12-2019

The Feedback-Related Negativity is a Time-Dependent Brain Mechanism that Facilitates Aversive Learning: Implications for the Reinforcement Learning FRN Hypothesis

Eric Rawls
University of Arkansas, Fayetteville

The Feedback-Related Negativity is a Time-Dependent Brain Mechanism that Facilitates Aversive Learning: Implications for the Reinforcement Learning FRN Hypothesis

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Psychology

by

Eric Rawls
Tulane University
Bachelor of Science in Mathematics, Psychology, 2012
University of New Orleans
Master of Science in Biopsychology, 2016

December 2019
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council.

_____
Connie Lamm, Ph.D.
Dissertation Director

_____
Jenn Veilleux, Ph.D.
Committee Member

_____
Douglas Rhoads, Ph.D.
Committee Member

Abstract

Organisms encode rewarding and aversive experiences through reinforcement learning, capitalizing on prediction errors (PEs), which adapt action strategies over time. Computational theories are explicit that PE signals should update action weights continuously over the course of a behavioral task, an important time-dependent variation that is eschewed in traditional neuroscience studies that average over large numbers of trials. I examined variation in reaction times and feedback-locked cortical activity over time as a function of PE to critically examine theories indicating that PE signals drive time-dependent learning. We recorded EEG while participants completed a novel reinforcement task that varied prediction error on a trial-by-trial basis. I applied a computational framework that modeled reaction time changes over the task as a function of prediction error and time. In positive reinforcement conditions, reaction times improved over the course of the task regardless of the PE. For negative reinforcement, learning effects were moderated by PE. For better than expected outcomes, more positive prediction errors (further from expectation) drove faster reaction times over the course of the task, and for worse than expected outcomes, more negative prediction errors (further from expectation) drove faster reaction times over the course of the task. Behavioral analyses were supplemented by single-trial robust regression of feedback-locked EEG. The feedback-related negativity (FRN), a mediofrontal ERP component thought to convey a PE signal, showed robust changes in activation over time but did not respond to trial-by-trial magnitude of prediction errors. This time-dependent change was evident only for reward delivery and aversive stimulus delivery, which represent on average the most salient outcomes in the task. Mediofrontal brain activity during this same time window and at the same scalp location drove subsequent reaction time improvements over the course of the task following aversive stimulus delivery. I suggest that the standard approach of examining the ERP as an average across conditions obscures important adaptation effects of the FRN that reflect reinforcement learning as outcomes are learned.

Acknowledgments

First and foremost, I must express my sincerest gratitude to my advisor Dr. Connie Lamm for her constant support, patience, and guidance over the five years I have been a doctoral student in her lab. Through her support I learned to critically examine existing theories on the science of the brain, how to craft and defend an argument, and how to think like a scientist. Her guidance has been invaluable to me throughout my graduate career, and without her careful supervision I would not be in the position I am to pursue my long-term career goals.

I also must thank my colleague, coauthor, and teacher, Dr. Vladimir Miskovic, for his careful attention and efforts on the several manuscripts he has helped me to prepare for publication.

I would also like to thank Dr. Jenn Veilleux and Dr. Douglas Rhoads for serving on my dissertation committee. I am grateful for their support that has motivated me and enabled me to complete my dissertation. I am grateful to Dr. Nathan Parks as well for serving on my candidature committee.

Finally, a special thanks is deserved by the many students in Dr. Connie Lamm's Developmental, Cognitive, and Affective Psychophysiology Laboratory for their invaluable assistance in collecting the data this dissertation is built from. Special among this group are (in alphabetical order): Carroll Bentley, Zachary Levy, Stephanie Long, Morgan Middlebrooks, Danielle Smith, Josh Upshaw, and Summer Webers, for supervising data collection sessions.

Table of Contents

List of Figures

1 Introduction

1.1 Computational perspectives on reinforcement learning

Through our interactions with the environment around us, we learn about the causes and effects of our actions, as well as what actions we should take in the future to maximize positive outcomes. The term reinforcement was first introduced by Pavlov (1903) to describe the associative pairing of a conditioned predictor stimulus with an unconditioned reinforcing stimulus. This initial conception used reinforcement to describe stimulus-stimulus learning, that is, to describe how an organism learns to associate a non-reward stimulus with a reward stimulus. The term has since become used more often to describe stimulus-response learning; that is, how an organism learns to associate an environmental stimulus with a behavioral response. The strength of stimulus-response associations depends on which responses most often lead to desired outcomes. This concept was described in Thorndike's (1905) law of effect, which suggested that responses which are followed by reward will be more strongly associated with the environmental situation that spurred them. When that situation re-occurs, a response which was closely followed by reward will be more likely to re-occur.

From a computational perspective, the reinforcement learning problem must map environmental situations onto those behaviors that maximize positive outcomes and minimize negative outcomes (Sutton & Barto, 1998). In computational approaches to reinforcement learning, a key concept revolves around the prediction error term (PE). The PE term represents the deviation of outcomes from expectations, guiding optimal learning from environmental occurrences. A large PE term indexes an outcome that was far from the prediction. This should spur changes in future action strategies, in order to bring future outcomes more in line with predictions. Likewise, a small or nonexistent PE term indexes outcomes that were properly predicted; this should translate to slower or no learning rates when outcomes are close to what was expected. One classic model of reinforcement learning is the Rescorla-Wagner model (Rescorla & Wagner, 1972). In this model, a stronger association between stimuli means that

one stimulus predicts the other (i.e. low prediction error) and a weaker association means that one stimulus does not accurately predict the other (i.e. high prediction error). This model conceives of associations as carrying a signed value – specifically, if a stimulus predicts a reward, it carries a positive association, and if a stimulus predicts an aversive outcome it carries a negative association. Therefore, the prediction error term in the Rescorla-Wagner model carries information about how strongly a stimulus predicts an outcome, as well as whether the stimulus predicts a better-than-expected or worse-than-expected outcome. This type of PE term is called either a signed or value PE. These terminologies will be used interchangeably in this review.

The Pearce-Hall error learning model (Pearce & Hall, 1980), like the Rescorla-Wagner model, conceived of learning in terms of associations between conditioned and unconditioned stimuli. However, the prediction error term in this model takes a different approach to that of the Rescorla-Wagner model. In the Pearce-Hall implementation, the strength of association between the unconditioned stimulus and the conditioned stimulus depends on the attention paid to the conditioned stimulus. Since both rewarding and aversive events are attention-grabbing, the PE term in this model does not explicitly code for whether an outcome was better or worse than expected. Instead, outcomes which are both salient (better-than-expected or worse-than-expected) and unexpected are associated more strongly with the predictive stimulus. This type of prediction error is called either an unsigned or salience PE. These terminologies are used interchangeably in the remainder of this review.

Previous theoretical perspectives on reinforcement learning suggest that the most important computational term underlying learning is the prediction error term (Glascher, Daw, Dayan, & O'Doherty, 2010; Glimcher, 2011; Pearce & Hall, 1980; Rescorla & Wagner, 1972; Sutton & Barto, 1998). However, theories have disagreed on the exact form this term should take. The Rescorla-Wagner model suggested that the PE term would follow a value function, that linearly increased with increasing reward value and decreased for increasingly aversive

stimuli (Rescorla & Wagner, 1972). The Pearce-Hall error learning model instead used a salience function for the prediction error term, which increased with increasing salience of rewards or punishments but was agnostic to the “value” (good or bad) of the outcome (Pearce & Hall, 1980). See Figure 1 for a depiction of signed vs. unsigned prediction error signaling.

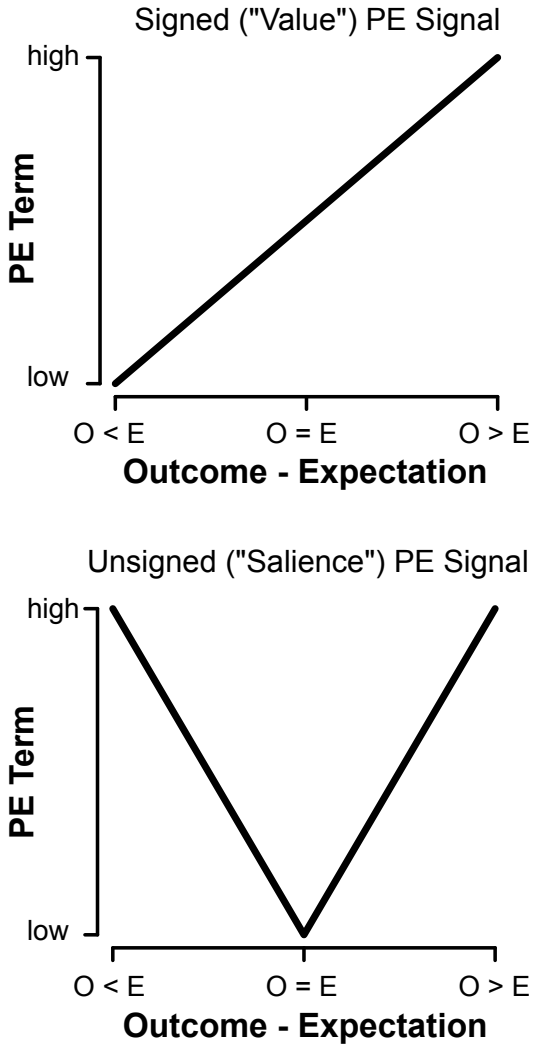


Figure 1. Comparison of Signed (“Value”) and unsigned (“Salience”) Prediction Errors. The main difference is in how the terms behave when outcomes are worse than expected (O < E); a signed prediction error will decrease with worse outcomes, whereas an unsigned prediction error will scale with increasing salience regardless of the value of an outcome. From Rawls et al. (*under review*).

A third class of models, temporal difference (TD; Sutton & Barto, 1987; Sutton, 1988) models, provide an encompassing theory and computational approach to studying how past outcomes are integrated in learning stimulus-response associations. TD models posit that both the subjective "value" of outcomes and the subjective "salience" of outcomes are critical in determining learning. Outcomes with positive value will stimulate approach behavior and outcomes with negative value will stimulate avoidance behavior. Meanwhile, outcomes with higher salience (better-than-expected or worse-than-expected) will result in more rapid alteration of associative weights, thereby speeding learning. This perspective on reinforcement learning does not assume importance of only a value or salience PE, but instead point to the importance of both terms in determining reinforcement learning outcomes. Most importantly, and most neglected in the current literature, is the implicit notion that the signaling of PEs should change over time as outcomes and task contingencies are learned.

1.2 Dopamine implicated in reinforcement learning through PE signals

There is substantial disagreement in neuroscience literature about how reinforcement learning computations, and in particular PE terms, are reflected in neural architecture. Most discussion around the neural implementation of reinforcement learning posits that dopamine is a key factor in representing the reinforcing properties of outcomes with regard to expectations. There are two main regions in the brain where dopamine cell bodies reside – the ventral tegmental area (VTA) and the substantia nigra pars compacta (SN). These pathways differentially contribute to the main functions of dopamine in the brain. The dopamine system originating in the substantia nigra modulates many of the motor functions of dopamine, while the dopamine system originating in the VTA modulates the reinforcing impacts of dopamine. Therefore, I will focus on the VTA dopamine system. The primary dopaminergic projections of the VTA are the mesocortical and mesolimbic pathways, which transmit to the prefrontal and motor cortices, and the nucleus accumbens, respectively. These dopaminergic pathways play an important role in reinforcement learning. In particular, dopaminergic signals might convey

prediction error signals in the brain. Sparse VTA dopaminergic projections also reach the amygdala and the hippocampus, which are implicated in emotional modulation of reinforcement and memory of reinforcement, respectively. In our review of the mesolimbic reward system, I will primarily focus on dopaminergic projections from VTA to NAcc, in line with most reinforcement literature.

1.3 Dopamine value PE coding in VTA neurons

Early research on dopamine in reinforcement suggested that dopamine neurons encoded a value (signed) prediction error (Schultz, Dayan, & Montague, 1997; Hollerman & Schultz, 1998). This study recorded firing rates of dopamine neurons in the macaque midbrain during delivery of expected reward, omission of expected reward, and delivery of unexpected reward. Results indicated that dopamine firing rates closely approximated what would be expected of a value prediction error signal. Specifically, firing rate increased from baseline for unexpected reward delivery (reward > expectation), firing rate remained at baseline for expected reward delivery (reward = expectation), and firing rate fell below baseline for unexpected reward omission (reward < expectation). This result has since been replicated many times and is one of the mainstays in reinforcement research.

Steinberg, Keiflin, Boivin, Witten, Deisseroth, & Janak (2013) causally inferred the link between positive reward prediction error signals in dopaminergic neurons and reward learning using optogenetic techniques. This experiment used a blocking procedure, in which a novel cue is co-presented with a preconditioned cue that predicted reward. Usually, this would not result in the organism learning to associate reward with the novel cue. However, when VTA dopamine neurons were stimulated while the cue stimuli were presented, this caused rats to learn to associate reward with the novel stimulus. This provides causal evidence that mimicking a positive prediction error in VTA causes associative learning that would normally be blocked. Importantly, this study only investigated whether a positive prediction error (simulating better than expected outcomes by increasing dopamine firing in VTA) drove reinforcement learning.

Dopamine neurons have a very low basal spike output, meaning that negative reward prediction errors cannot slow dopamine output very much. For this reason, negative reward prediction error signaling might not have the same causal behavioral effects as positive reward prediction error signaling, due to a lack of variance in the signal. To examine whether this is the case, Chang, Esher, Marrero-Garcia, Yau, Bonci, & Schoenbaum (2016) conditioned rats to associate one pellet of food reward with two separate cues. In this over-expectation procedure, the subsequent co-presentation of both cues at once causes the rat to expect two food pellets of reward (one for each conditioned stimulus). If only one food pellet is then given, a negative prediction error is generated, which would lower associative value of both predictive stimuli. However, Chang et al. (2016) presented two food pellets (no prediction error), while instead simulating a negative prediction error by optogenetically silencing dopamine neurons in VTA. Although the expected reward was delivered, therefore generating no prediction error, follow-up testing demonstrated a reduced ability of the conditioned stimuli to generate motivated responses, as though the rat had received a lesser reward than expected. Therefore, even though the outcome did not generate a prediction error, simulation of a negative prediction error by silencing output of VTA dopaminergic neurons was sufficient to decrease associative weight.

1.4 Dopamine salience PE coding in VTA neurons

Since the initial discovery that dopamine indexed a signed prediction error-like signal (Schulz et al., 1997), further research has indicated that this interpretation is far from clear-cut. While Schulz et al. (1997) and Hollerman and Schultz (1998) initial sets of findings indicated that dopamine neurons convey signed prediction errors, this interpretation is hard to reconcile with evidence that dopamine neurons in VTA respond to a multitude of potentially important stimuli. For example, it has been known for decades that VTA dopaminergic neurons respond to unconditioned sensory stimuli in both auditory and visual modalities (Chiodo, Antelman, Caggiula, & Lineberry, 1980; Horvitz, Stewart, & Jacobs, 1997). Furthermore, the finding that aversive stimuli increase phasic activation of VTA dopamine neurons has been replicated

numerous times (Brischoux, Chakraborty, Brierley, & Ungless, 2009; Guarraci & Kapp, 1999; Mantz, Thierry, & Glowinski, 1989; Matsumoto & Hikosaka, 2009). Thus, if dopaminergic prediction errors encode only linear, value-based prediction errors, then a dopamine increase for noxious stimuli is perplexing.

There is mounting evidence that dopamine neurons convey different signals depending on anatomical location. Brischoux, Chakraborty, Brierley, & Ungless (2009) investigated this issue further, and found that separate dopamine neurons seemed to respond to aversive and rewarding stimuli within the VTA. Specifically, these results showed that neurons in the ventral portion of VTA were excited by foot shock, while neurons in the dorsal VTA were inhibited by foot shock. This suggests that neurons in ventral portions of VTA might code for outcome salience over value (since a foot shock is salient but not valuable). A portion of the neurons that were unresponsive to or inhibited by foot shock were excited by release of (escape from) foot shock. This shows that these neurons meet theoretical criteria for a value PE signal for aversive stimuli, since escape or release from a noxious stimulus is reinforcing. Aversive stimuli therefore discriminate neurons in the VTA which signal a value PE from those that signal a salience PE. This suggests that there might be separate populations of dopamine neurons conveying signed and unsigned PEs in the brain (Hikosaka & Matsumoto, 2009). Some recent evidence even indicates that individual dopamine neurons might convey both salience and value PEs at different time-scales. In a recent review Schultz (2016) suggested that dopamine neurons might convey a rapid salience signal, followed by a more gradual value signal.

1.5 PE signals in dopamine release in nucleus accumbens

One way to tease apart the neural pathway-specific mechanisms by which dopamine impacts reinforcement is through careful examination of dopamine release in downstream targets of VTA dopamine cell bodies. The main target of dopaminergic projections from the VTA in the mesolimbic pathway, colloquially known as the “reward pathway,” is the nucleus accumbens (NAcc). The nucleus accumbens is comprised of two main regions, the core and the

shell. Hart, Rutledge, Glimcher, & Phillips (2014) used fast-scan cyclic voltammetry (an invasive real-time electricity-based method to measure neurotransmitter concentration) in combination with principal components analysis (PCA) to assess reward-evoked real-time dopamine release in rat nucleus accumbens. They concluded that a principal component of dopamine release in the rat nucleus accumbens follows a bidirectional value signal, coding symmetrically for negative and positive signed PEs. This study also identified an earlier principal component of dopamine release in NAcc that did not satisfy requirements of a signed PE. The authors suggested that this component might correspond to an early indicator of salience, perhaps a response to a compound stimulus predicting the end of the waiting period and the beginning of the reward period. Importantly, this early dopamine component seems to correspond to the early, nondiscriminant activity of dopamine neurons described in Schulz (2016), suggesting that both early and late phasic responses of dopamine neurons are accurately reflected in separable principal components of NAcc dopamine release, and that NAcc dopamine release might encode an early salience signal followed by a later value signal.

Different regions of the nucleus accumbens might differentially encode reinforcing or salient aspects of stimuli. Budygin, Park, Bass, Grinevich, Bonin, and Wightman (2012) used fast-scan cyclic voltammetry to assess real-time dopamine release in regions of the rat midbrain during processing of aversive stimuli. This study showed that tail pinch (an aversive stimulus commonly used in rat research) resulted in fast increases in dopamine concentration in nucleus accumbens core but did not change dopamine concentrations in nucleus accumbens shell. Meanwhile, release of tail pinch rapidly increased dopamine concentrations in nucleus accumbens shell. Therefore, core dopamine concentrations seem to encode delivery of aversive stimuli, while shell regions seem to code for negative reinforcement due to escape from aversive stimuli. This result is supported by Wightman et al. (2007) in a study in which rats were conditioned to fear a tone. Tone presentation decreased dopamine release in the core region, but increased dopamine transmission in the shell region. Since the unconditioned stimulus (foot

shock) was not presented, the authors suggested that increases in dopamine concentration in the NAcc shell region might code for negative reinforcement due to successful avoidance of aversive outcomes. This is consistent with a signed dopaminergic PE signal in NAcc shell. However, dopamine release in the NAcc core increases for several different forms of negative stimuli in the manner expected of a salience signal. The increase in DA release for aversive stimuli seems to facilitate active avoidance behavior. Olesen, Gentry, Chioma, & Cheer (2012) measured real-time dopamine release in NAcc core during an active avoidance paradigm. The authors found increases in accumbal dopamine release prior to successful avoidance, which seems linked to dopamine increases signaling the potential to avoid aversive stimuli. Accumbal dopamine release decreased prior to unsuccessful avoidance responses, which the authors linked to decreases in dopamine release signaling delivery of aversive stimuli. Based on these results, dopamine release seems to carry different information content in the nucleus accumbens core vs. shell. Specifically, dopamine release in the shell of the nucleus accumbens seems to carry a traditional dopaminergic value PE, while dopamine release in the core of the nucleus accumbens instead seems to carry a dopaminergic salience signal.

1.6 PE signals in dopamine release in prefrontal cortex

A separate group of dopaminergic cell bodies project from the VTA to the prefrontal cortex, forming the mesocortical dopamine pathway. Critically, these neurons differ in many functional properties from more conventional dopamine neurons. Conventional dopamine neurons exhibit broad action potentials (action potentials occur over a relatively long period of time), low frequency tonic spiking with high-frequency burst firing capacity, and strong post-inhibitory rebound spiking (cells spike when inhibition is released, without need for additional excitation). Lammel, Hetzel, Hackel, Jones, Liss, & Roeper (2008) demonstrated that, contrary to prior belief, only neurons in the dorsolateral VTA projecting to the nucleus accumbens shell displayed these conventional properties of dopamine cells. Using these conventional properties to define dopamine cells might have led to an oversampling of dopamine neurons projecting to

the nucleus accumbens shell in past studies of prediction error signaling. Importantly, dopamine release in the nucleus accumbens shell seems to display a signed dopaminergic PE, in line with many studies which have observed signed PE signaling in groups of dopamine neurons selected using conventional electrophysiological properties.

In contrast, Lammel et al. (2008) showed that dopamine neurons in ventromedial regions of the VTA, projecting to medial prefrontal cortex and nucleus accumbens core, have very different features. More specifically, these neurons do not show the conventional criteria and therefore may not have been selected in previous studies examining prediction errors. These neurons fire persistently at much higher frequencies than conventional dopamine neurons and can be effectively silenced by inhibition (that is, these neurons do not exhibit post-inhibitory rebound spiking). In contrast, inhibition does not silence conventional dorsolateral VTA dopamine neurons, which exhibit strong post-inhibitory rebound properties. Furthermore, dopamine neurons respond to reinforcement in different ways, depending on where they project to. Lammel, Ion, Roeper, & Malenka (2011) showed that rewarding events change excitatory strength of dopamine neurons projecting to nucleus accumbens shell, but not those projecting to prefrontal cortex. In contrast, aversive stimuli changed the excitatory weights of dopamine neurons projecting to prefrontal cortex. These results suggest that a population of conventional dopamine neurons, which are likely overrepresented in studies of dopamine signaling, project to the nucleus accumbens shell and carry a signed prediction error signal. Meanwhile, populations of unconventional dopamine neurons project to the nucleus accumbens core and prefrontal cortex, and carry salient, primarily aversive, information. This distinction was shown experimentally for the first time by Lammel et al. (2012). Optogenetic stimulation of dopamine neurons projecting to nucleus accumbens shell induced conditioned place preference, while optogenetic stimulation of dopamine neurons projecting to prefrontal cortex induced conditioned place aversion. This study causally implicated separate populations of dopamine neurons in the

control of reward and aversion signaling, mediated by signed and unsigned prediction error signals respectively.

Traditionally, dopamine neurons have been selected using a priori identified conventional criteria. Recent evidence suggests that these properties are only true of a distinct population of dopamine neurons projecting to the nucleus accumbens shell. Importantly, these studies have generally identified selected dopamine neurons as carrying a value PE, and dopamine release patterns in the nucleus accumbens shell are largely reflective of a value PE as well. However, populations of dopamine cells exhibiting less conventional properties preferentially project to the nucleus accumbens core and prefrontal cortex. Critically, these neurons are excited by aversive stimuli and produce strong conditioned place aversions when optogenetically stimulated. This suggests that the population of dopamine neurons projecting to prefrontal cortex does not encode value PEs, but instead is maximally sensitive to salient (primarily aversive) stimuli. Furthermore, the lack of post-inhibitory rebound in this population of dopamine cells suggests that these neurons can be effectively silenced by inhibition. Inhibitory GABAergic drive from prefrontal neurons might therefore be capable of inhibiting mesocortical dopamine output, therefore exerting top-down control over aversive learning.

1.7 Imaging Studies of PE Signaling in Humans

Research methods used in functioning human brains are limited by the need to be non-invasive and must be augmented using animal studies to draw neurobiologically relevant conclusions. For example, direct subsecond measure of dopamine release in the human brain cannot be assessed. Furthermore, electrical recording of human subcortical brain activity is impossible, except for certain patient populations with implanted recording electrodes. Despite these shortcomings, investigating the signaling of prediction errors in human brain activity is essential for an algorithmic understanding of human reinforcement learning. Here I review fMRI studies of subcortical and cortical oxygenated blood flow, which correlates strongly with neural activity in specific brain regions.

While research has identified human BOLD activity during reward anticipation and delivery (Knutson, Adams, Fong, & Hommer, 2001; Knutson & Cooper, 2005), targeting individual brainstem nuclei such as the VTA is a difficult proposition in fMRI research. The VTA is the size of a mere two (2) voxels using standard 3T fMRI acquisition templates. In one of the first investigations to successfully target human brainstem nuclei (VTA) using 3T fMRI, D'Ardenne, McClure, Nystrom, & Cohen (2008) imaged BOLD responses in the VTA during reinforcement processing. This study did not use high-field fMRI to acquire high-resolution images but instead made clever changes to image acquisition routines that allowed VTA imaging, including taking oblique slices that were chosen to include as much of the brainstem as possible, using altered normalization routines that have been shown to work better than normal routines for small brainstem nuclei, and coupling the pulse sequence to the subject's heart rate to minimize movement artifact from nearby blood vessels. This study showed that the BOLD response in VTA was significantly modulated by positive reward prediction errors, but not negative reward prediction errors. Therefore, oxygenated blood flow increased for rewards greater than expected, but did not decrease from baseline for rewards that were lower than expected. The authors note that this likely reflects the "restriction of range" issue noted above – most dopamine neurons in VTA fire at such low basal levels that a decrease from baseline is not detectable. This study did not identify any VTA correlate of negative reward prediction errors, or any VTA response to aversive events. The authors suggest that the number of cells in VTA signaling aversive events (approximately 30%) does not cause enough change in BOLD signal to be detected by fMRI imaging. Schott et al. (2008) conducted an innovative study examined fMRI imaging of brain activity and PET imaging of [11C]raclopride (a dopamine receptor agonist that is displaced by dopamine binding) displacement in human nucleus accumbens during reward anticipation. Therefore, this study measured oxygenated blood flow in VTA and dopamine binding in NAcc. Note that these two sessions were necessarily conducted separately. Furthermore, because of the slow dynamics of [11C]raclopride binding and

displacement, this study was unable to examine real-time dopamine binding dynamics. However, VTA activity correlated positively with decrease in [11C]raclopride binding in nucleus accumbens, suggesting that the same neural mechanisms underlying VTA/NAcc interactions in animal models are observable in human brain imaging studies.

fMRI studies examining negative or aversive prediction errors have largely supported evidence from animal models. Seymour et al. (2005) examined appetitive and aversive signaling using a negative reinforcement task. As predicted based on rat models, avoidance of pain gave rise to a positive prediction error-like signal in midbrain striatal regions. Furthermore, delivery of aversive stimuli resulted in a negative prediction error-like signal in orbitofrontal and anterior cingulate cortices. This corresponds to previous animal evidence in suggesting that mesolimbic dopamine projections from VTA might carry positive prediction error information to striatum, while mesocortical dopamine projections might carry negative or aversive prediction error signals to prefrontal cortex. As further support of this theory, Metereau & Dreher (2012) showed that activity in human ACC and anterior insula covaried with a salience prediction error rather than a value prediction error using appetitive and aversive juice. This pattern was confirmed in a recent meta-analysis of neuroimaging studies. Garrison, Erdeniz, & Done (2013) used activation likelihood analysis to examine studies of human prediction error signaling. This study showed that positive (reward) prediction errors was largely confined to striatal areas, as expected based on the anatomy of the brain mesolimbic dopamine pathway (e.g., NAcc shell). Furthermore, negative (aversive) prediction errors were largely associated with activity in prefrontal cortex and habenula (which projects to ventral VTA and carries information about aversive events).

Numerous fMRI studies have implicated striatal BOLD signaling in reward prediction error signaling in the human brain. Most of these studies do not have sufficient spatial resolution to image individual subcortical nuclei. However, these findings generally agree in showing subcortical BOLD increase in line with a value prediction error, as expected based on previous evidence from rat brain studies (NAcc core). Furthermore, human prefrontal cortex oxygenated

blood flow is strongly related to aversive stimuli and reflects negative prediction errors. These findings agree with previous animal research and suggest that separate populations of dopamine neurons originating in VTA and projecting to striatum and prefrontal cortex code reward and aversive prediction errors respectively.

1.8 Electrophysiological Studies of PE Signaling in Humans

While human electrophysiology does not have the spatial resolution of fMRI or the ability to image subcortical structures, EEG and MEG have been used extensively to study reinforcement learning and prediction error signaling, as well as to make inferences about dopamine activity. The general approach that is used in this line of research is to measure prefrontal neural activation during a task or activity that has a known effect on subcortical dopaminergic nuclei, such as reward anticipation and delivery. In this section I review EEG and MEG studies which record electromagnetic activity from the surface of the scalp, and therefore measure signals on the same millisecond time scale as invasive recording of dopamine neuron electrical activity.

Most EEG studies of reinforcement have examined event-related potentials (ERPs; averaged EEG), with a particular focus on the error-related negativity (ERN) and feedback-related negativity (FRN). These components of the event-related potential occur following behavioral errors and valenced feedback, respectively, and are dominant over prefrontal cortical regions. Early theories of the ERN and FRN (known as reinforcement learning FRN theory, or RL-FRN) posited that these deflections in the ERP were due to dopamine activity and reflected reward prediction errors. Holroyd & Coles (2002) posited that dopamine chronically inhibits the apical dendrites of motor neurons in anterior cingulate cortex, and that changes in these negative ERP deflections could be due to a pause in the baseline release of dopamine from VTA. This pause in baseline release of dopamine would be expected if VTA conveys a signed negative prediction error to anterior cingulate cortex. Importantly, Nieuwenhuis, Yeung, Holroyd, Schurger, & Cohen (2004) showed that the amplitude of the FRN depends on the difference

between expected and actual outcomes and was most negative for unexpected poor outcomes in positive reinforcement conditions (reward omission), in line with a reward prediction error signal. However, neither of these studies included aversive conditions. A later study by Holroyd's team (Holroyd, Pakzad-Vaezi, & Krigolson, 2008) reinterpreted the RL-FRN hypothesis under the idea that the potential difference between correct and error outcomes might actually represent a positivity in the waveform induced by delivery of reward, leading some researchers to refer to this potential instead as the Reward Positivity. This change in the theory did not require a complete overhaul of the dopaminergic hypothesis of the FRN, as the authors still contend this signal results from a reward signal conveyed by conventional dopamine neurons. Notably, this interpretation requires physiological evidence that DA neurons forming the mesocortical pathway actually transmit reward signals, which is lacking in animal studies reviewed above.

Recent research has contradicted the RL-FRN hypothesis. Talmi, Atkinson, & El-Deredy (2013) recorded EEG while participants completed a combined positive and negative reinforcement task. For positive reinforcement conditions, a secondary reinforcer was used (money) and for negative reinforcement a primary reinforcer was used (pain). As expected, FRN amplitude was greater for omission of reward than for reward delivery, since dopamine neuron firing in VTA decreases from baseline for omission of an expected reward. Unexpectedly, FRN amplitude was greater for omission of an aversive painful stimulus than for avoidance of the aversive stimulus. If the FRN represents a cortical correlate of a dopaminergic prediction error, then this would suggest that the dopamine neurons projecting to prefrontal cortex increase firing for delivery of an aversive stimulus compared to avoidance of a negative stimulus. This is in line with rat research outlined above, which suggests that the major dopaminergic inputs to prefrontal cortex are preferentially excited by salience prediction errors. Therefore, both reward and punishment delivery would be encoded by an increase in unconventional VTA neuron firing which is conveyed to prefrontal cortex. The results from Talmi, Atkinson, & El-Deredy (2013), as

well as basic animal research detailed above, fundamentally contradict the idea that dopamine release in PFC should follow a value function, an idea that is central to the RL-FRN hypothesis.

Several studies have since replicated and extended the results of Talmi, Atkinson, & El-Deredy (2013). Huang & Yu (2014) examined the FRN in a prediction paradigm which instructed participants of the likelihood of a win, then showed the outcome of that trial (no response was required). In line with Talmi et al. (2013), FRN amplitude was greater for omission of reward than for reward delivery but was also greater for punishment delivery than punishment omission. Recently, our lab replicated and extended these results (Rawls, Miskovic, Moody, Lee, Shirtcliff, & Lamm, *under review*). This study included a control condition, which primed the participant with an expectation of zero reinforcement value and where the outcome was always zero. As expected, there was no difference in FRN amplitude for correct and error feedback when the feedback did not include a prediction error. This result suggests that the FRN is driven by a conveyed dopaminergic salience prediction error and is not merely the result of error monitoring. Sambrook & Goslin (2014) conducted a study where they parametrically modulated prediction error and reported significant salience prediction errors in early FRN time windows and mixed evidence for value prediction errors across multiple time windows. Furthermore, Sambrook & Goslin (2015) conducted a meta-analysis of published ERP studies using great grand averages, a technique which measured published ERP waveforms directly and averaged waveforms for conditions across many individual studies. These results suggest influences of salience prediction errors across much of the ERP waveform including the FRN time period, as well as influences of value signals which sometimes overlapped with salience signals. These more recent results are in agreement with known neurobiology and so represent a step forward from early theories suggesting the FRN represents a signed dopaminergic prediction error, e.g. Holroyd & Coles, 2002. However, the relationship between dopamine neurotransmission and FRN prediction error signaling is at this point only theoretical and has never been critically examined.

1.9 The Neglected Role of Time in Dopamine Reinforcement Signaling

While Sutton & Barto's (1987) computer science focused class of TD-RL theories is a dominant perspective in the study of reinforcement learning, the idea of time, as well as trial-by-trial variations in brain activity, is completely eschewed in traditional neuroscience studies which rely on the averaging of many trials to obtain an accurate measure of brain activity related to reinforcement delivery or omission. Despite the omission of time in traditional neuroscience, studies have demonstrated a strong role of the passage of time on dopamine release, which changes the weights of learning outcomes over time. For example, Niv, Daw, Joel, and Dayan (2006) generate an average reinforcement learning model in which subjects choose both what action to perform and what latency to perform it at based on a combination of previous task outcomes weighted by time. This important application of DA theories to free-response tasks specifically implicates a change in learning signals both over time and momentarily (trial-by-trial), and makes a strong case for the separation of DA signals into tonic (slow) and phasic (momentary) components. Specifically, DA concentrations are predicted to change slowly over time as an organism learns, in addition to momentary changes in DA concentration following individual reinforcing events. Furthermore, TD-RL models suggests that a PE signal is generated whenever an organism's expectation of reward changes (Sutton & Barto, 1987), which requires a slow methodical change in DA signaling over time to achieve reinforcement learning.

The idea that DA signals should change slowly over time as an organism learns (tonic) is supported by studies completed by Montague, Dayan, Person, & Sejnowski (1995) and Montague, Dayan, & Sejnowski (1996) focusing on the insect equivalent of DA, octopamine. These results indicated that in insect models, octopamine release closely matched the predictions of the TD-RL model. Specifically, octopamine seems to signal phasic (momentary) changes in outcome expectation based on previous outcomes that over time progress to slow (tonic) changes in octopamine levels as insects learn what to expect. Niv (2007) integrates TD-

RL theories of dopamine in learning to suggest that slow (tonic) changes in DA concentration integrate cost-benefit action analyses in determining the speed of responses, necessarily neglected in all studies that examine only the response an organism gives (correct or incorrect) but not the speed of that response. Specifically, phasic shifts in DA concentration immediately following reinforcing outcomes might change learning rates, while over time, tonic changes in DA concentration might facilitate learned aspects of task performance including reaction times.

Indeed, Beeler, Daw, Frazier, & Zhuang (2010) show important evidence linking slow changes in tonic DA concentrations to the exploitation of reward learning. Beeler et al. (2010) examined the behavior of hyperdopaminergic knockout mice in a semi-naturalistic instrumental learning task. This analysis demonstrated that hyperdopaminergic mice displayed normal learning from recent outcomes. That is, knockout mice correctly learned to associate cues and outcomes (measured as correct and incorrect responses on lever presses). However, these knockout mice instead showed a diminished ability to exploit that learning over time (measured as a slower adaptation of responses following learning). Specifically, this predicts that tonic changes in DA over the course of learning might shift how well or rapidly an organism exhibits learned behaviors over the course of a task, rather than whether or not the organism learns the correct response. These separable aspects of DA, or the idea that momentary (phasic) changes in DA concentration should change learning of task parameters while slow (tonic) changes in DA concentration should alter how rapidly an organism uses the information it has learned, have yet to be examined together in human subjects.

A computational model of midbrain medium spiny neurons (Guthrie, Myers, & Gluck, 2009) that incorporates tonic levels of DA, therefore considering changes in DA level over time rather than momentary or phasic dopamine shifts based solely on outcome, supports hypotheses suggesting that for learning to occur, DA levels must change tonically over time. This model was able to reproduce 1) the behavior of organisms with normal levels of DA, 2) the altered learning of DA deficient Parkinsonian patients, and critically 3) the improvement in DA

dependent learning experienced by medicated Parkinsonian patients, providing strong evidence of the importance of tonic DA levels in reinforcement learning over time. Parker, Alberico, Miller, & Narayanan (2013) demonstrate that expectancy-based reaction time shifts over the course of a task critically depend on the binding of DA in PFC, providing further evidence that response time, and not just choice, is heavily influenced by outcome expectancy effects. Importantly, Bryden, Johnson, Tobia, Kashtelyan, & Roesch (2011) demonstrated that spiking rate of neurons in the anterior cingulate cortex of rats signals PEs, but also the need for enhanced neural resources following unexpected outcomes, specifying a role of prefrontal DA mechanisms in enhancing learning-based activity following outcomes which are beyond those expected. This generates the hypothesis, critical to the current work, that cingulate-generated potentials should weight learning on future trials as a function of task outcomes. This change in activity should compound over time as learning improves and tonic DA levels in ACC change. This work also suggested that the most salient or attention-grabbing outcomes should weight future learning the most.

Specific to learning outcomes dependent on PFC activation, Rinaldi, Mandillo, Oliverio, and Mele (2007) showed that DA antagonists administered to PFC selectively impaired spatial learning that occurs over time in mice. This type of learning, which depends on time but not on PEs, therefore seems to selectively depend on tonic DA concentration in PFC. Wood, Simon, Koerner, Kass, & Moghaddam (2017) showed a role of real-time action strategy deployment in pursuit of a reward in VTA, contrary to the popular belief that real-time action selection depends solely on PFC-basal ganglia interactions. This suggests, among other things, that VTA DA release provides a crucial computational substrate that allows DA neurons to function in the real-time control of behavioral output. A convincing account of the distinction between tonic and phasic DA release in PFC is shown by Ellwood et al. (2017). This study shows that tonic (long-lasting) stimulation of VTA dopaminergic neurons projecting to PFC stimulates maintenance of previous action strategies, while phasic (momentary) DA release in PFC instead stimulates

behaviors deviating from previously learned sequences. Stopper, Maric, Montes, Wiedman, & Floresco (2014) demonstrate that causal override of phasic DA signals redirects action selection during decision making, rather than cementing previously learned outcomes in behavior. Furthermore, Howe, Tierney, Sandberg, Phillips, & Graybiel (2013) demonstrate that prolonged (tonic) DA signaling in striatum calculates both the temporal proximity and the expected outcome of rewards to be given in the future. This result is critical for the current study, as in human reinforcement learning studies, rewards are not generally administered until the completion of the study, therefore constituting a distant, rather than proximal outcome. Together, this set of intriguing results generates a hypothesis that tonic stimulation of DA neurons projecting to PFC facilitates repetition of previously learned action sequences with increased speed, while phasic stimulation of DA projections to PFC facilitates slower but more exploratory behaviors in an environment. Note that the distinction of tonic vs. phasic activity is separate from the distinction of signed vs. unsigned PE. Rather, knowing that DA projections to PFC carry salient, and primarily aversive, information allows the hypothesis that salient aversive stimuli will increase both momentary (phasic) DA levels in PFC, as well as slowly changing tonic DA levels in PFC over time. Meanwhile, rewarding outcomes should stimulate change in DA release (both momentary, and compounding over time) in striatal regions. More specifically, the distinction between tonic and phasic shifts in DA concentration is about the time scale of signaling, not the PE content of the signal.

1.10 Current study: prefrontal cortical modulation of learning

The topic of whether and how the FRN reflects a reinforcement learning signal is hotly debated. The classical theory of the reinforcing properties of the FRN (Holroyd & Coles, 2002) suggest that the FRN is influenced by a signed prediction error generated by phasic dopamine release. This theory is generally vague about the biological underpinnings of these signals. Recent evidence suggests that neurons generating signed prediction errors do not project to prefrontal cortex (Lammel et al., 2008; Lammel et al., 2011), which is in line with recent

evidence that the FRN represents a salience prediction error (Huang & Yu, 2014; Rawls et al., *under review*; Sambrook & Goslin, 2014; Sambrook & Goslin, 2015; Talmi et al., 2013). However, despite the neurobiological plausibility of the theory that the FRN reflects salience signals transmitted from VTA via the mesocortical dopamine pathway, the relationship of FRN PE signaling and dopamine neurotransmission has never been examined. More importantly, previous research on the reinforcement learning significance of the FRN completely disregards the importance of changes in tonic (slow) DA release in PFC, as the averaging of trials does not allow us to examine tonic (slow) changes in PFC activation over time as a human learns. Averaging over trials, and therefore nullifying the change in DA levels over time, allows only an examination of momentary (phasic) DA signaling in PFC but averages over time and obscures slow changes in DA concentration. Specifically, by averaging over all like trials (for example reward omission, or reward delivery), tonic (slow) changes in DA are averaged over, leaving only the phasic (momentary) DA signal immediately following reinforcing events. To examine how slow (tonic) changes in PFC DA concentration might influence the course of reinforcement learning, different methods are necessary. I suggest that a new approach is necessary to investigate the extent to which tonic (slow) changes over time of DA in PFC. Instead of the standard approach in EEG studies, which involves averaging of brain responses over many trials, I instead fit computational models to single trials of feedback-related cortical brain signal that allow for the passage of time to be included as a quantitative predictor in brain activity.

1.10 Hypotheses

Our hypotheses are that 1) the FRN will be sensitive to both reinforcement type (positive or negative) and prediction error. I expect a pattern of results indicating a salience prediction error – FRN activation should be less negative (more inhibited) for unexpectedly bad and unexpectedly good results compared to expected results, in both positive and negative reinforcement. Furthermore, based on the aversion specificity of dopamine neurons projecting to prefrontal cortex, I expect less negative (more inhibited) FRN activation for negative

reinforcement than for positive. Most importantly, I hypothesize that a single-trial analysis using trial number (i.e. the passage of time across the duration of the task) as a predictor of the FRN will reveal 2) an effect of change over time (trial number) on the FRN, which would concur with evidence suggesting that tonic changes in DA concentration influence the PFC representation of reinforcement learning. This slow change in PFC signal should reflect learning over time and be preferentially driven by salience, rather than value, of delivered outcomes. Finally, I hypothesize that 3) time-dependent changes in the FRN will drive faster responding over the course of the task, while phasic violations of outcome expectation will instead facilitate slower responding. This is in line with animal studies detailed above demonstrating that tonic changes in DA level in PFC cement previously learned action strategies and contributes to faster repetition of learned actions, while phasic shifts in PFC DA levels instead facilitate slower and more exploratory behavior. These hypotheses are consistent with research relating slow changes in tonic levels of DA to faster responses in animals.

1.11 Importance and theoretical impact of proposed study

There is debate about whether the FRN reflects a salience or value PE, and whether this prediction error is influenced by dopamine or not. Notably, this debate has not yet extended to whether this signal responds primarily to tonic or phasic DA release in PFC. Previous results from our lab indicate that the FRN reflects a salience PE, but these results were derived in the standard way, i.e. by averaging over trials and therefore disregarding the passage of time. Salience PEs and value PEs are signaled by dopamine in a projection-specific manner (VTA to PFC = salience; VTA to NAcc shell = value), and these signals are impacted by slow changes in tonic DA concentration over time. Specifically, I expect that both momentary (phasic) and slow (tonic) changes in level of DA in PFC should depend on salience (not value) of delivered reinforcing outcomes. This prediction utilizes basic neuroscience research showing that dopamine neurons projecting to PFC signal primarily aversive salience, and is in line with our hypothesis that the FRN reflects a dopaminergic salience prediction error. Furthermore, I test

the relationship between the passage of time (trial number) and FRN signaling and hypothesize, based on basic animal research detailed above, the FRN (as a measure of PFC activation) should respond not only to phasic (trial-by-trial fluctuations in DA level) but also to tonic changes in DA level (which occur over the course of time). I suggest that both tonic and phasic task-based information is integrated in cortical DA signaling, which should speed the acquisition of reinforced behaviors while also facilitating action change when needed. I predict that this learning effect will be strongest following aversive outcomes, again in accordance with basic studies demonstrating that the mesocortical DA pathway primarily carries aversive information. Understanding the representation of expectancy violation and time-based learning in the prefrontal cortex is key to understanding many disorders revolving around reinforcement and dopaminergic dysfunction, such as addiction. Furthermore, these results will inform basic theories about how reinforcement signals are communicated through the brain.

2 Methods

2.1 Participants

Participants were 59 undergraduate students (37 female, mean age 19.2 [SD 2.06], 2 left handed) who completed the study after giving informed consent. Participants were excluded from signing up for the study if they had a self-reported current psychiatric diagnosis, uncorrected visual impairments, or were currently using psychoactive medication. All participation exclusion criteria were listed on the online recruitment platform used by the University of Arkansas Department of Psychological Sciences, but were not verified via clinical interview or otherwise. All procedures described in this study were approved by the University of Arkansas Institutional Review Board (Protocol # 1708016049). After recruitment, one participant was excluded due to failure to understand task instructions and two additional participants were excluded from data analyses due to having too few trials in one or more conditions to analyze (< 30 trials). All data analyses were conducted on a sample of 56 participants. Participants

received course credit for their time in the laboratory (6 credit hours, 8 total required for General Psychology students). Additionally, participants were informed they would be paid a cash sum based on how many points they accrued throughout the reinforcement task. Participants received one US dollar ($1) for every 2000 points they accrued, and on average participants received a bonus of $7 at the end of the study (minimum $2, maximum $12).

2.2 Reinforcement learning task

Participants completed a reinforced flanker which was a modification of the flanker task (Eriksen & Eriksen, 1972) using arrow stimuli (Figure 2). At the beginning of every trial, a fixation cross was presented. Following the initial fixation, participants were shown a black-and-white shape, signifying whether the trial was positive or negative reinforcement. Negative and positive reinforcement were presented in pseudo-random order, allowing us to examine both changes over time in behavior and brain function as well as the phasic shifts due to unexpected reinforcement outcomes in a trial-by-trial fashion. This shape disappeared after 500 ms and was replaced with a fixation cross (+) that lasted 500 – 700 ms. A set of congruent (< < < < < or > > > > > >) or incongruent (< < > < < or > > < > >) arrows were then shown for 100 ms, followed by a fixation cross lasting from 900 – 1100 ms. Valence feedback (correct or incorrect) was shown for 500 ms, followed by a fixation cross lasting 500 – 700 ms. Finally, point feedback was shown for 1000 ms.

Importantly, the average return for correct positive reinforcement trials was 50 points, but the actual return varied linearly from 20 points to 80 points on a trial-by-trial basis. Likewise, the average return for correct negative reinforcement trials was zero points, but the actual return varied from -30 to 30 points. Correct answers always resulted in better outcomes than incorrect answers. Before participants began, they completed 50 trials for practice. During practice trials, the outcome was always as expected (i.e. for positive reinforcement correct answers resulted in a gain of 50 points, and for negative reinforcement correct answers resulted in no loss of points). Participants could not move on to the main task until they demonstrated understanding

of the task during the practice, measured as an accuracy rate of 80% or above during practice. One participant was excluded from the study at this point for failure to understand task instructions (could not pass the practice after 3 attempts or 150 trials). Four additional participants (out of the 56 included) required a second attempt on the practice block in order to achieve 80% or greater accuracy. Participants responded using their left thumb (for left arrow targets) or their right thumb (for right arrow targets). Participants completed 960 trials of this task, divided into 16 blocks of 60 trials each. Completion of the reinforcement task required approximately 90 minutes. In the interest of presenting a minimal number of a priori defined analyses and hypotheses, only the EEG time-locked to the final point feedback is considered as a dependent variable in this manuscript. I do not analyze EEG time-locked to any stimuli other than the final point feedback. The current analysis was further restricted to only correct trial feedback, in order to remove the confound of error / performance monitoring on PE signaling. Note that typical reinforcement learning studies of the FRN require correct answers to provide “good” outcomes, and incorrect answers to provide “bad” outcomes, leading to the confounding of reinforcement processing and performance monitoring. Since many of the brain regions implicated in reinforcement signaling also respond to performance errors, the typical design of these studies severely restricts the interpretation of observed brain activity as a reinforcement learning signal. By creating an expectation of average outcome in the practice and then systematically providing more or less points than expected on correct trials, I separate the effect of PE (was it better or worse than expected?) from the effect of error monitoring (was the outcome correct or incorrect?).

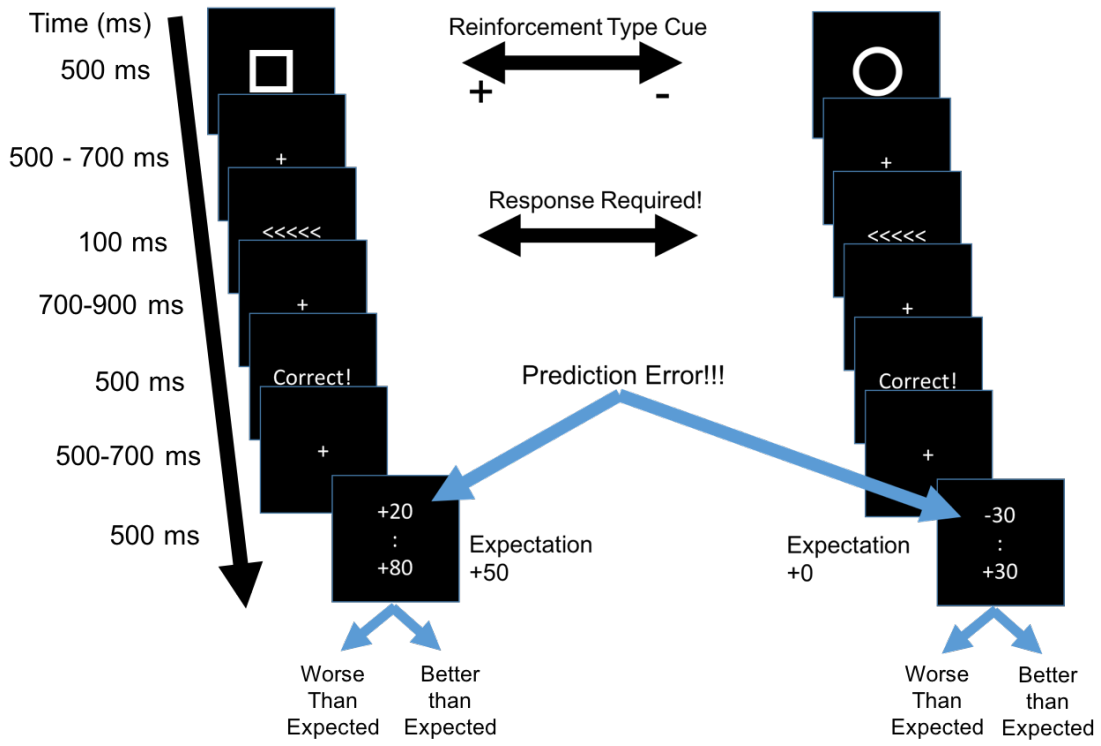


Figure 2. Task diagram for the modified reinforcement flanker paradigm. Participants were cued as to whether the current trial was to be positive or negative reinforcement with a white square or a white circle, respectively. Participants then had to respond to a flanker arrow stimulus which was either congruent or incongruent (congruent and incongruent trials were averaged over for the current manuscript). Only correct trials were analyzed, and participants were shown correct feedback on every trial prior to the point feedback being shown. Finally, participants were given some amount of points which was on average +50 for positive reinforcement and +0 for negative reinforcement. The crucial task manipulation was that every trial, the amount of points given deviated slightly from the overall mean expectation, generating outcomes which were worse-than-expected or better-than-expected.

2.2 EEG processing

EEG were sampled at 1000 Hz using a 129-channel EGI sensor array referenced to vertex (Philips EGI, Inc.). Recording began after impedances were reduced below 50 kΩ. Data were processed using EEGLAB 15 (Delorme & Makeig, 2004) and MATLAB 2017 (Mathworks). Continuous data were downsampled to 125 Hz with anti-aliasing, low-pass filtered at 30 Hz using a zero-phase FIR filter, and high-pass filtered at 0.1 Hz using a zero-phase FIR filter. Bad channels were removed using built-in eeglab functions if the channel SD was 4 or greater. Copies were made of each dataset, which were high-pass filtered at 1 Hz using a zero-phase

FIR filter, in preparation for computing independent components analysis (ICA; Makeig, Jung, Bell, & Sejnowski, 1995). All data were epoched into 3 second windows surrounding point feedback onset (from -1000 ms before to 2000 ms after). Infomax ICA was computed on the 1 Hz filtered dataset (Debener et al., 2010). Bad channels were not interpolated before running ICA. Artifactual independent components were detected using the SASICA plugin (Chaumon, Bishop, & Busch, 2015), which detected artifact components using a combination of three methods: 1) autocorrelation statistics, 2) focal component activity, and 3) routines from the ADJUST plugin (Mognon, Jovicich, Bruzzone, & Buiatti, 2011). Selected components were verified by visual inspection and removed from the data. ICA weights and artifact components calculated in the 1 Hz high-pass filtered dataset were copied to the 0.1 Hz high-pass filtered dataset; all further analyses were performed on the 0.1 Hz filtered dataset. Epochs containing fluctuations with voltage exceeding ±125 μV were detected and removed as well. Removed channels were interpolated using spherical splines and data were rereferenced to the montage average. Two subjects were excluded from all analyses due to having fewer than 30 clean trials in one or more conditions.

2.3 Conventional trial-averaged analyses

In order to render the results of this study comparable with the bulk of previous research, which used ERPs averaged over trials, I present initially a trial-averaged analysis. I split trials according to reinforcement type (positive or negative) and PE type (better or worse than expected). I analyzed the average behavior (RTs) after trials of each type in order to examine task-related response adaptation (learning) effects. I analyze the averaged ERP following the actual outcome (point gain/loss) cue for trials of each of the four trial types (positive/negative, better/worse) as well. The ERP was baseline corrected for the 200 ms preceding feedback presentation and was examined as the mean voltage from 250 – 350 ms at sensor FCz in line with standard FRN analysis procedure. All analyses were conducted in SPSS 26 using 2 (reinforcement type: positive or negative) X 2 (outcome: better or worse than expected)

repeated-measures ANOVAs. Any significant interactions were examined using the SPSS emmeans command. Since there were no covariates in this analysis this command provides the same result as if serial t-tests were used to examine the simple main effects. No multiple correction procedure was used in SPSS as SPSS does not apply a correction for any factors with fewer than three levels in an ANOVA (i.e. the results would be exactly the same if a Bonferroni correction was applied within SPSS).

2.4 Single-trial analysis

2.4.1 Single-trial behavioral analysis: changes in behavior over time

Single-trial analysis of behavior was initially conducted within each subject to examine task-related shifts in reaction times corresponding to reinforcement outcomes. This analysis was specifically intended to isolate effects of tonic vs. phasic task-related changes in learning rate. That is, within individual subjects, I sought to examine shifts in reaction time over the course of the task (tonic), following various magnitudes of prediction error (phasic). The specific task manipulation concerned PE signaling in positive and negative reinforcement conditions, during better or worse than expected outcomes, as a function of continuous prediction error. Finally, differences in learning were incorporated by use of a regressor for trial number (i.e. time on task). Therefore, I examined the reaction time in the trial following prediction error presentation in the task as a function of task constraints.

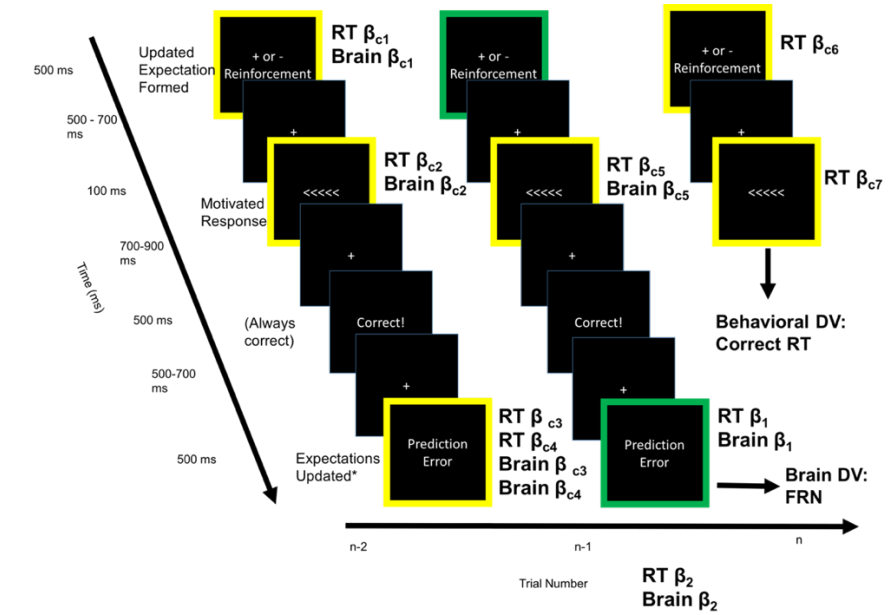
I began by controlling for sources of conflict monitoring or error monitoring that could influence task performance. I removed incorrect trials and trials following incorrect trials from consideration due to the high performance rate in the task and the fact that errors influence learning (performance monitoring control). Next, I fit single-trial robust regression (O'Leary, 1990) to within-subject RTs for each of four task conditions (positive reinforcement better-than-expected outcome, positive reinforcement worse-than-expected outcome, negative reinforcement better-than-expected outcome, and negative reinforcement worse-than-expected outcome), as the task conditions are categorical (Cohen & Cavanagh, 2011). This regression

equation was used to remove the influence of flanker congruency for the two flanker stimuli preceding the RT, to control for the influence of conflict monitoring & conflict adaptation on the ERP. This equation also included the continuous PE value returned on the previous trial, as well as the reinforcement cue from the previous trial (positive or negative reinforcement cue) to control for the context-dependent nature of the FRN. The studentized residuals of this model were returned for each subject and single-trial behavioral analysis was completed for each subject using a robust regression model fit to these residuals (RTs controlled for monitoring influences).

The main regression model of interest was a within-subjects model predicting reaction times within each of four conditions (positive reinforcement better-than-expected outcome, positive reinforcement worse-than-expected outcome, negative reinforcement better-than-expected outcome, and negative reinforcement worse-than-expected outcome). This allows further model comparison between conditions via examination of beta weights. The regression used a linear combination of PE (continuous linear variable), and trial number (i.e. time, ranging from 1 – 960), as well as an interaction of the two terms. Time and PE were z-scored before entering them in the regression so that the magnitude of trial number did not obscure that of PE (maximum 960 vs. maximum 80). All single-trial PE values were orthogonalized by the actual outcome value to ensure that any detected effects correspond to a true effect of PE. Note that the ability to fit precise models such as this is a key feature of single-trial within-subject regression analysis as this level of specificity and control is impossible to achieve in a trial-averaged analysis.

This model was fit to predict studentized residual RTs in the trial immediately following the prediction error outcome, in order to separate influences of trial number over the task and expectancy violations on response times. Beta weights from within-subject robust regression can be assumed to be Gaussian by the central limit theorem, and therefore were subjected to t-tests against a null hypothesis mean of zero, i.e. no relationship between the regressor and

reaction times. Each t-test proceeded with 53 degrees of freedom. Figure 3 indicates the task design and the equations used for single trial analyses.



Step 1: Control

$$RT_n = intercept + (\boldsymbol{\beta_{c1\text{-}7}})$$
$$FRN_{n-1} = intercept + (\boldsymbol{\beta_{c1\text{-}5}})$$

Return studentized residuals

Step 2: Test

$$RT_n = intercept + \boldsymbol{\beta_1} + \boldsymbol{\beta_2}$$
$$FRN_{n-1} = intercept + \boldsymbol{\beta_1} + \boldsymbol{\beta_2}$$

Add FRN to model

Step 3: Test

$$RT_n = intercept + \mathrm{FRN} + \boldsymbol{\beta_1} + \boldsymbol{\beta_2}$$

Figure 3. Regression model diagram for the reinforcement task. Single-trial modeling proceeded separately for positive and negative reinforcement, and worse-than expected vs. better-than-expected trials, resulting in four models for the RT and four models for the ERP. I used robust regression to model out all unwanted effects in a first control step while returning studentized residuals of RT and ERP. Control effects are shown with a yellow border. A second, meaningful analysis was conducted by using robust regression to fit all parameters of interest to both the single-trial RT and ERP. Finally, RTs were examined using stepwise regression as a function of task constraints, and as as a function of the ERP. This third step formally tests whether the ERP predicts reinforcement learning effects in this task.

2.4.2 Single-trial EEG analysis framework

For each participant, the single-trial FRN was measured at a cluster of mediofrontal sensors (radius 6 cm) centered on FCz (sensor 6 on the EGI cap). I restricted single-trial analyses to a cluster of neighboring prefrontal sensors based on standard measurement locations for the FRN. Since trial number resulted in positive beta weights at occipital sensors, this prevented the spurious selection of occipital sensors for FRN analyses. The single-trial FRN values were submitted to the same control analysis as the RT values described above, with the exception that control values following the FRN were excluded as they had not occurred yet (trial n). I then applied the regression model of interest to the single-trial ERPs from 100 – 600 ms post-feedback. For each subject, the mediofrontal sensor showing the greatest cumulative R2 was selected for further analysis as in (Bieniek, Frei, & Rousselet, 2013; Rousselet, Husk, Pernet, Gaspar, Bennett, & Sekuler, 2009; Rousselet, Gaspar, Pernet, Husk, Bennett, & Sekuler, 2010). The regression model applied to the EEG yields b values for each time point that can be assumed to be Gaussian by the central limit theorem and were therefore tested across subjects using two-tailed t-tests at each time point, followed by correction for multiple comparisons using Benjamini & Yekutieli's (2001) false discovery rate with an alpha of .05. I tested for condition differences in significant regression results using a series of 2 X 2 repeated measures ANOVAs (one for each time point from 96 – 600 ms, sampled at 125 Hz, therefore 64 ANOVAs). Significant ANOVA effects representing condition differences (due to reinforcement type or better vs. worse outcomes) were further examined using paired t-tests with 53 degrees of freedom.

2.4.3 Prediction of reinforcement learning over the course of the task

I fit a final model that predicted residualized RTs (calculated as described above) using a stepwise regression to assess whether mediofrontal activity following feedback presentation predicts reaction time in the following trial. The first step of this regression was the same as for the behavioral analysis, that is, the equation predicted controlled RTs as a function of the

previous PE and the previous reinforcement type. The second step of this equation entered the single-trial feedback-locked EEG as a regressor, to determine if mediofrontal activity predicts following-trial response time over and above the task itself. Again, this regression model applied to predict RT using the EEG yields b values for each time point that can be assumed to be Gaussian by the central limit theorem and were therefore tested across subjects using two-tailed t-tests at each time point, followed by correction for multiple comparisons using Benjamini & Yekutieli's (2001) false discovery rate with an alpha of .05.

3 Results

3.1 Conventional (trial-averaged) results

3.1.1 Trial-averaged analysis of reaction times

Reaction times were grouped according to which of the four different trial conditions they followed, producing four mean RTs for each participant (RTs following positive reinforcement better-than-expected outcomes, RTs following positive reinforcement worse-than-expected outcomes, RTs following negative reinforcement better-than-expected outcomes, and RTs following negative reinforcement worse-than-expected outcomes). Mean RTs following each of four trial types were analyzed according to a 2 (reinforcement type: positive or negative) X 2 (outcome: better or worse than expected) ANOVA. Results indicated a main effect of outcome (better or worse), $F(1,53) = 9.56$, $p = .003$, $\eta p2 = .15$, which was subsumed by a significant interaction between reinforcement type and outcome, $F(1,53) = 15.86$, $p = .0002$, $\eta p2 = .23$. Results of post hoc analyses confirmed that, for better-than-expected outcomes, RTs were faster following positive reinforcement trials compared to negative reinforcement trials, $p = .009$. The opposite was true worse-than-expected outcomes: RTs were faster following negative reinforcement trials compared to positive reinforcement trials, $p = .003$. Within reinforcement type, only positive reinforcement showed a significant difference. RTs following better-than-expected outcomes were faster than RTs following worse than expected outcomes, $p = 4e\text{-}6$.

This difference was not significant for negative reinforcement trials, p = .31. Graphical depiction of RT effects are shown in figure 4.

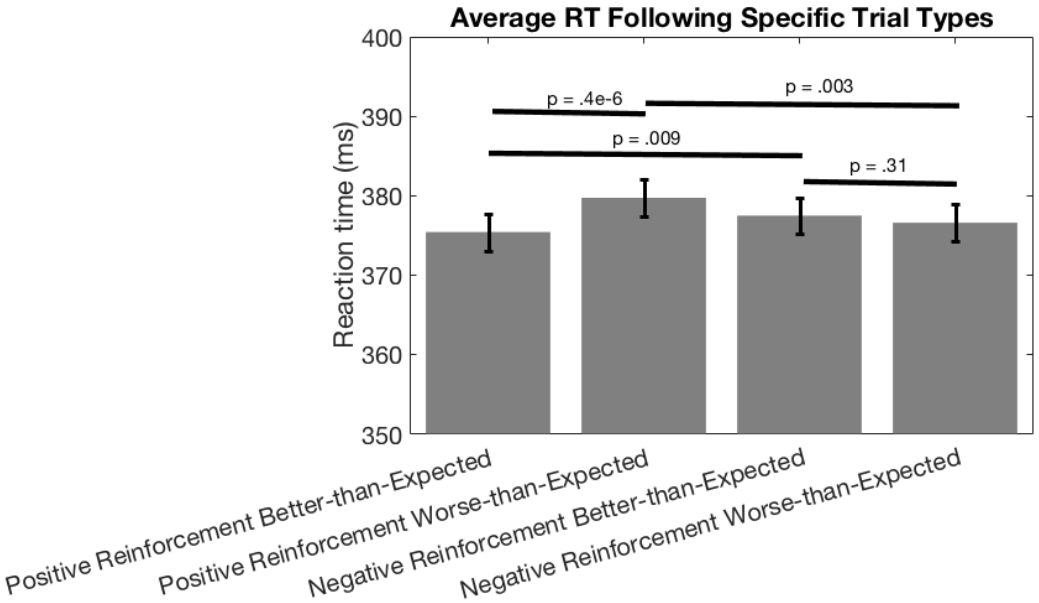


Figure 4. Trial averaged reaction times. Results indicated that, following better-than-expected outcomes, RTs were faster following positive reinforcement trials than following negative reinforcement trials. This effect was reversed following worse-than-expected outcomes. Finally, RTs were slower following worse-than-expected outcomes than following better-than-expected outcomes in positive reinforcement conditions only.

3.1.2 Conventional analysis of FRN amplitude

Similar to the RT analysis, FRN activations were grouped according to which of the four different trial conditions they followed and analyzed according to a 2 (reinforcement type: positive or negative) X 2 (better-than-expected or worse-than-expected outcome) ANOVA. Results indicated a main effect of outcome (better-than-expected or worse-than-expected), $F(1,53) = 14.05$, $p = .0004$, $\eta p2 = .21$, which was subsumed by a significant interaction between reinforcement type and outcome, $F(1,53) = 13.72$, $p = .001$, $\eta p2 = .21$. Results of post hoc analyses showed that the influence of outcome was only apparent for negative reinforcement outcomes. Negative reinforcement outcomes that were better-than-expected resulted in a more negative FRN compared to worse-than-expected, p = 5e-5. This indicates a reverse effect of PE

on FRN amplitude for negative reinforcement (that is, the FRN is more negative for better-than-expected than worse-than-expected outcomes in negative reinforcement), which is in line with the majority of recent research on the FRN. For worse-than-expected outcomes, FRN amplitude was more negative for positive compared to negative reinforcement trials, p = .0006. Furthermore, the effect of PE on the FRN was non-significant, which I attribute to the unique task manipulation which allows for a worse-than-expected reward to be delivered, as opposed to standard task designs which conflate full reward omission with worse-than-expected outcomes. The waveforms and difference topographic plots for this analysis are detailed in figure 5.

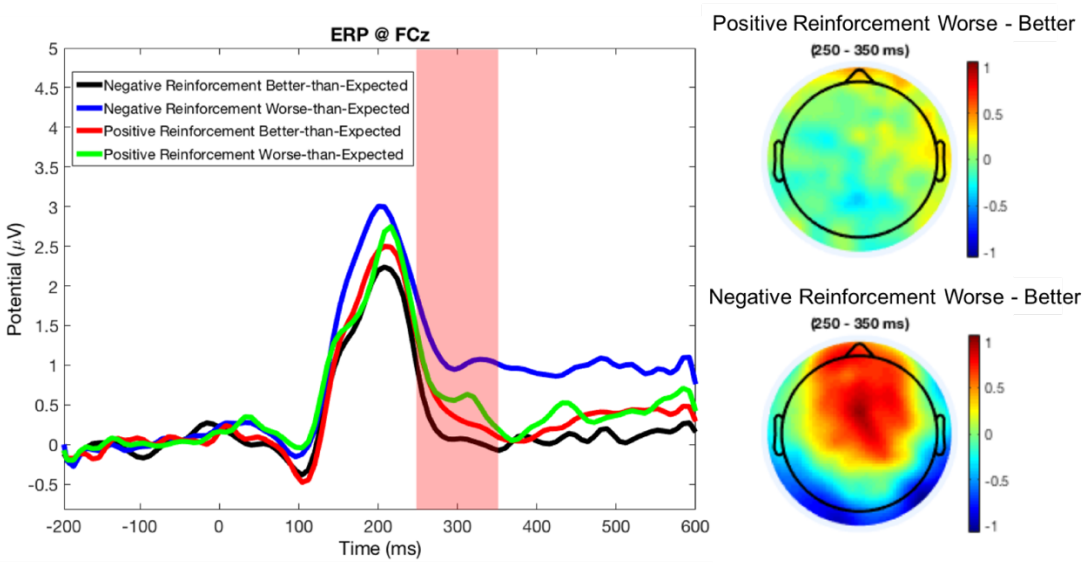


Figure 5. Graphical depiction of traditional ERP results. The left panel depicts the feedback-locked ERP for each of four conditions. Negative is plotted downward and the red shaded area indicates the time period used for analysis of ERP amplitudes (250 – 350 ms). Topographic plots to the right indicate the scalp distribution of the worse-than-expected minus better-than-expected difference. Topoplots are plotted with a standard hot-cold color template where red colors indicate positivities in the difference waveform and blue colors indicate negativities in the difference waveform. Both topographic plots are shown with a symmetric colormap to facilitate comparison of results. The colormap limits are indicated on the bar to the right of the topographic plots.

3.2 Single-Trial analysis of brain and behavior

3.2.1 Changes in reaction times are modulated by reinforcement salience

Rather than analyzing differences in mean RT, in this next section I use robust regression analysis to examine how single-trial RTs change over the course of the task (denoted as trial number from now on). For positive reinforcement conditions with better-than expected outcomes, only a main effect of trial number on RT was significant, mean beta weight = -.057, $t(53) = -3.39$, $p = .001$, which indicated that correct RTs became faster as trial number increased but this effect did not depend on the value of the prediction error. For positive reinforcement conditions with worse-than-expected outcomes, a mean effect of trial number was significant, mean beta weight = -.103, $t(53) = -2.49$, $p = .015$. This effect again indicated that correct RTs became faster as trial number increased but this effect did not depend on the value of the prediction error. This indicates an interesting null effect of continuous PE on reaction times in positive reinforcement. This is likely due to the fact that our task manipulation always resulted in some level of reward for positive reinforcement – correct trials, unlike task designs which conflate worse-than-expected outcomes with error commission.

Results for RTs following better-than-expected negative reinforcement outcomes indicated only a significant interaction between continuous PE value and trial number, mean beta weight = -.111, $t(55) = -2.983$, $p = .004$. Results for worse-than-expected negative reinforcement outcomes indicated significant main effects of both continuous PE, mean beta weight = .111, $t(55) = 2.772$, $p = .007$, and trial number, mean beta weight = -.06, $t(55) = -3.892$, $p = .0003$, as well as a significant interaction between continuous PE value and trial number, mean beta weight = .088, $t(55) = 3.005$, $p = .004$. These interaction effects were examined in more detail by utilizing the suggestions of Aiken, West, and Reno (1991), Barron and Kenny (1986) and Dawson (2014) in probing the impact of trial number on following reaction times at low (-1 SD) and high (+1 SD) levels of prediction error value.

For better-than-expected negative reinforcement outcomes, I found that more positive prediction error values (+1 SD, i.e. further from expectation) resulted in increased faster RTs over the course of the task, mean beta weight = .18, $t(55) = 2.021$, $p = .04$, but less positive PE values (-1 SD, i.e. closer to expectation) resulted in slower RTs over the course of the task, mean beta weight = -.04, $t(55) = -2.16$, $p = .03$. For worse-than-expected outcomes negative reinforcement outcomes, I found that more negative PE values (-1 SD, i.e. further from expectation) resulted in faster RTs over the course of the task, mean beta weight = -.15, $t(55) = -5.34$, $p = 1.97e\text{-}6$, but less negative PE values (+1 SD, i.e. closer to expectation) did not result in changes in response time as trial number increased ($p = .47$).

These effects demonstrate the influence of prediction errors and trial number on learning following negative reinforcement (by examining the next trial's behavior). Responses become quicker over trial number following unexpectedly better-than-expected negative reinforcement outcomes, and following worse-than-expected negative reinforcement outcomes responses become slower over the course of the task when outcomes are nearer to the expectation. Meanwhile, these results show that following positive reinforcement conditions, responses become quicker over the course of the task (trial number) and do not depend on the value of prediction errors. This suggests that in this task, participants learning was modulated by prediction errors in negative, but not positive, reinforcement conditions.

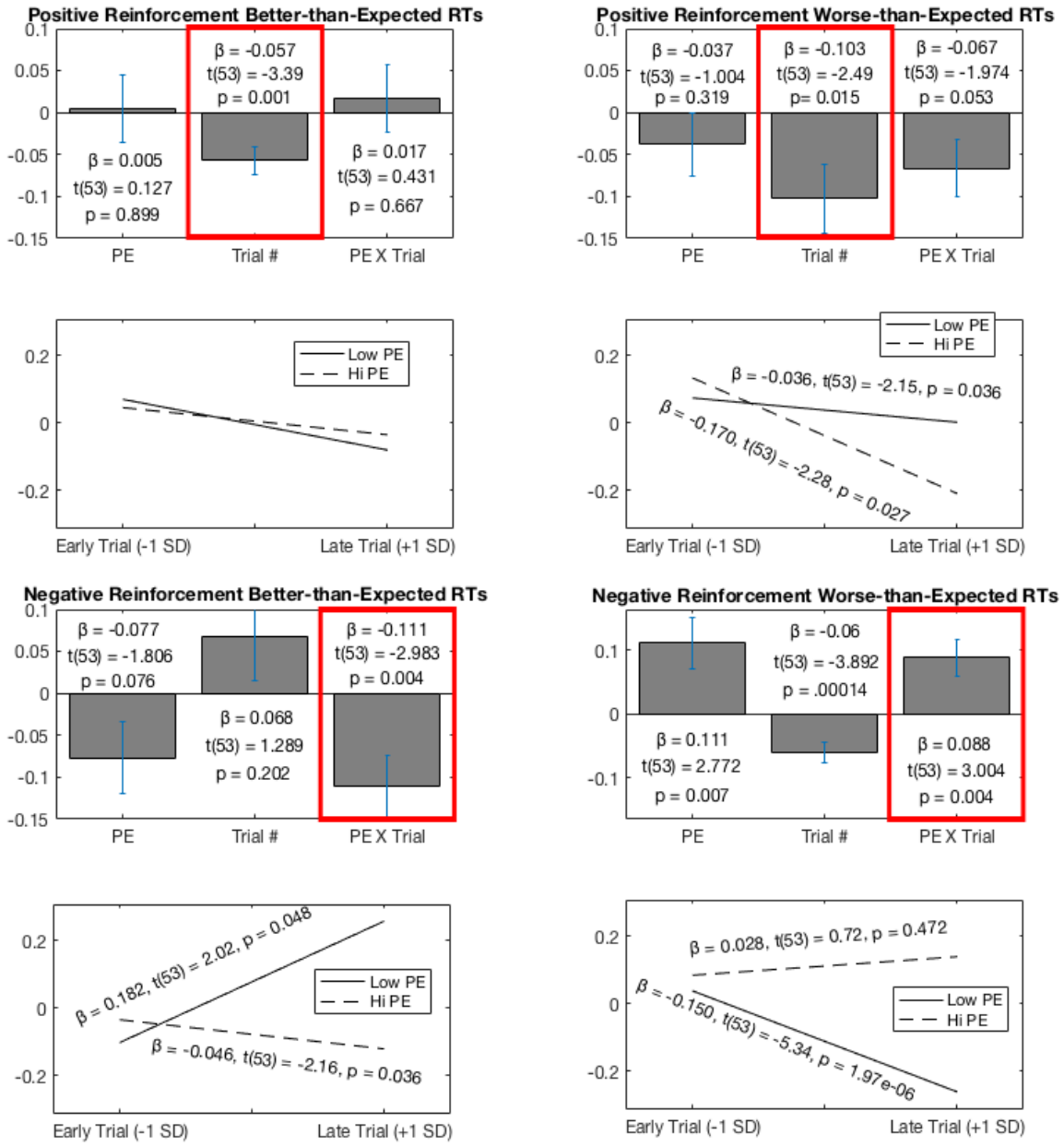


Figure 6. Single-trial reaction time analysis. I investigated the impact of reinforcement type (positive or negative), PE type (worse or better than expected), signed PE, and trial number by using separate single-trial robust regression of reaction times following worse-and-better than expected positive and negative reinforcement outcomes as a function of signed PE, trial number, and their interaction in order to specifically examine the influence of prediction error values on task learning, measured using RTs. Bar graphs depict mean between-subject regression results and t-tests for differences from zero; line graphs depict moderation analyses of significant interaction effects using the standard method of testing moderation effects at -1 SD and +1 SD from the mean. Important effects for further analysis are outlined in red for readability. There was only a main effect of trial for positive reinforcement conditions, but an interaction of trial and continuous PE for negative reinforcement trials.

3.2.2 FRN reflects processing of reinforcement and trial number

I analyzed the single-trial raw EEG by using robust regression in the same way as RTs described in the preceding section. This model was fit to every post-stimulus time point from 96 to 600 ms, yielding 64 regressions which were corrected for multiple comparisons using the false discovery rate (FDR; Benjamini & Yekutieli, 2001). Interestingly, single-trial ERP regression results indicated only significant main effects of trial number on the feedback-locked ERP; no effects of continuous PE were present (all $p > .016$, did not meet FDR cutoff). The effect of trial number on the ERP was not significant in all task conditions. Specifically, an early mediofrontal effect (maximal 140 ms) likely corresponding to trial number modulation of the frontal N1 was significant in positive reinforcement better-than-expected trials, positive reinforcement worse-than-expected outcome trials, and negative reinforcement worse-than-expected trials, while a later mediofrontal effect of trial number corresponding to the FRN was significant only for positive reinforcement better-than-expected trials and negative reinforcement worse-than-expected trials (i.e. the most salient outcomes). For N1 effects, results indicated that over increasing trial numbers the N1 grew more positive, i.e. habituated. For FRN effects, results instead indicated that the FRN grew more negative over the course of the task, likely corresponding to task learning effects which occur as trial number increases. Interestingly, the effect of trial number on the FRN was only significant for positive reinforcement trials with better-than-expected rewards and negative reinforcement with worse-than-expected aversive stimuli. All statistics for this analysis are presented graphically in figure 7.

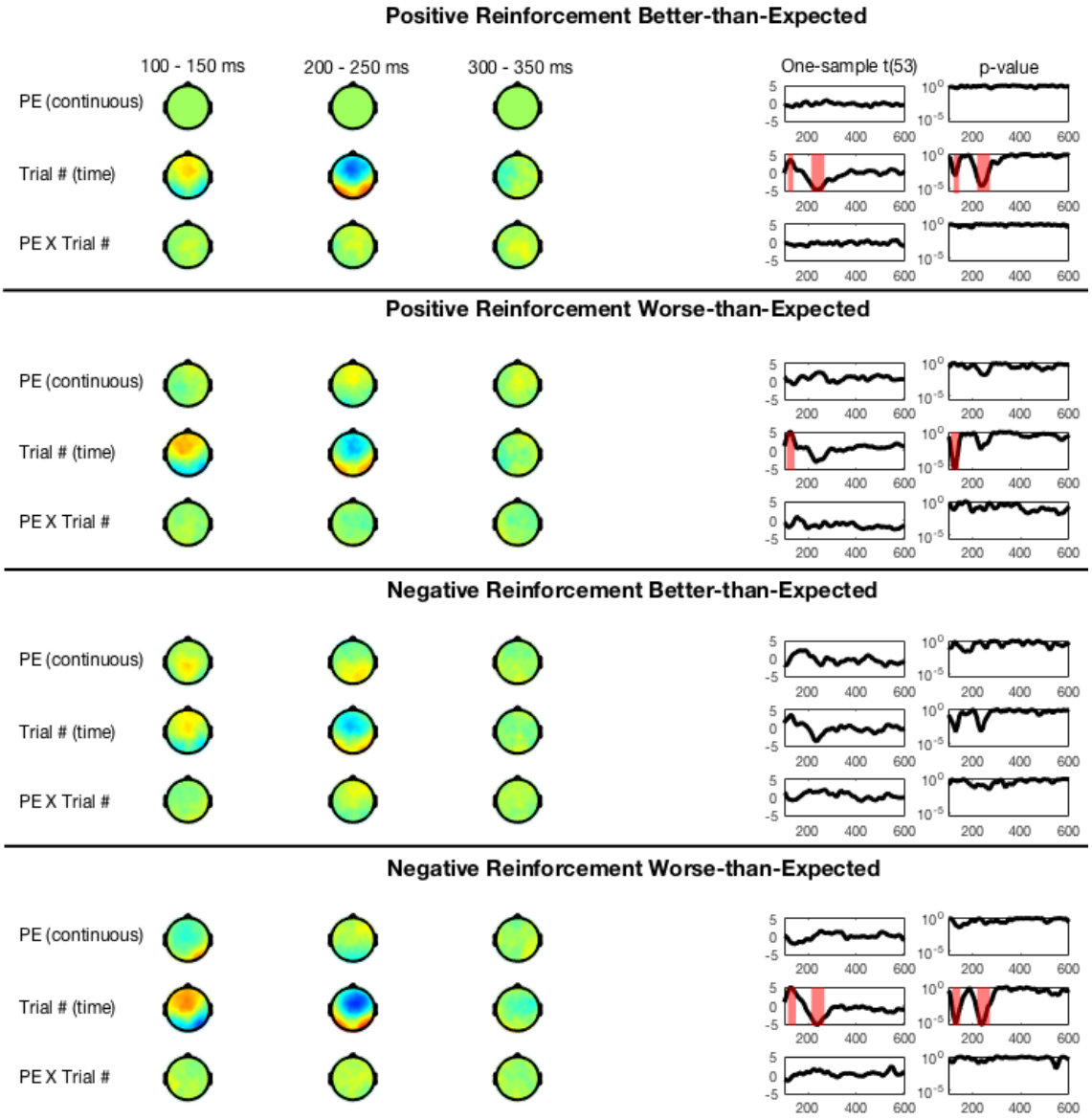


Figure 7. Single-trial ERP regression results. Results indicated no effect of continuous PE on the mediofrontal ERP. Instead, the main effect of trial number was significant for the FRN only when the outcome was salient (i.e. positive reinforcement better-than-expected outcomes, negative reinforcement worse-than-expected outcomes), in line with previous evidence that the FRN primarily reflects a binary decision of outcome-expectation match. Interestingly, there was also a main effect of trial number on the mediofrontal N1, which indicated that the N1 grew less negative (i.e. habituated) as trial number increased. This effect is line with the interpretation of the N1 as an early sensory potential, which is expected to habituate as a subject becomes familiar with task stimuli and requirements. Topographic plots are included to demonstrate the spatial extent of the observed effects, and do not indicate significance or lack thereof in themselves. Topographic plots are shown with a standard hot-cold color scheme where red indicates positive regression coefficients and blue indicates negative regression coefficients. Red shades in line plots indicate regions of significant single-trial regressions. Significance for within-subject regression coefficients was determined using one-sample t-tests, which test against a null hypothesized sample mean of zero.

3.2.3 FRN trial number effects differ by reinforcement and PE

While the previous results indicated that the regression effect of trial number was significant for positive reinforcement better-than-expected trials and negative reinforcement worse-than-expected trials (i.e. for only the most salient trials or the trials furthest from the average outcome over all conditions), this does not directly indicate whether the representation of trial number is different between different types of trials. That is, the effect of trial number might have been barely non-significant in one or more conditions, but not significantly different from other conditions. Therefore, I conducted a series of 2 X 2 repeated-measures ANOVAs on the beta coefficient for waveform trial number at each point from 96 – 600 ms. The statistical results of this set of ANOVA analyses, including F-values, p-values, and η_p^2 (a measure of ANOVA effect size), are presented graphically in Figure 8. Results after correction for multiple corrections showed no main effects of reinforcement type or outcome type on mediofrontal beta weights for trial number, but a significant interaction between reinforcement type (positive or negative) and outcome (worse-than-expected or better-than-expected) emerged from 200 – 250 ms indicating that the neural representation of trial number differed between the four conditions.

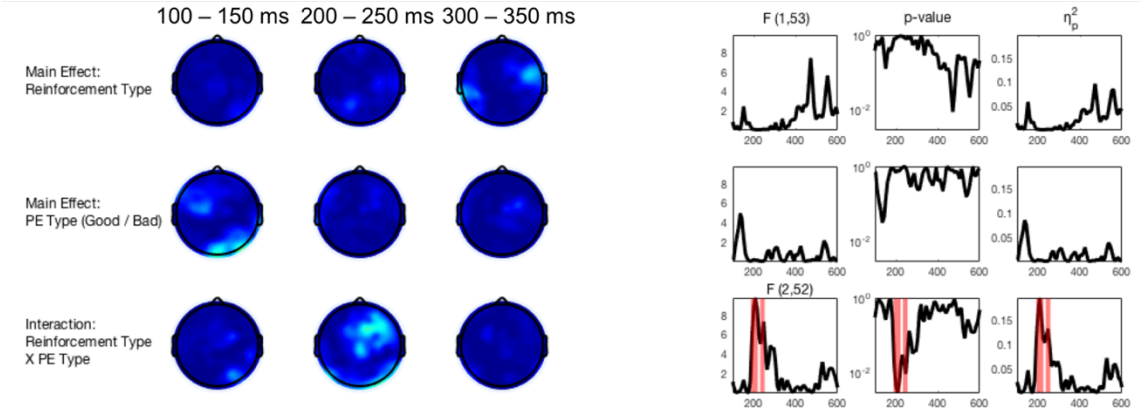


Figure 8. ANOVA testing of single-trial regressions. I tested whether regression weights for trial number differed by condition using a series of 2 X 2 repeated-measures ANOVAs. Shaded areas of line plots indicate regions of significance over time. Results indicated a significant interaction between the factors of Reinforcement Type (positive or negative) and Outcome Type (better-than-expected or worse-than-expected) during the same time period as the FRN has

been identified in the literature and the same time as the significant mediofrontal effects of trial number in the previous analysis, confirming that the brain representation of trial number differs between conditions. Topographic plots do not indicate significance or lack thereof but are instead shown only to indicate the relative spatial extent of the observed effects. Topographic plots are shown with the same standard hot-cold color scheme where red indicates high ANOVA F-values and blue indicates low ANOVA F-values. Note that the dominance of blue (low F) shows the relatively low ANOVA F-tests for nearly all points; the mediofrontal cluster that is lighter blue is the only region where the ANOVA tests were near significance. Due to individual differences in which sensor provided the best model fit and was therefore selected for further analysis, the actual sensors selected for analysis showed a significant interaction effect while the topos only show a region that is closer to significance than other points.

I quantified the results of this ANOVA interaction by using serial t-tests to determine the simple effects of reinforcement type and outcome type on the brain representation of trial number. Note that this procedure returns the same values as typical post hoc analyses computed in e.g. SPSS using either a LSD or Bonferroni correction, as this software does not apply any correction for factors with less than three levels; therefore, these results are equivalent to any corrected post hoc procedure which might be employed in SPSS. Results indicated that, as expected based on the previous results (section 3.2.2) mediofrontal beta weights for trial number were greater for positive reinforcement better-than-expected outcome trials and for negative reinforcement worse-than-expected outcome trials than for positive reinforcement worse-than-expected outcome trials or negative reinforcement better-than-expected outcome trials. All statistical results from this analysis, including t-values and p-values, are presented graphically in Figure 9.

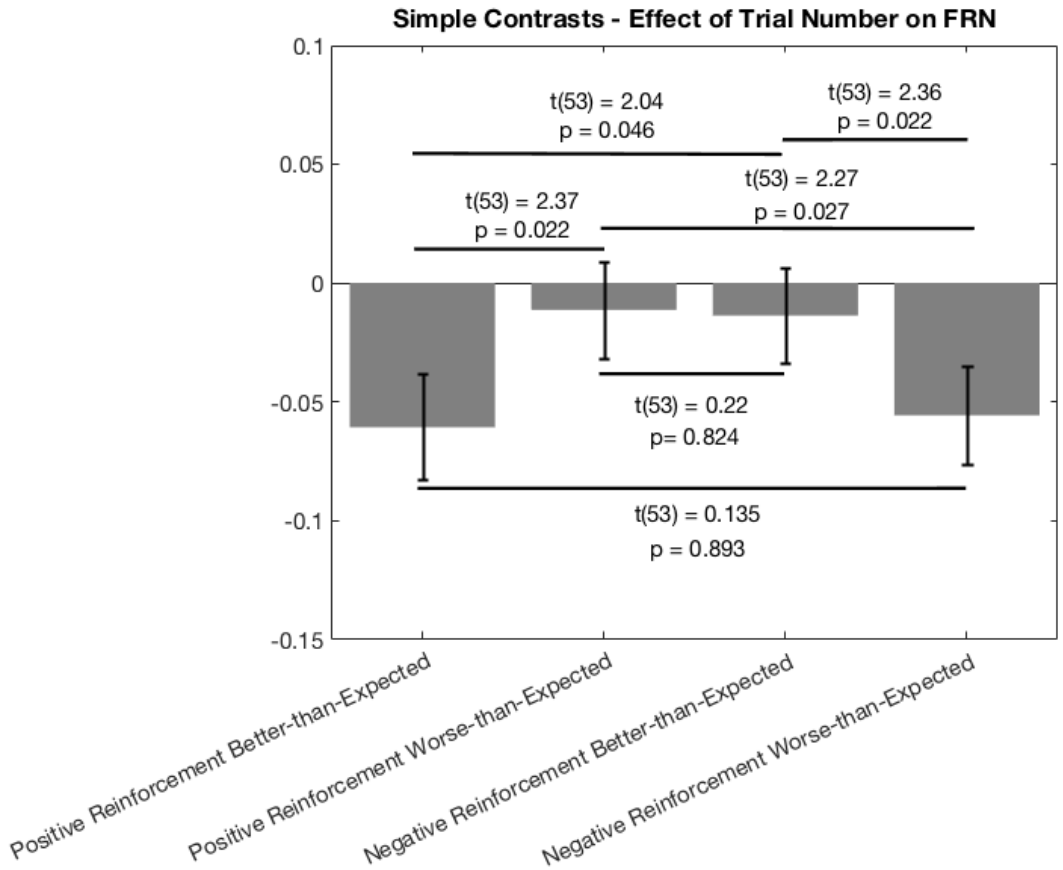


Figure 9. Post hoc testing of single-trial ANOVAs. I used serial t-tests to determine the source of identified interaction effects from repeated-measures ANOVA. Results indicated that for the most salient outcomes, i.e. the outcomes furthest from the overall task average (positive reinforcement better-than-expected, negative reinforcement worse-than-expected), the impact of trial number on the mediofrontal potential was strongest, while it was significantly lower for the less salient outcomes, i.e. the outcomes closest to the overall task expectation (positive reinforcement worse-than-expected, negative reinforcement better-than-expected).

3.2.4 Mediofrontal ERP predicts aversive reinforcement learning

To determine whether the mediofrontal ERP predicts learning in the following trial, I used the approach and matlab program described in (Fischer & Ullsperger, 2013). That is, I fit a stepwise regression model to the RT where the first steps of the model were the same as described above in Section 2.1. In the second step of the regression, I used single trials of feedback-locked EEG to predict RTs on the following trial, therefore testing whether neural processing of reinforcing events predicts adaptation of reaction times over and above the behavioral impact of the task. Critically, feedback-locked EEG robustly predicted RT on the

following trial, but only for negative reinforcement conditions with aversive (i.e. worse than expected) outcomes. This is in line with prior evidence suggesting that mediofrontal involvement in reinforcement, as well as the release of DA via the mesocortical tract, is primarily modulated by aversive experiences. Therefore, it is to be expected that mediofrontal brain activity should predict reaction times only following delivery of aversive stimuli. By contrast, positive reinforcement conditions never deliver aversive stimuli – the worst outcome in positive reinforcement conditions was an outcome of +20 points, which while lower than expectations, cannot be said to be an aversive stimulus. All statistical results from this analysis are illustrated graphically in Figure 10.

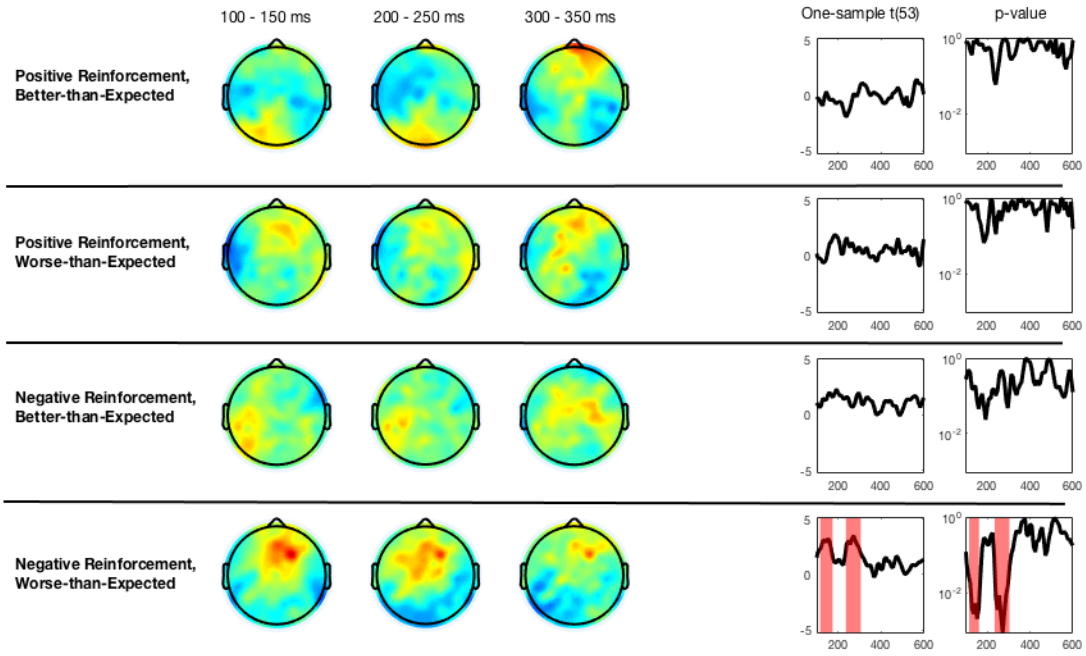


Figure 10. Single-trial brain-RT regression results. I used stepwise regression to determine whether feedback-locked brain activity in any of the four task outcome conditions predicted RT on the following trial. Topographic plots indicate the beta weights of EEG predicting reaction times using a standard hot-cold color template where blue indicates more negative beta coefficients while red indicates more positive beta coefficients. Results indicated that mediofrontal activity only predicted future behavioral adaptation in negative reinforcement conditions that were paired with delivery of an aversive stimulus. Notably, the effect of brain activity on following trial RTs also concentrated over mediofrontal electrodes (red clusters indicating positive beta weights). More specifically, the FRN exhibited a positive relationship with RT, such that more negative FRN amplitude over the course of the task predicted faster RTs, in line with an interpretation of the effect of trial number on the FRN as a learning signal.

This is in line with the suggestion that mediofrontal cortex is sensitive to release of DA according to evidence from Lammel (2015), who indicated that activity of the mesocortical DA tract selectively drives aversive reinforcement learning.

4 Discussion

In this study, I examined reinforcement learning effects in behavior (RTs) and the relationship of reinforcement learning to the mediofrontal ERP using a reinforcement learning task that avoids many of the pitfalls inherent in standard FRN analyses. Specifically, this task generated trial-by-trial expectancy-outcome discrepancies (prediction errors, PEs) in a way such that participants did not necessarily pair "incorrect" response feedback with "worse-than-expected" outcomes and "correct" feedback with "better-than-expected" outcomes. Furthermore, I used a computational approach to examine the ERP at the level of single trials. This is necessary in order to actually test neural implementations of reinforcement learning algorithms, as all extant learning algorithms specify the update of expectancies either trial-by-trial (Rescorla & Wagner, 1972) or continuously over time (Sutton & Barto, 1998). In standard studies of reinforcement learning and the feedback-locked ERP, this theoretically mandated change over time is nullified by the approach of averaging brain potentials over the entire task. Finally, I examined behavioral output in the same trial-by-trial manner as the ERP. This allows me to theoretically ground my behavioral results in animal literature indicating outcome- and trial number-dependent changes in reaction times, as well as to make putative conclusions about the differential impacts of tonic vs. phasic shifts in DA concentration released via the mesocortical pathway.

4.1 Comparison of Results to Standard ERP Studies of Reinforcement

There is a long-standing debate about whether the feedback-related negativity (FRN) signals a salience prediction error or a value prediction error (Figure 1). While there is a general agreement that we primarily learn the importance of environmental stimuli based on rewarding

and aversive experiences (Dayan & Balleine, 2002; Sutton & Barto, 1998), there is substantial disagreement in the ERP literature about how these signals are displayed in the cortex.

4.1.1 The Reinforcement Learning Feedback Related Negativity Hypothesis

Since the initial inception of the reinforcement learning FRN hypothesis (RL-FRN; Holroyd & Coles, 2002) a substantial body of evidence has been built suggesting that the FRN conveys a dopaminergic (DA) signed (value) PE signal. Holroyd & Coles (2002) argue for the existence of two separate neural systems that together influence adaptive behaviors, i.e., a mesencephalic reinforcement learning system and a “generic” error monitoring system in the cingulate cortex. Through combined computational modeling and psychophysiological experimentation, they suggest that the error-related negativity (ERN), a negative-going ERP component peaking soon after the commission of behavioral errors, might be elicited when an error-monitoring system detects that outcomes are worse than expected, and that this error signal is used to train motor systems in line with reinforcement learning principles. Citing prior evidence that delayed error feedback also produces an ERN-like deflection (Miltner, Braun, & Coles, 1997), the authors argue that this potential is not generated merely by the commission of an error, but instead by detection of the error, or perhaps by the use of the behavioral error to guide future behavior. They tested this hypothesis by having subjects complete a probabilistic reinforcement learning task utilizing certain and uncertain reward predictors, and correct or incorrect feedback. Change over time in the feedback-related negativity was assessed by binning ERPs into ten-trial running averages. Results indicated tonic shifts in the amplitude of the FRN over the task, which depended on the delivered outcome and the probability of that outcome. In general, since this initial study the impact of time on the feedback-locked brain potential has been discarded in favor examining the specific type of PE term contained in the ERP. Citing evidence from Schulz’s group, Holroyd & Coles (2002) contend that any DA reinforcement learning signal should follow a value function, and therefore suggest that the FRN must be influenced by the value discrepancy between the outcome and expectation. At the time

Holroyd and Cole published their work on the RL-FRN hypothesis, the value hypothesis concerning DA response was the dominant perspective in the neurophysiological literature examining reinforcement learning. The possibility that this cingulate-generated signal should reflect a value signal has been supported by many studies since their initial publication but is not in line with known neurophysiology of the DA system that has come to light since this hypothesis was conceived.

4.1.2 FRN Results Part 1 (Standard ERP Studies): Lack of Positive Reinforcement Coding in the FRN

Initially I conducted a standard ERP analysis (averaging across trials) of the feedback-related negativity to render these results comparable to the bulk of the literature on the FRN. My results indicated that the FRN only differentiated between worse-than-expected and better-than-expected outcomes in negative reinforcement conditions. Furthermore, the FRN was more negative following better-than-expected negative reinforcement outcomes than following worse-than-expected outcomes. My results did not indicate any difference in the mediofrontal ERP between worse-than-expected and better-than-expected positive reinforcement outcomes.

This null effect is contrary to the bulk of the FRN literature. For example, the initial set of studies on which the RL-FRN hypothesis was founded (Holroyd & Coles, 2002) indicated that the FRN was more negative for reward omission than for reward delivery. This result was replicated and the RL-FRN hypothesis extended by the results of Holroyd, Pakzad-Vaezi, and Krigolson (2008). While Holroyd et al. (2008) indicated again that the FRN was more pronounced for reward omissions than reward deliveries, they reinterpreted the basis of these results as a correct-related positivity in the waveform. It is important to note at this point that the interpretation of the mediofrontal potential as a reward positivity (as opposed to an aversion negativity) relies only on the order in which a difference wave is computed, as difference waves have been the method of choice in most of the past literature examining the feedback-locked mediofrontal potential. As our study does not use difference waves, we refer to the potential by

the original and standard name in the literature, in line with inspection of our grand-average waveforms indicating a negative-going peak during the time period of 250-350 ms post-feedback. The result describing a more negative mediofrontal potential following reward omission than reward delivery has been replicated many times in the literature. For example, Cooper, Duke, Pickering, and Smillie (2014) recorded EEG during an associative learning task that used only positive reinforcement and produced results clearly indicating that the FRN was most pronounced for unexpected reward omission, in line with an aversion account of the FRN. A study by Potts, Martin, Kamp, and Donchin (2011) introduced the task design that the Cooper et al. (2014) study was based on and provided essentially equivalent results, namely that the FRN was most negative for unexpected reward omissions in line with an aversion PE.

However, the current task design differs from previous task designs in one critical aspect. Note that the previously cited works have all compared the delivery of a reward to the omission of a reward, following the Skinnerian definition of positive reinforcement. However, these studies shared a common flaw that appears to be largely unrecognized in the literature. Namely, mediofrontal activity is highly sensitive to knowledge of error commission and performance monitoring. While a discussion of the effect of error and performance monitoring is largely outside the scope of this manuscript, we refer the reader to a recent review by Gehring, Goss, Coles, Meyer, & Donchin (2018) covering the error-related brain potential. Indeed, work demonstrating the influence of errors and error feedback on mediofrontal ERPs is far more numerous than work examining the response of the mediofrontal potential to prediction errors. For example, a PubMed search combining the terms using the search terms (“FRN” OR “feedback-related negativity” OR “feedback negativity” OR “reward positivity”) AND (“prediction error”) conducted over the last ten years returns 89 studies, while a search for ("ERN" OR "error-related negativity" OR "error negativity") returns 1019 studies in the last ten years.

So why is the fact that mediofrontal cortex is sensitive to the commission and recognition of errors important for the current work? Interestingly, the previous studies all had something in

common. Namely, “worse” outcomes, or reward omissions, always occurred on incorrect trials, while “better” outcomes, or reward deliveries, always occurred on correct trials. I propose that this inherently confounds the mediofrontal response to errors and the mediofrontal response to PEs. This confound is thoroughly engrained in the literature examining whether the FRN reflects a PE. In fact, in the original RL-FRN paper by Holroyd and Coles (2002) the FRN was called the feedback error-related negativity. This confound is notably not present in the original work documenting dopaminergic PEs by Schulz’ research group (1997, 1998). In this work, monkeys were simply cued to create an expectation of reward, which was then delivered (expected) or not delivered (unexpected). Therefore, while the animal research the RL-FRN hypothesis was initially built on did not confound performance monitoring signals with PE signaling (the monkey did not perform), most of the human PE research has carried this confound.

Our task design notably does not contain this confound in any analysis. I analyzed only correct trials, and more specifically only trials where participants already knew they were correct in their response (see Figure 1). Therefore, this study design allows for the systematic violation of reward expectations, while controlling for error monitoring effects (that is, there were none because all trials analyzed are correct). Our finding of a null difference between worse-than-expected and better-than-expected outcomes in positive reinforcement therefore suggests that in the absence of error monitoring, mediofrontal potentials might not signal reward prediction errors.

4.1.3 FRN Results Part 2 (Standard ERP Studies): Negative Reinforcement (Aversive) Coding in the FRN

Our results for negative reinforcement conditions tell a different story, indicating that the FRN is primarily responsive to aversive outcomes, rather than graded positive outcomes. Specifically, our analysis of the ERP shows that the FRN is more negative for better-than-expected outcomes in negative reinforcement (that is, those outcomes where the subject expected zero points but instead received a small reward) compared to worse-than-expected

outcomes in negative reinforcement conditions (that is, those conditions where the subject expected zero points but instead lost a small number of points). These results cannot be attributed to any form of error monitoring as the participants were already informed that their response was correct (see task design section for more info on this; see Figure 1). Importantly, the absolute deviation of rewards and punishments from the average expectation was equivalent between positive and negative reinforcement in our study, indicating that the FRN is indeed more sensitive to aversive expectancy violation than positive expectancy violation. A number of studies support a view of the FRN as carrying primarily aversive information or negative PEs.

For example, among the (admittedly small) number of studies that have examined the FRN in negative reinforcement conditions, the finding of a more negative potential following better-than-expected outcomes compared to worse-than-expected outcomes is well established. This effect was initially described in a seminal study by Talmi, Atkinson, and El-Deredy (2013) using blocks of positive reinforcement with money as the reinforcer, and negative reinforcement with electric shock as the reinforcer. This study was the original study to show an effect that when both positive and negative reinforcement are considered, the FRN flips sign between positive and negative reinforcement conditions, indicating a salience signal rather than a value signal. Specifically, the FRN was more negative for omitted positive and negative reinforcement outcomes than for delivered positive or negative reinforcement outcomes. Our results uphold the results of Talmi et al.'s analysis for negative reinforcement conditions, indicating that the FRN is more negative for better-than-expected negative reinforcement conditions than worse-than-expected negative reinforcement conditions. Huang and Yu (2014) demonstrate convincing convergent evidence for this effect in a task that includes both positive and negative reinforcement. This task generated the result that FRN amplitudes were more negative for reward omissions compared to reward deliveries, but more negative for punishment omissions than for punishment deliveries. Since these types of outcome are opposite in valence this must

be interpreted as salience coding. Sallet, Camille, and Procyk (2013) demonstrate a similar effect using multiple levels of positive and negative PE. They indicate that the FRN is most negative for the worst positive PEs, but is most negative for the best negative PEs.

This flip in sign between positive and negative modalities (i.e. the FRN is most negative for reward omissions, but most negative for escape from aversive outcomes as well) strongly supports the notion of salience coding in the FRN. Hird, El-Deredy, Jones, and Talmi (2018) demonstrate convergent results using a paradigm that delivered appetitive and aversive tastes. They observed a "typical FRN" (more negative for reward omission) in the appetitive condition, but a "reverse FRN" for aversive tastes (that is, the FRN was more negative for omission of aversive outcomes than for delivery of aversive outcomes), in line with a salience account. While this study also demonstrated aversive PEs later in the waveform, there were notably no effects of reward PE, in opposition to RL-FRN theory. Pfabigan et al. (2015), using a Monetary Incentive Delay (MID) paradigm, demonstrate that the FRN is more negative for reward omission compared to reward delivery, but is reversed for negative reinforcement, being more negative for avoidance of aversive stimulus delivery than for delivery of aversive stimuli. These "reverse FRN" results were replicated in a study I completed last year (Rawls, Miskovic, Moody, Lee, Shirtcliff, & Lamm, under review), again using an MID paradigm. In a critical step, Pfabigan et al. (2015) demonstrated that activity localized to cingulate cortex (putative generator of the FRN) showed the same effects that were observed at the scalp level. This is notable, because it suggests that indeed, as predicted by neurobiology, the cingulate cortex computes salience, not value, PEs. In one of the most convincing source-localized FRN studies, Hauser et al. (2014) simultaneously recorded EEG and fMRI during a probabilistic reinforcement learning task. This study rigorously examined the expression of signed and unsigned PEs in the post-feedback waveform while using the superior spatial information provided by fMRI to localize the FRN definitively to the cingulate cortex. This study orthogonalized signed and unsigned PEs (which

are often correlated), and demonstrated that only unsigned PEs project to the FRN topography using dynamic causal modeling.

It is not surprising to the author that more studies support the view of the FRN as being responsive to salient, primarily aversive, outcomes, as it is in line with basic neurobiology of the DA system, which is known to carry primarily aversive information along the mesocortical pathway in animal models (Lammel et al., 2008; Lammel et al., 2011). Our trial-averaged results fully support previous findings that the FRN is more negative for better-than-expected outcomes than for worse-than-expected outcomes when negative reinforcement is cued. Original RL-FRN hypotheses suggest that dopamine release might inhibit the apical dendrites of pyramidal cells in mediofrontal cortex, resulting in a more positive potential (since the FRN is a negative-going potential following feedback). Our results are well in line with this long-standing way of thinking. Specifically, based on known neurobiology, we would hypothesize that dopamine release in cortex should increase as aversive outcomes are delivered, which could explain the less-negative mediofrontal potential (FRN) following worse-than-expected outcomes compared to better-than-expected outcomes for negative reinforcement.

4.2 Single-Trial Analysis of Reinforcement Learning

4.2.1 Single-Trial Results Part 1: Influence of Trial number on The Feedback-Related Negativity

If hypotheses regarding the biological underpinnings of the FRN are true, then there should be a change in the ERP amplitude as outcomes are learned. While the original RL-FRN paper (Holroyd & Coles) indicated that the FRN became more negative as trial number increased, and this activation instead propagated to the outcome cue over time in line with expectations of a DA modulated signal, relatively few studies have examined the hypothetically mandated shift in FRN amplitude over the course of the task. A notable exception is Krigolson, Hassal, and Handy (2014), who convincingly demonstrated that the aversion PE in the FRN became more negative over time while the corresponding potential in response to reinforcement

cues became less negative over time. Our results show the very same effect – specifically, that the FRN grows more negative as trial number increases, i.e. as the task is learned. Furthermore, I indicate the novel results that this trial-by-trial shift in the negativity of the waveform is gated by salient reinforcement outcomes – specifically, the effect of trial number on increased FRN negativities is driven entirely by better-than-expected rewards and worse-than-expected aversive outcomes (most salient outcomes). This suggests that the FRN should become more negative as trial number increases following salience PEs – which is not in line with the interpretation of the FRN as a "reward positivity." Indeed, it is notable that more recent studies of differences in dopaminergic projections have shown that only mesolimbic DA projections carry reward signals. Additionally, recent animal neurobiology results have shown that mesocortical DA projections carry primarily aversive PEs, while some mesocortical DA neurons carry salience information. Our results are in line with the interpretation of the FRN as primarily responsive to salient outcomes (better-than-expected rewards and worse-than-expected outcomes), as well as meeting basic predictions that the FRN should scale with learning over time (and replicating Krigolson et al., 2014 who demonstrated that the FRN grows more negative over time). Put another way, the FRN rapidly becomes more negative as task contingencies are learned. A Pearce-Hall salience signal (Pearce & Hall, 1980) is often referred to as a "learning rate," suggesting that learning should increase with more salient outcomes. Since the FRN becomes more negative with learning, it is no surprise that the impact of trial number on the FRN is only significant for the most salient outcomes (reflecting increased learning following more salient, compared to less salient, outcomes). Specifically, the FRN becomes more negative over the course of the task, which reflects a learning effect, but this learning effect is most pronounced following salient outcomes, in line with the notion that the FRN reflects a Pearce-Hall salience PE. This can be viewed in light of the hypothesis that the FRN is impacted by a dopaminergic salience signal. As task outcomes are learned over the course of time, dopamine release dwindles. That is, dopamine release is expected to decrease

proportionally to how well or thoroughly an organism learns the requirements of a task. Since dopamine primarily reflects a learning signal, dopamine release is increased when the organism needs to learn task constraints but will decrease over time as task constraints are cemented behaviorally (the organism no longer needs to learn, as it already knows how to complete the task). Thus, as the organism learns and dopamine release dwindles, FRN amplitudes increase. Finally, our interpretation of the FRN as a reinforcement learning signal is cemented by our single-trial brain-behavior regression results indicating that more negative FRN amplitude following worse-than-expected aversive stimulus delivery spurred more adaptive responding in the following trial. This provides empirical evidence that negative shifts in the FRN over the course of the task following aversive outcomes are indeed predictive of future reinforcement learning, rather than merely reflecting habituation.

4.2.2 Single-Trial Results Part 2: Influence of Trial Number and Prediction Errors on Reaction Times

Many, if not most, studies of the FRN information content are notably without a corresponding analysis of behavior evoked during the task. However, in order to fully understand the computations the cortex is undergoing during reinforcement processing, behavior should also be analyzed. Specifically, there is a large amount of prior evidence, primarily from animal studies, that shows definite impacts of reinforcement and dopamine levels on the speed of responses (in animal literature referred to as vigor). Bryce and Floresco (2019) trained rats in an effort-discounting task where rats chose between low-effort, low-reward scenarios and high-effort, high-reward scenarios. They then tested these effects while D2 receptors in the nucleus accumbens were stimulated. Results indicated that excessive stimulation of DA receptors resulted in longer response latencies (reduced response vigor) and lower willingness to expend effort to obtain a reward. In interpreting this result it is key to note that the nucleus accumbens is targeted by the mesolimbic set of DA projections, and therefore carries primarily reward information – as the authors perfused NAcc with DA, this likely resulted

in the rats already feeling a sense of subjective reward that would be expected to lower response vigor (slower RTs) in the pursuit of additional reward. Our results for the outcome-related shift of reaction times were confined to negative reinforcement and indicated that following better-than-expected outcomes participants were slower in their next responses, but that participants increased reaction speed following unexpectedly large punishments. This is well in line with Bryce and Floresco's (2019) study, which indicated that following DA perfusion in NAcc, rats were already satiated and did not pursue rewards as readily. We interpret our results indicating that participants slowed responses following better-than-expected negative reinforcement outcomes in light of the aforementioned results and suggest that following better-than-expected negative reinforcement outcomes, our participants were already "satiated" and had less motivation to respond quickly.

Bryce and Floresco (2016) indicated that this tradeoff between effort and reward seeking was altered by infusion of a stress factor (corticotropin-releasing factor; Bryce & Floresco 2016). This indicated that infusion of a stress agonist reduced the willingness of animals to pursue high-effort options. While this might seem at odds with our results for reaction times in negative reinforcement (in which responses became faster following aversive outcomes) it is important to note that there was no aversive learning occurring in the aforementioned animal study since CRH was administered by the investigator. I instead suggest that when the opportunity to escape further punishments is made possible by responding more quickly (as in our task), the expected reinforcing effects of aversive stimulus delivery are seen – that is, in an effort to avoid further aversive stimuli, humans increase response speed following punishment.

Niv, Daw, Joel, & Dayan (2007) extend reinforcement learning models to operant tasks, and indicate that tonic levels of DA shift the willingness of an agent to engage in more effortful behavior. Specifically, results indicate that higher tonic levels of DA facilitate slower but more exploratory responding. Our examination of RTs following positive reinforcement, where participants were given a graded level of reward on every trial, support the model by Niv et al.

(2007). As the average reward was learned, and indeed as participants learned that positive reinforcement correct feedback would never be followed by an omission of reward or a loss, tonic levels of DA would be required by neurobiological theories to decrease, facilitating monotonic effects of increased reaction speeds over the course of the task as we demonstrate (main effect of trial number on single-trial RTs for positive reinforcement). This particular hypothesized tonic effect was reflected in neural activity following rewarding outcomes, as we demonstrate that the FRN becomes more negative over time following better-than-expected rewards. Interpreting the FRN through theories suggesting it is influenced by release of dopamine in cortex, this potential would be expected to grow more negative over time as dopamine release in cortex decreases. We found precisely this effect using single-trial analysis. While the previous mentioned study was a theoretical model, this result was empirically demonstrated in a human sample by Guitart-Masip, Beierholm, Dolan, Duzel, & Dayan (2011), who indicated that the average reaction time was partially explained by the average rate of reward in a task, which changed slowly but systematically. This study did not examine the influence of immediate reward on RTs as our study did, but I also note that in positive reinforcement conditions I found null effects of immediate reward but instead only a tonic effect where reaction time speeded as trial number increased. This is in line with the findings of Guitart-Masip et al., and indicates that as subjects learned the average rate of reward in the task their responses became faster.

In a novel study of changes in human reaction times combined with fMRI scanning, Evers, Stiers, and Ramaekers (2017) probed the effect of tonic and phasic shifts in DA on RT. Notably, this study indicated that tonic DA levels in striatum indicate an average reward signal, and predicted faster RTs. Meanwhile, tonic shifts in DA decreased the striatal response to gains and losses in a task (phasic shifts in DA). This is in line with our positive reinforcement results for RTs, suggesting that reduced phasic DA release as rewards are learned facilitates faster and less exploratory responding. The present study appears to be the first attempt to extend the

analysis of reinforcement-modulated reaction time learning to negative reinforcement conditions, and therefore it remains to be seen what this effect might look like in an animal model, or indeed in a standard reinforcement learning task. However, taken together, the aforementioned results show that reaction time shifts to faster responding as task outcomes and contingencies are learned. This result fits well with our results indicating faster responding and greater learning immediately following aversive negative reinforcement outcomes, which is in line with theories holding that organisms must learn from negative outcomes to avoid those same outcomes in the future.

4.3 Limitations of the Current Study

While the study described in this manuscript remedies many shortcomings and confounds of previous studies of the FRN, it is not without limitations. Perhaps the most limiting factor of this study is the novel design of the task, which means that replication of this study design and of these effects are necessary before strong conclusions can be drawn from this study. Furthermore, the inspiration for this study comes primarily from invasive animal research examining dopamine neurobiology. While our results are largely in line with what must be expected of dopaminergic activity in cortex, we are not able to make any claims that the observed effects are rooted in dopamine neurotransmission. It is possible that these effects are not tied to dopamine neurotransmission at all but are instead the result of endogenous computations undertaken completely in the cortex (Cavanagh & Frank, 2014; Cavanagh, Zambrano-Vazquez, & Allen, 2012). Since serotonergic neurons were also recently demonstrated to compute salience PEs (Matias, Lottem, Dugue, & Mainen, 2017), it is even possible that the observed brain potential effects are conveyed to cortex from neurons in the raphe nucleus. However, previous examinations of the FRN have repeatedly suggested that this potential is the result of cortical dopamine release, and so we interpret our results through this existing theoretical lens while incorporating more recent neurobiological examinations that were not known when the RL-FRN hypothesis was formulated. Finally, while our result controls for

the influence of error monitoring, it is still possible given the task design that the FRN is primarily responsive only to the valence of outcomes, and not to PEs. That is, the observed effects for negative reinforcement might be due simply to the fact that these outcomes were either gains or losses, while in positive reinforcement all outcomes were gains. However, under an axiomatic definition of a PE as merely reflecting a deviation from expected outcomes, the actual valence (gain or loss of points) should not be a determining factor. That is, a signed PE is decoupled from the valence of the outcome itself, and merely reflects whether the outcome was better or worse than expected, so this should not be a problem for the described study.

4.4 Conclusion

The present study utilized a unique manipulation to separate the influence of error and performance monitoring from the influence of worse- or better-than-expected outcome signaling. Previous studies using Skinnerian definitions of positive reinforcement (rewards are either delivered or omitted) have demonstrated that the FRN is more negative following omitted rewards than delivered rewards. However, previous studies have also largely confounded the influences of error monitoring in an effort to understand whether the FRN reflects a signed PE. In a task design that decouples the influence of error monitoring from outcome processing, we do not find any influence of worse-than-expected outcomes compared to better-than-expected outcomes in positive reinforcement conditions. This is in line with more recent animal neurobiology results, indicating that the mesocortical dopamine pathway carries primarily information about aversive salience rather than reward information (which is mostly localized along the mesolimbic dopamine pathway). In this controlled design, we still found the expected effect for negative reinforcement that better-than-expected outcomes resulted in greater mediofrontal negativity (FRNs; suggesting less mesocortical DA activity) compared to worse-than-expected outcomes (suggesting more mesocortical DA activity). This is in line with neurobiological evidence that dopamine neurons projecting to cortex appear to increase firing rate following aversive stimuli, which by existing theories should inhibit the FRN (less negative

activation). If the FRN signaled value PEs, the effect we found (FRN is more negative following better-than-expected rather than worse-than-expected negative reinforcement outcomes) would necessarily have to be reversed. More specifically, if the FRN was sensitive to value PEs it would be more inhibited by DA activation for better-than-expected outcomes (rather than more inhibited for worse-than-expected outcomes). As dopamine neurons projecting to the cortex are known to become excited by aversive stimulus delivery, this might still be in line with Holroyd & Coles (2002) original notion that dopamine inhibits the apical dendrites of cortical pyramidal neurons.

In a single-trial analysis of the FRN, we replicate previous evidence that the FRN grows more negative over time as task conditions are learned. This corresponds to a hypothesized decrease over time in dopamine levels in cortex, which by existing theories should disinhibit cortical pyramidal cells. Notably, we extended these results by showing that this change over time is only evident following salient outcomes, which according to extant theories are the outcomes that are most valuable to learn from. We showed that the FRN grows more negative over time following better-than-expected rewards and worse-than-expected aversive stimulus delivery, which are the most salient outcomes delivered in the task and therefore the most critical outcomes for a subject to learn from. We verify that this is indeed an effect of task learning, rather than an effect of mere habituation, by showing that following negative reinforcement worse-than-expected outcomes, more negative FRN amplitudes drive faster responding on the next consecutive trial. Taken together, our trial-averaged and single-trial analysis of the FRN indicates that the FRN is most sensitive to aversive outcomes and drives behavioral learning effects following aversive outcomes as well. This represents an important challenge to the dominant RL-FRN theory and provides neurobiologically-inspired evidence against the interpretation of the mesocortical dopamine tract as carrying a signed PE to mediofrontal cortex.

5 References

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. Sage.

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, *51*(6), 1173.

Beeler, J. A., Daw, N. D., Frazier, C. R., & Zhuang, X. (2010). Tonic dopamine modulates exploitation of reward learning. *Frontiers in behavioral neuroscience*, *4*, 170.

Benjamini, Y., & Yekutieli, D. (2001). The control of the false discovery rate in multiple testing under dependency. *The annals of statistics*, *29*(4), 1165-1188.

Bieniek, M. M., Frei, L. S., & Rousselet, G. A. (2013). Early ERPs to faces: aging, luminance, and individual differences. *Frontiers in psychology*, *4*, 268.

Brischoux, F., Chakraborty, S., Brierley, D. I., & Ungless, M. A. (2009). Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. *Proceedings of the National Academy of Sciences*, *106*(12), 4894-4899.

Bryce, C. A., & Floresco, S. B. (2016). Perturbations in effort-related decision-making driven by acute stress and corticotropin-releasing factor. *Neuropsychopharmacology*, *41*(8), 2147.

Bryce, C. A., & Floresco, S. B. (2019). Alterations in effort-related decision-making induced by stimulation of dopamine D 1, D 2, D 3, and corticotropin-releasing factor receptors in nucleus accumbens subregions. *Psychopharmacology*, 1-14.

Bryden, D. W., Johnson, E. E., Tobia, S. C., Kashtelyan, V., & Roesch, M. R. (2011). Attention for learning signals in anterior cingulate cortex. *Journal of Neuroscience*, *31*(50), 18266-18274.

Budygin, E. A., Park, J., Bass, C. E., Grinevich, V. P., Bonin, K. D., & Wightman, R. M. (2012). Aversive stimulus differentially triggers subsecond dopamine release in reward regions. *Neuroscience*, *201*, 331-337.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in cognitive sciences*, *18*(8), 414-421.

Cavanagh, J. F., Zambrano-Vazquez, L., & Allen, J. J. (2012). Theta lingua franca: A common mid-frontal substrate for action monitoring processes. *Psychophysiology*, *49*(2), 220-238.

Chang, C. Y., Esber, G. R., Marrero-Garcia, Y., Yau, H. J., Bonci, A., & Schoenbaum, G. (2016). Brief optogenetic inhibition of dopamine neurons mimics endogenous negative reward prediction errors. *Nature neuroscience*, *19*(1), 111.

Chaumon, M., Bishop, D. V., & Busch, N. A. (2015). A practical guide to the selection of independent components of the electroencephalogram for artifact correction. *Journal of neuroscience methods*, *250*, 47-63.

Chiodo, L. A., Antelman, S. M., Caggiula, A. R., & Lineberry, C. G. (1980). Sensory stimuli alter the discharge rate of dopamine (DA) neurons: evidence for two functional types of DA cells in the substantia nigra. *Brain research*.

Cohen, M. X., & Cavanagh, J. F. (2011). Single-trial regression elucidates the role of prefrontal theta oscillations in response conflict. *Frontiers in psychology*, *2*, 30.

Cooper, A. J., Duke, É., Pickering, A. D., & Smillie, L. D. (2014). Individual differences in reward prediction error: contrasting relations between feedback-related negativity and trait measures of reward sensitivity, impulsivity and extraversion. *Frontiers in Human Neuroscience*, *8*, 248.

D'Ardenne, K., McClure, S. M., Nystrom, L. E., & Cohen, J. D. (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science*, *319*(5867), 1264-1267.

Dawson, J. F. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology*, *29*(1), 1-19.

Dayan, P., & Balleine, B. W. (2002). Reward, motivation, and reinforcement learning. *Neuron*, *36*(2), 285-298.

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of neuroscience methods*, *134*(1), 9-21.

Ellwood, I. T., Patel, T., Wadia, V., Lee, A. T., Liptak, A. T., Bender, K. J., & Sohal, V. S. (2017). Tonic or phasic stimulation of dopaminergic projections to prefrontal cortex causes mice to maintain or deviate from previously learned behavioral strategies. *Journal of Neuroscience*, *37*(35), 8315-8329.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & psychophysics*, *16*(1), 143-149.

Evers, E. A., Stiers, P., & Ramaekers, J. G. (2017). High reward expectancy during methylphenidate depresses the dopaminergic response to gain and loss. *Social cognitive and affective neuroscience*, *12*(2), 311-318.

Fischer, A. G., & Ullsperger, M. (2013). Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron*, *79*(6), 1243-1255.

Garrison, J., Erdeniz, B., & Done, J. (2013). Prediction error in reinforcement learning: a meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*, *37*(7), 1297-1310.

Gehring, W. J., Goss, B., Coles, M. G., Meyer, D. E., & Donchin, E. (2018). The error-related negativity. *Perspectives on Psychological Science*, *13*(2), 200-204.

Gläscher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*, *66*(4), 585-595.

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, *108*(Supplement 3), 15647-15654.

Guarraci, F. A., & Kapp, B. S. (1999). An electrophysiological characterization of ventral tegmental area dopaminergic neurons during differential pavlovian fear conditioning in the awake rabbit. *Behavioural brain research*, *99*(2), 169-179.

Guitart-Masip, M., Beierholm, U. R., Dolan, R., Duzel, E., & Dayan, P. (2011). Vigor in the face of fluctuating rates of reward: an experimental examination. *Journal of cognitive neuroscience*, *23*(12), 3933-3938.

Guthrie, M., Myers, C. E., & Gluck, M. A. (2009). A neurocomputational model of tonic and phasic dopamine in action selection: a comparison with cognitive deficits in Parkinson's disease. *Behavioural brain research*, *200*(1), 48-59.

Hart, A. S., Rutledge, R. B., Glimcher, P. W., & Phillips, P. E. (2014). Phasic dopamine release in the rat nucleus accumbens symmetrically encodes a reward prediction error term. *Journal of Neuroscience*, *34*(3), 698-704.

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.

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.

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

Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychological review*, *109*(4), 679.

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.

Horvitz, J. C., Stewart, T., & Jacobs, B. L. (1997). Burst activity of ventral tegmental dopamine neurons is elicited by sensory stimuli in the awake cat. *Brain research*, *759*(2), 251-258.

Howe, M. W., Tierney, P. L., Sandberg, S. G., Phillips, P. E., & Graybiel, A. M. (2013). Prolonged dopamine signalling in striatum signals proximity and value of distant rewards. *nature*, *500*(7464), 575.

Huang, Y., & Yu, R. (2014). The feedback-related negativity reflects “more or less” prediction error in appetitive and aversive conditions. *Frontiers in neuroscience*, *8*, 108.

Knutson, B., & Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Current opinion in neurology*, *18*(4), 411-417.

Knutson, B., Adams, C. M., Fong, G. W., & Hommer, D. (2001). Anticipation of increasing monetary reward selectively recruits nucleus accumbens. *Journal of Neuroscience*, *21*(16), RC159-RC159.

Krigolson, O. E., Hassall, C. D., & Handy, T. C. (2014). How we learn to make decisions: rapid propagation of reinforcement learning prediction errors in humans. *Journal of cognitive neuroscience*, *26*(3), 635-644.

Lammel, S., Hetzel, A., Häckel, O., Jones, I., Liss, B., & Roeper, J. (2008). Unique properties of mesoprefrontal neurons within a dual mesocorticolimbic dopamine system. *Neuron*, *57*(5), 760-773.

Lammel, S., Ion, D. I., Roeper, J., & Malenka, R. C. (2011). Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. *Neuron*, *70*(5), 855-862.

Lammel, S., Lim, B. K., Ran, C., Huang, K. W., Betley, M. J., Tye, K. M., ... & Malenka, R. C. (2012). Input-specific control of reward and aversion in the ventral tegmental area. *Nature*, *491*(7423), 212.

Makeig, S., Bell, A. J., Jung, T. P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. In *Advances in neural information processing systems*(pp. 145-151).

Mantz, J., Thierry, A. M., & Glowinski, J. (1989). Effect of noxious tail pinch on the discharge rate of mesocortical and mesolimbic dopamine neurons: selective activation of the mesocortical system. *Brain research*, *476*(2), 377-381.

Matias, S., Lottem, E., Dugue, G. P., & Mainen, Z. F. (2017). Activity patterns of serotonin neurons underlying cognitive flexibility. *Elife*, *6*, e20552.

Matsumoto, M., & Hikosaka, O. (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature*, *459*(7248), 837.

Metereau, E., & Dreher, J. C. (2012). Cerebral correlates of salient prediction error for different rewards and punishments. *Cerebral Cortex*, *23*(2), 477-487.

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.

Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*, *48*(2), 229-240.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of neuroscience*, *16*(5), 1936-1947.

Montague, P. R., Dayan, P., Person, C., & Sejnowski, T. J. (1995). Bee foraging in uncertain environments using predictive hebbian learning. *Nature*, *377*(6551), 725.

Nieuwenhuis, S., Yeung, N., Holroyd, C. B., Schurger, A., & Cohen, J. D. (2004). Sensitivity of electrophysiological activity from medial frontal cortex to utilitarian and performance feedback. *Cerebral Cortex*, *14*(7), 741-747.

Niv, Y. (2007). Cost, benefit, tonic, phasic: what do response rates tell us about dopamine and motivation?. *ANNALS-NEW YORK ACADEMY OF SCIENCES*, *1104*, 357.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507-520.

O'Leary, D. P. (1990). Robust regression computation using iteratively reweighted least squares. *SIAM Journal on Matrix Analysis and Applications*, *11*(3), 466-480.

Oleson, E. B., Gentry, R. N., Chioma, V. C., & Cheer, J. F. (2012). Subsecond dopamine release in the nucleus accumbens predicts conditioned punishment and its successful avoidance. *Journal of Neuroscience*, *32*(42), 14804-14808.

Parker, K. L., Alberico, S. L., Miller, A. D., & Narayanan, N. S. (2013). Prefrontal D1 dopamine signaling is necessary for temporal expectation during reaction time performance. *Neuroscience*, *255*, 246-254.

Pavlov, I. P. (1903, April). The experimental psychology and psychopathology of animals. In *14th international medical congress. Madrid, Spain*.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological review*, *87*(6), 532.

Pfabigan, D. M., Seidel, E. M., Paul, K., Grahl, A., Sailer, U., Lanzenberger, R., ... & Lamm, C. (2015). Context-sensitivity of the feedback-related negativity for zero-value feedback outcomes. *Biological Psychology*, *104*, 184-192.

Potts, G. F., Martin, L. E., Kamp, S. M., & Donchin, E. (2011). Neural response to action and reward prediction errors: Comparing the error-related negativity to behavioral errors and the feedback-related negativity to reward prediction violations. *Psychophysiology*, *48*(2), 218-228.

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*, *2*, 64-99.

Rinaldi, A., Mandillo, S., Oliverio, A., & Mele, A. (2007). D1 and D2 receptor antagonist injections in the prefrontal cortex selectively impair spatial learning in mice. *Neuropsychopharmacology*, *32*(2), 309.

Rousselet, G. A., Gaspar, C. M., Pernet, C. R., Husk, J. S., Bennett, P. J., & Sekuler, A. B. (2010). Healthy aging delays scalp EEG sensitivity to noise in a face discrimination task. *Frontiers in psychology*, *1*, 19.

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*, 209.

Sambrook, T. D., & Goslin, J. (2014). Mediofrontal event-related potentials in response to positive, negative and unsigned prediction errors. *Neuropsychologia*, *61*, 1-10.

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.

Schott, B. H., Minuzzi, L., Krebs, R. M., Elmenhorst, D., Lang, M., Winz, O. H., ... & Düzel, E. (2008). Mesolimbic functional magnetic resonance imaging activations during reward anticipation correlate with reward-related ventral striatal dopamine release. *Journal of Neuroscience*, *28*(52), 14311-14319.

Schultz, W. (2016). Dopamine reward prediction-error signalling: a two-component response. *Nature Reviews Neuroscience*, *17*(3), 183.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593-1599.

Seymour, B., O'doherty, J. P., Koltzenburg, M., Wiech, K., Frackowiak, R., Friston, K., & Dolan, R. (2005). Opponent appetitive-aversive neural processes underlie predictive learning of pain relief. *Nature neuroscience*, *8*(9), 1234.

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature neuroscience*, *16*(7), 966.

Stopper, C. M., Maric, T. L., Montes, D. R., Wiedman, C. R., & Floresco, S. B. (2014). Overriding phasic dopamine signals redirects action selection during risk/reward decision making. *Neuron*, *84*(1), 177-189.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine learning*, *3*(1), 9-44.

Sutton, R. S., & Barto, A. G. (1987, July). A temporal-difference model of classical conditioning. In *Proceedings of the ninth annual conference of the cognitive science society*(pp. 355-378).

Sutton, R. S., & Barto, A. G. (1998). *Introduction to reinforcement learning* (Vol. 2, No. 4). Cambridge: MIT press.

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.

Thorndike, E. L. (1905). *The elements of psychology*. New York: A. G. Seiler.

Wightman, R. M., Heien, M. L., Wassum, K. M., Sombers, L. A., Aragona, B. J., Khan, A. S., ... & Carelli, R. M. (2007). Dopamine release is heterogeneous within microenvironments of the rat nucleus accumbens. *European Journal of Neuroscience*, *26*(7), 2046-2054.

Wood, J., Simon, N. W., Koerner, F. S., Kass, R. E., & Moghaddam, B. (2017). Networks of VTA neurons encode real-time information about uncertain numbers of actions executed to earn a reward. *Frontiers in behavioral neuroscience*, *11*, 140.

6 Appendices

Office of Research Compliance
Institutional Review Board

December 21, 2016

MEMORANDUM

TO:	Eric Rawls Connie Lamm
FROM:	Ro Windwalker IRB Coordinator
RE:	New Protocol Approval
IRB Protocol #:	16-12-304
Protocol Title:	*Neural Mechanisms Underlying Conflict Monitoring in Smokers and Non-Smokers*
Review Type:	☐ EXEMPT ☒ EXPEDITED ☐ FULL IRB
Approved Project Period:	Start Date: 12/19/2016 Expiration Date: 12/18/2017

Your protocol has been approved by the IRB. Protocols are approved for a maximum period of one year. If you wish to continue the project past the approved project period (see above), you must submit a request, using the form *Continuing Review for IRB Approved Projects*, prior to the expiration date. This form is available from the IRB Coordinator or on the Research Compliance website (https://vpred.uark.edu/units/rscp/index.php). As a courtesy, you will be sent a reminder two months in advance of that date. However, failure to receive a reminder does not negate your obligation to make the request in sufficient time for review and approval. Federal regulations prohibit retroactive approval of continuation. Failure to receive approval to continue the project prior to the expiration date will result in Termination of the protocol approval. The IRB Coordinator can give you guidance on submission times.

This protocol has been approved for 120 participants. If you wish to make *any* modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

If you have questions or need any assistance from the IRB, please contact me at 109 MLKG Building, 5-2208, or irb@uark.edu.

109 MLKG • 1 University of Arkansas • Fayetteville, AR 72701-1201 • (479) 575-2208 • Fax (479) 575-6527 • Email irb@uark.edu

UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

March 27, 2017

MEMORANDUM

TO:	Eric Rawls Connie Lamm
FROM:	Ro Windwalker IRB Coordinator
RE:	PROJECT MODIFICATION
IRB Protocol #:	16-12-304
Protocol Title:	*Neural Mechanisms Underlying Conflict Monitoring in Smokers and Non-Smokers*
Review Type:	☐ EXEMPT ☒ EXPEDITED ☐ FULL IRB
Approved Project Period:	Start Date: 03/22/2017 Expiration Date: 12/18/2017

Your request to modify the referenced protocol has been approved by the IRB. **This protocol is currently approved for 120 total participants.** If you wish to make any further modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

Please note that this approval does not extend the Approved Project Period. Should you wish to extend your project beyond the current expiration date, you must submit a request for continuation using the UAF IRB form "Continuing Review for IRB Approved Projects." The request should be sent to the IRB Coordinator, 109 MLKG Building.

For protocols requiring FULL IRB review, please submit your request at least one month prior to the current expiration date. (High-risk protocols may require even more time for approval.) For protocols requiring an EXPEDITED or EXEMPT review, submit your request at least two weeks prior to the current expiration date. Failure to obtain approval for a continuation *on or prior to* the currently approved expiration date will result in termination of the protocol and you will be required to submit a new protocol to the IRB before continuing the project. Data collected past the protocol expiration date may need to be eliminated from the dataset should you wish to publish. Only data collected under a currently approved protocol can be certified by the IRB for any purpose.

If you have questions or need any assistance from the IRB, please contact me at 109 MLKG Building, 5-2208, or irb@uark.edu.

UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

April 12, 2017

MEMORANDUM

TO:	Eric Rawls Connie Lamm
FROM:	Ro Windwalker IRB Coordinator
RE:	PROJECT MODIFICATION
IRB Protocol #:	16-12-304
Protocol Title:	*Neural Mechanisms Underlying Conflict Monitoring in Smokers and Non-Smokers*
Review Type:	☐ EXEMPT ☒ EXPEDITED ☐ FULL IRB
Approved Project Period:	Start Date: 04/07/2017 Expiration Date: 12/18/2017

Your request to modify the referenced protocol has been approved by the IRB. **This protocol is currently approved for 120 total participants.** If you wish to make any further modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

Please note that this approval does not extend the Approved Project Period. Should you wish to extend your project beyond the current expiration date, you must submit a request for continuation using the UAF IRB form "Continuing Review for IRB Approved Projects." The request should be sent to the IRB Coordinator, 109 MLKG Building.

For protocols requiring FULL IRB review, please submit your request at least one month prior to the current expiration date. (High-risk protocols may require even more time for approval.) For protocols requiring an EXPEDITED or EXEMPT review, submit your request at least two weeks prior to the current expiration date. Failure to obtain approval for a continuation *on or prior to* the currently approved expiration date will result in termination of the protocol and you will be required to submit a new protocol to the IRB before continuing the project. Data collected past the protocol expiration date may need to be eliminated from the dataset should you wish to publish. Only data collected under a currently approved protocol can be certified by the IRB for any purpose.

If you have questions or need any assistance from the IRB, please contact me at 109 MLKG Building, 5-2208, or irb@uark.edu.

UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

May 2, 2017

MEMORANDUM

TO:	Eric Rawls Connie Lamm
FROM:	Ro Windwalker IRB Coordinator
RE:	PROJECT MODIFICATION
IRB Protocol #:	16-12-304
Protocol Title:	*Neural Mechanisms Underlying Conflict Monitoring in Smokers and Non-Smokers*
Review Type:	☐ EXEMPT ☒ EXPEDITED ☐ FULL IRB
Approved Project Period:	Start Date: 04/28/2017 Expiration Date: 12/18/2017

Your request to modify the referenced protocol to add an additional questionnaire has been approved by the IRB. **This protocol is currently approved for 120 total participants.** If you wish to make any further modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

Please note that this approval does not extend the Approved Project Period. Should you wish to extend your project beyond the current expiration date, you must submit a request for continuation using the UAF IRB form "Continuing Review for IRB Approved Projects." The request should be sent to the IRB Coordinator, 109 MLKG Building.

For protocols requiring FULL IRB review, please submit your request at least one month prior to the current expiration date. (High-risk protocols may require even more time for approval.) For protocols requiring an EXPEDITED or EXEMPT review, submit your request at least two weeks prior to the current expiration date. Failure to obtain approval for a continuation *on or prior to* the currently approved expiration date will result in termination of the protocol and you will be required to submit a new protocol to the IRB before continuing the project. Data collected past the protocol expiration date may need to be eliminated from the dataset should you wish to publish. Only data collected under a currently approved protocol can be certified by the IRB for any purpose.

If you have questions or need any assistance from the IRB, please contact me at 109 MLKG Building, 5-2208, or irb@uark.edu.

UNIVERSITY OF ARKANSAS

Office of Research Compliance
Institutional Review Board

August 17, 2017

MEMORANDUM

TO:	Eric Rawls Connie Lamm
FROM:	Ro Windwalker IRB Coordinator
RE:	PROJECT MODIFICATION
IRB Protocol #:	1708016049 (previously 16-12-304)
Protocol Title:	*Neural Mechanisms Underlying Conflict Monitoring in Smokers and Non-Smokers*
Review Type:	☐ EXEMPT ☒ EXPEDITED ☐ FULL IRB
Approved Project Period:	Start Date: 08/14/2017 Expiration Date: 12/18/2017

Your request to modify the referenced protocol has been approved by the IRB. **This protocol is currently approved for 120 total participants.** If you wish to make any further modifications in the approved protocol, including enrolling more than this number, you must seek approval *prior to* implementing those changes. All modifications should be requested in writing (email is acceptable) and must provide sufficient detail to assess the impact of the change.

Please note that this approval does not extend the Approved Project Period. Should you wish to extend your project beyond the current expiration date, you must submit a request for continuation using the UAF IRB form "Continuing Review for IRB Approved Projects." The request should be sent to the IRB Coordinator, 109 MLKG Building.

For protocols requiring FULL IRB review, please submit your request at least one month prior to the current expiration date. (High-risk protocols may require even more time for approval.) For protocols requiring an EXPEDITED or EXEMPT review, submit your request at least two weeks prior to the current expiration date. Failure to obtain approval for a continuation *on or prior to* the currently approved expiration date will result in termination of the protocol and you will be required to submit a new protocol to the IRB before continuing the project. Data collected past the protocol expiration date may need to be eliminated from the dataset should you wish to publish. Only data collected under a currently approved protocol can be certified by the IRB for any purpose.

If you have questions or need any assistance from the IRB, please contact me at 109 MLKG Building, 5-2208, or irb@uark.edu.

UNIVERSITY OF ARKANSAS | Research & Economic Development
Research Compliance

August 18, 2017

MEMORANDUM

TO:	Dr. Connie Lamm
FROM:	Ines Pinto, Biosafety Committee Chair
RE:	New Protocol
PROTOCOL #:	18012
PROTOCOL TITLE:	Neural Mechanisms Underlying Conflict Monitoring in Smokers and Non-Smokers
APPROVED PROJECT PERIOD:	Start Date August 17, 2017 Expiration Date August 16, 2020

The Institutional Biosafety Committee (IBC) has approved Protocol 18012, "Neural Mechanisms Underlying Conflict Monitoring in Smokers and Non-Smokers". You may begin your study.

If modifications are made to the protocol during the study, please submit a written request to the IBC for review and approval before initiating any changes.

The IBC appreciates your assistance and cooperation in complying with University and Federal guidelines for research involving hazardous biological materials.