Reward, salience and agency in event related potentials for appetitive and aversive contexts.

Harry Stewardson*

Thomas D. Sambrook

* Corresponding author

School of Psychology, University of East Anglia,

Norwich Business Park, NR4 7TJ, United Kingdom

Tel +44 1603 597753

Fax +44 1603 597756

Abstract

Cognitive architectures tasked with swiftly and adaptively processing biologically important events are likely to classify these on two central axes: *motivational salience*, i.e. those events' importance and unexpectedness, and *motivational value*, the utility they hold, relative to that expected. Because of its temporal precision, electroencephalography provides an opportunity to resolve processes associated with these two axes. A focus of attention for the last two decades has been the feedback related negativity (FRN), a frontocentral component occurring 240–340 ms after valenced events that are not fully predicted. Both motivational salience and value are present in such events and competing

claims have been made for which of these is encoded by the FRN. The present study suggests that motivational value, in the form of a reward prediction error, is the primary determinant of the FRN in active contexts, while in both passive and active contexts, a weaker and earlier overlapping motivational salience component may be present.

Keywords: ERP, FRN, RewP, motivational salience, reward prediction error,

Introduction

Adaptive behaviour in an intelligent organism requires neural processes capable of extracting the full gamut of information associated with motivationally relevant events in the environment. Some of these processes will be 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 will be concerned with adjusting longer term expectations about the future profitability of the environment, and of particular actions pursued within it. A challenge for cognitive neuroscientists is to isolate the neural signatures associated with these functionally distinct processes.

Foremost of concerns for any motivated organism is valence, i.e., has something good or bad happened? 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 reward prediction error. Reward prediction errors (RPEs) represent a powerful computational means for learning appropriate actions in response to motivational cues. The sign of the RPE (positive or negative) respectively strengthens or weakens the propensity for future selection of an action, and its magnitude determines the degree of strengthening or weakening. By these means, RPEs can produce very effective machine learning (Sutton and Barto 1998), and they represent a computationally parsimonious 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 error (Kamin 1968). Midbrain dopamine neurons have been shown to encode reward prediction errors (Schultz, et al. 1997).

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. Demonstrating the presence (or indeed absence) of such encoders would improve our understanding of the computational architecture underlying motivated behaviour. In identifying such an encoder, temporal precision is at a premium, since even a fully modular RPE encoder would likely be swiftly relaying computational input and output to non–RPE systems. For this reason, electrophysiology represents an appropriate technique for the isolation of RPEs.

One candidate for such an encoding lies in the interval associated with the feedback related negativity, or FRN (Holroyd and Coles 2002; Miltner, et al. 1997), sometimes known as the reward positivity, or RewP (Proudfit 2015). 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 positive reward prediction error (+RPE, i.e. better than expected outcome) and a negative reward prediction error (–RPE, i.e. worse than expected outcome). 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 and Goslin 2015), suggesting it may encode RPE size in addition to sign, thus fully coding for RPE utility in the manner of an axiomatic RPE encoder (Caplin and Dean 2008). This is supported by studies manipulating RPE utility as a continuous variable (Cavanagh 2015; Gu, et al. 2020; Sambrook and Goslin 2014; Sambrook and Goslin 2016). When the generation of prediction errors (and consequent learning) is prevented in a blocking procedure, the FRN is attenuated, suggesting it may play an obligatory role in the transmission of RPEs (Luque, et al. 2012).

Other studies have claimed that activity in the temporospatial interval of the FRN is better interpreted as a response to *motivational salience* rather than the *motivational value* embodied by RPE sign. Motivational salience consists in an outcome's ability to elicit attention due to its motivational relevance (Schultz 2016) and is high for *both* appetitive and aversive events, and low for neutral events (Berridge and 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 (juice, 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 (e.g. Fujiwara, et al. 2009, O'Doherty, et al. 2003) or motivational salience signal (e.g. Jensen, et al. 2007, Metereau and 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 and Potts 2018; Soder, et al. 2020; Talmi, et al. 2013) have made the case that the FRN comprises a motivational salience signal, 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 component. 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.

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 differencing of bad and good outcomes produced a positive, rather than negative-going FRN. This is parsimoniously explained by the FRN simply coding motivational salience, i.e. the delivery rather than omission of a motivating, valenced stimulus, but with no encoding of its actual valence.

In stark contrast, two other studies using pain as a primary aversive reinforcer (Heydari and Holroyd 2015; Mulligan and Hajcak 2018) found domain to have no effect on the polarity of the FRN, supporting the RPE 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 encoding employed active, instrumental designs. Since RPEs are understood to be computational terms in instrumental learning, we should be unsurprised to find 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 and Goslin 2015; Walsh and Anderson 2012). Additionally, the four passive studies reported somewhat earlier and smaller FRN peaks in both appetitive and aversive conditions, raising the possibility that the component observed was revealed by attenuation 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.

The interval occupied by the FRN lies from 240 to 340 ms post feedback according to a meta-analysis of fifty-five studies (Sambrook and Goslin 2015). If this interval is occupied by an RPE 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 encoding and motivational salience effects 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 component, 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.

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 agency (i.e. active vs. passive) and domain (appetitive vs. aversive) within a single experiment to establish whether an active context instated a coding of motivational value in the FRN interval.

Our primary aim was thus a resolution of motivational salience vs. value accounts by manipulation of agency. As a secondary aim, we wished to further characterise the active-context FRN, or any other observed active-context motivational value signal in terms of its capacity to carry a continuous measure of RPE utility, in a fashion consistent with an

axiomatic RPE encoder (Caplin and Dean 2008) and with that frequently used in reinforcement learning algorithms (Sutton and Barto 1998).

Materials and Methods

Participants

Sixty-four students of 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 neurological 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) 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.

Experimental Design

Three factors were orthogonally manipulated. *Domain* consisted in whether appetitive (money) or aversive (noise) outcomes were at stake, *RPE sign* consisted in +RPEs (money delivery, noise omission) vs. –RPEs (noise delivery, money omission) and *agency* consisted in 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 domain and 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 the FRN, 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 domain and RPE sign. 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 encoder of RPE utility 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 encoder is indicated, discriminating the utility for one sign 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, domain should have no effect.

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, 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 (i.e. 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.

Before the first block of each condition, participants were shown a pair of symbols and instructed which indicated delivery and which indicated omission of the reward or punishment at hand for that condition. The mapping of symbols to outcome or delivery 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.

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.

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 to 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 to 700 ms. Other non-specific artefacts in the interval -200 to 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).

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 thus 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.

Computational Modelling of RPE utility

While some studies of the FRN build a continuous measure of utility directly into their design (e.g. 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 coding of prediction error size, it can reveal an encoding of RPE sign, which is sufficient to address the motivational salience vs. value 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 and 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 (2011), 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 (Wills, et al, 2018) and is available at: https://github.com/thomasdsambrook/Q

Results

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 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.

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 =. 001950), for the active aversive condition between 270 and 320 ms (Monte Carlo p = 0.00095) and the passive appetitive condition between 220 and 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 to 260 ms) were entered into a 2 x 2 domain x agency analysis of variance. This revealed no significant effect of domain, but an effect of agency ($F = 4.66$, $p = .036$, $\sigma^2 = .096$) and a significant interaction ($F = 5.47$, $p = .024$, $\sigma^2 = .11$). The effect of 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 component, 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 (i.e. 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 and 260 ms revealed a significant difference ($t(44) = 2.58$, $p = .013$) as predicted by a motivational salience encoding.

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 $p = 0.00018$), 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 $p = 0.0008$, with decreasing utility associated with positive voltage). As such, no encoder of motivational salience, active in both 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.

Discussion

There has been ongoing debate as to whether activity in the FRN interval codes for motivational value or motivational salience. This study found that, in active conditions, the FRN was same-signed for both appetitive and aversive stimuli, suggesting the FRN encodes motivational value, in the form of a reward prediction error, during instrumental learning. Previous studies showing a reversal of FRN polarity in the aversive domain may have failed

to elicit this RPE encoder as a consequence of employing a passive design, and so revealed an earlier overlapping motivational salience encoder. 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 and 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 value (Gu, et al. 2020).

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 domains and the latency differences were potentially attributable to component overlap as described earlier. Future studies applying source localisation will further assess the evidence for such a common encoding of RPEs over appetitive and aversive domains.

As a secondary aim, the study explored coding of continuous RPE utility. It found evidence that, in a later interval, RPE utility was coded in a continuous rather than simply dichotomous fashion, as shown by a positive correlation between +RPE size and voltage, regardless of domain, Notably, the size of +RPEs in the aversive domain was strongly

represented, despite this “outcome” 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 and 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., 2011b, Holroyd et al., 2008, Proudfit, 2015, though see Gu et al., 2020 where –RPE utility is encoded). Notably, encoding of continuous RPE utility did not occur in the interval associated with the FRN however, suggesting the possibility of separate processes for categorical and continuous measures of prediction error, evidence for which has been presented elsewhere (Fouragnan, et al. 2015; Philiastides, et al. 2010).

This study was designed in order to assess the *relative* evidence for two competing hypotheses of activity in the FRN interval: whether this represents motivational value or motivational salience. While we argue that motivational value 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. A component responding to passive RPEs is also implicated by this study and such a component 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 and 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 and Schoenbaum 2008). Whether

scalp activity associated with active and passive RPEs is based on the same or different generators 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 generators, 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, i.e. 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 (Foti et al., 2011b, Foti et al., 2011a, though see 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.

Another encoding claimed to lie in the feedback-locked waveform is unsigned prediction error (Hauser, et al. 2014). Sometimes referred to as "simple surprise", this constitutes another form of salience that is distinct from motivational salience due to its having a V-shaped relationship with RPE utility within each domain, treating unexpected omissions and deliveries as equally salient. Gu, et al. (2020) were able to reverse the polarity of the FRN for monetary losses by task instructions that stressed predictive accuracy (unsigned prediction error), and then reinstate the normal polarity when instructing

participants to focus on reward (RPE). A further property that may be encoded is domain. While we have made the case for motivational value encoding in the FRN being at last partially domain-independent, a component coding for 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 and Dayan 2011).

In short, even very ambitious multifactorial designs will struggle to unconfound all the computational terms present in feedback processing. The inevitability of component overlap compounds 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 motivational value and salience are computed by 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 sign, domain and actor agency are critical determinants of the FRN, and that when they are experimentally controlled, an encoding of motivational value is revealed.

References

Balleine BW, Delgado MR, Hikosaka O. 2007. The role of the dorsal striatum in reward and decision-making. J Neurosci. 27:8161-8165.

Berridge KC, Robinson TE. 1998. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Res Rev. 28:309-369.

Boureau YL, Dayan P. 2011. Opponency Revisited: Competition and Cooperation Between Dopamine and Serotonin. Neuropsychopharmacol. 36:74-97.

Bromberg-Martin ES, Matsumoto M, Hikosaka O. 2010. Dopamine in Motivational Control: Rewarding, Aversive, and Alerting. Neuron. 68:815-834.

Brown PL, Jenkins HM. 1968. Auto-Shaping of Pigeons Key-Peck. J Exp Anal Behav. 11:1-&.

Byrd RH, Lu PH, Nocedal J, Zhu CY. 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. Curr Opin Neurobiol. 18:197-202.

Carlson JM, Foti D, Mujica-Parodi LR, 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. 57:1608-1616.

Cavanagh JF. 2015. Cortical delta activity reflects reward prediction error and related behavioral adjustments, but at different times. Neuroimage. 110:205-216.

Cohen MX, Cavanagh JF, Slagter HA. 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. Hum Brain Mapp. 32:2270-2271.

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

Daw ND. 2011. Trial-by-trial data analysis using computational models. In: Delgado MR, Phelps EA, Robbins TW, editors. Decision making, affect, and learning: Attention and performance XXIII Oxford: Oxford University Press. p 3-38.

Foti D, Weinberg A, Dien J, Hajcak G. 2011. Event-Related Potential Activity in the Basal Ganglia Differentiates Rewards From Nonrewards: Response to Commentary. Hum Brain Mapp. 32:2267-2269.

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. Hum Brain Mapp. 32:2207-2216.

Fouragnan E, Retzler C, Mullinger K, Philiastides MG. 2015. Two spatiotemporally distinct value systems shape reward-based learning in the human brain. Nat Commun. 6.

Gu Y, Liu T, Zhang X, Long Q, Hu N, Zhang Y, Chen A. 2020. The Event-Related Potentials Responding to Outcome Valence and Expectancy Violation during Feedback Processing. Cereb Cortex.

Hassall CD, Hajcak G, Krigolson OE. 2019. The importance of agency in human reward processing. Cogn Affect Behav Ne. 19:1458-1466.

Hauser TU, 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.

Heydari S, Holroyd CB. 2015. Feedback Related Negativity: Reward Prediction Error or Salience Prediction Error? Psychophysiology. 52:S111-S111.

Hird EJ, El-Deredy W, Jones A, Talmi D. 2018. Temporal dissociation of salience and prediction error responses to appetitive and aversive taste. Psychophysiology. 55.

Holroyd CB, Coles MGH. 2002. The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. Psychol Rev. 109:679-709.

Holroyd CB, Pakzad-Vaezi KL, Krigolson OE. 2008. The feedback correct-related positivity: Sensitivity of the event-related brain potential to unexpected positive feedback. Psychophysiology. 45:688-697.

Jensen J, Smith AJ, Willeit M, Crawley AP, Mikulis DJ, Vitcu I, Kapur S. 2007. Separate brain regions code for salience vs. valence during reward prediction in humans. Hum Brain Mapp. 28:294-302.

Kamin L, J. 1968. “Attention-like” processes in classical conditioning. In: Jones MR, editor. Miami Symposium on the Prediction of Behavior 1967: Aversive Stimulation Coral Gables, FL: University of Miami Press. p 9-31.

Luque D, Lopez FJ, Marco-Pallares J, Camara E, Rodriguez-Fornells A. 2012. Feedback-related Brain Potential Activity Complies with Basic Assumptions of Associative Learning Theory. J Cognitive Neurosci. 24:794-808.

Maris E, Oostenveld R. 2007. Nonparametric statistical testing of EEG- and MEG-data. J Neurosci Meth. 164:177-190.

Matsumoto M, Hikosaka O. 2009. Two types of dopamine neuron distinctly convey positive and negative motivational signals. Nature. 459:837-U834.

Metereau E, Dreher JC. 2013. Cerebral Correlates of Salient Prediction Error for Different Rewards and Punishments. Cereb Cortex. 23:477-487.

Miltner WHR, Braun CH, Coles MGH. 1997. Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a "generic" neural system for error detection. J Cognitive Neurosci. 9:788-798.

Mulligan EM, Hajcak G. 2018. The electrocortical response to rewarding and aversive feedback: The reward positivity does not reflect salience in simple gambling tasks. Int J Psychophysiol. 132:262-267.

Niv Y, Schoenbaum G. 2008. Dialogues on prediction errors. Trends Cogn Sci. 12:265-272.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston KJ, Dolan RJ. 2004. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science. 304:452-454.

O'Doherty JP, Dayan P, Friston KJ, Critchley H, Dolan RJ. 2003. Temporal difference models and reward-related learning in the human brain. Neuron. 38:329-337.

Pedroni A, Langer N, Koenig T, Allemand M, Jancke L. 2011. Electroencephalographic Topography Measures of Experienced Utility. J Neurosci. 31:10474-10480.

Perrin F, Pernier J, Bertrand O, Echallier JF. 1989. Spherical Splines for Scalp Potential and Current-Density Mapping. Electroen Clin Neuro. 72:184-187.

Philiastides MG, Biele G, Vavatzanidis N, Kazzer P, Heekeren HR. 2010. Temporal dynamics of prediction error processing during reward-based decision making. Neuroimage. 53:221-232.

Proudfit GH. 2015. The reward positivity: From basic research on reward to a biomarker for depression. Psychophysiology. 52:449-459.

Sambrook TD, Goslin J. 2014. Mediofrontal event-related potentials in response to positive, negative and unsigned prediction errors. Neuropsychologia. 61:1-10.

Sambrook TD, Goslin J. 2015. A Neural Reward Prediction Error Revealed by a Meta-Analysis of ERPs Using Great Grand Averages. Psychol Bull. 141:213-235.

Sambrook TD, Goslin J. 2016. Principal components analysis of reward prediction errors in a reinforcement learning task. Neuroimage. 124:276-286.

Sambrook TD, Hardwick B, Wills AJ, Goslin J. 2018. Model-free and model-based reward prediction errors in EEG. Neuroimage. 178:162-171.

Schmidt B, Kessler L, Hecht H, Hewig J, Holroyd CB, Miltner WHR. 2019. What you give is what you get: Payment of one randomly selected trial induces risk-aversion and decreases brain responses to monetary feedback. Cogn Affect Behav Ne. 19:187-196.

Schultz W. 2016. Dopamine reward prediction-error signalling: a two-component response. Nature Reviews Neuroscience. 17:183-195.

Schultz W, Dayan P, Montague PR. 1997. A neural substrate of prediction and reward. Science. 275:1593-1599.

Soder HE, Potts GF. 2018. Medial frontal cortex response to unexpected motivationally salient outcomes. Int J Psychophysiol. 132:268-276.

Soder HE, Suchting R, Potts GF. 2020. Electrophysiological responses to appetitive and aversive outcomes: A comparison of college drinkers and non-drinkers. Neurosci Lett. 714.

Sutton RS, Barto AG. 1998. Introduction to reinforcement learning: MIT Press.

Talmi D, Atkinson R, El-Deredy W. 2013. The Feedback-Related Negativity Signals Salience Prediction Errors, Not Reward Prediction Errors. J Neurosci. 33:8264-8269.

Walsh MM, Anderson JR. 2011. Learning from delayed feedback: neural responses in temporal credit assignment. Cogn Affect Behav Ne. 11:131-143.

Walsh MM, Anderson JR. 2012. Learning from experience: Event-related potential correlates of reward processing, neural adaptation, and behavioral choice. Neurosci Biobehav R. 36:1870-1884.

Weinberg A, Luhmann CC, Bress JN, Hajcak G. 2012. Better late than never? The effect of feedback delay on ERP indices of reward processing. Cogn Affect Behav Ne. 12:671-677.

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.

Yeung N, Holroyd CB, Cohen JD. 2005. ERP correlates of feedback and reward processing in the presence and absence of response choice. Cereb Cortex. 15:535-544.

Zaghloul KA, Blanco JA, Weidemann CT, McGill K, Jaggi JL, Baltuch GH, Kahana MJ. 2009. Human Substantia Nigra Neurons Encode Unexpected Financial Rewards. Science. 323:1496-1499.

Figure Captions

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

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.

Figure 3. RPE encoding in the feedback-locked waveform. (A) shows simple waveforms by domain, agency and RPE sign. (B) shows FRNs (dashed lines) by domain and 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 RPE 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.

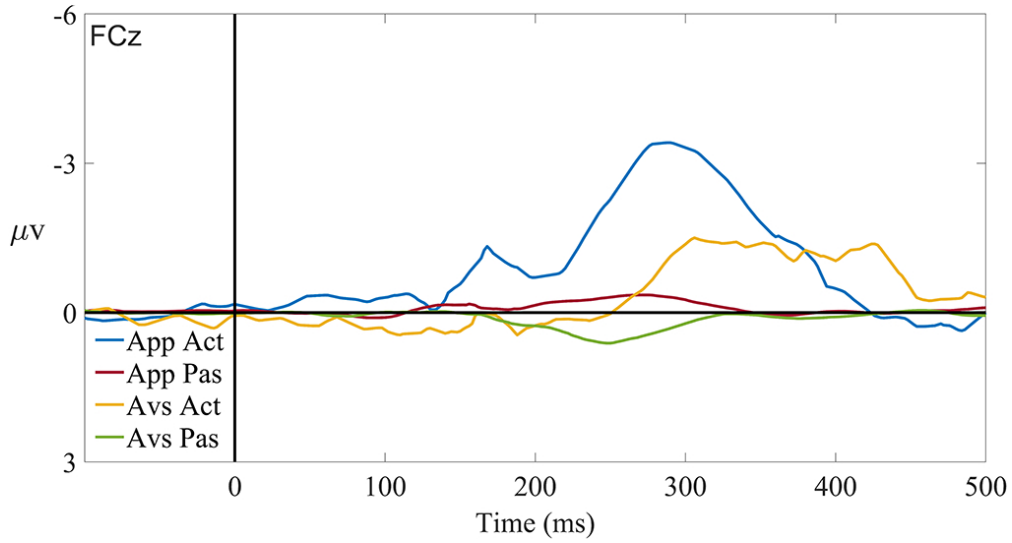

Figure 1. Average FRNs from six studies manipulating domain x agency. Weighted averages are shown, based on experiments' sample size. In designs manipulating RPE size as well as sign, the largest available RPE sizes were used. See Sambrook and Goslin (2015) for a description of the averaging technique.

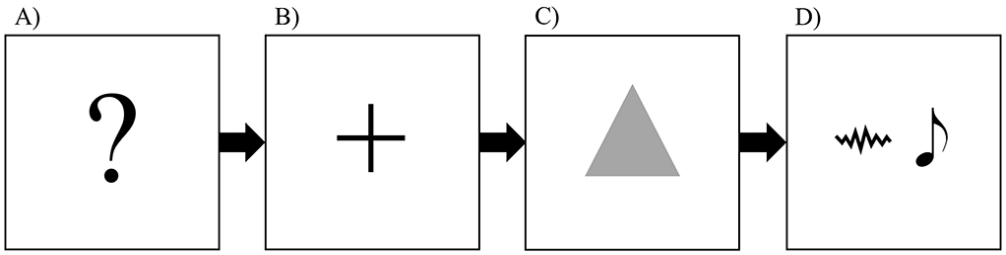

Figure 2. One trial of the experiment. A) 1000 ms in passive condition or until keypress (max 1000 ms) in active conditions B) 600–700 ms fixation C) 1000 ms truth cue D) 700 ms noise, cash-till sound or silence.

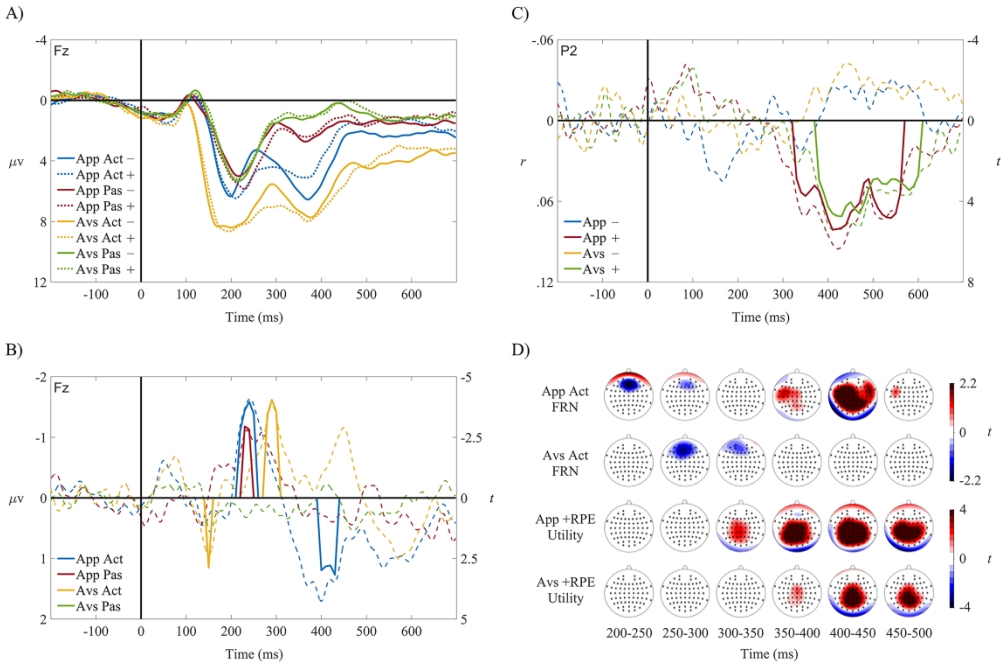

Figure 3. RPE encoding in the feedback-locked waveform. (A) shows simple waveforms by domain, agency and RPE sign. (B) shows FRNs (dashed lines) by domain and agency, obtained by differencing +RPE and –RPE simple waveforms in (A). Bold lines indicate significance under a t-test (see right hand axis) in intervals retained after cluster randomisation. (C) shows the correlation, r, of voltage and RPE 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.

562x372mm (300 x 300 DPI)