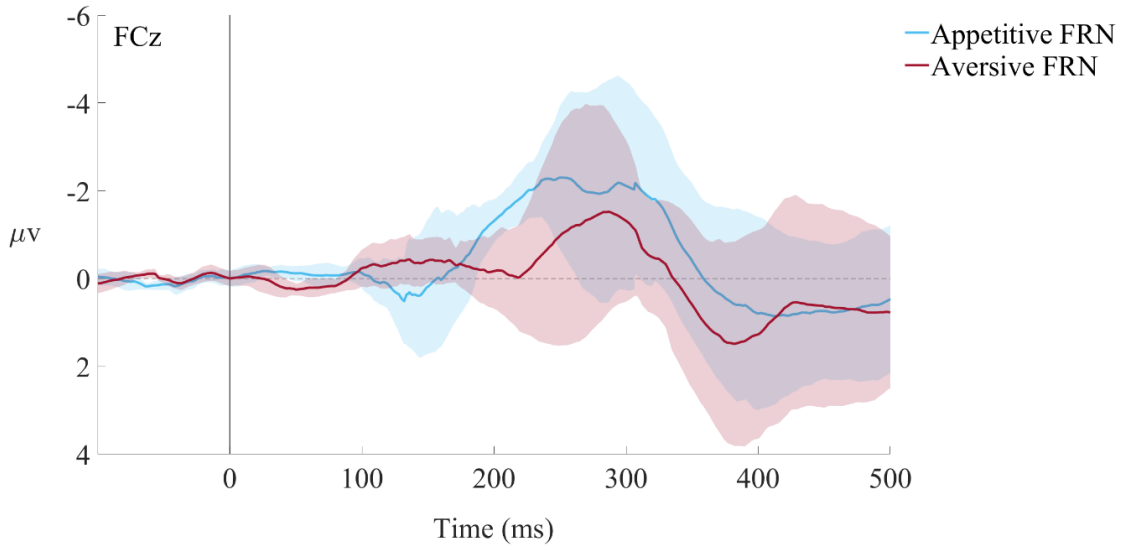

Figure 24. The FRN (worse outcomes – better outcomes on each trial) for studies with abstract cues, separated by outcome domain. Shading shows two standard errors of the mean.

4.9 Discussion

4.9.1 Results summary

In this study we used exhaustive Bayesian contrasts to examine the feedback-locked waveform in greater detail than what is achievable through the use of standard operationalisations. This is different to past chapters in that the standard operationalisations which we have used thus far do not consider null effects to be predictive of certain types of encoding. We aimed to use this previously unharnessed information to more accurately identify types of RPE and motivational salience encoding across the waveform. Furthermore, we included experiment form as an additional variable necessary for the correct identification of RPE-sign encoding, and thus could better disambiguate it from RPE-utility and motivational salience.

4.9.2 Evidence for motivational salience encoding

4.9.2.1 Parietal encoding at 360 ms

Of the three key encoders we tested for, the strongest evidence found was for motivational salience. Evidence for this encoding was first found at 160 ms, however the supporting analyses failed to provide substantial support for this with only a small effect of size found limited to size form studies, and no PCA component explaining the effect was identified. Rather, the primary evidence for motivational salience encoding was found at around 360 ms, largely supportive of past research investigating the P3, a parietal component occurring 250 – 500 ms after the presentation of task-relevant information that has been associated strongly with processing a motivational salience response (Bellebaum, Polezzi, et al., 2010; Kreussel et al., 2012; Toyomaki & Murohashi, 2005; Wu & Zhou, 2009). Unlike in our initial exhaustive Bayesian contrasts, we found a lack of evidence for motivational salience encoding at around 300 ms and onwards in our frontal – parietal Bayesian contrasts, suggesting that this encoding originates parietally.

4.9.2.2 Frontal encoding at 280 ms

As we demonstrated in our investigations of RPE valence, we also found the FRN to be likely confounded by a temporally similar motivational salience response at around 280 ms. This finding supports our earlier results of an overlapping motivational salience encoder found following PCA, the standard operationalisation of the effect of size shown in Figure 18, and our previous chapters. Considering that motivational salience is not an RPE encoder per se, rather it is responsive to the motivational importance or prominence of an event, it is unsurprising that such a process may overlap temporally with RPE processing. While this is a largely unavoidable artefact of the EEG approach, we were able to not only identify, but reduce the effect of size culpable for the overlapping motivational salience response through manipulating blocking type and cue type, revealing an underlying general-purpose RPE.

The process of manipulating these variables allowed us to rule out the possibility of the RPE-sign encoder being neglectful of aversive outcomes. A great deal of studies investigating the FRN have considered such neglect as a possibility (Cohen et al., 2007; Eppinger et al., 2009; San Martín et al., 2010), and considering our meta-analytic approach, it could be that aversive neglect was not the case in some of these studies at all, but rather an *apparent* aversive neglect caused by overlapping motivational salience encoding. As such, we suggest that future research considers the influence of blocking type and cue type on the encoders present in the waveform, and makes use of these variables where possible in analyses to filter out unwanted overlap and reduce the possibility of misdiagnosing encoders.

4.9.3 Evidence for RPE-sign encoding

Another finding from the present study is the evidence for RPE-sign encoding found at around 280 ms, demonstrated by our exhaustive Bayesian contrasts, Bayesian repeated-measures ANOVAs, and standard operationalisations of domain and domain*size. We must note that while the predicted results across these multiple analyses were largely consistent with RPE-sign encoding, the overall relative evidence for this encoding indeed being present at this interval remains relatively weak. What is perhaps of more interest is that this evidence, although fairly weak, supports the case for RPE-sign more so than RPE-utility, counter to recent findings in the literature suggesting that RPE-utility is the more dominant contender for encoding represented at this time point (Gu et al., 2021; Sambrook & Goslin, 2015, 2016). Up to this point, the case for RPE-utility has largely been made with only the standard operationalisation of the domain*size effect; the same as was used in this study in the supporting standard operationalisation of domain*size shown in Figure 19. In fact, looking at just this test, we support the case for RPE-utility with significant domain*size effects for both domain and size form studies. However, the purpose of this study was not to rely on standard operationalisations alone, but instead to gain both relative and actual insight through multiple analysis methods.

The inclusion of experiment form as a variable allowed us to better disambiguate RPE-sign from RPE-utility; a domain*size effect is predicted for RPE-utility encoding in both experiment forms, but only in domain form studies for RPE-sign encoding. Furthermore, an effect of domain is also predicted for RPE-utility encoding in both experiment forms, but only in size form studies for RPE-sign encoding. While we indeed found RPE-utility evidence in the form of a domain*size effect for both experiment forms, this effect was noticeably stronger for domain form studies, indicative instead of RPE-sign. Furthermore, evidence for RPE-utility encoding was seldom found in our exhaustive Bayesian contrasts or Bayesian repeated-measures ANOVAs, and again provided greater support for RPE-sign at around 280 ms. As such, it is likely the case that the RPE-utility encoding suggested to be represented in the waveform by past research may in fact more likely be RPE-sign encoding; only through using multiple analysis methods and inferring from both the exhaustive relative evidence *and* the specific evidence provided by operationalised effects were we able to better diagnose the waveform.

4.9.4 Chapter conclusions

With the addition of experiment form as an experimental variable, this chapter has demonstrated how RPE-sign, RPE-utility and motivational salience can be better disambiguated in the waveform through a meta-analysis of their unique neural responses. Multiple, exhaustive analyses revealed an RPE-sign encoder at around 280 ms post-feedback, supporting the case made in past research, and our previous chapters, that this interval indeed houses an RPE. Furthermore, investigating RPE valence at this interval revealed an apparent aversive neglect which, through further analysis of blocking type and cue type, was found to likely

be due to overlapping motivational salience. Removing this overlap rectified the reduced aversive FRN, supporting the case that the RPE-sign response is fitting of a general-purpose RPE encoder.

5. Evidence for Parietal Reward Prediction Errors using Great Grand Average Meta-analysis

This chapter is based on a published paper:

Stewardson, H. J., & Sambrook, T. D. (2020). Evidence for parietal reward prediction errors using great grand average meta-analysis. International Journal of Psychophysiology, 152, 81-86.

5.1 Chapter Abstract

As a basic principle within the economics of decision-making, reinforcement learning dictates that individuals strive to repeat behaviour that elicits reward, and avoid behaviour that elicits punishment. Alongside the frontocentral FRN, which has been the central component of interest for this thesis thus far, the more parietal P3 component is also implicated in outcome processing. However, whether this component demonstrates a potential for RPE encoding is currently unclear. Much alike the case for the FRN, one reason for this lack of clarity is inconsistent quantification of the component in the literature. A recent meta-analysis that directly quantified published waveforms rather than using reported effect sizes found strong evidence that the FRN encodes an RPE; in the current study, such a meta-analysis was performed on parietal waveforms to establish whether the P3, or parietal areas generally, are sensitive to RPEs as well. A strong effect was found, both of RPE-utility encoding and simple valence sensitivity at a latency associated with the P3.

5.2 Introduction

According to reinforcement learning, an agent seeks to repeat behaviours that are rewarding, and avoid behaviours that are punishing (Thorndike, 1898), and with an effective learning mechanism, rewards and punishments are not valued dichotomously, but instead as RPEs (Sutton & Barto, 1998). As we have covered in the previous chapters, distinguishing RPE valence is crucial for reinforcement learning, however optimal learning should also be influenced by the size of an RPE: big or unexpected outcomes should drive greater learning than small or expected outcomes. Since the direction of this learning (whether an action is more or less likely to be repeated) is determined by valence, with large RPEs having opposite effects depending on whether they are positive or negative, the signature response of a neural RPE-utility encoder consists in a sign x size interaction (Caplin & Dean, 2008).

Thus far we have attempted to resolve ongoing debate in the literature regarding the underlying RPE encoding capabilities represented in the FRN. This chapter aims to further this goal of resolution, with the addition of investigating a second component elicited in outcome processing: the P3. The next subsection shall summarise our knowledge of the FRN up to this point and it's standing within reinforcement learning research, before doing the same for the P3.

5.2.1 The FRN

The FRN is a frontocentral component occurring approximately 240 – 340 ms after feedback received in a task involving either rewards or punishments. As we have discussed throughout this thesis, it has been proposed as demonstrating an RPE-sign encoder (Holroyd & Coles, 2002) and is typically quantified by a difference wave created by subtracting the voltage of +RPEs from -RPEs. This results in the FRN's characteristic negative peaked difference wave when, in the simple waveforms underlying it, +RPEs produce a relatively positive-going voltage compared to -RPEs. The term 'reward positivity' may refer to this relative positive-going voltage of the simple waveforms but also to the claim that the component is sensitive specifically to rewards rather than non-rewards (Proudfit, 2015).

A number of authors however have asserted that this component reflects not RPE-sign, but motivational salience (Garofalo et al., 2014; Hauser et al., 2014; Talmi et al., 2012). This claim predicts a flat difference wave for the negative – positive contrast (provided each outcome is equally likely) and instead a difference wave for a small – large contrast. A meta-analysis by Sambrook and Goslin (2015) instead showed both these effects to be present in the interval associated with the FRN. Importantly, there existed also the signature sign x size interaction indicating the encoding of RPE-utility, a signed quantitative term complying with RL-theory (Sutton & Barto, 1998).

5.2.2 The P3

The P3, or P300, is a parietally distributed, positive going ERP, occurring 250 – 500 ms after presentation of task-relevant information. The P3 has been widely studied in decision-making research, and is associated with engagement of attention, the processing of novelty and expectation, and cognitive workload (Polich, 2003). As with the FRN, its status with regard to RPE encoding has been unclear. An influential "independent coding model" (Yeung & Sanfey, 2004) has claimed that the P3 is sensitive only to RPE size, with the FRN sensitive only to valence. It is certainly the case that the P3 is usually represented by larger amplitudes following larger outcomes regardless of their valence, (Bellebaum, Kobza, et al., 2010; Gu et al., 2011; Kreussel et al., 2012; Toyomaki & Murohashi, 2005), and unexpected outcomes likewise (Bellebaum & Daum, 2008; Hajcak et al., 2005, 2007; Wu & Zhou, 2009) and, as such, may be indicative of UPE or motivational salience encoding. This does not entail encoding for valence, however. In fact, in a review of the effect of valence on the P3, San Martín (2012) found that sensitivity to valence was also typically found, though the direction of effect was not consistent. Similarly, a review (Glazer et al., 2018) found equivocal evidence for the P3's status as representing an RPE-sign encoder. In contrast to these main effects, there has been a dearth of research on the P3's status as showing RPE-utility encoding, capturing the full RPE sign x size interaction.

5.2.3 Summary

The study in this chapter addresses the above gap in knowledge. It uses the great grand average methodology of ERP meta-analysis discussed in the previous chapter, in which published effect sizes are disregarded in favour of direct quantification of published waveforms. The method has been extensively validated (Sambrook & Goslin, 2015) and in the case of the FRN was shown to provide superior estimates of effect size than conventional meta-analysis. We have previously argued the benefits of this approach as a solution to inconsistencies in the latency at which ERP components are identified, and was indeed put to such use in the previous chapter. The present study demonstrates a further advantage of the method, which is its capacity to use data for purposes other than those to which it was originally put. The meta-analysis provided here is based on thirty-three studies, of which only a small number specifically addressed the P3. In this study, the principal effect under investigation is an RPE-utility encoding at parietal sites. However, we also present effect sizes for encoding of RPE-sign and for motivational salience. We additionally perform a moderator analysis on the RPE-utility effect to establish if it is sensitive to the degree of control over the outcome that the task appears to afford.

5.3 Method

5.3.1 Inclusion and Exclusion Criteria

For inclusion, a study needed to manipulate the independent variables of valence (either positive or negative) and RPE size (either small or large). Experiments were included provided that both of these variables were manipulated in a 2 x 2 factorial design. RPE size could be manipulated by adjusting either the magnitude of outcomes or their likelihood. Experiments were included if they had intermediary levels of RPE size, although these were ignored in favour of the most extreme levels to maximise contrasts. Experiments had to provide, as a dependent variable, simple waveforms corresponding to the design above, or difference waves appropriate for the analysis as described below. Importantly, provided that such voltage x time waveforms were presented, no statistics associated with these needed to be reported. Waveforms had to be presented parietally (Pz, CPz or POz) but also frontocentrally (FCz, Fz, Cz) in order to distinguish parietal effects from those of the FRN. The feedback-locked epoch needed to run, at minimum, from -100 – 500 ms.

Regarding participants, experiments either needed a population of healthy adults or a healthy control group if an alternative experimental group was used. Any studies including participants who had been selected based on a screening process were excluded.

Experimental tasks were restricted to those offering monetary stakes (either wins or losses) but could otherwise vary, including passive tasks in which participants knew they could not influence outcomes, guessing tasks in which participants were encouraged to increase the likelihood of desirable outcomes

despite no control being really available, and rule implementation tasks in which outcomes were probabilistically linked to a choice rule that participants had to learn. These three kinds of task also constituted the three levels of the moderator variable “control over outcome” operationalised at three levels: “passive”, “guessing” and “rule implementation”, following Sambrook and Goslin (2015). Criteria for inclusion and exclusion were identical to those used by Sambrook and Goslin (2015) with the exception of the additional requirement of a parietal waveform.

5.3.2 Search Strategies

English language journals and books were searched using following databases: PsycInfo, PsycArticles, ERIC, PubMed and Web of Science. Results were compiled in Zotero. The search covered journal titles, abstracts and keywords. Articles were gathered based on three searches. First, articles used in the Sambrook and Goslin (2015) meta-analysis of the FRN were used if they featured parietal waveforms. This yielded twenty studies. Second, the principal search string used in that meta-analysis was used once again but in the time frame 2015 to present, in order to capture articles published subsequently. This string was “feedback negativity” OR “feedback-related negativity” OR “feedback error-related negativity” OR “reward positivity” OR “feedback correct related positivity”. This search was conducted on the basis that much of the RPE literature is focussed on the FRN and likely to be associated with this term, but that parietal waveforms might nevertheless be provided. This search provided twelve more studies. Third, a mutually exclusive search was performed on articles targeting the P3 but with limiting terms owing to the extensive P3 literature. The search string was ("P3" OR "P300") AND ("learning" OR "motivational significance" OR “prediction error”) AND NOT ("feedback negativity" OR "feedback-related negativity" OR "feedback error-related negativity" OR "reward positivity" OR "feedback correct related positivity"). This search yielded one more study. In total, these searches gave a final study sample of thirty-three studies, shown below in Table 6. The set of studies overlapped heavily with the studies used to meta-analyse the frontal, FRN component in Sambrook and Goslin 2015. Given so few additional non-FRN studies of RPEs were found by our literature search, we thought it highly unlikely that there would be a significant number of obtainable unpublished studies that had not been located in the generation of that paper.

Table 6. Studies used for the meta-analysis.

Experiment	N	Modulator	Frontal Site/s	Parietal Site/s	Moderator Level
Bellebaum and Daum (2008)	17	Likelihood	FC3, FCz, FC4	P3, Pz, P4	Rule Implementation
Bellebaum et al. (2011)	18	Likelihood	Cz	Pz	Rule Implementation
Hajcak et al. (2005). Expt 1	17	Likelihood	Fz	Pz	Guessing
Hajcak et al. (2005). Expt 2	12	Likelihood	Fz	Pz	Guessing
Hajcak et al. (2007)	17	Likelihood	Fz	Pz	Guessing
Hu et al. (2018)	22	Likelihood	FCz	CPz	Guessing
Kreussel et al. (2012)	24	Likelihood	Fz	Pz	Rule Implementation
Liao et al. (2011)	15	Likelihood	Fz	Pz	Rule Implementation
Pfabigan et al. (2011)	20	Likelihood	FCz	Pz	Guessing
Salim et al. (2015)	39	Likelihood	Fz	Pz	Passive
Sambrook and Goslin (2016)	42	Likelihood	FCz	Pz	Guessing
Sambrook and Goslin Unpublished	48	Likelihood	FCz	Pz	Guessing
Walentowska et al. (2016)	30	Likelihood	Fz, FCz	CPz, Pz, P1, P2	Guessing
Wu and Zhou (2009)	16	Likelihood	FCz	Pz	Guessing
Bellebaum et al. (2010)	15	Magnitude	Fz	Pz	Rule Implementation
Banis and Lorist (2012)	32	Magnitude	FCz	Pz	Guessing
Gu et al. (2011)	24	Magnitude	Fz	CPz	Guessing
Kamarajan et al. (2009)	48	Magnitude	FCz	Pz	Guessing
Kreussel et al. (2012)	24	Magnitude	Fz	Pz	Rule Implementation
Luo and Qu (2013)	18	Magnitude	FCz	Pz	Guessing
Meadows et al. (2016)	19	Magnitude	FCz	Pz	Guessing
Pfabigan et al. (2015)	31	Magnitude	Fz	Pz	Rule Implementation
Sato et al. (2005)	18	Magnitude	Fz	Pz	Guessing
Schuermann et al. (2012)	20	Magnitude	FCz	CPz	Passive
Sambrook and Goslin (2014)	55	Magnitude	Fz	Pz	Passive
Sambrook and Goslin (2016)	45	Magnitude	FCz	Pz	Guessing
Van den Berg et al. (2011)	42	Magnitude	Fz	Pz	Guessing
Wei et al. (2018)	22	Magnitude	FCz	Pz	Guessing
Wischnewski and Schutter (2018)	20	Magnitude	Fz, FC1, FC2, Cz	Cz, CP1, CP2, Pz	Guessing
Wu and Zhou (2009)	16	Magnitude	FCz	Pz	Guessing
Yu and Zhou (2006)	20	Magnitude	Fz	Pz	Guessing
Yu and Zhou (2008)	14	Magnitude	Fz	Pz	Guessing
Zheng and Liu (2015)	43	Magnitude	FCz	Pz	Guessing

5.3.3 Coding Procedures

The data collection process mimicked that used in Chapter 4 (See Coding Procedures subsection 4.2.3). In the case of one paper (Banis & Lorist, 2012), original data was used, previously obtained from the authors, since no parietal waveform was presented in the paper. Original data was also used from all papers authored by Sambrook.

5.3.4 Statistical Methods

A major strength of the great grand average method of meta-analysis is that it establishes average effect size at all points on the waveform, not merely those reported in individual papers, thus maximising the data used and eliminating idiosyncrasies in quantification in individual papers. It means however, that study-level variance is unknown, since this is rarely shown in the grand average waveforms we use here as data. Instead of standardised effect sizes, which are the norm in conventional meta-analysis, simple or "raw" effect sizes are used (see Sambrook & Goslin, 2015 for a discussion of the merits of each). Three simple effect sizes were computed, realised in each case as a waveform of effect size, based on differencing of the simple waveforms described in the design earlier. The *RPE valence effect* was computed by a difference wave (large -RPE – large +RPE). Large RPEs only were used in order to maximise the contrast. The *RPE size effect* was computed by the difference wave (small -RPE + small +RPE) – (large -RPE + large +RPE). Following Sambrook and Goslin (2015), the *valence*size effect* was computed by the difference wave (small -RPE – small +RPE) – (large -RPE – large +RPE), a difference of difference waves which implements an axiomatic test for RPE-utility encoding (Caplin & Dean, 2008). In the case of the FRN (operationalised as a difference wave), it predicts that the FRN for small outcomes will be of lower amplitude than the FRN for large outcomes, generating a difference of difference waves of non-zero amplitude. An equivalent rationale was used here to establish parietal RPE-utility encoding.

Study weighting in conventional meta-analysis is based either on a study's variance or its sample size. Since variance was unavailable, study size was used instead, implemented in the form of weighted *t* tests. The significance of effect sizes was established by submitting the appropriate difference waves to running univariate weighted one-sample *t* tests (test value = 0) at each 2 ms time point. All figures in the Results section, both of *t* tests and voltage plots show weighted effects. Publication bias was assessed by subjecting funnel plots to the trim and fill procedure of Duval and Tweedie (2000) implemented in R (R Core team) using the metafor package (Viechtbauer, 2010). Owing to a dearth of studies at the "passive" level of the moderator, moderator analysis was performed on just the "guessing" and "rule implementation" levels. This was implemented as a 2 x 2 (moderator x site) ANOVA on valence*size effect size at its peak (286 ms frontally and 348 ms parietally).

5.4 Results

5.4.1 Effect of RPE Valence

Figure 25a shows the RPE valence effect at parietal and frontal sites. A typical frontocentral FRN is found, peaking at 286 ms, and a later parietal effect peaking at 312 ms with an effect size of -2.33 μv. One-sample *t* tests were conducted on these waveforms and are shown in Figure 25b. Between 400 – 442 ms a significant RPE valence effect was found only at parietal sites. Results were similar when unweighted *t* tests

were performed and also when effects were examined separately for studies using likelihood and magnitude as the prediction error size modulators, and these are shown in Figure 26. Trim and fill applied to the parietal effect size at peak revealed evidence of publication bias. After this was removed with the addition of ten imputed studies, the effect size fell to -1.43 µv but remained highly significant ($z = -3.42$, $p < 0.01$). The funnel plot with imputed studies is shown in Figure 27. The main effect of RPE size (ignoring RPE valence) is shown in Figure 28 and simple waveforms for these effects are provided in Figures 29 and 30.

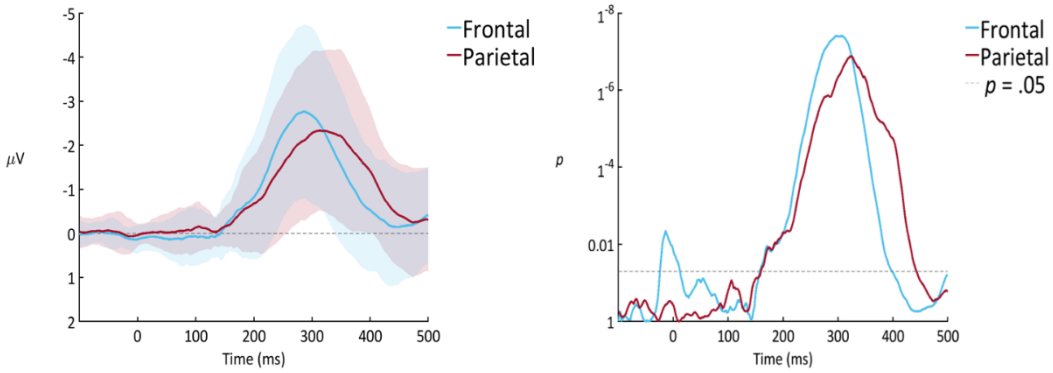

Figure 25. Effect of RPE valence (negative – positive difference wave) at frontal and parietal sites with standard deviations shown in shadow. a. Great grand average voltage. b. Significance of each difference wave under a one-sample t test over experiments.

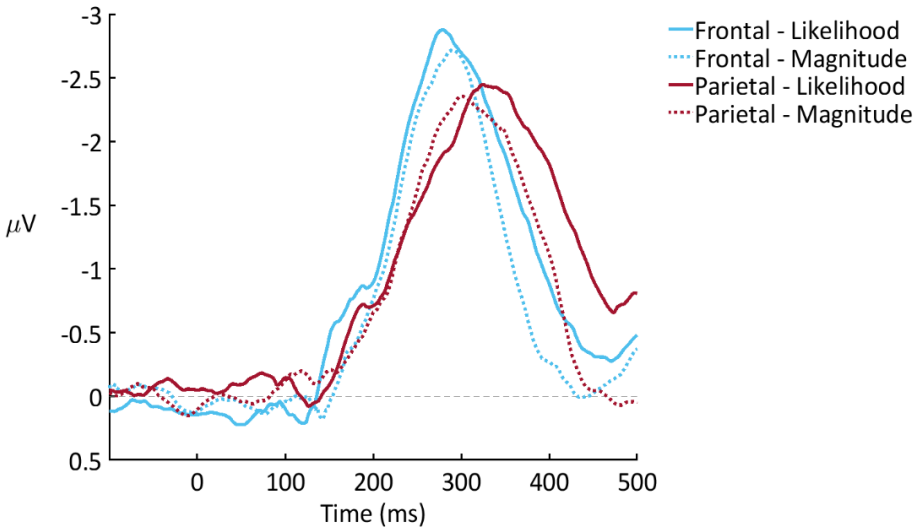

Figure 26. Effect of RPE valence (Figure 1a) broken down by whether likelihood or magnitude is used to modulate RPE size.

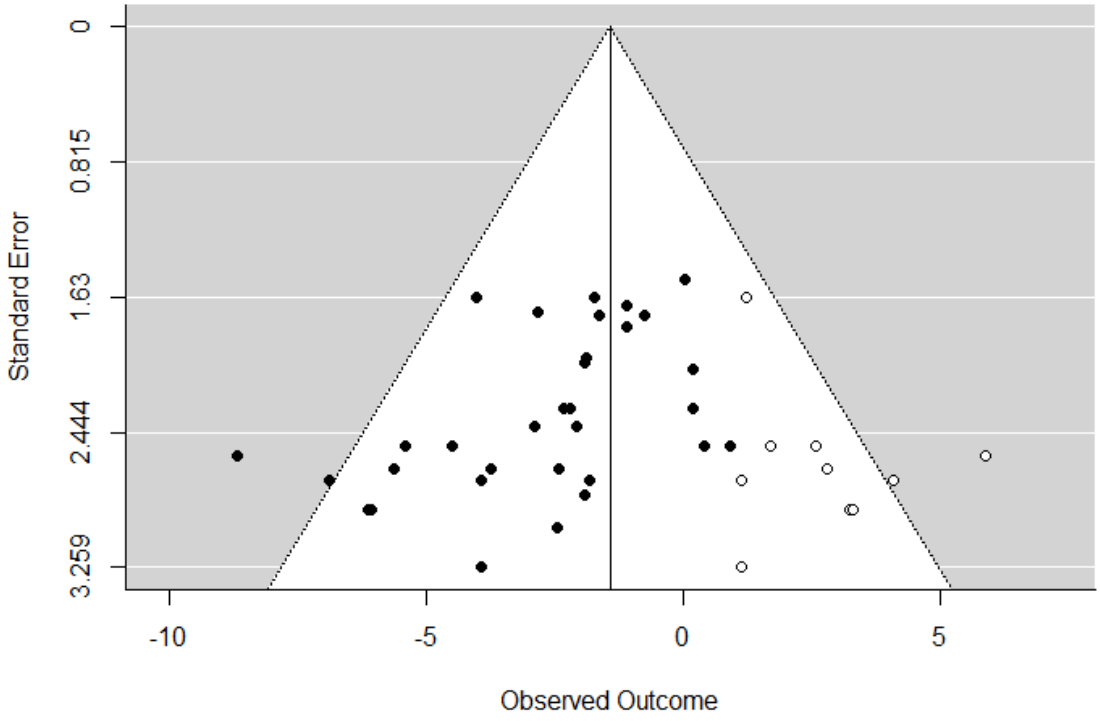

Figure 27. Funnel plot for the parietal effect of RPE valence at its peak of 312 ms. Imputed studies resulting from trim and fill are shown in white.

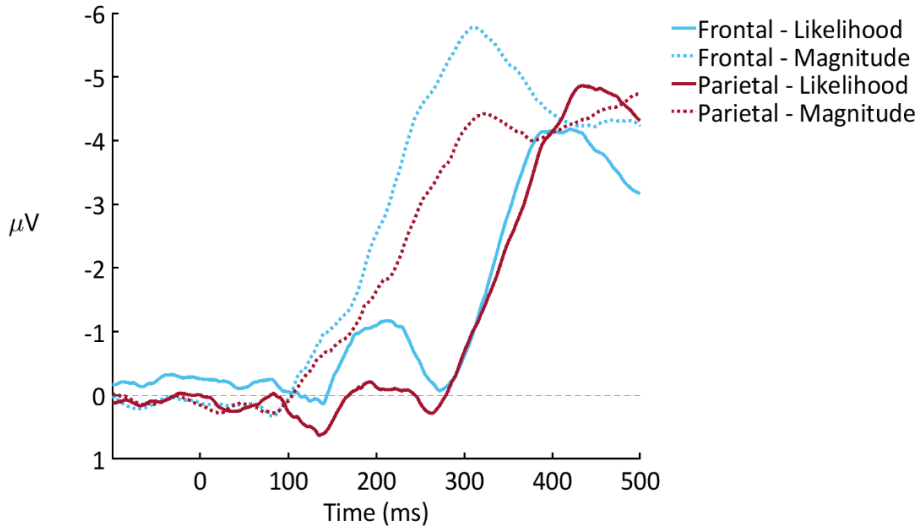

Figure 28. Effect of RPE size broken down by whether likelihood or magnitude is used as its modulator.

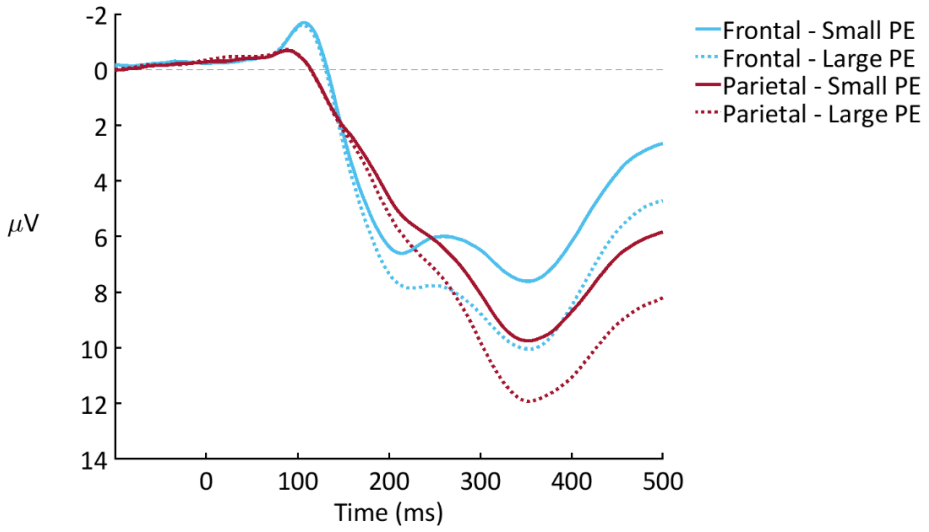

Figure 29. Simple ERPs for RPE size, for rewarding outcomes only.

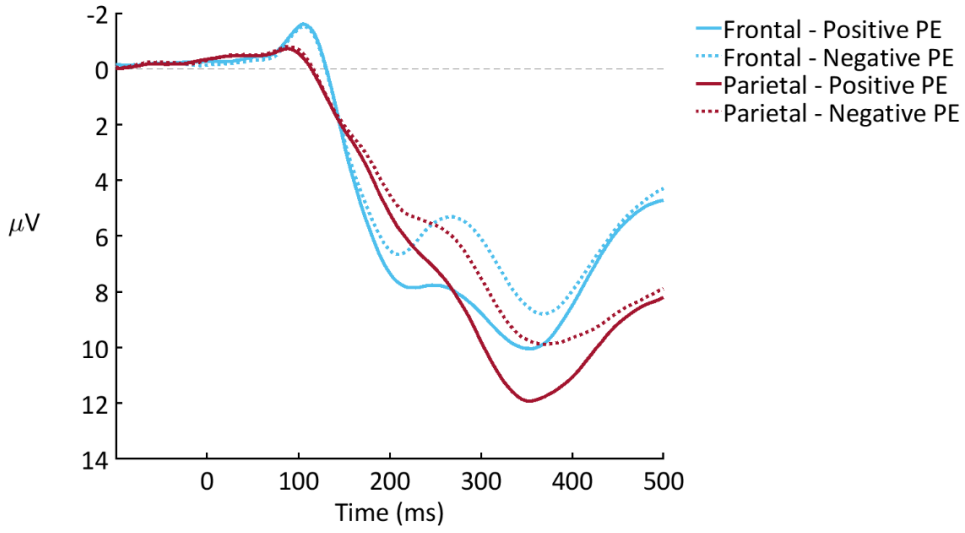

Figure 30. Simple ERPs for RPE valence, for large outcomes only.

5.4.2 Effect of Valence*Size

Figure 31a shows the valence*size effect at parietal and frontal sites. In keeping with Sambrook and Goslin (2015) a frontal valence*size effect is found at a latency associated with the FRN, peaking at 286 ms and a later parietal effect peaking at 348 ms with an effect size of -.89 μv. Figure 31b shows the significance of the effect at each site. In the interval 356 ms to 418 ms a significant effect was found only at parietal sites. Again, results were similar when using unweighted *t* tests or analysing likelihood or magnitude modulated

studies only (Figure 32). Trim and fill applied to the parietal effect size at peak revealed no evidence of publication bias, and this is shown in the funnel plot in Figure 33. Moderator analysis showed a main effect of moderator: as in Sambrook and Goslin (2015) the valence*size effect was stronger in rule implementation than in guessing ($F_{1,28}$ = 4.24 p = .031, σ^2 = .16), but no interaction between this and site ($F_{1,28}$ = .59 p = .45). Figure 34 shows moderator effects across the full waveform.

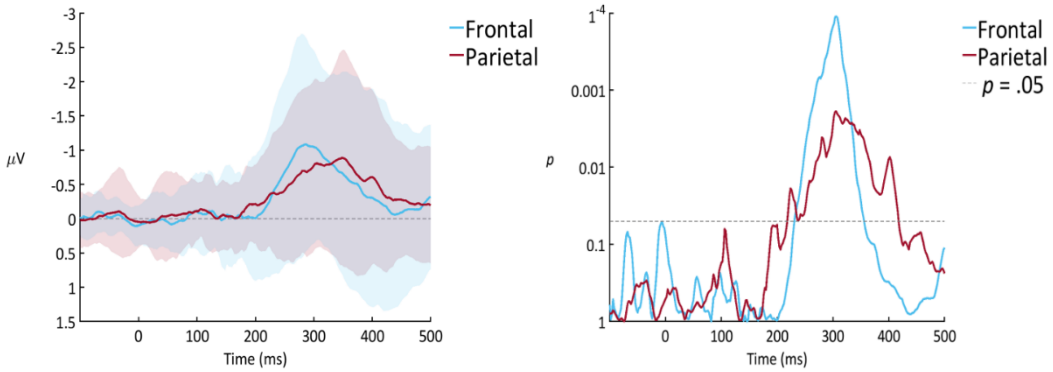

*Figure 31. Effect of valence*size (small negative – small positive difference wave) – (large negative – large positive difference wave) at frontal and parietal sites with standard deviations shown in shadow. a. Great grand average voltage. b. Significance of each difference wave under a one-sample t test over experiments.*

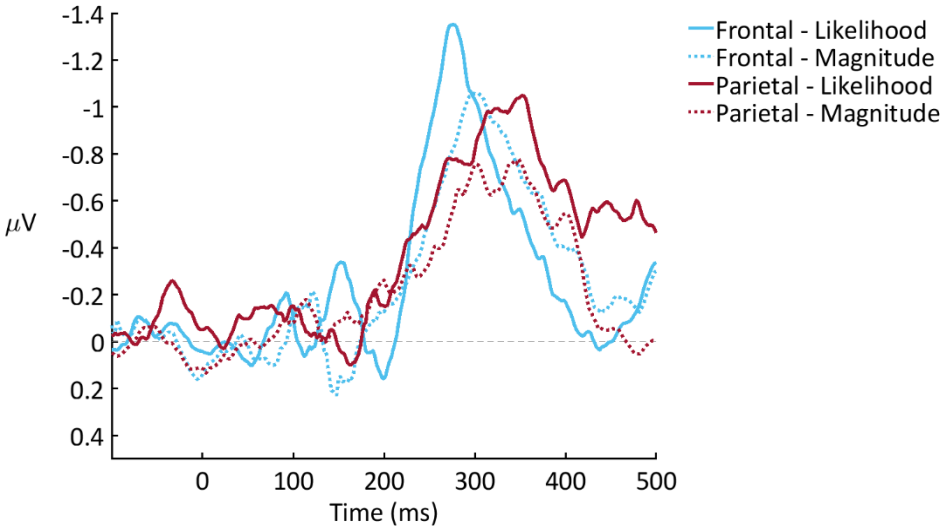

*Figure 32. Effect of valence*size (Figure 2a) broken down by whether likelihood or magnitude is used to modulate RPE size.*

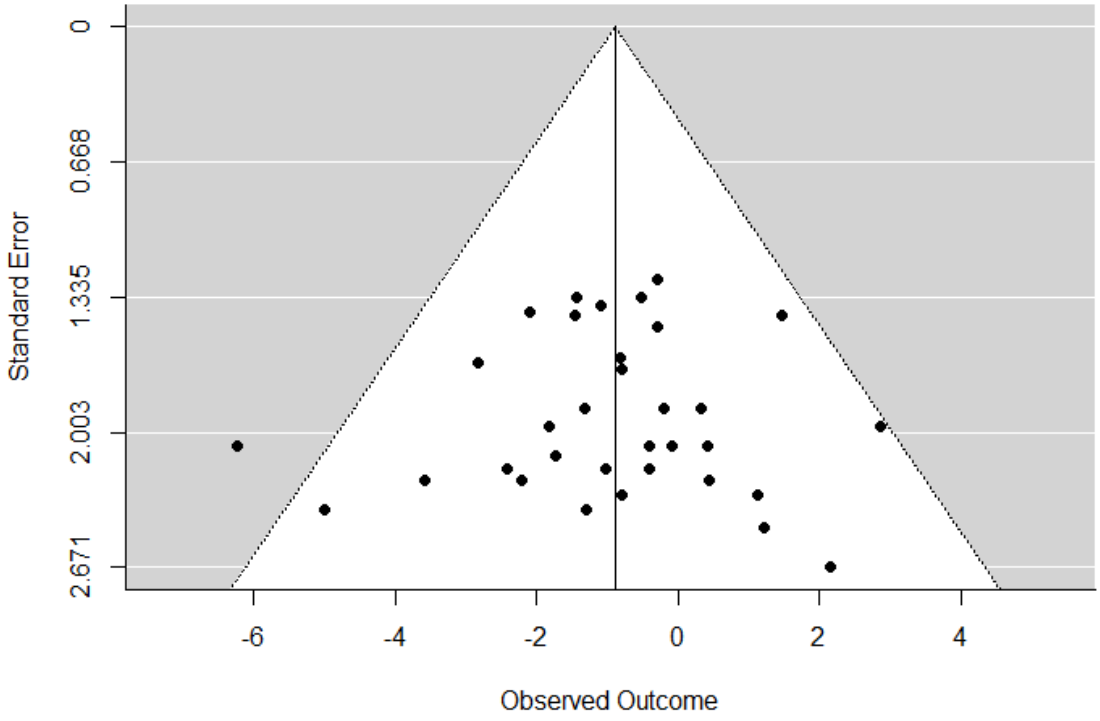

*Figure 33. Funnel plot for the parietal effect of valence*size at its peak of 354 ms. Trim and fill reveals no publication bias.*

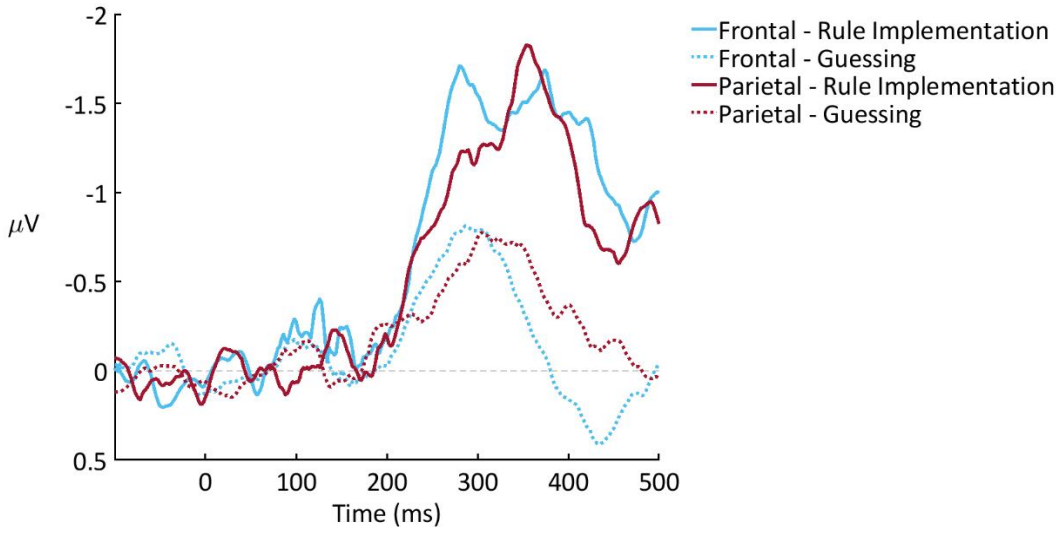

*Figure 34. Effect of valence*size (Figure 2a) broken down by moderator level.*

5.4.3 Independent Components vs. Information Relay

The parietal effect demonstrated does not necessarily imply an independent process since the underlying generator may simply be receiving information relayed from the generator responsible for the FRN. If this were the case we would expect a strong correlation, across experiments, of the two effects in the intervals in which they were significant. In contrast, a weak correlation between these effects, despite their being strong responses in their own right, may suggest independent processes. Such a correlation, performed across subjects, is commonly used within individual experiments for the purpose of demonstrating common vs. separate processes. It is equally valid when run across experiments, however. Fully independent RPE encoders are likely to capture different aspects of this property, which can take forms beyond its basic formulation (as we cover in the Discussion section below), and different tasks would thus be expected to variably elicit one or other component if those components are independent.

As an indicator of non-independent processes, correlated activity needs to be taken in the context of the strong temporal and spatial correlations present in any ERP. If an effect at a parietal site constitutes the arrival of information previously held only at a frontal site then we would expect the correlation between parietal and frontal sites, at this time lag, to be greater than the correlation between the parietal site and itself over the same lag. The heat map in Figure 35 shows the difference between these correlations, expressed as a signed Z score. Two masks have been applied; firstly the parietal-frontal correlation must be greater than the parietal-parietal correlation (since this is signature of a relayed signal described above), and secondly the parietal-frontal correlation must be significant in its own right (r = .343, N = 33). The rectangle indicates the temporal co-ordinates at which both frontal and parietal sites show a significant valence*size effect (as portrayed in Figure 31b), which is where we would expect relatively large parietal – frontal correlations if information were being relayed. These are largely absent, suggesting independent frontal and parietal components.

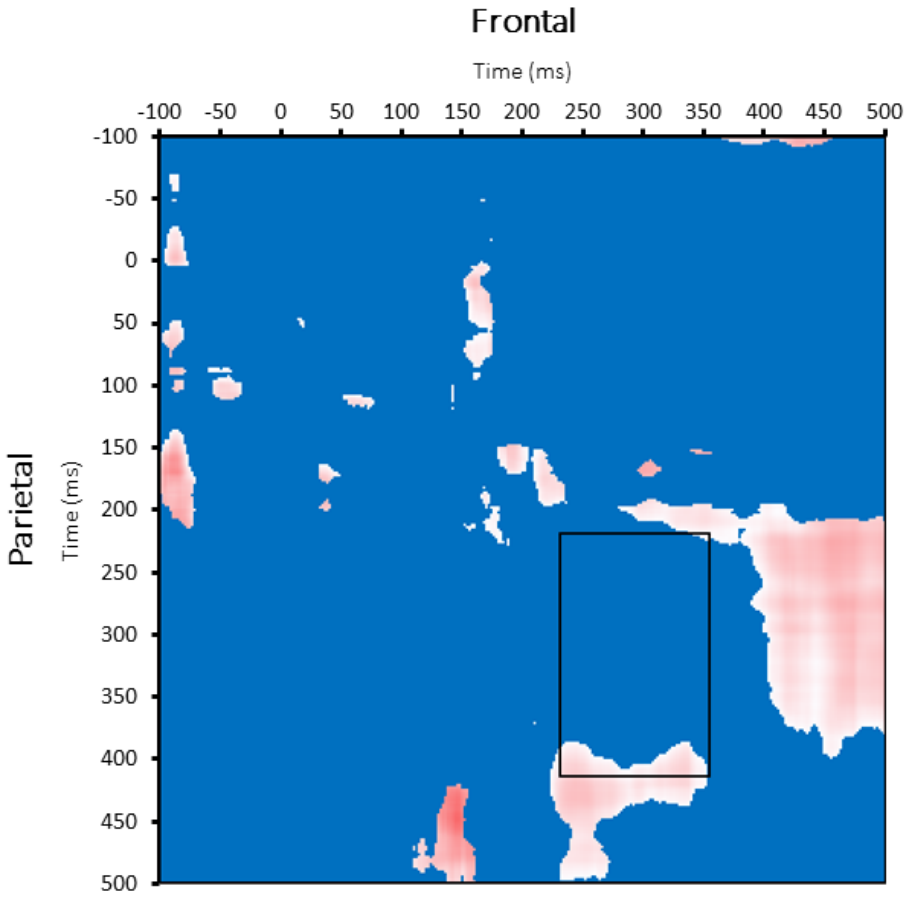

Figure 35. Heat map depicting superiority of parietal-frontal correlations over parietal-parietal correlations at all possible temporal co-ordinates. Regions where parietal-parietal correlations are superior are masked with blue (for example the diagonal from top left to bottom right where the parietal-parietal correlation must be 1). Regions where the parietal-frontal correlation is non-significant are also masked out. The rectangle encloses co-ordinates where both frontal and parietal effect are in operation (see Figure 2): if a signal is relayed between sites, correlations should be strong at these co-ordinates.

5.5 Discussion

5.5.1 Results summary

This meta-analysis has established the reality of a parietal encoding not just of RPE-sign, but RPE-utility, with better outcomes in both cases associated with a relative positivity in voltage. The existence of this parietal effect has been unresolved in the literature, with the comprehensive review of San Martín (2012), concluding a parietal valence effect was likely, but that its direction was unclear. Unfortunately, such a conclusion is very possible when attempting to collate effects taken from inconsistent intervals of the feedback-locked waveform since it comprises a series of peaks and troughs and small shifts in the

measurement window can reverse the polarity of effects. The great grand averaging method used here reveals the underlying direction of effect and its latency.

One benefit of a GGA meta-analysis is that it may indicate an appropriate interval in which to measure a component in the future, as Sambrook and Goslin (2015) provided for the FRN. This is somewhat problematic in the present case because the parietal effect of valence*size overlaps closely with the much stronger, and same-signed, frontal effect. Indeed, we cannot rule out the possibility that the earlier portion of the parietal effect is entirely due to volume conduction from the frontal effect. However, this cannot be the case in the interval in which the parietal response is stronger. On a pragmatic basis, we therefore recommend the parietal RPE-utility encoding component be quantified as mean activity in the interval 350 – 420 ms, this corresponding to the interval in which our meta-analysis found the parietal effect to be stronger than the frontal effect and still significant (as shown in Figure 2b). Future studies that better separate the two components will be used to refine this.

5.5.2 Limitations of difference waves

We should clarify that the difference wave approach we have used does not allow us to conclude whether the RPE valence effect arises from sensitivity (in opposing directions) to both good and bad outcomes, or from sensitivity to just one valence. It also does not inform us whether the valence*size effect arises from a sensitivity to RPE size in both -RPEs and +RPEs, in other words reflecting a full bivalent encoding across the range of possible RPEs. The significant valence*size effect could be generated by an encoder that is sensitive to the size of +RPEs but not -RPEs or vice versa. Difference waves indicate that the brain performs discriminations along the lines of an experiment's conditions, but the observed wave is not necessarily representative of neural activity. This is true whether difference waves are built from great grand averages as in the present case, or from the grand averages found in a single experiment. Establishing the contribution of +RPEs or -RPEs to the observed valence*size effect depends on the interpretation of simple, single condition waveforms, however this is inherently problematic owing to the multiple components contributing to such waveforms (Luck, 2014) and problems arising from interpreting single condition waveforms in the specific context of the FRN have been illustrated by Sambrook and Goslin (2016). While this question can be addressed using methods for decomposing component overlap, these methods lie beyond the current study which thus restricts itself to difference waves.

5.5.3 Evidence of parietal RPE encoding

The parietal effects, particularly in the case of the valence*size effect, lie within the range of activity typically ascribed to the P3 component. An influential model (Yeung & Sanfey, 2004) claims that in reinforcement learning tasks such as those used in this study, the P3 reflects an encoding of only motivational salience, and the FRN reflects and encoding of only RPE-sign. This claim is not supported by

the present study. Nevertheless, we would be cautious in ascribing to the P3 the role of encoding all three properties of RPE valence, valence*size, and RPE size. First, it must be noted that there is too little spatial information in our meta-analysis to know whether the observed parietal effect is centred at a site representative of the P3 (e.g., Pz). Second, it must be borne in mind that the P3 is a large sustained potential evoked by many tasks. It appears to be broadly implicated in a cluster of related operations: context updating, surprising (e.g., oddball) events and salient and task relevant information processing (Courchesne et al., 1975; Donchin et al., 1975; Donchin & Coles, 1988). In practice however, given its large amplitude, the P3 probably has widespread sources (Lutzenberger, Elbert, & Rockstroh. 1987), and so multiple components doubtless occupy the P3 interval. For this reason, not all effects observed in that interval need be attributed to that component. In fact, it would be highly problematic to label the P3 solely as an RPE-utility encoder since, as shown in this study, (Figure 28), by Sambrook and Goslin (2015), by Yeung and Sanfey (2004) and by many others (San Martín, 2012) there is a strong parietal encoding of motivational salience, and this property is well aligned with the P3's sensitivity to oddball and salient stimuli. Since motivational salience is orthogonal to both RPE-sign and RPE-utility it is not possible for all these encodings to be carried out by a single neural generator. The temporally overlapping presence of all these encoders at the scalp implies temporally overlapping but spatially separate neural generators of the effect, and we thus limit our conclusions to there being a 'parietally expressed RPE-utility encoder'. Note that we can be confident that the valence*size effect does not simply arise from overlapping components encoding RPE-sign and motivational salience separately because these would sum rather than interact in the manner characteristic of an axiomatic RPE-utility encoder, as we detailed in the previous chapter.

5.5.4 Multiple RPE encoders

Notwithstanding the difficulty of assigning experimental effects to existing components, an important finding is the unequivocal demonstration of RPE-utility encoding beyond the FRN. Since this does not appear to be merely the relaying of earlier frontocentral activity, it raises the question of what a second RPE-utility encoder achieves. Previous studies have highlighted a role for the P3 in behavioural change (Chase et al., 2011; Zhang et al., 2013). This is the ultimate purpose of RPE computation: to update action values in the light of experience and promote optimal choice. It is thus possible that the parietal effect is a representation of RPEs in the context of action value updating. If this were the case however, it might be expected that the parietal RPE signal would be more dependent than the frontal RPE signal on the task being clearly controllable. However, no site x controllability interaction was found.

Alternatively, the parietal effect may reflect the encoding of a different kind of RPE. There has been a recent growth in studies investigating the neural correlates of model-free vs. model-based reinforcement learning. Model-free learning, as indexed in the tasks represented in this meta-analysis, merely entails the actor maintaining action values and updating them with RPEs. In model-based learning, the actor explicitly models environmental contingencies, updating these in the light of unexpected outcomes. RPEs are not

generally needed for model-based reinforcement learning. Nevertheless, research has provided evidence for neural encoding of model-based RPEs (Daw et al., 2011; Sambrook et al., 2018), and the parietal effect shown here may constitute such an encoding. It is not possible to test this hypothesis without a design that incorporates opportunities for both kinds of learning and which is furthermore sensitive to which kind of learning a participant is engaged in. In fact, the experiments used here do not support model-based learning insofar as they are "structureless" and a model-free process of tracking action values could not be improved upon. Nevertheless, we might expect model-based RPE encoding to occur insofar as participants are likely to have generated ad-hoc models to explain the pattern of outcomes they witness, using these to generate expectations and consequent RPEs.

A further possibility that must be considered is that the parietal RPE-utility encoder (and indeed the frontocentral one) is not an RPE in the strict computational sense at all, but some other construct that correlates with this quantity. Model-free learning can be thought of as a simple mathematical model of propositional reasoning (Mitchell et al., 2009). Alternatively, RPE effects may be better explained by episodic memory formation (Gershman & Daw, 2017; Vikbladh et al., 2017). In both these cases, computational models employing quite different psychological constructs can lead to the same predictions as those from reinforcement learning, making them difficult to distinguish. Rather than reflecting methodological limitations, this may aptly describe the underlying processes however, since model-based reinforcement learning is likely to be continuous with general cognition (Chater, 2009).

5.5.5 Chapter conclusions

The present study presents evidence for the parietal encoding of RPEs. It demonstrates sensitivity not merely to whether outcomes are good or bad, but the degree to which this is so, thus complying with an axiomatic test for RPE-utility encoding. It uses the great grand averaging technique to repurpose scientific articles that had, as their principal aim, the study of the FRN. While it is not the first study to assess encoding of parietal RPEs (and certainly not the first to assess the P3's response to RPE valence), it enjoys the benefits of a large sample size and incorporates sufficient spread in tasks and other experimental details to ensure the demonstrated effect is robust. By clarifying the presence and time-course of parietal RPE-utility encoding it is hoped that future studies will be able to use this information to better interpret feedback-locked ERPs to begin to reveal the circuitry underlying human reinforcement learning.

6. General Discussion

Recent literature investigating the neural foundation of human reinforcement learning widely supports the case for RPEs, but varying approaches to study methodologies, and the problems of disambiguating overlapping electrophysiological components, leaves the specific manner in which RPEs are encoded as a mystery. Throughout this thesis, we aimed to address these issues in order to answer the specific question of if there is a general-purpose RPE encoder, consistently responsive for all manner of domains, reinforcers and RPE size modulators in the reinforcement learning waveform. This chapter looks to finally answer this question, providing a breakdown on both where the current literature leaves us in regard to the possibility of general-purpose RPE encoding, and the additional insight this thesis grants through the research and discussions within. Moreover, we discuss the findings of additional encoders suggested to be present within the waveform, with particular consideration to the roles they have alongside the general-purpose RPE encoder we suggest being present.

6.1. Evidence of General-purpose RPE Encoding

6.1.1 RPE or salience

As stated at the beginning of this thesis, Thorndike's law of effect dictates that decision-making is driven by adaptive control, and recent neuroscience has provided strong grounds to believe that a key part of this adaptive control, reinforcement learning, is driven by the encoding of RPEs (Schultz et al., 1997; Sutton, 1988). Pivotal research by Holroyd and Coles (2002) demonstrated evidence for such a process in the FRN, a component seemingly representative of RPE encoding that originates in the ACC. In light of our own research, we largely demonstrate support for the case that the FRN is best attributed to the encoding of a neural RPE, and that this encoding most likely takes the form of RPE-sign, rather than RPE-utility or mere motivational salience encoding. The following subsections clarify this viewpoint through the insight we have gained specifically from studying participant agency and the manipulation of primary or secondary reinforcers.

6.1.1.1 Insights from agency

The level of control an agent has over an outcome has been a prominent topic in the reinforcement learning literature (Hassall et al., 2019; Yeung et al., 2005). While we expect RPEs to be largely tied to instrumental learning, it is generally understood that humans and other animals also have the ability to learn from others' actions rather than their own, and from the absence of action altogether (Bellebaum et al., 2010; Rescorla & Wagner, 1972). Learning from such situations may be handled by separate systems, and this

view has some research support (Martin & Potts, 2011; Walsh & Anderson, 2012). However, it may also be that there is a single system insensitive to the level of agency an individual has, and this is indeed expected for a general-purpose RPE encoder. This thesis largely supports the view that participant agency has a profound effect on the generation of RPEs. This is shown in chapter 2, where we revealed a significant signal consistent with RPE encoding in the interval of the FRN during active conditions and a much smaller signal during passive conditions, similar to the previous findings demonstrated by Yeung et al. (2005).

If the weaker passive RPE signal we identified was to represent a single general-purpose RPE encoder insensitive to agency, rather than a separate RPE encoder specifically responsive to passive RPEs, it raises the important question of why this signal was smaller for passive conditions. This is a particularly interesting question considering that our results from chapters 3 and 4 point largely towards this encoder being that of RPE-sign, which is an encoding partially defined by a lack of response to RPE size and thus not varying in FRN amplitude. In this manner, an RPE-sign encoder should produce a signal of equal amplitude for a +RPE of 50p, for example, regardless of whether the participant was active or passive in a decision-making task. The attenuation of the passive condition signal may thus be due to additional features of reinforcement learning. As an example, motivation has been found to contribute to the amplitude of the FRN, with the difference in participants' interest in tasks being predictive of the difference in amplitude of their FRNs for those tasks (Yeung et al., 2005). This effect ties into our previous predictions about the FRN amplitude in chapter 2, where RPE encoding can be expected to be reduced or absent in passive events due to a lack of instrumental learning. In the case of our own results, it is understandable that participants may have had a reduced interest in the task when they were simply passive observers, and it cannot be ruled out that this may impact the amplitude of the FRN, and indeed be why past studies employing passive designs have failed to support the RPE case for the FRN (Talmi et al., 2013; Hird et al., 2018; Soder & Potts, 2018). A reduction in the amplitude of the signal produced by the RPE-sign encoder is then expected to consequentially lead to an apparent inflation of overlapping motivational salience effects that are expected for all motivating events; a result that is indeed seen in the aforementioned research as well as in our findings in chapter 2. Overall, this effect, and the similarity in latency of the appetitive and aversive FRNs also observed in chapter 2, leads us to suggest that there is indeed a single RPE encoder that is responsive to events regardless of participant agency, and that while still present, the amplitude of the FRN is only reduced in passive designs because of the lack of motivation in the absence of instrumental learning.

6.1.1.2 Insights from primary/secondary reinforcement

Another requirement for a general-purpose RPE encoder is that it should be insensitive to whether reinforcers are primary or secondary. Studies of midbrain dopamine neuron responses have found phasic firing to increase or dip for positive and negative RPEs respectively following both primary and secondary reinforcement (Matsumoto & Hikosaka, 2009; Sutton & Barto, 1998), and this behaviour should be observable in the EEG waveform if there is a general-purpose RPE encoder responsible for the signal. This

thesis generally supports this view, with the results of chapter 3 revealing a significant RPE-sign response in the interval of the FRN when locked to the delivery of primary reinforcement, supporting the RPE-sign response typically observed in the wider literature that predominantly uses secondary reinforcement. The signal in chapter 3 peaked earlier than this typical cue-locked FRN however, indicating some level of difference between primary and secondary reinforcement learning. As we stated in chapter 3, there should not be such a difference in latency for primary or secondary reinforcement learning in line with TD learning theory, as RPE signals initialised at the receipt of any primary reinforcement propagate back to any conditioned stimuli that predict them.

It is possible that this difference could be due to preliminary processes, or indeed the addition of later cue-based processes, also needed for cue-based learning that lie outside of the general-purpose RPE encoder present for both primary and secondary reinforcers. For example, it may be that cue-based tasks require the additional processing of cue associations stored in working memory prior to RPE generation (McDougle et al., 2020). Such an explanation is plausible, and reinforces the case for a single, general-purpose RPE encoder being present insofar as it is not the encoder itself that is differing in latency, merely it is overlain by additional, non-RPE oriented processes within cue-based learning tasks. Furthermore, the fact that the FRN we observed is topographically similar to that found in the other chapters of this thesis and across the cue-based literature is indicative of a single encoder responsive to both primary and secondary reinforcers.

6.1.2 Insensitivity to domains

An RPE is defined as the difference between an outcome's actual value and its expected value prior to receipt and, as such, its valence is independent of outcome domain. For a general-purpose RPE, it is thus expected that domain should have no effect on the waveform. However, at the forefront of debate, and of interest to this thesis, is the demonstration of the FRN representing increased sensitivity to appetitive domains over aversive domains (Walsh & Anderson, 2012), as well as the finding that studies investigating in only the aversive domain have found the FRN to represent motivational salience encoding rather than RPE encoding (Talmi et al., 2013; Hird et al., 2018; Soder & Potts, 2018). In spite of these findings, the current research supports consistent RPE encoding across both appetitive and aversive domains in the time interval of the FRN.

Firstly, in chapter 2 we demonstrated a similar -RPE – +RPE difference wave for both white noise and monetary reward RPEs, countering the claim that the FRN represents an encoder that is more sensitive to the appetitive domain. Importantly, this difference wave was also of the same polarity for the aversive white noise trials and the appetitive monetary trials, consistent with RPE-sign encoding rather than motivational salience encoding. Secondly, our meta-analysis findings in chapter 4 also demonstrated an RPE-sign encoder that is insensitive to outcome domain. Initially, the FRN was found to be reduced in the aversive domain compared to the appetitive domain, with this being explained by either the RPE encoder being neglectful of

aversive domain RPEs, or instead because of an overlapping motivational salience encoder summing with the appetitive FRN, but cancelling out the aversive FRN. The possibility of RPE encoding being neglectful of the aversive domain was ruled out following the manipulation of blocking type, whereby no difference between the amplitudes of the appetitive FRN – aversive FRN difference wave for blocked and unblocked designs was found. Following this, the reduced aversive FRN effect was more likely due to a larger, overlapping motivational salience signal in studies using concrete cues compared to a smaller signal in studies using abstract cues; a finding we observed following analyses investigating the effect of cue type the post-feedback waveform. This effect is due to the positivity generated from motivational salience encoding being greater for concrete cues relative to abstract cues, leading to a larger cancelling of the relative negativity generated by RPE-sign encoding in aversive domains and ultimately leading to the reduced aversive FRN we originally observed. Lastly, in chapter 3 we revealed a correctly-signed FRN following primary aversive reinforcement from electric shocks, contrary to the reversed FRN found in previous electric shock research (Talmi et al., 2013). This is also supported by our results in chapter 2, where the motivational salience account can be largely attributed to the use of a passive agency design rather than as a result of aversive reinforcement.

As it stands the reduction, or indeed absence, of aversive RPE encoding is likely due to external factors, and thus we propose the presence of a general-purpose RPE encoder that is correctly insensitive to outcome domain. However, that does not mean outcome domain should not be considered as a variable in future studies. Rather, we express the manipulation of outcome domain as a necessity for the investigation of potential general-purpose RPE encoding for two key reasons. Firstly, while not strictly required for the operationalisation of RPEs, studying the response across both appetitive and aversive domains allows for the confirmation of a null effect of domain when it is predicted, as in the instance of general-purpose RPE encoding. It cannot be known that an encoder is correctly insensitive to outcome domain unless tests across both domains are run and amplitudes are found to be equal. Secondly, we have shown in chapter 4 that the inclusion of domain as a variable allows for better disambiguation of RPE-sign and utility encoders based on their context dependencies. Unlike with RPE-utility encoding, the valence of events for RPE-sign encoding is affected by experiment form, but outcome domain remains consistent regardless of this, making it an ideal variable for isolating context-free and context-dependent encoders.

6.1.3 Insensitivity to RPE size modulation

One characteristic of RPEs is that they can occur not only from changes in outcome magnitude, but also changes in outcome expectancy. Considering this, a final rule for general-purpose RPE encoding is that it should be insensitive to these outcome size modulators, whereby it is responsive to RPEs consisting of both changes in outcome magnitude and expectancy. The literature has been generally supportive of the case that the FRN is responsive to changes in outcome expectancy, but the case for magnitude has proven inconclusive (San Martín, 2012). In line with recent meta-analysis findings (Sambrook and Goslin, 2015)

and the classic RL theory, the results of this thesis provide support that the RPE encoder in the interval of the FRN does indeed encode RPEs modulated by changes in magnitude as well as expectancy, and such is insensitive in the manner expected of a general-purpose RPE encoder.

Our meta-analysis in chapter 4 revealed both an RPE-sign encoder, consistent with general-purpose RPE encoding, as well as a later RPE-utility encoder, from a sample of studies manipulating RPE size through magnitude only. Furthermore, our second meta-analysis in chapter 5 supported these findings and those of Sambrook and Goslin (2015) using a sample of studies including both magnitude and expectancy as possible RPE size modulators, with a significant RPE valence effect and significant valence*size effect both found for the frontal waveform. We must note that our results in support of general-purpose RPE encoding being present regardless of RPE size modulation are limited to these few findings, as lab studies manipulating this variable were not covered in this thesis due to a focus on largely outcome magnitude only. Furthermore, while we included a record of RPE size modulation in chapter 5, effects were not analysed as RPE size modulation was not of direct interest. In the wider literature however, studies employing designs including the manipulation of both outcome expectancy and magnitude have fared well in establishing the FRN as insensitive to such modulation (Luo & Qu, 2013; Wu & Zhou, 2009), giving further support to the claim that the underlying RPE-sign encoder we have identified at this interval may represent a general-purpose RPE encoder. We propose further research manipulating RPE size through both expectancy and magnitude in this manner will benefit our understanding of the effects of RPE size modulation on the proposed general-purpose RPE encoder.

6.2 Additional Encoding

The above discussion considers the wide evidence we have attained for making the case that a general-purpose RPE encoder, in the form of RPE-sign, is present in the time interval associated with the FRN. In addition to this however, we have also found consistent evidence of both RPE-utility and motivational salience encoders which, while not necessarily general-purpose, may support additional reinforcement learning processes within human decision-making.

6.2.1 Evidence of motivational salience

A common finding in the FRN literature is the presence of motivational salience encoding, which in some cases has been identified as the encoding most representative of the FRN itself (Talmi et al., 2013; Hird et al., 2018) and, more recently, as an additional overlapping encoder that introduces compounding effects on the waveform (Sambrook & Goslin, 2016). The findings in this thesis are supportive of the latter, with evidence of overlapping motivational salience found throughout our research. In chapter 2 we found an increased latency of the aversive domain FRN (occurring around 270 – 320 ms) compared to the appetitive domain FRN (occurring around 210 – 270 ms). This is an expected effect on the waveform in the event of

overlapping motivational salience summing with RPE-sign encoding in appetitive domains, but cancelling with it in aversive domains, leaving the true interval of encoding lying somewhere between. In chapter 4 we were able to identify such an instance of overlapping encoders and isolate them, with evidence for RPE-sign encoding and motivational salience encoding being found at around 280 ms post-feedback. This directly supports our findings in chapter 2 if we consider the true interval of RPE encoding to lie in between the appetitive and aversive FRNs, and indeed supports similar findings in the wider literature (Holroyd et al., 2008; Foti et al., 2011). Chapter 3 revealed no such evidence of possible motivational salience encoding, however the use of only the aversive domain makes it impossible to rule out this encoding as being present with the approach used; a design manipulating RPE size across both appetitive and aversive domains is required to distinguish between RPE encoding via voltage positivity and motivational salience encoding via voltage negativity.

The presence of motivational salience encoding is to be expected in the reinforcement learning waveform and is supported by a great deal of research (Garofalo et al., 2014; Hauser et al., 2014; Pfabigan et al., 2015; Sallet et al., 2013). Motivational salience has strong ties to reinforcement learning in that rewarding or punishing events have an essential motivational relevance tied to their occurrence (Schultz, 2016). It's presence however does remain an issue in the search for any underlying RPE processes, owing to the difficulty of disambiguating the waveform into its constituent components and the compounding effects they have. We thus recommend consideration of overlapping salience effects when studying the waveform in the future, particularly in designs using concrete or preconditioned cues, which we have shown to contribute to a considerable strengthening of the positivity produced by the overlapping motivational salience encoder in chapter 4. As such, it would be beneficial for future research to manipulate cue-type, or perhaps employ purely abstract cues, so that overlapping motivational salience encoding can be reduced at the analysis stage. Furthermore, the use of component isolation techniques such as PCA have proven to be highly valuable in this thesis and in the past (Holroyd et al., 2008; Foti et al., 2011), so we recommend the inclusion of such methods when attempting to decompose the reinforcement learning waveform.

6.2.2 Evidence of RPE-utility

When considering the adaptability of human decision-making there are also strong grounds for the neural encoding of RPE-utility, where both the valence and size of events are processed on a continuous scale. This has been shown in human behavioural adaptation, whereby the degree of behavioural adjustment has been found to correlate with the amplitude of the FRN (Chase et al., 2011; Luu et al., 2003). Such processing can be considered a necessity for RL-theory and the idea of a dopaminergic basis for RPEs whereby increasingly better outcomes should result in higher dopamine release (Caplin and Dean, 2008). While RPE-utility has rarely been identified as the dominant encoding represented by the typical FRN, a separate component responsible for RPE-utility has emerged numerous times in recent literature at around 320 – 340 ms post-feedback at the frontocentral scalp (Cavanagh, 2015; Sambrook & Goslin, 2016). The

results gathered in this thesis are largely supportive of such a component being present in the reinforcement learning waveform at this interval. Firstly, in chapter 4 we demonstrated a likely RPE-utility encoder at around 320 ms following Bayesian contrasts analysis, overlapping with the general-purpose RPE-sign encoder found earlier in the waveform, and the motivational salience encoder found later. Second, there is evidence of RPE-utility encoding found in chapter 5, where a significant RPE-utility effect was identified in the frontal waveform peaking at around 300 ms. Lastly, our lab study in chapter 2 revealed a possible RPE-utility response, albeit occurring later at around 370 ms and demonstrating a greater response to +RPEs compared to -RPEs.

The specific role of an overlaying RPE-utility encoder in reinforcement learning is of great interest, especially considering the already-present encoding of both RPE valence and size handled by the other encoders in the waveform. If RPE valence is encoded primarily by the general-purpose RPE-sign encoder in the interval of the FRN, and RPE size through the temporally overlapping motivational salience encoding, then the additional RPE-utility encoder must provide an additional role in reinforcement learning that is present for different events. Importantly there is the consideration of context-dependency on RPE encoding; as we covered in chapter 4, RPE-utility encoding is distinct from RPE-sign encoding not only in its response to RPE valence and size, but also in its response being context-free rather than context-dependent. Thus, it would be unsurprising for there to be a requirement of both RPE-sign and RPE-utility encoding in the brain for the more adaptive learning and behavioural adjustment discussed at the start of this section. Specifically, there is value in an agent understanding an outcome as being better or worse than expected in one given scenario, but also understanding it as being better or worse than alternative outcomes available to them in different scenarios. We can expect both RPE encoders to work in tandem, with the initial general-purpose RPE-sign encoder responding to the trial at hand, and the later RPE-utility encoder responding instead to the wider context.

6.3 Thesis conclusions

As we have covered throughout this chapter, extra considerations need to be made when studying RPE encoders in the waveform in order to avoid their misinterpretation, and we end this thesis by summarising the ways in which this can be done. Firstly, we propose the application of both Bayesian analysis and standard operationalisations in order to gain both relative and actual evidence for effects indicative of RPE encoding. The reasoning for this is that component overlap can lead to inaccuracy in the assignment of encoders that share effects when using only standard operationalisations, as typically used throughout the literature. While the relative evidence provided by Bayesian contrasts analysis is also limited in that legitimate encoders can appear diminished when they occur at the same interval as a stronger encoder, using both methods allows for both sets of evidence to support the case for each other, and overall strengthen the interpretation of results.

Secondly, future research should take heed of the many, and in some cases seldom-explored, additional variables we have found as relevant to disambiguating general-purpose RPE encoding; outcome domain, participant agency, primary reinforcement, blocking type, cue type, and perhaps most notably experiment form. All such variables can influence the amplitudes of different encoders in different ways, and researchers may wish to consider the manipulation or control of these variables when designing experiments to best filter the encoder, or encoders, under investigation. Once again, this difficulty of disambiguation stems largely from component overlap, and our results have shown how simple consideration of events that are, say, limited to a passive participant agency or aversive domain stimulus can lead to either identifying general-purpose RPE encoding in the waveform, or not identifying any kind of RPE at all.

In all, the present research has utilised a multitude of designs and analytical methods in an attempt to better disambiguate the underlying encoders that comprise reinforcement learning. The EEG literature surrounding this topic to date has been fraught with inconsistency and such has led to wide debate; an unfortunate but expected outcome considering the vast array of study methodologies available with which to study reinforcement learning discussed above, as well as the complicating omnipresence of overlapping components. We have sought through the use of multiple methods, and comparing results across the existing literature, for evidence of RPE encoding in the reinforcement learning waveform in order to answer the one key question of if there is the presence of a universal, or general-purpose, RPE encoder. When taken together, our results demonstrate an RPE-sign encoder, in the time interval and typical scalp site of the FRN, that is indeed consistent with such encoding.

7. References

Aarts, K., & Pourtois, G. (2012). Anxiety disrupts the evaluative component of performance monitoring: An ERP study. *Neuropsychologia*, *50*(7), 1286–1296. https://doi.org/10.1016/j.neuropsychologia.2012.02.012

Andreatta, M., Michelmann, S., Pauli, P., & Hewig, J. (2017). Learning processes underlying avoidance of negative outcomes. *Psychophysiology*, *54*(4), 578–590. https://doi.org/10.1111/psyp.12822

Andreatta, M., & Pauli, P. (2015). Appetitive vs. Aversive conditioning in humans. *Frontiers in Behavioral Neuroscience*, *9*. https://www.frontiersin.org/article/10.3389/fnbeh.2015.00128

Balleine, B. W., Delgado, M. R., & Hikosaka, O. (2007). The Role of the Dorsal Striatum in Reward and Decision-Making. *Journal of Neuroscience*, *27*(31), 8161–8165. https://doi.org/10.1523/JNEUROSCI.1554-07.2007

Banis, S., & Lorist, M. M. (2012). Acute noise stress impairs feedback processing. *Biological Psychology*, *91*(2), 163–171. psyh. https://doi.org/10.1016/j.biopsycho.2012.06.009

Bellebaum, C., & Daum, I. (2008). Learning-related changes in reward expectancy are reflected in the feedback-related negativity. *European Journal of Neuroscience*, *27*(7), 1823–1835. https://doi.org/10.1111/j.1460-9568.2008.06138.x

Bellebaum, C., Kobza, S., Thiele, S., & Daum, I. (2010). It was not MY fault: Event-related brain potentials in active and observational learning from feedback. *Cerebral Cortex*, *20*(12), 2874–2883. psyh. https://doi.org/10.1093/cercor/bhq038

Bellebaum, C., Kobza, S., Thiele, S., & Daum, I. (2011). Processing of expected and unexpected monetary performance outcomes in healthy older subjects. *Behavioral Neuroscience*, *125*(2), 241–251. https://doi.org/10.1037/a0022536

Bellebaum, C., Polezzi, D., & Daum, I. (2010). It is less than you expected: The feedback-related negativity reflects violations of reward magnitude expectations. *Neuropsychologia*, *48*(11), 3343–3350. https://doi.org/10.1016/j.neuropsychologia.2010.07.023

Berridge, K. C., & Robinson, T. E. (1998). What is the role of dopamine in reward: Hedonic impact, reward learning, or incentive salience? *Brain Research Reviews*, *28*(3), 309–369. https://doi.org/10.1016/S0165-0173(98)00019-8

Blumenthal, T. D., Cuthbert, B. N., Filion, D. L., Hackley, S., Lipp, O. V., & Van Boxtel, A. (2005). Committee report: Guidelines for human startle eyeblink electromyographic studies. *Psychophysiology*, *42*(1), 1–15. https://doi.org/10.1111/j.1469-8986.2005.00271.x

Boudewyn, M. A., Luck, S. J., Farrens, J. L., & Kappenman, E. S. (2018). How many trials does it take to get a significant ERP effect? It depends. *Psychophysiology*, *55*(6), e13049. https://doi.org/10.1111/psyp.13049

Boureau, Y.-L., & Dayan, P. (2011). Opponency Revisited: Competition and Cooperation Between Dopamine and Serotonin. *Neuropsychopharmacology*, *36*(1), 74–97. https://doi.org/10.1038/npp.2010.151

Bromberg-Martin, E. S., Matsumoto, M., & Hikosaka, O. (2010). Dopamine in Motivational Control: Rewarding, Aversive, and Alerting. *Neuron*, *68*(5), 815–834. https://doi.org/10.1016/j.neuron.2010.11.022

Brown, C. A., Seymour, B., Boyle, Y., El-Deredy, W., & Jones, A. K. P. (2008). Modulation of pain ratings by expectation and uncertainty: Behavioral characteristics and anticipatory neural correlates. *PAIN*, *135*(3), 240–250. https://doi.org/10.1016/j.pain.2007.05.022

Brown, D. R., & Cavanagh, J. F. (2018). Rewarding images do not invoke the reward positivity: They inflate it. *International Journal of Psychophysiology*, *132*, 226–235. https://doi.org/10.1016/j.ijpsycho.2018.02.012

Brown, P. L., & Jenkins, H. M. (1968). Auto-Shaping of the Pigeon's Key-Peck1. *Journal of the Experimental Analysis of Behavior*, *11*(1), 1–8. https://doi.org/10.1901/jeab.1968.11-1

Byrd, R. H., Lu, P. H., Nocedal, J., Zhu, C. Y., (1995). A Limited Memory Algorithm for Bound Constrained Optimization. Siam J Sci Comput. 16:1190-1208.

Caplin, A., & Dean, M. (2008). Axiomatic methods, dopamine and reward prediction error. *Current Opinion in Neurobiology*, *18*(2), 197–202. https://doi.org/10.1016/j.conb.2008.07.007

Cardinal, R. N., Parkinson, J. A., Hall, J., & Everitt, B. J. (2002). Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex. *Neuroscience & Biobehavioral Reviews*, *26*(3), 321–352. https://doi.org/10.1016/S0149-7634(02)00007-6

Carlson, J. M., Foti, D., Mujica-Parodi, L. R., Harmon-Jones, E., & Hajcak, G. (2011). Ventral striatal and medial prefrontal BOLD activation is correlated with reward-related electrocortical activity: A

combined ERP and fMRI study. *NeuroImage*, *57*(4), 1608–1616. https://doi.org/10.1016/j.neuroimage.2011.05.037

Cavanagh, J. F. (2015). Cortical delta activity reflects reward prediction error and related behavioral adjustments, but at different times. *NeuroImage*, *110*, 205–216. https://doi.org/10.1016/j.neuroimage.2015.02.007

Cazé, R. D., & van der Meer, M. A. A. (2013). Adaptive properties of differential learning rates for positive and negative outcomes. *Biological Cybernetics*, *107*(6), 711–719. https://doi.org/10.1007/s00422-013-0571-5

Chase, H. W., Swainson, R., Durham, L., Benham, L., & Cools, R. (2011). Feedback-related Negativity Codes Prediction Error but Not Behavioral Adjustment during Probabilistic Reversal Learning. *Journal of Cognitive Neuroscience*, *23*(4), 936–946. https://doi.org/10.1162/jocn.2010.21456

Chater, N. (2009). Rational and mechanistic perspectives on reinforcement learning. *Cognition*, *113*(3), 350–364. https://doi.org/10.1016/j.cognition.2008.06.014

Cohen, M. X., Cavanagh, J. F., & Slagter, H. A. (2011). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity: Commentary. *Human Brain Mapping*, *32*(12), 2270–2271. https://doi.org/10.1002/hbm.21358

Cohen, M. X., Elger, C. E., & Ranganath, C. (2007). Reward Expectation Modulates Feedback-Related Negativity and EEG Spectra. *NeuroImage*, *35*(2), 968–978. https://doi.org/10.1016/j.neuroimage.2006.11.056

Core Team R. 2020. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Courchesne, E., Hillyard, S. A., & Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. *Electroencephalography and Clinical Neurophysiology*, *39*(2), 131–143. https://doi.org/10.1016/0013-4694(75)90003-6

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-Based Influences on Humans' Choices and Striatal Prediction Errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Delgado, M. R., Phelps, E. A., & Robbins, T. W. (2011). *Decision Making, Affect, and Learning: Attention and Performance XXIII*. OUP Oxford.

Den Ouden, H., Kok, P., & De Lange, F. (2012). How Prediction Errors Shape Perception, Attention, and Motivation. *Frontiers in Psychology*, *3*. https://www.frontiersin.org/article/10.3389/fpsyg.2012.00548

Dien, J. (2010a). Evaluating two-step PCA of ERP data with Geomin, Infomax, Oblimin, Promax, and Varimax rotations. *Psychophysiology*, *47*(1), 170–183. https://doi.org/10.1111/j.1469-8986.2009.00885.x

Dien, J. (2010b). The ERP PCA Toolkit: An open source program for advanced statistical analysis of event-related potential data. *Journal of Neuroscience Methods*, *187*(1), 138–145. https://doi.org/10.1016/j.jneumeth.2009.12.009

Dien, J., Beal, D. J., & Berg, P. (2005). Optimizing principal components analysis of event-related potentials: Matrix type, factor loading weighting, extraction, and rotations. *Clinical Neurophysiology*, *116*(8), 1808–1825. https://doi.org/10.1016/j.clinph.2004.11.025

Dien, J., Khoe, W., Mangun, G.R., 2007. Evaluation of PCA and ICA of simulated ERPs: promax vs. infomax rotations. Hum. Brain Mapp. 28 (8), 742–763.

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(3), 357–374. https://doi.org/10.1017/S0140525X00058027

Donchin, E., Tueting, P., Ritter, W., Kutas, M., & Heffley, E. (1975). On the independence of the CNV and the P300 components of the human averaged evoked potential. *Electroencephalography & Clinical Neurophysiology*, *38*(5), 449–461. psyh. https://doi.org/10.1016/0013-4694(75)90187-X

Duval, S., & Tweedie, R. (2000). Trim and Fill: A Simple Funnel-Plot–Based Method of Testing and Adjusting for Publication Bias in Meta-Analysis. *Biometrics*, *56*(2), 455–463. https://doi.org/10.1111/j.0006-341X.2000.00455.x

Eppinger, B., Mock, B., & Kray, J. (2009). Developmental differences in learning and error processing: Evidence from ERPs. *Psychophysiology*, *46*(5), 1043–1053. https://doi.org/10.1111/j.1469-8986.2009.00838.x

Foti, D., Weinberg, A., Dien, J., & Hajcak, G. (2011). Event-related potential activity in the basal ganglia differentiates rewards from nonrewards: Temporospatial principal components analysis and source localization of the feedback negativity. *Human Brain Mapping*, *32*(12), 2207–2216. https://doi.org/10.1002/hbm.21182

Fouragnan, E., Retzler, C., Mullinger, K., & Philiastides, M. G. (2015). Two spatiotemporally distinct value systems shape reward-based learning in the human brain. *Nature Communications*, *6*(1), 8107. https://doi.org/10.1038/ncomms9107

Fujiwara, J., Tobler, P. N., Taira, M., Iijima, T., & Tsutsui, K.-I. (2009). Segregated and Integrated Coding of Reward and Punishment in the Cingulate Cortex. *Journal of Neurophysiology*, *101*(6), 3284–3293. https://doi.org/10.1152/jn.90909.2008

Garofalo, S., Maier, M. E., & di Pellegrino, G. (2014). Mediofrontal negativity signals unexpected omission of aversive events. *Scientific Reports*, *4*, 4816. https://doi.org/10.1038/srep04816

Gehring, W. J., & Willoughby, A. R. (2002). The Medial Frontal Cortex and the Rapid Processing of Monetary Gains and Losses. *Science*, *295*(5563), 2279–2282.

Gershman, S. J., & Daw, N. D. (2017). Reinforcement Learning and Episodic Memory in Humans and Animals: An Integrative Framework. *Annual Review of Psychology*, *68*(1), 101–128. https://doi.org/10.1146/annurev-psych-122414-033625

Geurts, D. E. M., Huys, Q. J. M., den Ouden, H. E. M., & Cools, R. (2013). Aversive Pavlovian Control of Instrumental Behavior in Humans. *Journal of Cognitive Neuroscience*, *25*(9), 1428–1441. https://doi.org/10.1162/jocn_a_00425

Glazer, J. E., Kelley, N. J., Pornpattananangkul, N., Mittal, V. A., & Nusslock, R. (2018). Beyond the FRN: Broadening the time-course of EEG and ERP components implicated in reward processing. *International Journal of Psychophysiology*, *132*(Part B), 184–202. psyh. https://doi.org/10.1016/j.ijpsycho.2018.02.002

Gu, R., Lei, Z., Broster, L., Wu, T., Jiang, Y., & Luo, Y. (2011). Beyond valence and magnitude: A flexible evaluative coding system in the brain. *Neuropsychologia*, *49*(14), 3891–3897. psyh. https://doi.org/10.1016/j.neuropsychologia.2011.10.006

Gu, Y., Liu, T., Zhang, X., Long, Q., Hu, N., Zhang, Y., & Chen, A. (2021). The Event-Related Potentials Responding to Outcome Valence and Expectancy Violation during Feedback Processing. *Cerebral Cortex*, *31*(2), 1060–1076. https://doi.org/10.1093/cercor/bhaa274

Hajcak, G., Holroyd, C. B., Moser, J. S., & Simons, R. F. (2005). Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology*, *42*(2), 161–170. https://doi.org/10.1111/j.1469-8986.2005.00278.x

Hajcak, G., Moser, J. S., Holroyd, C. B., & Simons, R. F. (2007). It's worse than you thought: The feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology, 44*(6), 905–912. https://doi.org/10.1111/j.1469-8986.2007.00567.x

Hammerstrom, M. R., Ferguson, T. D., Williams, C. C., & Krigolson, O. E. (2021). What happens when right means wrong? The impact of conflict arising from competing feedback responses. *Brain Research, 1761*, 147393. https://doi.org/10.1016/j.brainres.2021.147393

Hassall, C. D., Hajcak, G., & Krigolson, O. E. (2019). The importance of agency in human reward processing. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(6), 1458–1466. https://doi.org/10.3758/s13415-019-00730-2

Hauser, T. U., Iannaccone, R., Stämpfli, P., Drechsler, R., Brandeis, D., Walitza, S., & Brem, S. (2014). The feedback-related negativity (FRN) revisited: New insights into the localization, meaning and network organization. *NeuroImage*, *84*, 159–168. https://doi.org/10.1016/j.neuroimage.2013.08.028

Heydari, S. (2015). . *Negativity: Reward Prediction Error or Salience Prediction Error?* [Thesis]. https://dspace.library.uvic.ca/handle/1828/5956

Heydari, S., & Holroyd, C. B. (2016). Reward positivity: Reward prediction error or salience prediction error? *Psychophysiology*, *53*(8), 1185–1192. https://doi.org/10.1111/psyp.12673

Hird, E. J., El-Deredy, W., Jones, A., & Talmi, D. (2018). Temporal dissociation of salience and prediction error responses to appetitive and aversive taste. *Psychophysiology*, *55*(2), e12976. https://doi.org/10.1111/psyp.12976

Hollerman, J. R., Tremblay, L., & Schultz, W. (1998). Influence of Reward Expectation on Behavior-Related Neuronal Activity in Primate Striatum. *Journal of Neurophysiology*, *80*(2), 947–963. https://doi.org/10.1152/jn.1998.80.2.947

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review*, *109*(4), 679–709. https://doi.org/10.1037/0033-295X.109.4.679

Holroyd, C. B., Krigolson, O. E., Baker, R., Lee, S., & Gibson, J. (2009). When is an error not a prediction error? An electrophysiological investigation. *Cognitive, Affective, & Behavioral Neuroscience*, *9*(1), 59–70. https://doi.org/10.3758/CABN.9.1.59

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., & Cohen, J. D. (2003). Errors in reward prediction are reflected in the event-related brain potential. *NeuroReport*, *14*(18), 2481–2484.

Holroyd, C. B., Pakzad-Vaezi, K. L., & Krigolson, O. E. (2008). The feedback correct-related positivity: Sensitivity of the event-related brain potential to unexpected positive feedback. *Psychophysiology*, *45*(5), 688–697.

Horn, J.L., 1965. A rationale and test for the number of factors in factor analysis. Psychometrika 30 (2), 179–185.

Houk, J. C., Davis, J. L., & Beiser, D. G. (1995). *Models of Information Processing in the Basal Ganglia*. MIT Press.

Howell, D. C. (2016). *Fundamental statistics for the behavioral sciences*. Cengage Learning.

Hu, X., Xu, Z., Li, Y., & Mai, X. (2018). The impact of trust decision-making on outcome processing: Evidence from brain potentials and neural oscillations. *Neuropsychologia*, *119*, 136–144. psyh. https://doi.org/10.1016/j.neuropsychologia.2018.07.036

Jensen, J., Smith, A. J., Willeit, M., Crawley, A. P., Mikulis, D. J., Vitcu, I., & Kapur, S. (2007). Separate brain regions code for salience vs. Valence during reward prediction in humans. *Human Brain Mapping*, *28*(4), 294–302. https://doi.org/10.1002/hbm.20274

Kamarajan, C., Porjesz, B., Rangaswamy, M., Tang, Y., Chorlian, D. B., Padmanabhapillai, A., Saunders, R., Pandey, A. K., Roopesh, B. N., Manz, N., Stimus, A. T., & Begleiter, H. (2009). Brain signatures of monetary loss and gain: Outcome-related potentials in a single outcome gambling task. *Behavioural Brain Research*, *197*(1), 62–76. https://doi.org/10.1016/j.bbr.2008.08.011

Kamin, L. J. (1967). Attention-like processes in classical conditioning. In *Symp. On Aversive Motivation Miami* (No. TR-5).

Koch, M., Schmid, A., & Schnitzler, H. U. (1996). Pleasure-attenuation of startle is disrupted by lesions of the nucleus accumbens. *Neuroreport*, *7*(8), 1442–1446. https://doi.org/10.1097/00001756-199605310-00024

Kreussel, L., Hewig, J., Kretschmer, N., Hecht, H., Coles, M. G. H., & Miltner, W. H. R. (2012). The influence of the magnitude, probability, and valence of potential wins and losses on the amplitude of the feedback negativity. *Psychophysiology*, *49*(2), 207–219. psyh. https://doi.org/10.1111/j.1469-8986.2011.01291.x

Li, P., Han, C., Lei, Y., Holroyd, C. B., & Li, H. (2011). Responsibility modulates neural mechanisms of outcome processing: An ERP study. *Psychophysiology*, *48*(8), 1129–1133. https://doi.org/10.1111/j.1469-8986.2011.01182.x

Liao, Y., Gramann, K., Feng, W., DeáK, G. O., & Li, H. (2011). This ought to be good: Brain activity accompanying positive and negative expectations and outcomes. *Psychophysiology*, *48*(10), 1412–1419. psyh. https://doi.org/10.1111/j.1469-8986.2011.01205.x

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, J., ... & Wagenmakers, E. J. (2019). JASP: Graphical statistical software for common statistical designs. *Journal of Statistical Software*, *88*, 1-17.

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique*. MIT Press.

Luo, Q., & Qu, C. (2013). Comparison enhances size sensitivity: Neural correlates of outcome magnitude processing. *PloS One*, *8*(8), e71186. https://doi.org/10.1371/journal.pone.0071186

Luque, D., López, F. J., Marco-Pallares, J., Càmara, E., & Rodríguez-Fornells, A. (2012). Feedback-related Brain Potential Activity Complies with Basic Assumptions of Associative Learning Theory. *Journal of Cognitive Neuroscience*, *24*(4), 794–808. https://doi.org/10.1162/jocn_a_00145

Lutzenberger, W., Elbert, T. H., & Rockstroh, B. (1987). A brief tutorial on the implications of volume conduction for the interpretation of the EEG. *Journal of Psychophysiology*, *1*, 81-89.

Luu, P., Tucker, D. M., Derryberry, D., Reed, M., & Poulsen, C. (2003). Electrophysiological Responses to Errors and Feedback in the Process of Action Regulation. *Psychological Science*, *14*(1), 47–53. https://doi.org/10.1111/1467-9280.01417

Marco-Pallares, J., Cucurell, D., Münte, T. F., Strien, N., & Rodriguez-Fornells, A. (2011). On the number of trials needed for a stable feedback-related negativity. *Psychophysiology*, *48*(6), 852–860. https://doi.org/10.1111/j.1469-8986.2010.01152.x

Marco-Pallarés, J., Krämer, U. M., Strehl, S., Schröder, A., & Münte, T. F. (2010). When decisions of others matter to me: An electrophysiological analysis. *BMC Neuroscience*, *11*(1), 86. https://doi.org/10.1186/1471-2202-11-86

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–190. https://doi.org/10.1016/j.jneumeth.2007.03.024

Martin, L. E., & Potts, G. F. (2011). Medial frontal event-related potentials and reward prediction: Do responses matter? *Brain and Cognition*, *77*(1), 128–134. https://doi.org/10.1016/j.bandc.2011.04.001

Martín, R. S. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Frontiers in Human Neuroscience*, *6*. psyh.

https://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=2012-33137-001&authtype=sso&custid=s8993828&site=ehost-live

Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, *459*(7248), 837–841. https://doi.org/10.1038/nature08028

McDougle, S. D., Ballard, I. C., Baribault, B., Bishop, S. J., & Collins, A. G. (2020). Executive modulation of brain reward systems endows goals with value. *bioRxiv*, 2020-10.

Meadows, C. C., Gable, P. A., Lohse, K. R., & Miller, M. W. (2016). The effects of reward magnitude on reward processing: An averaged and single trial event-related potential study. *Biological Psychology*, *118*, 154–160. psyh. https://doi.org/10.1016/j.biopsycho.2016.06.002

Metereau, E., & Dreher, J.-C. (2013). Cerebral Correlates of Salient Prediction Error for Different Rewards and Punishments. *Cerebral Cortex*, *23*(2), 477–487. https://doi.org/10.1093/cercor/bhs037

Miltner, W. H., Braun, C. H., & Coles, M. G. (1997). Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a 'generic' neural system for error detection. *Journal of Cognitive Neuroscience*, *9*(6), 788–798. https://doi.org/10.1162/jocn.1997.9.6.788

Mirenowicz, J., & Schultz, W. (1996). Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature*, *379*(6564), 449–451. https://doi.org/10.1038/379449a0

Mitchell, C. J., Houwer, J. D., & Lovibond, P. F. (2009). The propositional nature of human associative learning. *Behavioral and Brain Sciences*, *32*(2), 183–198. https://doi.org/10.1017/S0140525X09000855

Mogenson, G. J., Jones, D. L., & Yim, C. Y. (1980). From motivation to action: Functional interface between the limbic system and the motor system. *Progress in Neurobiology*, *14*(2), 69–97. https://doi.org/10.1016/0301-0082(80)90018-0

Mogenson, G. J., & Yang, C. R. (1991). The Contribution of Basal Forebrain to Limbic—Motor Integration and the Mediation of Motivation to Action. In T. C. Napier, P. W. Kalivas, & I. Hanin (Eds.), *The Basal Forebrain: Anatomy to Function* (pp. 267–290). Springer US. https://doi.org/10.1007/978-1-4757-0145-6_14

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor 0.9.11-1*. Comprehensive R Archive Network.

Mulligan, E. M., & Hajcak, G. (2017). The electrocortical response to rewarding and aversive feedback: The reward positivity does not reflect salience in simple gambling tasks. *International Journal of*

Psychophysiology: Official Journal of the International Organization of Psychophysiology. https://doi.org/10.1016/j.ijpsycho.2017.11.015

Nieuwenhuis, S., Holroyd, C. B., Mol, N., & Coles, M. G. H. (2004). Reinforcement-related brain potentials from medial frontal cortex: Origins and functional significance. *Neuroscience & Biobehavioral Reviews*, *28*(4), 441–448. https://doi.org/10.1016/j.neubiorev.2004.05.003

Nieuwenhuis, S., Ridderinkhof, K. R., Talsma, D., Coles, M. G. H., Holroyd, C. B., Kok, A., & van der Molen, M. W. (2002). A computational account of altered error processing in older age: Dopamine and the error-related negativity. *Cognitive, Affective, & Behavioral Neuroscience*, *2*(1), 19–36. https://doi.org/10.3758/CABN.2.1.19

Niv, Y., & Schoenbaum, G. (2008). Dialogues on prediction errors. *Trends in Cognitive Sciences*, *12*(7), 265–272. https://doi.org/10.1016/j.tics.2008.03.006

O'Doherty, J., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable Roles of Ventral and Dorsal Striatum in Instrumental Conditioning. *Science*, *304*(5669), 452–454. https://doi.org/10.1126/science.1094285

Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: An integrative review of ERP findings. *Biological Psychology*, *77*(3), 247–265. https://doi.org/10.1016/j.biopsycho.2007.11.006

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2010). FieldTrip: Open Source Software for Advanced Analysis of MEG, EEG, and Invasive Electrophysiological Data. *Computational Intelligence and Neuroscience*, *2011*, e156869. https://doi.org/10.1155/2011/156869

Pedroni, A., Langer, N., Koenig, T., Allemand, M., & Jäncke, L. (2011). Electroencephalographic Topography Measures of Experienced Utility. *Journal of Neuroscience*, *31*(29), 10474–10480. https://doi.org/10.1523/JNEUROSCI.5488-10.2011

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and Clinical Neurophysiology*, *72*(2), 184–187. https://doi.org/10.1016/0013-4694(89)90180-6

Pfabigan, D. M., Alexopoulos, J., Bauer, H., Lamm, C., & Sailer, U. (2011). All about the money—External performance monitoring is affected by monetary, but not by socially conveyed feedback cues in more antisocial individuals. *Frontiers in Human Neuroscience*, *5*. psyh. https://doi.org/10.3389/fnhum.2011.00100

Pfabigan, D. M., Seidel, E.-M., Paul, K., Grahl, A., Sailer, U., Lanzenberger, R., Windischberger, C., & Lamm, C. (2015). Context-sensitivity of the feedback-related negativity for zero-value feedback outcomes. *Biological Psychology*, *104*, 184–192. psyh. https://doi.org/10.1016/j.biopsycho.2014.12.007

Philiastides, M. G., Biele, G., Vavatzanidis, N., Kazzer, P., & Heekeren, H. R. (2010). Temporal dynamics of prediction error processing during reward-based decision making. *NeuroImage*, *53*(1), 221–232. https://doi.org/10.1016/j.neuroimage.2010.05.052

Pietrock, C., Ebrahimi, C., Katthagen, T. M., Koch, S. P., Heinz, A., Rothkirch, M., & Schlagenhauf, F. (2019). Pupil dilation as an implicit measure of appetitive Pavlovian learning. *Psychophysiology*, *56*(12), e13463. https://doi.org/10.1111/psyp.13463

Polich, J. (2003). Theoretical Overview of P3a and P3b. In J. Polich (Ed.), *Detection of Change: Event-Related Potential and fMRI Findings* (pp. 83–98). Springer US. https://doi.org/10.1007/978-1-4615-0294-4_5

Proudfit, G. H. (2015). The reward positivity: From basic research on reward to a biomarker for depression. *Psychophysiology*, *52*(4), 449–459. psyh. https://doi.org/10.1111/psyp.12370

Rescorla, R. A., & Wagner, A. R. (1972). A Theory of Pavlovian Conditioning: Variations in the Effectiveness of Reinforcement and Nonreinforcement. *Classical conditioning II: Current research and theory*, 64–99.

Salim, M. A. M., van der Veen, F. M., van Dongen, J. D. M., & Franken, I. H. A. (2015). Brain activity elicited by reward and reward omission in individuals with psychopathic traits: An ERP study. *Biological Psychology*, *110*, 50–58. psyh. https://doi.org/10.1016/j.biopsycho.2015.07.001

Sallet, J., Camille, N., & Procyk, E. (2013). Modulation of feedback-related negativity during trial-and-error exploration and encoding of behavioral shifts. *Frontiers in Neuroscience*, *7*. https://www.frontiersin.org/article/10.3389/fnins.2013.00209

Sambrook, T. D., & Goslin, J. (2014). Mediofrontal event-related potentials in response to positive, negative and unsigned prediction errors. *Neuropsychologia*, *61*, 1–10. psyh. https://doi.org/10.1016/j.neuropsychologia.2014.06.004

Sambrook, T. D., & Goslin, J. (2015). A neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychological Bulletin*, *141*(1), 213–235. https://doi.org/10.1037/bul0000006

Sambrook, T. D., & Goslin, J. (2016). Principal components analysis of reward prediction errors in a reinforcement learning task. *NeuroImage*, *124*, 276–286. https://doi.org/10.1016/j.neuroimage.2015.07.032

Sambrook, T. D., Hardwick, B., Wills, A. J., & Goslin, J. (2018a). Model-free and model-based reward prediction errors in EEG. *NeuroImage*, *178*, 162–171. https://doi.org/10.1016/j.neuroimage.2018.05.023

Sambrook, T. D., Hardwick, B., Wills, A. J., & Goslin, J. (2018b). Model-free and model-based reward prediction errors in EEG. *NeuroImage*, *178*, 162–171. https://doi.org/10.1016/j.neuroimage.2018.05.023

Sambrook, T. D., Roser, M., & Goslin, J. (2012). Prospect theory does not describe the feedback-related negativity value function. *Psychophysiology*, *49*(12), 1533–1544. https://doi.org/10.1111/j.1469-8986.2012.01482.x

San Martín, R. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Frontiers in Human Neuroscience*, *6*, 304. https://doi.org/10.3389/fnhum.2012.00304

San Martín, R., Manes, F., Hurtado, E., Isla, P., & Ibañez, A. (2010). Size and probability of rewards modulate the feedback error-related negativity associated with wins but not losses in a monetarily rewarded gambling task. *NeuroImage*, *51*(3), 1194–1204. https://doi.org/10.1016/j.neuroimage.2010.03.031

Sato, A., Yasuda, A., Ohira, H., Miyawaki, K., Nishikawa, M., Kumano, H., & Kuboki, T. (2005). Effects of value and reward magnitude on feedback negativity and P300. *NeuroReport*, *16*(4), 407.

Schmidt, B., Keßler, L., Hecht, H., Hewig, J., Holroyd, C. B., & Miltner, W. H. R. (2019). What you give is what you get: Payment of one randomly selected trial induces risk-aversion and decreases brain responses to monetary feedback. *Cognitive, Affective, & Behavioral Neuroscience*, *19*(1), 187–196. https://doi.org/10.3758/s13415-018-00656-1

Schuermann, B., Endrass, T., & Kathmann, N. (2012). Neural correlates of feedback processing in decision-making under risk. *Frontiers in Human Neuroscience*, *6*. https://doi.org/10.3389/fnhum.2012.00204

Schultz, W. (2002). Getting Formal with Dopamine and Reward. *Neuron*, *36*(2), 241–263. https://doi.org/10.1016/S0896-6273(02)00967-4

Schultz, W. (2013). Updating dopamine reward signals. *Current Opinion in Neurobiology*, *23*(2), 229–238. https://doi.org/10.1016/j.conb.2012.11.012

Schultz, W. (2016). Dopamine reward prediction-error signalling: A two-component response. *Nature Reviews Neuroscience*, *17*(3), 183–195. https://doi.org/10.1038/nrn.2015.26

Schultz, W. (2017). Reward prediction error. *Current Biology*, *27*(10), R369–R371. https://doi.org/10.1016/j.cub.2017.02.064

Schultz, W., Dayan, P., & Montague, P. R. (1997). A Neural Substrate of Prediction and Reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593

Soder, H. E., & Potts, G. F. (2017). Medial frontal cortex response to unexpected motivationally salient outcomes. *International Journal of Psychophysiology: Official Journal of the International Organization of Psychophysiology*. https://doi.org/10.1016/j.ijpsycho.2017.11.003

Soder, H. E., Suchting, R., & Potts, G. F. (2020). Electrophysiological responses to appetitive and aversive outcomes: A comparison of college drinkers and non-drinkers. *Neuroscience Letters*, *714*, 134549. https://doi.org/10.1016/j.neulet.2019.134549

Stewardson, H. J., & Sambrook, T. D. (2020). Evidence for parietal reward prediction errors using great grand average meta-analysis. *International Journal of Psychophysiology*, *152*, 81-86.

Stewardson, H. J., & Sambrook, T. D. (2021a). Reward, Salience, and Agency in Event-Related Potentials for Appetitive and Aversive Contexts. *Cerebral Cortex*, *31*(11), 5006–5014. https://doi.org/10.1093/cercor/bhab137

Stewardson, H. J., & Sambrook, T. D. (2021b). Reward prediction error in the ERP following unconditioned aversive stimuli. *Scientific reports*, *11*(1), 1-10.

Stewardson, H. J., & Sambrook, T. D. (2022). Valence precedes value in neural encoding of prediction error. *Psychophysiology*.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine Learning*, *3*(1), 9–44. https://doi.org/10.1007/BF00115009

Sutton, R. S., & Barto, A. G. (1998). Reinforcement Learning: An Introduction. *IEEE Transactions on Neural Networks*, *9*(5), 1054. https://doi.org/10.1109/TNN.1998.712192

Tabachnick, B. G., & Fidell, L. S. (2013). Using multivariate statistics: International edition. *Pearson2012*.

Talmi, D., Atkinson, R., & El-Deredy, W. (2013). The Feedback-Related Negativity Signals Salience Prediction Errors, Not Reward Prediction Errors. *Journal of Neuroscience*, *33*(19), 8264–8269. https://doi.org/10.1523/JNEUROSCI.5695-12.2013

Talmi, D., Fuentemilla, L., Litvak, V., Duzel, E., & Dolan, R. J. (2012). An MEG signature corresponding to an axiomatic model of reward prediction error. *NeuroImage*, *59*(1), 635–645. https://doi.org/10.1016/j.neuroimage.2011.06.051

Talmi, D., Seymour, B., Dayan, P., & Dolan, R. J. (2008). Human Pavlovian–Instrumental Transfer. *Journal of Neuroscience*, *28*(2), 360–368. https://doi.org/10.1523/JNEUROSCI.4028-07.2008

Thorndike, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *The Psychological Review: Monograph Supplements*, *2*(4), i–109. https://doi.org/10.1037/h0092987

Toyomaki, A., & Murohashi, H. (2005). The ERPs to feedback indicating monetary loss and gain on the game of modified "rock–paper–scissors". *International Congress Series*, *1278*, 381–384. https://doi.org/10.1016/j.ics.2004.11.032

Tzovara, A., Korn, C. W., & Bach, D. R. (2018). Human Pavlovian fear conditioning conforms to probabilistic learning. *PLOS Computational Biology*, *14*(8), e1006243. https://doi.org/10.1371/journal.pcbi.1006243

Ungless, M. A., Magill, P. J., & Bolam, J. P. (2004). Uniform Inhibition of Dopamine Neurons in the Ventral Tegmental Area by Aversive Stimuli. *Science*, *303*(5666), 2040–2042. https://doi.org/10.1126/science.1093360

Van den Berg, I., Franken, I. H. A., & Muris, P. (2011). Individual differences in sensitivity to reward: Association with electrophysiological responses to monetary gains and losses. *Journal of Psychophysiology*, *25*(2), 81–86. https://doi.org/10.1027/0269-8803/a000032

Van den Berg, I., Shaul, L., Van der Veen, F. M., & Franken, I. H. A. (2012). The role of monetary incentives in feedback processing: Why we should pay our participants. *NeuroReport*, *23*(6), 347–353. https://doi.org/10.1097/WNR.0b013e328351db2f

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software*, *36*(3), 1–48.

Vikbladh, O., Shohamy, D., & Daw, N. (2017). *Episodic Contributions to Model-Based Reinforcement Learning*. 2.

Walentowska, W., Moors, A., Paul, K., & Pourtois, G. (2016). Goal relevance influences performance monitoring at the level of the FRN and P3 components. *Psychophysiology*, *53*(7), 1020–1033. psyh. https://doi.org/10.1111/psyp.12651

Walsh, M. M., & Anderson, J. R. (2011). Learning from delayed feedback: Neural responses in temporal credit assignment. *Cognitive, Affective, & Behavioral Neuroscience*, *11*(2), 131–143. https://doi.org/10.3758/s13415-011-0027-0

Walsh, M. M., & Anderson, J. R. (2012). Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice. *Neuroscience & Biobehavioral Reviews*, *36*(8), 1870–1884. https://doi.org/10.1016/j.neubiorev.2012.05.008

Wei, S., Zheng, Y., Li, Q., Dai, W., Sun, J., Wu, H., & Liu, X. (2018). Enhanced neural responses to monetary rewards in methamphetamine use disordered individuals compared to healthy controls. *Physiology & Behavior*, *195*, 118–127. psyh. https://doi.org/10.1016/j.physbeh.2018.08.003

Weinberg, A., Luhmann, C. C., Bress, J. N., & Hajcak, G. (2012). Better late than never? The effect of feedback delay on ERP indices of reward processing. *Cognitive, Affective, & Behavioral Neuroscience*, *12*(4), 671–677. https://doi.org/10.3758/s13415-012-0104-z

Weinberg, A., Riesel, A., & Proudfit, G. H. (2014). Show me the Money: The impact of actual rewards and losses on the feedback negativity. *Brain and Cognition*, *87*, 134–139. https://doi.org/10.1016/j.bandc.2014.03.015

Wills, A. J., Dome, L., Edmunds C. E., Honke, G., Inkster, A. B., Schlegelmilch, R., & Spicer, S. G. (2019). catlearn: Formal Psychological Models of Categorization and Learning. https://CRAN.R-project.org/package=catlearn. R package version 0.6.2.

Wischnewski, M., & Schutter, D. J. L. G. (2018). Dissociating absolute and relative reward- and punishment-related electrocortical processing: An event-related potential study. *International Journal of Psychophysiology*, *126*, 13–19. psyh. https://doi.org/10.1016/j.ijpsycho.2018.02.010

Wu, Y., & Zhou, X. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Research*, *1286*, 114–122. https://doi.org/10.1016/j.brainres.2009.06.032

Yang, Q., Gu, R., Tang, P., & Luo, Y.-J. (2013). How does cognitive reappraisal affect the response to gains and losses? *Psychophysiology*, *50*(11), 1094–1103. https://doi.org/10.1111/psyp.12091

Yeung, N., Holroyd, C. B., & Cohen, J. D. (2005). ERP Correlates of Feedback and Reward Processing in the Presence and Absence of Response Choice. *Cerebral Cortex*, *15*(5), 535–544. https://doi.org/10.1093/cercor/bhh153

Yeung, N., & Sanfey, A. G. (2004). Independent Coding of Reward Magnitude and Valence in the Human Brain. *The Journal of Neuroscience*, *24*(28), 6258–6264. psyh. https://doi.org/10.1523/JNEUROSCI.4537-03.2004

Yu, R., & Zhou, X. (2006). Brain responses to outcomes of one's own and other's performance in a gambling task. *NeuroReport*, *17*(16), 1747. https://doi.org/10.1097/01.wnr.0000239960.98813.50

Yu, R., & Zhou, X. (2008). To Bet or Not to Bet? The Error Negativity or Error-related Negativity Associated with Risk-taking Choices. *Journal of Cognitive Neuroscience*, *21*(4), 684–696. https://doi.org/10.1162/jocn.2009.21034

Zaghloul, K. A., Blanco, J. A., Weidemann, C. T., McGill, K., Jaggi, J. L., Baltuch, G. H., & Kahana, M. J. (2009). Human Substantia Nigra Neurons Encode Unexpected Financial Rewards. *Science*, *323*(5920), 1496–1499. https://doi.org/10.1126/science.1167342

Zhang, D., Gu, R., Wu, T., Broster, L. S., Luo, Y., Jiang, Y., & Luo, Y. (2013). An electrophysiological index of changes in risk decision-making strategies. *Neuropsychologia*, *51*(8), 1397–1407. psyh. https://doi.org/10.1016/j.neuropsychologia.2013.04.014

Zheng, Y., & Liu, X. (2015). Blunted neural responses to monetary risk in high sensation seekers. *Neuropsychologia*, *71*, 173–180. psyh. https://doi.org/10.1016/j.neuropsychologia.2015.04.002