The Effects of Self-Relevance on Neural Learning Signals Indexing Attention, Perception, and Learning

By

Mathew James Rocha Hammerstrom

B.Sc. University of Victoria, 2020

A thesis submitted in partial fulfilment of the requirements for the degree of

MASTER OF SCIENCE

in the Division of Medical Sciences

University of Victoria

We acknowledge and respect the ləkʷəŋən peoples on whose traditional territory the university stands and the Songhees, Esquimalt and W̱SÁNEĆ peoples whose historical relationships with the land continue to this day.

Supervisory Committee Page

The Effects of Self-Relevance on Neural Learning Signals Indexing Attention, Perception, and Learning

By

Mathew James Rocha Hammerstrom

B.Sc. University of Victoria, 2020

Supervisory Committee:

Dr. Olav Krigolson, Supervisor
Division of Medical Sciences

Dr. Jodie Gawryluk, Committee Member
Division of Medical Sciences

Dr. David Turk, Committee Member
Division of Medical Sciences

Abstract

Humans tend to preferentially process information relevant to themselves. For instance, in experiments where participants learn to manipulate stimuli referenced to themselves or someone else, participants exhibit larger reward processing signals for themselves. Additionally, attention and perception are biased not only towards one's self but those related to them. However, the aspect of processing information related to known-others has not been addressed in reward learning. Here, I sought to address this issue. Specifically, I recorded electroencephalographic (EEG) data from 15 undergraduate student participants who played a simple two-choice "bandit" gambling game where a photo presented before each gamble indicated whether it benefited either the participant, an individual they knew, or a stranger. EEG data from 64 electrodes on a standard 10-20 layout were analyzed for event-related potentials (ERPs) elicited by target photos and gambling outcomes. Post experiment, I examined the relationship between relatedness and the amplitude reward learning ERPs, namely the reward positivity and the P300, with one-way repeated measures analyses of variance. My results demonstrate that the amplitudes of reward learning ERPs are sensitive to the target of a gamble. A secondary goal of this research was to determine if these differences could be explained by attentional and perceptual responses to cues of who a given gamble was for. Indeed, stepwise linear regression analyses identified the P2, N2, and P3 indexed relevance to self as predictors of resultant reward signals. My findings provide further evidence that a reward learning system within the medial-frontal cortex is sensitive to others with varying self-relevance, which may be a function of biases in attention and perception.

Table of Contents

List of Tables

List of Figures

List of Abbreviations Used

EEG	Electroencephalography
ERP	Event-Related Potentials
RM ANOVA	Repeated-Measures Analysis of Variance
SFG	Superior Frontal Gyrus
tsPCA	Temporospatial Principal Components Analysis
vmPFC	Ventromedial Prefrontal Cortex
VS	Ventral Striatum
VTA	Ventral Tegmental Area

Acknowledgements

There are several individuals I would like to thank. I would like to acknowledge the guidance of my supervisor, Dr Olav Krigolson. Without his unwavering support and friendship from the moment I met him I would not be here today. Thank you to Chad Williams, who mentored me starting in my undergrad and has taught me most of the practical skills I have. I would also like to thank the other members of my supervisory committee, Dr David Turk and Dr Jodie Gawryluk, for their time, effort, and guidance. Thank you to our amazing research assistants, whose work keeps our laboratory running year-round.

Finally, thank you to my family and friends for not only encouraging me to do what I loved to do but for lending me the support I needed to do it.

"My dear and special friend

There's never a point at which we ever loved

The point at which I looked at you

Has always been part of my brain"

- Damon Albarn of *Blur*, on *Beachcoma*

Chapter 1: General Introduction

Humans tend to be biased about information that is relevant to them. This can be realized when someone calls out our name in a busy crowd, or even when making monetary decisions that affect those around us. In a range of cognitive functions, we will always process self-related information preferentially. However, our knowledge of how some of these functions interact when our sense of self is involved is limited. For example, while there is a wealth of evidence for biases in attention and perception on a range of references to the self, the same can not be said for reward learning. Another interesting source of inquiry is understanding how processing the attentional and perceptual cues of this self-related information may in turn affect that reward learning. As such, this thesis is primarily focused on understanding how reward learning changes when the target of learning changes on a scale of relatedness. The secondary focus will then be to determine if any changes can be explained by attentional and perceptual processing. Before these issues are addressed, a review of self-reference and cognition is warranted.

1. 1. Our Sense of Self

Humanity's first conceptuality of our sense of self stems from early Greek philosophy, psychology, and even poetry. Weaved throughout Homer's *Iliad* and *Odyssey* are debates on human and supernatural agency; are our actions determined by our own will or that of the gods (Seligman, 2020; B. Williams & Long, 2008)? Specifically, *Odyssey* posits that while gods may determine our ultimate fate, humans determine the actions in their lives via reasoning and imagination (B. Williams & Long, 2008). Following this, *psuchē* was introduced as an idea representing the animating principle of humanity, which is a central framework for our sense of self (Gill, 2017). For example, in Plato's *Republic*, he describes the *psuchē* as a determinant of human behaviour, resulting from our senses of reason (logical thinking), spirit (emotion), and appetite (desire; Gill, 2017). Aristotle built on this in *Nicomachean Ethics* by adding the idea of character: habits and ideals formed throughout development but rigid in adulthood. Within *psuchē,* humans have full agency over our decisions, but that agency stems from our character (Seligman, 2020). In the following century, as the Roman empire shifted more towards Christianity, the philosophy of the self too shifted towards total supernatural will again (Seligman, 2020). Saint Augustine of Hippo, who is considered the most influential philosopher of this era, returned to the idea that human agency is purely driven by the will of God. A theocratic view of our sense of self would prevail for the ensuing millennium until early modern approaches revisited the Greeks.

Early modern discussions of consciousness in the 1600s returned to agentic views of the self. Perhaps most importantly are Descartes's (1637; 1641) assertions, the most popular of which is “*cogito ergo sum*”, or “I think therefore I am” (see Cockburn, 2001 for an overview). Essentially, Descartes believed that to be aware of thought is proof of self-existence. Locke (1700) expanded this to affirm that we have an intuitive knowledge of our existence, and as such, we are conscious of our being. However, there were opponents to the idea that there is a perception of the self. Notably, Hume (1739;1740) and Kant (1781;1787) posited that our sense of self is simply a collection of momentary impressions and perceptions that we are aware of, and we are not actively perceiving our “selves”. Put simply, they shortened Descartes’ impressions of consciousness to “I am”. However, the aforementioned discussions of consciousness place the self as an object, rather than a subject (Fichte, 1794; Rundell, 2004). As such, Fichte posited that we experience and define our sense of self by merely existing, rather than separating it from ourselves. These debates on our sense of self-created frameworks for modern, psychological approaches to understanding the self.

Contemporary approaches have synthesized previous philosophical approaches to the self into practical frameworks. For example, Neisser (1988) posits that our sense of self is built based on the different types of information we perceive. Perhaps most importantly is the ecological self, built from perceptions of how we fit into our physical environment. In this model, we have a minimal sense of self which involves synthesizing our awareness of our actions, or sense of ownership, with the intention behind them, or sense of agency (Gallagher, 2000; Neisser, 1988; Wiggins, 1994). Indeed, there is evidence for neural correlates for maintaining our physical sense of self (Shimada et al., 2005). However, the minimal sense of self is not sufficient to explain the continuity we experience as part of consciousness (Gallagher, 2000). Extending this, some consider a narrative sense of self, where these constant, momentary impressions of experience are strung together by our imagination (Dennet, 1993; Gallagher, 2000; Gazzaniga, 1995; Neisser & Fivush, 1994), in line with Hume, Kant and Fichte's subjective self. In this model, our episodic memories, or memories of past events, are the foundation of our sense of self as they link our previous experiences to our current conscious experience (Gallagher, 2000). While the debate continues about how the minimal and narrative selves interact, it is clear they are strong models for modern approaches to understanding the self (Damasio, 1999; Gallagher, 2000). As such, our sense of self is still a powerful model for understanding quantitative aspects of behaviour and cognition (Gallagher, 2000).

1. 2. The Self in Cognition

1. 2. 1. Memory

Our sense of self, as described in Chapter 1.1., is reliant on episodic memories of our past experiences. To differentially process information relevant to ourselves, memories of what is self-relevant are required. Indeed, we retain memories based on our self goals and use them to bias the processing of incoming information (Conway, 2005; H. L. Williams et al., 2008), improving its recall if it is relevant to self (Cunningham et al., 2008, 2011; Rogers et al., 1977; van den Bos et al., 2010). Here, I'll review seminal work that has highlighted the bias towards the self in memory.

Early investigations of our sense of self in memory focused on the 'self-reference effect' (SRE). According to Rogers and colleagues (1977), information processed regarding the self is remembered better than information related to someone else (ie. the SRE). In this experiment, participants viewed a series of adjectives and were asked a series of questions about them that were self-related (ex. 'Are you shy'?) or lexical (ex. 'Does shy rhyme with truck?'; 'Does shy mean the same as blue?"). After participants completed the task, they were asked to record all the adjectives that they remember responding to. Results showed that recall of adjectives used in self-related questions was better than in non-related questions, suggesting that reference to self represents aids in memory encoding. This SRE has been attributed to the enhanced depth of processing, where self-relevant information is encoded in memory differently than other types of associative processing (Conway & Dewhurst, 1995; Craik & Lockhart, 1972; Cunningham & Turk, 2017; Symons & Johnson, 1997). The depth of processing model of memory indicates that, because more meaningful stimuli are perceived with greater depth, their memory traces are also encoded in memory with greater depth (Craik & Lockhart, 1972). In this case, the SRE would be dependent on the appraisal of self-relevant information as meaningful, as in this trait evaluation paradigm where self-related cues are explicitly linked to the items that participants are meant to recall (Symons & Johnson, 1997). This linking would lead to an enhanced encoding of that information, leading to better recall observed in follow-up memory tasks. As such, the SRE would not be exclusive to traits, as used in Rogers and Colleagues (1997). Indeed, any information associated with the self, like objects we own, would also enjoy this advantage.

Object ownership is also a powerful tool for studying how self-relevance impacts memory. Seminal work by Cunningham and Colleagues (2008) explored whether ownership of an item improved its recall. Participants sat next to a partner (a researcher posing as a participant) in front of a table with two coloured shopping baskets and were asked to sort shopping items into “their” or their partner’s basket. An experimenter showed an item with a coloured sticker and then gave it to one of the two partners to sort into the basket with matching colour. After the sorting task, participants completed a recognition task where they were shown items and asked if they are ‘Old’ (were in the sorting task) or ‘New’ (not in the sorting task). Results showed that participants recalled items sorted into their basket faster and more accurately than items sorted into their partner’s, regardless of whether the participant or their partner sorted it. This pattern of results suggests that objects we own elicit an SRE, but it is unclear if ownership is powerful enough in isolation to elicit this advantage. In other words, can the conditions of how we gain ownership over an object attenuate how we encode its related memories?

In a follow-up experiment, researchers further explored object ownership by manipulating choice in the shopping task (Cunningham et al., 2011). First, participants and their partners were able to choose where the items were sorted. In this case, results again showed an SRE for owned items, but only when the participant sorted the items. Next, participants assigned ownership to items before the sorting task by checking off squares on a grid that corresponded to items which would be sorted into their basket later. This allowed experimenters to create the illusion of choice over sorting, while actual ownership over items was attributed randomly. The same pattern of results observed when participants had control over sorting was observed. This pattern of results suggests that the SRE is based on how self-relevant the context of encoding is, as self-involvement or self-ownership may not be sufficient in isolation for biased memory encoding (Cunningham et al., 2008, 2011; Cunningham & Turk, 2017; van den Bos et al., 2010). This disputes previous accounts of a memory system that acts as a framework for biasing self-information (ex. Symons & Johnson, 1997). As participants did not have access to memories of the items they sorted in the second and third experiments, they could not weigh these items against existing self-relevant memories from a self-memory system (Cunningham et al., 2011). Instead, it was suggested that the combination of ownership and agency triggers low-level processes that in turn lead to enhanced memory. However, both ownership tasks and the trait evaluation paradigm rely on explicitly linking ourselves to the items to be remembered, which may not reflect everyday life where these links tend to be incidental (Cloutier & Neil Macrae, 2008). Would an SRE still be observed if we do not overtly link information to ourselves?

Turk and colleagues (2008) modified the trait evaluation paradigm to examine if explicit evaluative processing of self-relevance is necessary for enhanced memory encoding. In this experiment, participants were shown the face or full name of themselves or a well-known other (i.e., celebrity Angelina Jolie) paired with different traits. One group of participants completed the experiment similar to previous iterations of the trait evaluation paradigm and were asked to indicate whether the trait adjective described the person it was paired with. A second participant group was tasked with simply indicating whether or not the trait appeared above the referent cue. Both groups were tested with the same memory test where the trait adjectives are presented and participants are asked to indicate if they remember seeing them in the previous task. Results demonstrated that recall was better for participants who explicitly linked traits to referent cues compared to those that didn't. However, in both conditions, participants' recall was improved for items paired with themselves than with others, confirming that the presence of self is sufficient to elicit an SRE (Cunningham et al., 2011; Turk et al., 2008). If the SRE is not reliant on explicit pairing with self cues, then it can not be explained by enhanced depth of processing when items are linked to us (Cunningham & Turk, 2017; Turk et al., 2008). Instead, self-relevant information is likely encoded differently than other information, potentially due to a separate, self-memory system (Northoff et al., 2006; Turk et al., 2008).

Neuroimaging studies have identified a network of brain areas that encode self-relevant stimuli into long-term memory (Cunningham & Turk, 2017; Heinzel et al., 2011; Northoff et al., 2006; Turk, van Bussel, Waiter, et al., 2011). Specifically, Turk and colleagues (2011) used fMRI to examine the neural correlates of self-memory advantages using object ownership. In this experiment, participants sorted shopping items into self- and other-owned baskets and were later tested for object recognition of those items. Results from this study highlight a network of brain areas with unique roles in encoding self-relevant information. First, the self-advantage in memory was characterized by increased activation in the caudal anterior cingulate cortex (cACC) and medial superior frontal gyrus (SFG), areas typically involved in modulating attention to salient stimuli (Chiu et al., 2008), positive reward (Liu et al., 2007), and motivational stimuli (Winterer et al., 2002). Next, the regions of the anterior inferior parietal cortex had higher activity for self-owned objects, suggesting encoding of motor affordances (Handy et al., 2003). Finally, activity in the left insula, implicated in both emotion (Dupont et al., 2003; Lovero et al., 2009) and reward (Liu et al., 2007), was related to self-object processing but not memory. Temporal analyses of these activation patterns suggest a neural network that detects a self-object's relevance, activates relevant motor affordances and determines its value (i.e., reward; Turk, van Bussel, Waiter, et al., 2011).

In summary, humans bias encoding memories related to ourselves. Early work focusing on trait evaluation paradigms demonstrated that self-relevant information is remembered better than other information (Rogers et al., 1977; Symons & Johnson, 1997). This SRE even extends to objects we own (Cunningham et al., 2008, 2011; van den Bos et al., 2010). This bias in memory is likely due to a self-memory system that identifies self-relevance and biases memory encoding accordingly (Cunningham & Turk, 2017; Heinzel et al., 2011; Northoff et al., 2006; Turk, van Bussel, Waiter, et al., 2011). As such, the bias of self in cognition likely occurs at the onset of attention and subsequent perception of self-relevant information.

1. 2. 2. Attention and Perception

A large body of work examining differences in how humans process self-related information is focused on attention and perception. Visual stimuli are processed and identified more rapidly and accurately when they are associated with our sense of self (Keyes & Dlugokencka, 2014; Moray, 1959; Sui et al., 2006, 2012; Treisman, 1960; Turk, van Bussel, Brebner, et al., 2011) The own-name effect, popularly referred to as the “Cocktail Party Phenomenon”, is a robust psychological phenomenon in which hearing your name elicits selective attentional capture (Alexopoulos et al., 2012; Humphreys & Sui, 2015; Moray, 1959; Treisman, 1960). Given the complexity of attentional and perceptual processing and their role in subsequent cognition, reviewing this literature is necessary.

Hearing one’s name, even when not attending to it, can facilitate attention to subsequent stimuli. In the earliest example of this phenomenon, Moray (1959) had participants attend to a spoken phrase in one ear while an unrelated string of words was presented in the other, and later asked them to recall the unattended words. Participants were unable to reliably recall any of the words presented in the unattended ear. However, when the participant’s name was presented before a trial, their recall of unattended words and phrases improved. Moray termed this effect the “identification paradox”, where unattended information is blocked from conscious perception unless it includes subjectively important information. Similar advantages are observed with other self-relevant stimuli in the dichotic listening task, such as personality traits (Bargh, 1982). However, some criticized this experiment as a measurement of attention, as it is difficult to ensure that participants were not attending to the unattended ear (Lachter et al., 2004).

One’s name can also alter visual attention. Wolford and Morrison (1980) examined whether one’s name elicited similar effects to Moray in a visual setting. In this experiment, participants were tasked with determining whether two numbers presented laterally were of the same parity (both odd or even) or not while irrelevant words were presented between them. For some of the trials, the word presented was the participant’s last name. Results showed that participants were able to reliably detect the presence of their name, but had larger reaction times for these trials. As such, it seems that reading one’s name not only captures attention but may also take away attentional resources from the current task (Bargh & Pratto, 1986; Wolford & Morrison, 1980). In a more recent example, Alexopoulos and colleagues (2012) completed a series of experiments to fully examine attentional capture elicited from one’s name. In these experiments, researchers manipulated how a participant’s name is presented (ex. Inverted), the names presented (ex. Presenting familiar, other names), and task goals (ex. Suppress reaction to your name) to explore how involuntary this effect may be. Participants routinely react to their name faster than others, regardless of how it is presented, the names it is presented with, and at the cost of task performance, suggesting a powerful bias of attention for the self in visual attention. Although, an alternative explanation is that the familiarity of our name is sufficient to guide attention (Humphreys & Sui, 2015). Additionally, these studies do little to specifically address concerns that participants may be attending to information they are instructed not to (Lachter et al., 2004). As such, new approaches to studying attention to self-relevant stimuli are required.

To fully understand attention to self-relevant stimuli, direct measurement of biomarkers of attention is required. Yang and colleagues (2013) had participants view names with varying self-relevance while they recorded eye-tracking data. In each trial of the experiment, participants viewed six names and were tasked with determining if one of three target names, their own; their mother’s; or a celebrity’s, was present among them. Behavioural results showed that participants responded faster and more accurately when their name was the target, compared to similar performance between their mother's name and the celebrity's. Eye-tracking data matched this, as participants demonstrated quicker first eye movements and fewer saccades in response to their name, suggesting an automatic capture of attention (Hayhoe, 2004; Yang et al., 2013). Given that both behavioural and eye-tracking results did not dissociate the two known others, it is likely that familiarity with names can not explain how we attend to them. Instead, perhaps there is a unique pattern of brain activity for self-relevant stimuli.

Numerous accounts of self-biases in attention may point to a distinct attention system for the self. Humphreys & Sui (2015) propose the existence of a Self-Attention Network (SAN), a network of brain areas dedicated to responding to self-relevant stimuli. Studies using fMRI have demonstrated that the ventromedial prefrontal cortex (vmPFC) is necessary for modulating attention towards self-relevant stimuli, and is activated by the presence of self-relevant stimuli (Jenkins & Mitchell, 2011; Sui et al., 2013), while the left posterior superior temporal sulcus (LpSTS) is linked to attention in social contexts (DiQuattro & Geng, 2011). In the SAN, self-relevant information activates the vmPFC, which in turn activates the LpSTS which biases the top-down attentional control network (Humphreys & Sui, 2015). Further, in this framework, top-down attentional control can inhibit the activity of self-biased attention via inhibitory connections to the vmPFC (Humphreys & Sui, 2015; Sui et al., 2015). This may be beneficial for scenarios where rejecting self-related information is necessary for task performance. Overall, the SAN accounts for evidence of self-biases in attention and provides a framework for how that bias may guide perception and subsequent action.

While own-name studies are a good medium to study the modulation of attention to self-relevant information, own-face studies are well-suited for examining the same modulation of perception. Tong and Nakayama (1999) examined participants searching for their faces or unrelated observer (i.e., A stranger) faces amongst unrelated distractors. When faces are presented either upright or inverted, participants routinely recognize their face faster than the stranger's face, even after learning the stranger's face over time. This advantage is still observed when faces are presented with varying profile views and depth. The authors suggest that one’s face has a robust representation in the brain, where extensive visual experience leads to a decreased demand for attentional resources when viewing it (Tong & Nakayama, 1999). In a follow-up study, Kircher and colleagues (2001) examined the recognition of one’s face compared to that of their partner to understand if robust representations may extend to others. In this experiment, participants viewed a series of images where their face or their partner’s face was morphed with a similar-looking stranger’s, to create novel yet recognizable face representations. The participants were tasked with indicating whether the presented face was more similar to their own or their partners. Results showed no differences in response time between self- and partner-faces across the morphed spectrum, suggesting that one's face is not processed differently on a behavioural level when compared with another overlearned, emotionally salient face (Kircher et al., 2001). However, since then there has been contradictory evidence on the face perception of known others. For example, fMRI studies have identified the ventromedial prefrontal cortex’s (vmPFC) role in guiding attention to self-relevant stimuli (Sui et al., 2015, more on this later), yet its activity is different for known faces in American participants (Heatherton et al., 2006) but the same for known faces in Chinese participants (Zhu et al., 2007). As such, it may be necessary to consider the differences between self, known, and unknown faces with more depth.

In an interesting example of this, Sui & Humphreys (2013) examined how participants perceived their faces compared to known and unknown others by studying ex-Gaussian reaction time distributions. This approach assumes that categorizing each face yields different distributions of reaction times. The advantage of examining ex-Gaussian distributions is that they are thought to reflect different underlying processes contributing to reaction time, and can illuminate differences between fast, stimulus-driven reactions and slow, deliberative ones (Balota et al., 2008; Sui & Humphreys, 2013). In this study, participants were tasked with categorizing their own, their friends, and an unknown other’s face as rapidly and accurately as possible (Sui & Humphreys, 2013). The first experiment included three sub-experiments, representing three pairings of the three face categories (i.e., Self-Familiar, Familiar-Unfamiliar, and Self-Unfamiliar). Participants were shown faces and asked to respond with which of the two faces from the current pairing they observed. In the second set of experiments, participants were instead asked to group faces as known or unknown (i.e., Self & Familiar or Unknown), or self or non-self (i.e., Self or Familiar & Unfamiliar). Standard analysis of mean reaction times in all experiments showed that participants were faster to categorize their faces compared to others, but still faster for familiar than unfamiliar faces. Interestingly, while reaction times to self faces were stable between tasks, the distribution of reaction times to familiar faces was modulated by how it was categorized. Specifically, when compared to unfamiliar faces in the first experiments, the distribution of familiar face reaction times was closer to that of self faces. In the second set of experiments, when familiar faces were categorized with self faces, familiar and unfamiliar other faces exhibited similar reaction time distributions. In summary, their results suggest that perceiving our face is uniquely automatic while perceiving other faces is deliberative and thus, subject to context. While studying face perception has proved fruitful for understanding self biases in perception, the saliency of faces may make it difficult to draw conclusions about general perceptual processes (Sui et al., 2012).

Indeed, perceptual advantages for self-relevant stimuli can be observed with extraneous stimuli. Sui and colleagues (2012) examined how assigning self-relevance to shapes in a perceptual matching experiment affects performance. In this experiment, participants are taught to associate three geometric shapes to themselves, one of their friends, and a stranger, respectively (ex. “Mary is a circle, you are a square, and a stranger is a triangle”). Next, participants are shown pairings of the shapes and names of the three individuals and are tasked with indicating whether the pair matches the associations made previously. Participants were able to identify self-shape pairs faster than other-shape pairs but also were faster for the familiar- than stranger-shape pairs. Next, the experiment was modified so that the stranger is replaced with the word ‘none’ to act as a baseline, and a new pairing for the participant’s mother is introduced. As expected, perceptual matching performance was better for self- and familiar- associations compared to the neutral associations. Again, self-shape pairs are perceived more rapidly and accurately than both friend- and mother-shape pairs.

In summary, a wealth of evidence supports the idea that information relevant to our sense of self is given preferential treatment early in cognition. At the onset of attention, self-relevant stimuli elicit enhanced processing, which some propose is due to a network of brain areas dedicated to processing self-related stimuli (Humphreys & Sui, 2015; Sui et al., 2013). Additionally, both implicit and explicit links between ourselves and visual stimuli aid in perceptual processes (Keyes & Dlugokencka, 2014; Kircher et al., 2001; Sui et al., 2012; Sui & Humphreys, 2013; Tong & Nakayama, 1999). Nonetheless, this body of work does little to explain if and how these advantages may apply to higher orders of cognition.

1. 2. 3. Learning and Reward

It stands to reason humans would preferentially process information that leads to a reward for them. Some researchers have posited that self-biases arise because self-relevant information is inherently rewarding (Northoff & Hayes, 2011). This idea stems from fMRI studies on both self-referential processing and reward, which both highlight the ventral tegmental area (VTA), ventromedial prefrontal cortex (vmPFC), and ventral striatum (VS) as important brain areas for these functions (de Greck et al., 2008; Enzi et al., 2009; Ersner-Hershfield et al., 2009; Phan et al., 2004). As such, Northoff and Hayes (2011) suggest a model of self and reward stimuli where pertinent stimuli are processed in parallel (Northoff & Hayes, 2011). Consider the example of receiving a monetary reward. First, we may process this stimulus as self-specific, which requires rapidly detecting the stimuli as something specific to us (i.e., This money belongs to me now). At this point, the money has a relatively low reward value because all that is known is that it belongs to us. Next, we assign value to money, which is principally a reward function and not reliant on self-referential processing. This process may account for evidence from pathological gamblers with enhanced reward activity but normal self-referential processing (de Greck et al., 2010). Finally, we can attribute direct self-relevance to the reward and indicate what this money means to us (i.e., I can use this money to buy something; Northoff and Hayes, 2011). This model provides a strong framework to consider how our sense of self interacts with reward processing.

While this is an attractive idea, recent work has rejected reward as an explanation for self-biases and instead characterizes them as overlapping yet distinct representations in the brain (Sui & Humphreys, 2015a). In this experiment, researchers implemented the shape-association task (see Sui et al., 2012, *Chapter 1. 2. 2. Attention and Perception*), where participants learn associations between different shapes and with themselves, someone they know, or a stranger, and are tested with random pairings of these shapes and people. In this iteration of the task, self-related trials received no reward while friend- or stranger-related trials received either high or low rewards. This approach allowed researchers to examine whether self-relevance is inherently rewarding by comparing self-non-rewards to other rewards. Results showed that behavioural performance (response time and accuracy) was improved for self-related trials compared to both friend and stranger trials with low rewards, but similar to both trial types when they elicited high rewards. Further, improved performance was observed between friend and stranger trials dependent on which trial type was associated with high rewards. As such, differences in behaviour based on the target of rewards (bias to self-trials) were not related to differences based on reward value (bias for larger reward). While this suggests segregation of reward and self in the brain, more evidence is needed to understand how self-related and high-reward stimuli are represented in the brain.

Redundancy gains are performance improvements observed when multiple target stimuli appear in a visual search task and are a good method for separating reward from other concepts in information integration. Sui and Humphreys (2015b) added redundancy gains to the shape-association task to examine if self and reward biases lead to similar information integration processes. In this version of the task, participants only learned self- and friend-shape pairings and were then tasked with indicating whether a given single shape or shape-pair was associated with themselves or their friend. This approach allows comparison between single targets and multiple targets and thus explores the effects of redundancy gains in shape-pair trials. Again, visual search performance was similar for single shapes between self and high reward trials. In shape-pair trials with physically (ex. Two identical shapes) or conceptually identical shapes (ex. two shapes that are both associated with the same person), redundancy gains are larger for self-trials. However, reward value only modulated redundancy gains for conceptually related shapes, suggesting a modulation of information integration at a higher level. As such, perceptual integration is subject to self- but not reward-bias, in line with previous work on self-reference and perceptual matching (Sui et al., 2012). These results favour Northoff and Hayes' (2011) account of parallel processing of self and reward, where there are both overlapping (conceptual integration) and separate (perceptual integration) representations of these concepts in cognition.

1. 3. Neuroimaging

1. 3. 1. Electroencephalography and Event-Related Potentials

One method to measure the neural response to a given event is electroencephalography (EEG). EEG is a neuroimaging method where electrodes are placed on the participant's head to record electrical potentials from the scalp (Cohen, 2017). Specifically, large populations of neurons perpendicular to the cortical surface must fire post synaptic potentials in synchrony to produce electrical dipoles that can be measured as voltages. These dipoles are produced in high volume from various cortical structures, and their summed activity is reflected at a given electrode. For example, a single dipole generated in medial, subcortical structure could be captured in EEG data with varying polarity and intensity across numerous electrodes. As such, this approach does not explicitly produce a signal that directly maps signals to specific neural circuitry. This makes it difficult to both 1) make conclusions about specific neuroanatomical sources of activity, and 2) dissociate potentials between electrodes. Further, the signal recorded from an EEG electrode is also subject to electrical artifacts from nearby, non-neural sources such as movements in the face and neck (especially eyeblinks; Luck, 2014). However, EEG has numerous advantages over other neuroimaging methods. For example, compared to fMRI, which measures blood flow through the brain over one-to-two second intervals, EEG directly probes electrical brain activity with high temporal resolution (i.e., one-to-two millisecond intervals). Further, there are methods to overcome the aforementioned challenges and still produce a signal reflective of brain activity.

A common way to use EEG to probe neural activity is the event-related potential (ERP) technique. This method involves using time-locked epochs of EEG activity yoked to “events” (i.e., visual stimuli, participant responses, etc.) during the performance of psychological tasks and reflects the neural response that the event elicits (Luck, 2014). ERPs are essentially the averaged response to a stimulus, so high volumes of trials with these stimuli are required to reduce the signal-to-noise ratio. The resultant waveforms can be examined for components that represent underlying neural processes. For example, the P300 ERP component is commonly measured as a positive deflection around 300-400 milliseconds after a stimulus, and is thought to represent context updating in perception (Duncan-Johnson, 1981; Patel & Azzam, 2005; Polich, 2003, 2007; Zhou et al., 2010). In sum, ERP experiments may produce a myriad of components to examine based on their amplitude, latency, topography, and the stimuli that elicited them. Here, I will review a history of ERPs important for the current enquiry, and highlight examples from research on self-referential processing.

1. 3. 2. ERPs for Attention and Perception

Background

The high temporal resolution of ERPs is well-suited for examining attention and perception. In visual-attention experiments, components reflecting attentional processing can be observed less than 100 ms after stimulus onset (the N1 and P1; Woodman, 2010; Zhang & Luck, 2008). Early attentional components typically reflect spatial attention (Zhang & Luck, 2008), with subsequent signals indexing feature-encoding (Woodman, 2010). For example, the N170 is a negative-going component that is elicited by the processing of faces (Luck, 2014; Morgan et al., 2008). Given that these ERPs index attention and perception, they are especially sensitive to the characteristics of stimuli, and are vulnerable to physical stimuli confounds (Woodman, 2010). Further, temporal and spatial overlap between attentional and perceptual ERPs, and the cognitive processes they index, can make it difficult to dissociate them. Nonetheless, these methodological concerns can largely be addressed by borrowing from existing research that probes similar processes.

Evidence in Self-Referential Processing

ERPs have successfully been employed to probe differences in attention and perception resultant from the self-reference effect. Object ownership experiments have demonstrated decreases in the P1 (Turk, van Bussel, Brebner, et al., 2011) and increases in the P2 for self-owned objects (Krigolson et al., 2013). In face-perception experiments, viewing one's face versus that of others yields a larger N170 (Keyes et al., 2010; Keyes & Dlugokencka, 2014).

Similar to behavioural investigations, there is electrophysiological evidence of differential processing of one's face. Keyes & Dlugokencka (2010) explored participants' electroencephalography (EEG) responses to their faces and other faces known to them. Participants were shown their face, the face of someone they know personally (ex. A roommate), and the face of an unfamiliar other, and had to respond according to their orientation. At around 250 ms after the presentation of self-faces, the amplitude of the P2 and N2 were reduced compared to other faces. Differences between friend and stranger faces were not apparent until around 400 ms after face presentation. As such, it seems that viewing own faces elicits automatic attentional processing while dissociating familiarity with faces is a later process (see also Sui et al., 2006). Importantly, the timings used to quantify these components were not guided by local maxima or minima, but instead by different-sized intervals set throughout the ERP time course. Generally, this method for quantifying ERP components is not encouraged, as it can bias affects and reduces applicability to other contexts (Luck, 2014; Luck & Gaspelin, 2017). More careful consideration of component topography and timing is required, especially in the context of attention and perception (Woodman, 2010).

Borrowing from the aforementioned behavioural studies, name-reading elicits components that index self-referential processing. Fan and colleagues (2013) examined a series of ERP components elicited by viewing names on a scale of relevance to self. In this experiment, a modified version of the Oddball paradigm (more on this in *Chapter 1. 3. 4. ERPs for Learning: The P300*), where participants viewed a series of lexical and non-lexical phrases, which served as standard and target stimuli, respectively.

Participants were instructed to respond to target stimuli and ignore all others. Syllable-matched names with varying self-relevance were mixed into these trials as distractors. Results showed that self-relevant names elicited the largest P2 and P3 components and smallest N2, but a similar N1 component to other names. This suggests that detecting the existence of a name happens early in the information processing stream, evident from similar N1 between conditions. As such, the P2 is the first indication of a name's reference to the self, and the N2 and P3 reflect subsequent perceptual processing of self-reference. Importantly, results also indicated these components had frontocentral topographies, in line with fMRI results highlighting midline areas important for self-referential processing (de Greck et al., 2008, 2010; Enzi et al., 2009; Humphrey et al., 1994; Jenkins & Mitchell, 2011; Northoff et al., 2006; Sui et al., 2013).

Finally, the ERP technique has also been applied to study attention and perception in perceived ownership (see *Chapter 1. 2. 1. Memory*). Similar to previous iterations of the ownership task, Turk and colleagues (2011b) presented participants with supermarket items and used colourized borders to signify if the object "belonged" to the participant or an unknown other while EEG data were recorded. ERPs yoked to the presentation of ownership cues showed a reduced P1 amplitude for self-ownership cues compared to items owned by others, suggesting attention is narrowed to items shortly after taking ownership. Additionally, increases in the P3 for self-cues were observed, indicating enhanced attentional and perceptual processing. In contrast to Fan and colleagues (2013), current results showed differences in scalp location for the P3 (Turk, van Bussel, Brebner, et al., 2011). Again, this highlights the need for a careful selection of topographies and timings used to quantify components. This will be addressed in *Chapter 1. 3. 5. Principal Components Analysis for ERP Research*.

1. 3. 3. ERPs for Learning: The Reward Positivity

Background

Importantly, ERPs provide a means to examine the neural processes that underlie reward learning/feedback processing. The earliest example of this comes from Miltner, Braun, & Coles (1997), who identified an ERP elicited by performance feedback called feedback error-related negativity (fERN). In this experiment, participants completed a time estimation task where they were instructed to respond when they thought one second had passed and were given feedback on the accuracy of their assessment. Results showed that in comparison to correct feedback, error-related feedback elicited a negative deflection in the ERP peaking at about 230-330 ms following the feedback stimulus. While the fERN is the basis for a significant portion of modern reward learning ERP studies, the neural process it indexes was the subject of much debate in the following years.

Some have suggested that the fERN captures the expectedness of feedback, rather than its valence. This stemmed from numerous accounts of variance in this signal depending on feedback context (see Niewenhaus et al., 2004 for a review). Specifically, Holroyd and colleagues (2008) addressed this by comparing the fERN, now termed the feedback-related negativity (FRN), with the N2, a component that measures conflict monitoring. Previously, the N2 had been associated with responses to infrequent stimuli in visual oddball tasks and has similar timing and topography to the FRN (Donchin & Coles, 1988; Folstein & van Petten, 2008). As such, Holroyd and colleagues (2008) had participants complete a visual oddball and time estimation task so that these waveforms could be compared. Results showed that error feedback in the time estimation task simply elicits an instance of the N2, which generally signals unexpected task-relevant events. Additionally, correct feedback elicited a positive deflection in the N2 that was postulated to be caused by inhibition of N2 signalling or super-position of a unique component that cancels out the N2. At the time, researchers designated this component as the feedback correct-related positivity, but the following work generally accepts the modern title of reward positivity (see Krigolson, 2018 for a review).

The reward positivity is commonly used as the principal measure in modern ERP studies on reward learning. This component is typically present around 250 ms after the onset of performance feedback (Proudfit, 2015). This component is maximal over frontocentral electrode sites and has previously been source-localized to the anterior cingulate cortex (ACC; Holroyd et al., 2004). The reward positivity is thought to predict a positive prediction error-a signal that feedback is better than expected-in reinforcement learning tasks (Krigolson, 2018; Luck, 2014; Proudfit, 2015; C. C. Williams et al., 2021). For example, a participant may receive a ‘win’ after selecting an option that previously resulted in a ‘loss’, and as such, the outcome was better than expected and elicits a reward positivity. As the participant continues the task and learns how likely it is that a given option will provide a ‘win’, they get better at predicting the outcomes of their decisions. Thus, the magnitude of the prediction error, and the amplitude of the reward positivity, diminishes as we learn and become less dependent on feedback for performance evaluation (Krigolson et al., 2009, 2014; Walsh & Anderson, 2012; Williams et al., 2019).

The reward positivity is commonly measured in reinforcement learning paradigms where participants have to learn how to make correct or high-value decisions: namely time-estimation (Holroyd et al., 2005; Holroyd & Krigolson, 2007; C. C. Williams et al., 2017a), a multi-armed bandit (Hammerstrom et al., 2021; Hassall, Hajcak, et al., 2019; Hassall, Mcdonald, et al., 2019), and gambling tasks (Hassall et al., 2013; C. C. Williams et al., 2016; Yeung et al., 2005; Zhou et al., 2010). Importantly, these tasks are typically designed in a manner where participants will produce an approximately equal volume of wins and losses. In the example of multi-armed bandit tasks, researchers can set the win probabilities of the bandits to 60% and 10%, so that even if participants frequently select the better choice, they will receive nearly equal amounts of win and loss feedback (Colino et al., 2020; Hammerstrom et al., 2021; Howse et al., 2018; Krigolson et al., 2017). This component of the paradigm ensures that frequency contamination (higher volume of one condition) does not drive any changes in the surrounding frequency-sensitive ERP components that may alter the reward positivity (Holroyd, 2004; Krigolson, 2018).

Evidence from Self-Referential Processing

Neuroimaging evidence exists in support of a self-reference effect in reward learning (Hassall et al., 2016; Krigolson et al., 2013). Specifically, Krigolson and colleagues (2013) had participants complete a modified version of the ownership task from Turk et al. (2011a), where borders signified targets of gambles while EEG was recorded. Participants were instructed to press a button with a probability of providing a reward, and if they were successful the reward would be provided to the target by the border (either the participant or an unknown other). Results showed that the amplitude of the P2 yoked to the target cue reflected whether the subsequent gamble had utility for the participant (i.e., larger P2 for self targets). The reward positivity was also diminished when participants gambled for others. Together, these results suggest reward learning in the brain is preferentially activated by stimuli that can lead to rewards for ourselves. However, as mentioned in *Chapter 1. 2. 3. Learning and Reward*, links between self-reference and reward are sensitive to reward value (Sui & Humphreys, 2015a, 2015b).

Hassall and colleagues (2016) modified the paradigm used in Krigolson et al. (2013) so that different blocks of trials would yield either high or low rewards. Importantly, participants were told that rewards gained on trials for "Others" would be given to other participants in the study. Results showed that the reward positivity was only larger for high reward trials when participants gambled for themselves, indicating that participants are not tracking the value of rewards when gambling for others. This is contrary to the aforementioned attempts to dissociate the self and reward (Sui & Humphreys, 2015a, 2015b), where participants preferentially tracked reward value when familiar and unfamiliar others were involved. The principal difference between the theoretical basis of these studies (i.e., ignoring the presence of neuroimaging methods) is how self-relevance is portrayed. In the former example, self-relevance is binary (i.e., self or other), and its effects on reward processing are determined by who receives the reward (Hassall et al., 2016). In the latter, self-relevance is on a continuum (i.e., self, friend or stranger), and its effect on reward processing are based on reward value, regardless of who receives it (Sui & Humphreys, 2015a, 2015b). To fully understand how self-relevance affects reward processing, a synthesis of these approaches is required.

1. 3. 4. ERPs for Learning: The P300

Background

The P300 is an endogenous ERP elicited in response to certain visual stimuli. This ERP is a relatively large positivity typically recorded at the 300ms-400ms offset (Patel & Azzam, 2005).

According to Donchin and Coles (1988), the P300 is elicited by stimuli that are awaited in highly repetitive and structured tasks. Perhaps the most common task used to elicit the P300 is the Visual Oddball (Patel & Azzam, 2005; Polich, 2007). Generally, this task involves viewing a series of coloured circles where one colour (the Oddball) appears less frequently than the other (the control; ex. 75% of circles are blue and 25% are green). ERP waveforms yoked to control circles (the common stimuli) are subtracted from Oddball waveforms (the uncommon stimuli), typically resulting in positivity.

But what does the P300 represent? While it has been hypothesized that the P300 (or P3) is the brain's response to surprising stimuli (S. Sutton et al., 1965, 1967), it has also been proposed to reflect the process of context updating (Donchin, 1981; Duncan-Johnson & Donchin, 1977) and context closure (Donchin & Coles, 1988; Polich, 2003, 2007). This awaited stimuli, or the "surprising" stimuli, is perceived as a different type than the common, or "repeated", stimuli and thus the difference elicits a large positivity. This difference is a characteristic of both the degree of the uncertainty of sensory modality in the stimulus, as well as the ability to correctly anticipate the stimulus. The magnitude of the P300 is inversely proportional to the expectancy, not the frequency, of the unexpected event. This means that highly surprising events elicit large P300 components, and vice versa (Patel & Azzam, 2005).

In addition to its role in visual perception, the P300 can also be elicited by performance feedback. In the aforementioned reinforcement learning tasks, the P300 can be elicited following the reward positivity in response to win and loss feedback (Hajcak et al., 2005, 2007; Hammerstrom et al., 2021; Leng & Zhou, 2010; Wu & Zhou, 2009; Zhou et al., 2010). Many posit that evaluating feedback in these tasks involves an initial encoding of the outcome (indexed by the reward positivity; Sambrook & Goslin, 2015; Walsh & Anderson, 2012) and a subsequent appraisal of the reward/outcome magnitude (indexed by the P300; (Hajcak et al., 2005, 2007; Leng & Zhou, 2010; Sato et al., 2005; Wu & Zhou, 2009; Yeung et al., 2005). As such, to fully examine reward learning in these paradigms, measurement of both of these components is warranted (Krigolson, 2018). To date, there are no examples of a P300 in response to feedback in a self-referential learning task. The current study will attempt to address this.

1. 3. 5. Principal Components Analysis for ERP Research

As previously mentioned, ERPs are not often distinct in which electrodes they are present, and how soon after a stimulus they appear (see *Chapter* 1. 3. 1. *Electroencephalography and Event-Related Potentials*). To overcome the challenges of overlapping timings and topography, researchers can implement principal components analysis (PCA). For example, early implementations of PCA allowed researchers to separate two highly overlapping P300 components into two signals with distinct functions and source localizations (Dien et al., 2003). Of particular interest is temporospatial PCA, where EEG data are first subjected to temporal dimension reduction yielding a set of temporal factors by using time points in ERP epochs as variables (Dien, 2012). From this set, researchers can select factors based on timing (i.e., similar timing to the ERP components of interest) or how much variance in the data they account for. The selected temporal factor can be submitted to spatial PCA with electrode locations as variables, yielding a set of spatial factors (topographies). Selecting a spatial factor consistent with our ERP of interest results in a factor score that can be submitted to inferential statistics to 1) confirm the existence of this ERP in the dataset, and 2) weigh components between experimental conditions for a better representation of its activity. In practice, this method is successful in dissociating ERPs in studies of perception (Brier et al., 2008; Daffner et al., 2015; Foti et al., 2011; Mott et al., 2014; Schomaker & Meeter, 2014; Tenke et al., 2010), emotion (Foti et al., 2009; MacNamara et al., 2019), and reward (Foti et al., 2011; Hammerstrom et al., 2021). Here, this method can help separate and confirm the existence of the numerous ERP components elicited by self-reference cues and performance feedback.

1. 4. The Current Study

The exact neural mechanisms for how humans process rewards for individuals familiar and unfamiliar to them remain unclear. The top-down modulation of attention and perception for the self, familiar, and unfamiliar others is a robust phenomenon. However, how these biases interact with reward learning is yet to be fully explored. This thesis consists of an experiment designed to determine how humans process rewards for themselves compared to familiar and unfamiliar others, and if differences in attention and perception can explain that processing.

In this experiment, participants were tasked with making rewarding gambling decisions that benefitted different individuals while EEG data were recorded. The gambling task has participants select one of two coloured squares resulting in a win or a loss. Before each decision, the participants are shown who would receive the reward; themselves, someone they know personally, or someone they do not know. The goal of this experiment was to examine whether reward processing on a range of relatedness to the individual is affected by self-biases and if these effects can be explained by the implicit processing of self-related information. Specifically, the primary hypothesis was that the reward positivity and P300 would decrease linearly when gambling with friends and strangers compared to one's self. In other words, the amplitude of these components would be largest for 'You'-targeted gambles, but larger for 'Known-Other'-targeted gambles than 'Stranger'-targeted gambles. This would indicate that reward learning systems are sensitive to a continuum of self-relevance, rather than a binary dissociation of self and all others. The secondary hypothesis is that the attentional and perceptual ERPs yoked to relevance cues, the P1, P2, N2, and P3, would predict changes in the reward components. In this case, step-wise linear regression analyses would identify the amplitude of cue-locked ERPs as significant contributors to an equation predicting the amplitude of the reward positivity and P300. This pattern of results would show that differences in reward learning caused by self-relevance can be explained by how we process cues of that relevance.

Chapter 2: Experiment One - The Neural Correlates of Gambling for Yourself and Others

2. 1. Introduction

Humans tend to be biased towards information relevant to themselves. One potential explanation for this is that we are inherently self-preservative, and our behaviour reflects the desire to maximize utility (Cunningham et al., 2013; Mill, 1863). Another possibility is that we find self-relevant information inherently rewarding, and bias cognition accordingly (Northoff & Hayes, 2011). Either way, stimuli related to our sense of self are differentially processed throughout the brain, including attention (Alexopoulos et al., 2012; Humphreys & Sui, 2015; Keyes & Dlugokencka, 2014; Turk, van Bussel, Brebner, et al., 2011), perception (Sui et al., 2012, 2014; Sui & Humphreys, 2015b), memory (Cunningham et al., 2008, 2011), and learning (Hassall et al., 2016; Krigolson et al., 2013). However, little is known about how these advantages interact between cognitive functions. Can biased attention and perception of self-related information explain how we process it at a higher level?

Biased attentional and perceptual processing of self-relevant information is a robust psychological phenomenon. Both auditory and visual stimuli are attended to more rapidly when they are associated with our sense of self (Bargh, 1982; Moray, 1959; Sui et al., 2006; Wolford & Morrison, 1980; Yang et al., 2013). Further, self-relevant stimuli yield similar advantages in perceptual processing (Keyes et al., 2010; Keyes & Dlugokencka, 2014), even when self-relevance is associated with novel stimuli (Sui et al., 2012). This is also evident in studies employing electroencephalography (EEG), where event-related potentials (ERPs) indexing these processes are sensitive to stimuli’s relatedness to self (Fan et al., 2013; Turk, van Bussel, Brebner, et al., 2011). Cues that an object belongs to somone elicit a smaller P1 (Turk, van Bussel, Brebner, et al., 2011) and larger P2 (Krigolson et al., 2013) than unowned objects. Attention to names with high self-relevance elicits enhanced N2, P2, and P3 components (Fan et al., 2013). Taken together, evidence suggests that self-related information is preferentially attended to and perceived for importance, perhaps to the benefit of later cognitive processes.

Another cognitive process affected by self-relevance is reward processing. For example, reward positivity is an ERP component that is maximal when feedback indicates an outcome is better than expected (Krigolson, 2018; Proudfit, 2015; C. C. Williams et al., 2017b). In experiments where perceived ownership is weighed against reward learning, the reward positivity is larger for self-related rewards than other-related rewards (Hassall et al., 2016; Krigolson et al., 2013). Also evoked by feedback, the P300[1] is an ERP component prominent after stimulus onset that reflects the functional significance of feedback stimuli (Hajcak et al., 2005, 2007; Leng and Zhou, 2010; Zhou et al., 2010; Yeung et al., 2005; Sato et al., 2005; Wu and Zhou, 2009). Together, the reward positivity and P300 provide an effective means for studying feedback processing.

The goal of the present experiment was to extend previous work on self-relevance and reward learning. Specifically, I sought to determine if reward learning systems are sensitive to familiar and unfamiliar others, suggesting they track self-reference on a continuum. To accomplish this, I had participants play a gambling task where each gamble benefitted themselves, someone they know personally, or a stranger. In this task, participants select one of two coloured squares resulting in a win or a loss, where one square has a higher probability of resulting in a win while EEG was recorded. Before each decision, a photo and the name of either the participant, their known other, or the unknown other was shown before the decision to indicate who would benefit from the subsequent gamble. ‘Win’ feedback on a decision would result in a monetary reward for that person. This approach yields two areas of analysis: attentional and perceptual ERPs yoked to the presentation of the target cue (P1, P2, N2, P3) and reward learning ERPs linked to the presentation of performance feedback (reward positivity and P300).

---

[1] In this thesis, the P300 is used as a measure of both perceptual processing and reward. As such, for clarity, the component elicited by target cues (indexing perception) is referred to as the P3, while the reward component elicited from feedback is referred to as the P300.

It was predicted that the amplitude of the ERP analogues of reward feedback processing would be related to the target of decisions in the gambling task. First, I hypothesized that the amplitudes of ERPs linked to feedback (the reward positivity and P300) would follow a linear relationship with the gambling targets, in that self-gambles would yield the largest reward signals, but known-other gambles would yield larger reward signals than unknown-other gambles. This hypothesis stems from previous work showing reward learning signals were larger for self-gambles than for other gambles. (Hassall et al., 2016; Krigolson et al., 2013). I predicted a similar linear pattern in the amplitudes of ERPs linked to the presentation of faces, (P1, P2, N2 and P3), based on attention and perception experiments where familiar others also enjoy cognitive advantages (Fan et al., 2013; Keyes & Dlugokencka, 2014; Sui et al., 2006, 2012). As such, I also anticipated the predicted relationship between reward signals and gambling targets would be explained by the visual processing of target faces. Specifically, I hypothesized that a linear combination of attentional ERPs (P1 and P2 amplitude) and perceptual ERPs (N2 and P3 amplitude) would predict reward learning ERPs (reward positivity and P300 amplitude). The predicted pattern of results would indicate reward processing is sensitive to different categories of self-relevance, rather than binary comparisons between ourselves and all others, and that this bias arises from the visual processing of relatedness cues.

2. 2. Methods

2. 2. 1 Participants and Materials

Fifteen undergraduate students (4 male, 11 female, 0 non-binary; mean age 22.5 years old [95%CI: 20.7 years old, 24.3 years old]) from the University of Victoria were recruited to participate in this study. Participants volunteered either through the University of Victoria's online research participation system or through paper ads posted around the University of Victoria campus. Participants were compensated with either psychology course credits or cash, respectively. Before commencing the experiment, all participants provided informed consent in agreeance with the guidelines established by the University of Victoria Human Research Ethics Board (Ethics Protocol Number: 20-1512) and followed the ethical standards specified in the 1964 Declaration of Helsinki.

Recruitment

Before participants arrived in the laboratory to complete the experiment, they were asked to provide information about themselves and someone they know. I asked that the participant receive verbal consent from the known-other before providing information about them for the study. Participants provided photographs of themself and the person they identified as a known-other, that individual's name, and the participant's relationship to that individual (ex. friend, significant other, family member, etc.). Participants were given a sample photo from a photo database used in the experiment to match their expression and position (neutral expression portrait, showing the top ¼ of the torso up to the top of the head). Finally, I asked for the email address of the participant's colleague to obtain their consent to be included in the study and reimburse them after the study.

To assess relatedness between the participant and the colleague, I used the Inclusion of the Other in the Self scale (IOtS scale; Aron et al., 1992; see Appendix B). The IOtS scale has participants select from a set of seven pairs of circles with varying overlap (the 1st pair is not overlapping at all; the 7th pair is almost entirely overlapped). Participants were asked "Which picture best describes your relationship with [the colleague's name]?", and to input the corresponding number.

2. 2. 2. Apparatus and Procedures

Participants were first prompted with the name and photograph of the colleague they identified to ensure their correctness. Next, they were asked to complete a computerized version of the IOtS scale before starting the experiment.

Following this, participants completed a modified computer version of a two-armed bandit task (Sutton & Barto, 2018, see Figure 1) in a sound-dampened room in which EEG data was recorded. During the performance of the two-armed bandit task participants completed a series of gambles by selecting between one of two coloured squares presented on a standard 19” LCD monitor. Participants were given two blocks of practice trials with no stakes and altered probabilities to familiarize themselves with the task. In a key manipulation, in this task participants were making gambles for one of three individuals with varied relatedness to them. Before each trial, a photo and name appeared indicating who the participant was gambling for. Participants were instructed that each ‘winning’ decision would yield a monetary reward for that individual. For example, if a trial began with a ‘known-other’ stimulus (ex. In *Figure 1,* the known-other is someone named “Isaiah”) and the participant won, the participant’s known-other would receive money. As such, I instructed participants that their goal was to win as much money as possible for each of the three people in the experiment (themselves, their known-other, and the stranger).

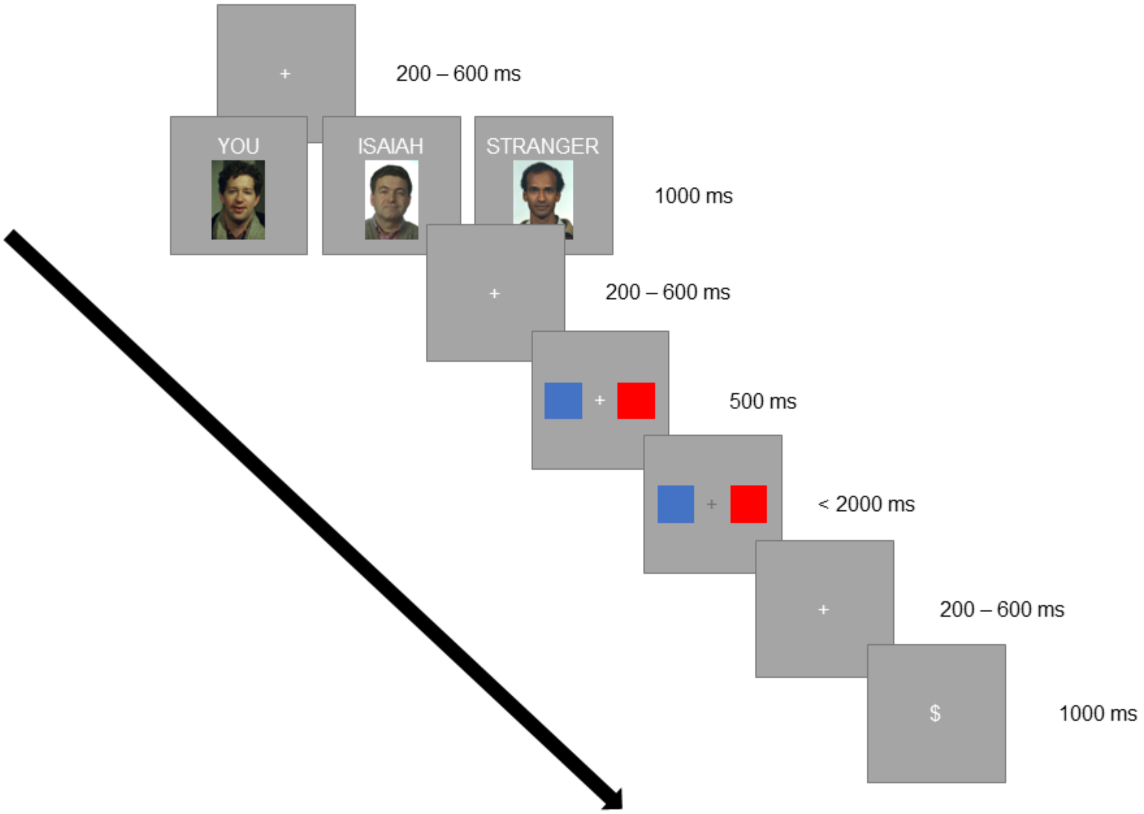

Figure 1. An example of a single trial of the modified bandit task. In this example, the outcome of the trial was a 'win', so a "$" symbol appeared at the end of the trial.

Each trial began with a fixation cross, followed by the presentation of a photo and name corresponding to one of the three gambling targets. In the 'You'-targeted trials, the photo of the participant appeared in the centre of the screen with the word 'YOU' above it. On the 'Known-Other' trials, the photo of the colleague identified by the participant appeared in the centre of the screen with the colleague's name above it (ex. 'ISAIAH'). Finally, in 'Stranger' trials, a photo of an individual was randomly selected from the Color FERET Database (National Institute of Standards and Technology (NIST), 1996), with the word 'STRANGER' above it. The target stimulus appeared for 1000ms. Next, a white fixation cross was presented, followed by two-coloured squares that appeared onscreen for an additional 500 ms. Next, a change in the colour of the fixation cross from white to grey (RGB: [128 128 128]) signalled participants to select one of the two squares using a computer gamepad. Participants had up to 2000 ms from this point to make a selection. Following the selection of a square, both of the squares disappeared and a white fixation cross appeared. After this, a feedback symbol indicating the outcome of the trial ('$' for a win, '0' for a loss) appeared in place of the fixation cross for 1000 ms. Fixation crosses that appeared before and after face presentation, and after a square was chosen were jittered from 200 – 600 ms. The next experimental trial began immediately after the offset of the feedback stimulus. Based on previous work in our laboratory (Colino et al., 2020; Hammerstrom et al., 2021; Howse et al., 2018; Krigolson et al., 2017), one square colour had a higher probability of winning than the other square colour (60% versus 10%) to ensure an approximate number of win and loss trials per participant. Importantly, these percentages also make the task learnable (i.e., one square won more often than the other so there was an optimal choice on each trial) and ensure a win/loss rate of roughly 50% to avoid frequency contamination of the N2 (see Holroyd, 2004).

Unbeknownst to participants, each win earned 10 cents. At the end of the experiment, the participant was given cash based on their performance in the 'You' trials (M = $4.88, [95% CI: $4.50, $5.26]), and their colleague was sent an electronic gift card based on the participant's performance in 'Friend' trials (M = $5.01, [$4.54, $5.47]). After the completion of the experiment, I debriefed the participants by informing them that the 'Stranger' was a fictional entity created for the study and that they would not receive money.

The experiment consisted of 15 blocks of 20 trials, with each trial corresponding to one of the three conditions. The three conditions were randomized trial-to-trial but equally distributed throughout the experiment so that there were 100 trials for each condition. The task was programmed in MATLAB (Version 9.6, Mathworks, Natick, USA) using the Psychophysics Toolbox extension (Brainard, 1997) and participants used a ResponsePixx (VPixx Technologies) button box to make their selection. The code for this experiment can be found on this project's OSF page (https://osf.io/a7h94/?view_only=3c0c412d265a4ee9b6ab7323b1e80a64).

2. 2. 3. Data Acquisition

Response time (ms), outcome (win or loss), and accuracy (selection of the square with the higher win probability) were recorded by the MATLAB (Version 9.6, Mathworks, Natick, USA) experiment script. EEG data were collected from 64 active electrodes, mounted in a 10-20 layout fitted cap (ActiCAP, Brain Products GmbH, Munich, Germany), using Brain Vision Recorder software (Version 1.21, Brain Products GmbH, Munich, Germany). All electrodes were referenced to electrode AFz during recording, and impedances were maintained below 20 kΩ at all times. EEG data were recorded at a sampling rate of 500 Hz, amplified (ActiChamp, Revision 2, Brain Products GmbH, Munich, Germany), and filtered through an antialiasing low-pass filter of 245 Hz.

2. 2. 4. Data Analysis

2.2.4.1. Behavioural Data analysis

For all three conditions, I computed each participant's mean response time, outcome percentage, and accuracy. The response time was calculated as the mean time it took for participants to select one of the squares using the button box after the fixation cross turned grey, signalling when the participant was permitted to respond in each block. The outcome percentage was calculated as the percentage of all valid trials (excluding early responses or responses made with an invalid button) within a condition ('You', 'Friend', or 'Stranger') that resulted in 'win' feedback for the participant. Finally, I calculated accuracy as the proportion of trials in each condition that the participant selected the square with the higher probability of winning.

2.2.4.2. EEG Data Pre-Processing

All EEG data were processed using MATLAB (Version 9.6, Mathworks, Natick, USA), using the EEGLAB open-source toolbox (Delorme & Makeig, 2004) and custom software developed in the

Krigolson Laboratory, available here: https://github.com/neuro-tools. First, channels were visually inspected for noisy data and removed accordingly (M = 0.8 channels, [0.3, 1.3]). Continuous EEG data were then re-referenced to mastoid channels (TP9, TP10). A dual-pass phase-free Butterworth filter with a band-pass of 0.1 Hz to 30 Hz and a 60 Hz notch filter was applied to the re-referenced EEG data. To identify and remove ocular artifacts, an independent component analysis (ICA) was conducted on the filtered data to identify components associated with ocular artifacts (Delorme & Makeig, 2004). Visual examination of the ICA factor loadings, as well as cross-correlations between EEG data and ICA component activations, guided the selection of components that contained eye blinks and movements so that they could be removed. Following the removal of components that were associated with ocular artifacts, the EEG data were reconstructed from the remaining ICA components. At this point, channels that were removed at the beginning of the analysis were re-interpolated using the method of spherical splines. All data were segmented by condition and feedback outcome (win or loss) into shorter epochs spanning from -200ms to 600ms after the stimulus presentation. The segments were then baseline-corrected using the 200ms window before feedback onset. Segments were next examined for artifacts and segments of data containing a gradient larger than 10 µV/ms or segments with an absolute difference of more than 150 µV were removed which resulted in an average of 20.1% [11.6%, 29.7%] of data being lost across participants. In line with open science practices, the pre-processing script for this experiment and the raw data can be found at the OSF page for this experiment (https://osf.io/a7h94/?view_only=3c0c412d265a4ee9b6ab7323b1e80a64).

2.2.4.3. EEG Component Analysis

Following artifact rejection, ERPs were created by averaging the segments across all trials for each of the three experimental conditions (‘You’, ‘Known-Other’, or ‘Stranger’) for the face cues (presentation of gambling target) and the two gamble outcomes (win or loss) resulting in nine ERP waveforms. The ERP components of interest for this experiment are listed in *Table 1*.

Table 1

ERP Components of Interest

Elicited by:	Component	Latency (ms post-stimulus)	Topography	Example
Attentional and Perceptual Cues	P1	130-150	Lateral, Occipital	(Handy et al., 2001)
	P2	180-240	Midline, Occipital/Parietal	(Krigolson et al., 2013)
	N2	270-370	Midline, Central	(Fan et al., 2013)
	P3	400-500	Midline, Occipital/Parietal	(Turk, van Bussel, Brebner, et al., 2011)
Reward Feedback	Reward Positivity	225-325	Midline, Central	(Hassall et al., 2016)
	P300	300-600	Midline, Central/Parietal	(Hammerstrom et al., 2021)

To provide further evidence of the topography and timing of the aforementioned components, I submitted the ERP data to a tempo-spatial Principal Components Analysis (tsPCA: see Dien, 2012; Dien et al., 2003; Foti et al., 2009) using custom MATLAB code and the EP Toolkit (Dien, 2010). The temporal data were submitted to a PCA with an Infomax rotation. Visual inspection of the resulting temporal factors guided the selection of a factor with latencies matching our ERPs of interest. The data for these temporal factors were then reshaped with topography (channel) as the independent variable and were submitted to a spatial PCA with Promax rotation. The channel where the resulting spatial loadings were maximal were selected as channels for individual ERP analysis.

For attentional and perceptual ERPs, I examined components evoked by the face cue. To ensure I did not bias component timings based on conditional effects, I collapsed face waveforms across conditions to find component timings. I visually inspected the collapsed waveforms for peak timings consistent with the P1, P2, N2, and P3 at the aforementioned channels identified by tsPCA, and found the maximum amplitude in a 50 ms window centred on that time. These components were then quantified as the mean ERP amplitude ± 25 ms from the timing of that maximal peak (P1:112 ms, P2: 238 ms, N2: 284 ms, P3: 398 ms) in conditional waveforms.

For the reward positivity and feedback-evoked P300, I examined components locked to feedback onset. To ensure I did not bias our component timing based on conditional effects, I used the overall difference (win-loss) waveforms to find our component timing and calculated ERP amplitudes based on their respective channels. Overall difference waveforms for each participant were created by subtracting the average loss waveform from the average win waveform collapsed across both conditions ([you win + known-other win + stranger win] – [you loss + known-other loss + stranger loss]). Finally, grand average condition and difference ERPs were generated by averaging the respective individual ERP waveforms.

I found that Channel FCz was maximal at 229 ms in the overall difference waveform, so the reward positivity was then quantified for each participant and condition from this latency. In a large-sample study involving the two-armed bandit task, C. C. Williams and colleagues (2021) found that, when using the mean-peak detection method, a window size of 92 ms is the best practice for quantifying the reward positivity. As such, I quantified the reward positivity as the mean component amplitude at channel FCz ± 46 ms of 229 ms post-stimulus.

I also found that Channel Pz was maximal at 251 ms, which is not in line with typical P300 findings. As such, I followed similar processes to face cue analysis and assessed tsPCA analysis linked to feedback to determine if a component consistent with the P300, a positive deflection 300-400ms after stimulus presentation (Patel & Azzam, 2005), was present. This analysis yielded a component with similar topography and timing as the P300, and thus I used a 92 ms window centred on its peak to quantify the P300. The limitations of this approach are discussed in *Chapter 3. 2. Limitations*.

2.2.4.4. Statistical Analysis

I submitted our behavioural measures (response time, outcome percentage, and accuracy) to one-way (condition: 'You', 'Known-Other', Stranger') repeated-measures ANOVAs.

Similarly, I compared the ERP components of interest across conditions by submitting mean component amplitudes to one-way repeated-measures ANOVAs. Mauchly's Test for Sphericity was computed for all ANOVAs, and if sphericity was violated Greenhouse-Geisser values are reported. Further, I examined planned comparisons between 'You' and 'Known-Other' trials and 'Known-Other' and 'Stranger' trials with pairwise t-tests. ANOVAs were analysed post-hoc with Bonferroni corrections to compare the 'You' and 'Stranger' trials. Next, to determine if the reward positivity scales with relatedness, I compared component amplitudes in the 'known-other' condition and to the IOtS score the participant gave for their friend at the beginning of the experiment with Pearson's R correlation.

Finally, I submitted condition face-locked ERP amplitudes (P1, P2, N2, and P3) to stepwise linear mixed-effects models. This approach starts by using one of the face-locked ERPs as a predictor variable in a linear regression model predicting a given feedback-locked ERP, and an Akaike Information Criteria (AIC) value is generated indicating the goodness of fit of that model. Next, the other predictors are added one at a time and the AIC values of the resultant models are compared. Predictors will remain in the model if the AIC of the current model is lower than the previous iteration (indicating a better fit). At each step where there are multiple predictors in the model, the next model will remove predictors and test the resulting models as well. Once all combinations are tested, the result is a combination of predictors with the best-fitting model for the outcome variable. All statistical operations were conducted in R via R Studio (R Core Team, 2020), the source code for this can be found on the OSF page for this experiment (https://osf.io/a7h94/?view_only=3c0c412d265a4ee9b6ab7323b1e80a64).

2. 3. Results

2. 3. 1. Behavioral Results

IoTS Scale. Results from the IoTS scale presented to participants, separated by relationship category, are presented in Table 2.

Table 2

Descriptive Statistics for IoTS scale scores

	M	SD
Friend	4.25	1.14
Significant Other	6.00	1.28
Family Member	6.00	1.10
All	5.53	1.38

Performance. Participants selected the square with a higher probability of providing a 'Win' an average of 74.8% [68.2%, 81.5%] of all experimental trials. Further, I found that participants were selecting the higher probability square more often than chance, $t(14) = 7.97$, = $p < .001$, indicating that they did learn the task. A one-way repeated measures ANOVAs with three levels (condition: You, Known-Other, Stranger) were applied to all behavioural measures.

Response Time. Mauchly's test indicated the assumption of sphericity had been violated for response time, so Greenhouse-Geisser corrected tests were reported ($\epsilon = 0.63$). Response time did not differ between conditions, $F(1.26, 17.64) = 0.07$, $p > .05$, $\eta_g^2 = 0.00$ (Figure 2).

Outcome. The repeated measures ANOVA examining outcome rate revealed a main effect of condition, $F(2, 28) = 4.24$, $p < .01$, $\eta_g^2 = 0.09$ (Figure 2, middle). Planned comparisons demonstrated no difference between 'You' and 'Known-Other' trials, $M_D = 1.27\%$ [-3.13%, 5.66%], $t(14) = 0.62$, $p > .05$, but a difference between 'Known-Other' and 'Stranger' trials, $M_D = 6.60\%$ [1.71%, 11.49%], $t(14) =$

2.89, $p < .05$. A Bonferroni post-hoc test revealed no differences in accuracy between 'You' and 'Stranger' conditions, $p = .239$.

Accuracy. Mauchly's test indicated the assumption of sphericity had been violated for accuracy data, so Greenhouse-Geisser corrected tests were reported ($\epsilon = 0.60$). Finally, the repeated measures ANOVA examining accuracy revealed a main effect of condition, $F(1.20, 16.8) = 7.24$, $p < .01$, $\eta_g^2 = 0.15$ (Figure 2, right). Planned comparisons demonstrated no difference between 'You' and 'Known-Other' trials, $M_D = 2.33\%$ [-1.38%, 6.05%], $t(14) = 1.35$, $p > .05$, but a difference between 'Known-Other' and 'Stranger' trials, $M_D = 13.53\%$ [4.59%, 22.48%], $t(14) = 3.24$, $p < .01$. Again, Bonferroni post-hoc tests revealed no difference between 'You' and 'Stranger' conditions, $p = .104$.

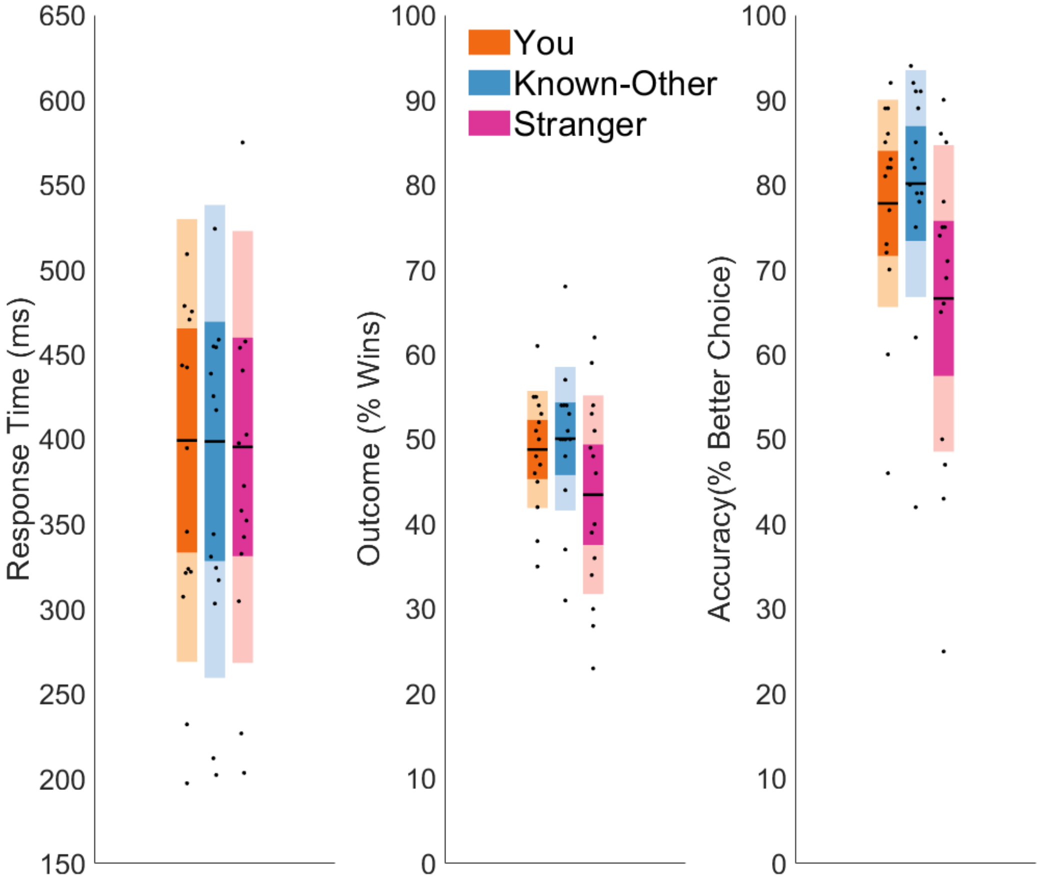

Figure 2. Behavioural performance outcomes for the three conditions. The black lines correspond to the mean, and inner and outer boxes correspond to one standard deviation and 95% within-subjects confidence intervals, respectively.

2. 3. 2. ERP Results

2.3.2.1. Target-Locked ERPs

tsPCA analysis revealed factors matching the timing and topography of the aforementioned components linked to target cues (Figure 3). Spatial loadings were analyzed for maximal activations, which guided channel selection for quantifying these ERPs (P1: Oz, P2: Pz, N2: Cz, P3:Pz).

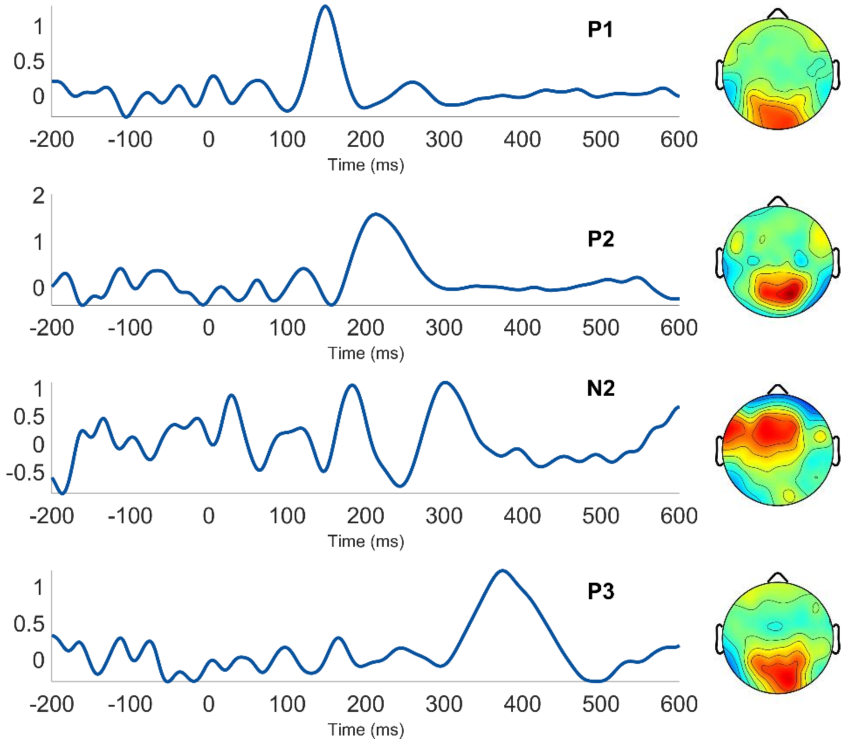

Figure 3. tsPCA factors for components matching the timing and topography of the aforementioned ERPs yoked to target cues.

Component analysis linked to the condition stimuli yielded components similar in timing and topographies as the P1, N2, P2, and P3 (Figure 4, Figure 5). The mean component amplitudes were found at the channels identified by tsPCA and are plotted in Figure 5.

P1. A repeated-measures ANOVA on P1 amplitude at channel Oz revealed no effect of condition, $F(2, 28) = 3.20$, $p > .05$, $\eta_g^2 = 0.02$.

P2. Similar analysis revealed an effect of condition on P2 amplitude at channel Pz, $F(2,28) = 3.86$, $p < 0.05$, $\eta_g^2 = 0.04$. Planned comparisons demonstrated no difference in P2 amplitude between ‘You’ and ‘Known-Other’ trials, M_D = 1.05 µV [-0.73 µV, 2.84 µV], $t(14) = 1.27$, $p > .05$, or between ‘Known-Other’ and ‘Stranger’ trials, M_D = 1.36 µV [-0.20 µV, 2.92 µV], $t(14) = 1.86$, $p > .05$. A Bonferroni post-hoc test revealed no difference in P2 amplitude between ‘You’ and ‘Stranger’ conditions, $p = .10$.

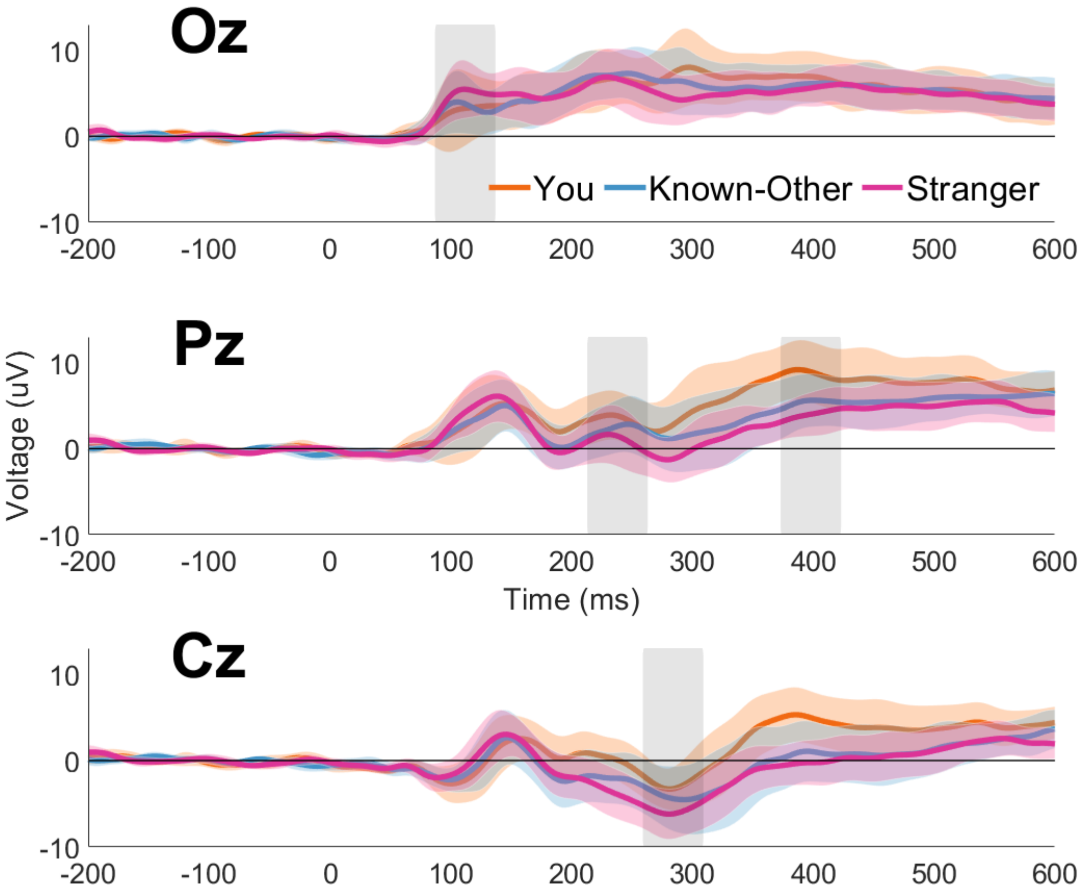

Figure 4. Grand average waveforms yoked to conditional stimuli for the P1 (**Oz**), P2 (**Pz**), N2 (**Cz**) and P3 (**Pz**). The grey rectangles represent the time ranges used to quantify ERPs, and waveform shading reflects 95% within-subjects confidence intervals.

N2. The repeated-measures ANOVA revealed an effect of condition on N2 amplitude at channel Cz, $F(2, 28) = 3.77$, $p < .05$, $\eta_g^2 = 0.05$. Planned comparisons demonstrated no difference in N2 amplitude between ‘You’ and ‘Known-Other’ trials, $M_D = 1.53$ μV [-0.92 μV, 3.97 μV], $t(14) = 1.34$, $p > .05$, or between ‘Known-Other’ and ‘Stranger’ trials, $M_D = 1.60$ μV [-1.36 μV, 4.55 μV], $t(14) = 1.16$, $p > .05$. A Bonferroni post-hoc test revealed a difference in N2 amplitude between ‘You’ and ‘Stranger’ conditions, $p = .0062$.

P3. The repeated-measures ANOVA revealed an effect of condition on P3 amplitude at channel Pz, $F(2, 28) = 14.50$, $p < .001$, $\eta_g^2 = 0.13$. Planned comparisons demonstrated a larger P3 amplitude for ‘You’ trials than ‘Known-Other’ trials, $M_D = 3.33$ μV [1.55 μV, 5.10 μV], $t(14) = 4.02$, $p < .01$, but no difference between ‘Known-Other’ and ‘Stranger’ trials, $M_D = 1.38$ μV [-0.53 μV, 3.29 μV], $t(14) = 1.55$, $p > .05$. A Bonferroni post-hoc test revealed a difference in P3 amplitude between ‘You’ and ‘Stranger’ conditions, *p* = .00078.

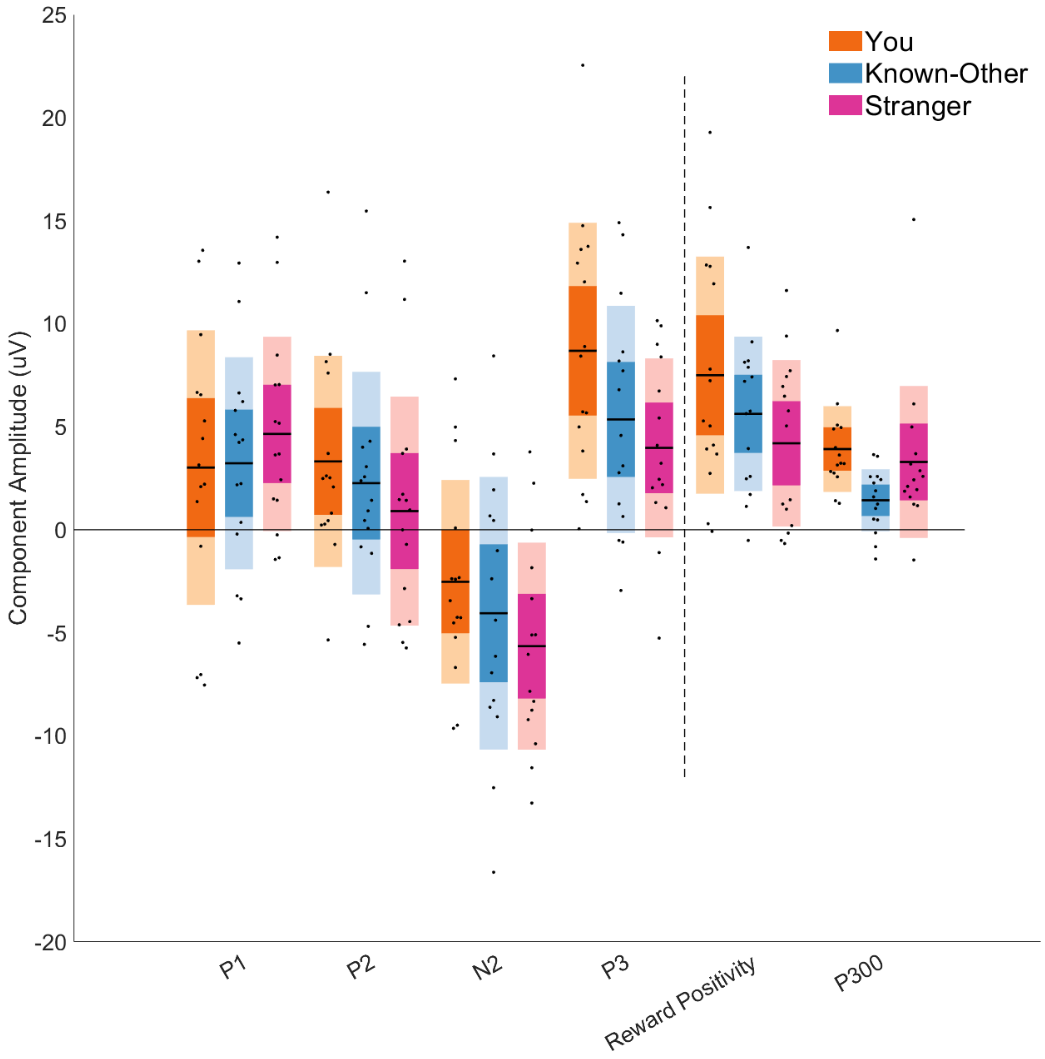

Figure 5. ERP component amplitudes for the three experimental conditions. The black lines correspond to the mean, and inner and outer boxes correspond to one standard deviation and 95% confidence intervals, respectively. The dotted line separates target-locked ERPs from feedback-locked ERPs.

2.3.2.2. Feedback-Locked ERPs

tsPCA analysis yoked to feedback stimuli revealed factors matching the timing and topography of the reward positivity and the P300 (Figure 6).

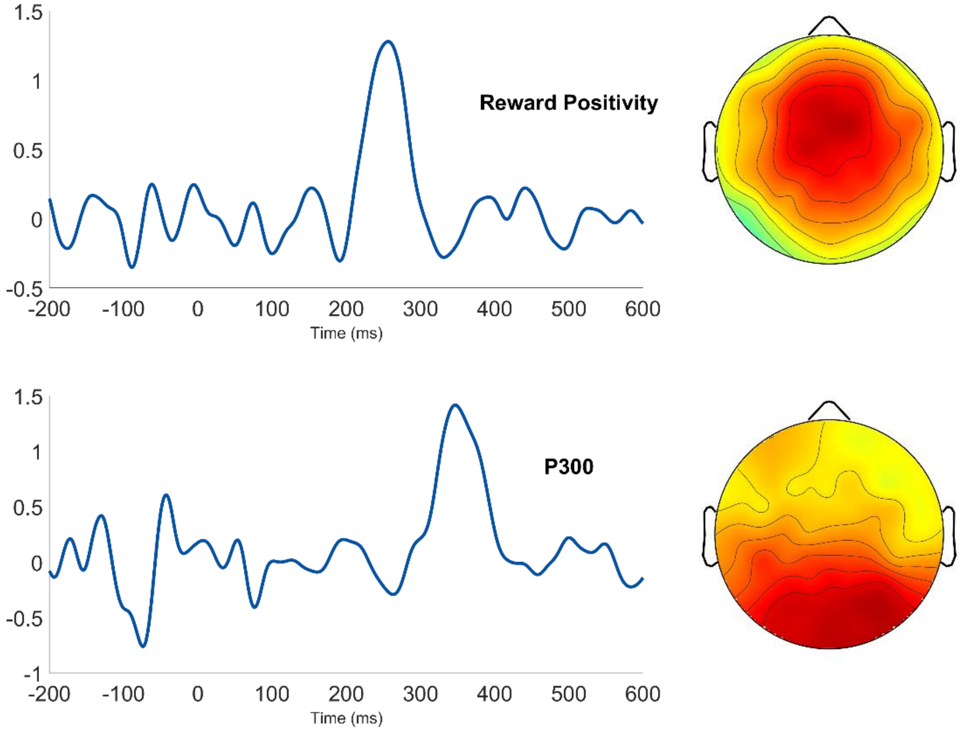

Figure 6. tsPCA factors for components matching the timing and topography of the aforementioned ERPs yoked to feedback.

Reward positivity. ERP analysis yoked to feedback stimuli yielded ERP components similar in timing and topography to the reward positivity (Conditional Waveforms: Figure 7, Difference Waves: Figure 9). Statistical analysis revealed an effect of condition, $F(2, 28) = 6.91$, $p < .001$, $\eta_g^2 = 0.08$ (Figure 5). Planned comparisons demonstrated no difference in reward positivity amplitude between ‘You’ and ‘Known-Other’ trials, MD = 1.87 µV [-0.12 µV, 3.87 µV], t(14) = 2.01, p > .05, or between ‘Known-Other’ and ‘Stranger’ trials, MD = 1.43 µV [-0.43 µV, 3.28 µV], t(14) = 1.65, p > .05. A Bonferroni post-hoc test revealed reward positivity amplitude was larger for ‘You’ trials than ‘Stranger’, p = .0062. Reward positivity amplitude in the “Known-Other” condition did not correlate with relatedness score on the IOtS scale, $t(13) = 1.45$, $r = 0.37$, $p > .05$.

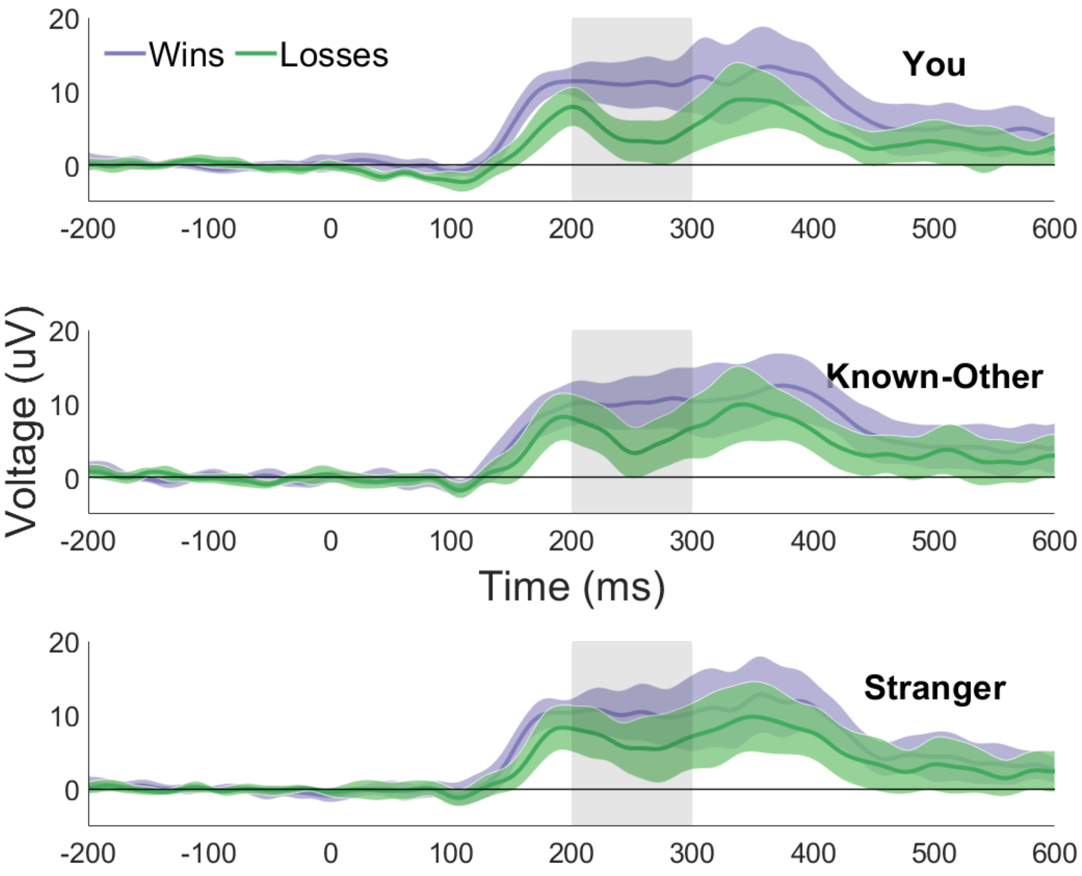

Figure 7. Feedback-locked Win and Loss waveforms for the three experimental conditions at channel FCz. The rectangles represent the ranges used to quantify the reward positivity, and waveform shading reflects 95% within-subjects confidence intervals.

P300. ERP analysis yoked to feedback stimuli yielded ERP components similar in timing and topography to the P300 (Conditional Waveforms: Figure 8, Difference Waves: Figure 9). Mauchly's test indicated the assumption of sphericity had been violated for P300 amplitude, so Greenhouse-Geisser corrected tests were reported ($\epsilon = 0.62$). The repeated-measures ANOVA revealed an effect of condition on P300 amplitude at channel Pz, $F(1.24, 17.36) = 4.78$, $p < .5$, $\eta_g^2 = 0.15$. Planned comparisons demonstrated a larger P300 amplitude for 'You' trials than 'Known-Other' trials, $M_D = 2.49$ µV [1.11 µV, 3.87 µV], $t(14) = 3.86$, $p < .01$, but no difference between 'Known-Other' and 'Stranger' trials, M_D = -

1.86 µV [-4.26 µV, 0.54 µV], *t*(14) = -1.66, *p* > .05. A Bonferroni post-hoc test revealed no difference in P300 amplitude between ‘You’ and ‘Stranger’ conditions, *p* = 1.00.

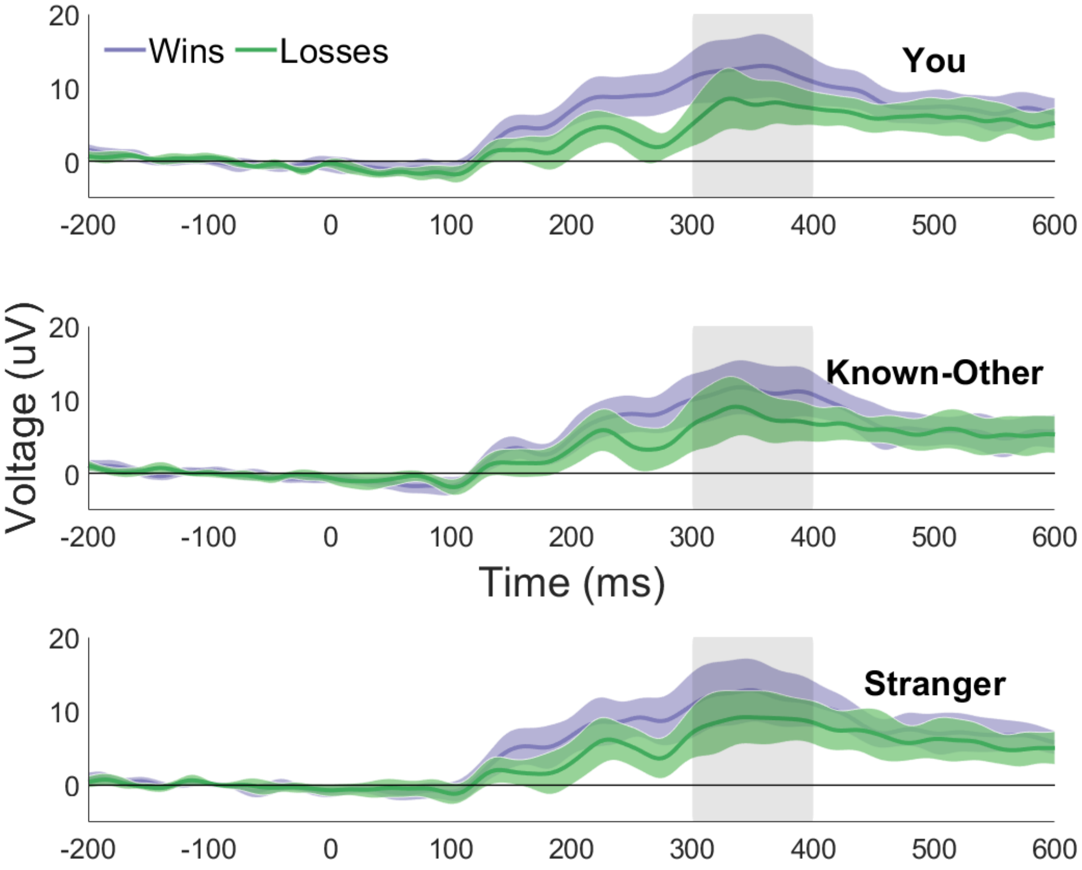

Figure 8. Feedback-locked Win and Loss waveforms for the three experimental conditions at channel Pz. The rectangles represent the time range used to quantify the P300, and waveform shading reflects 95% within-subjects confidence intervals.

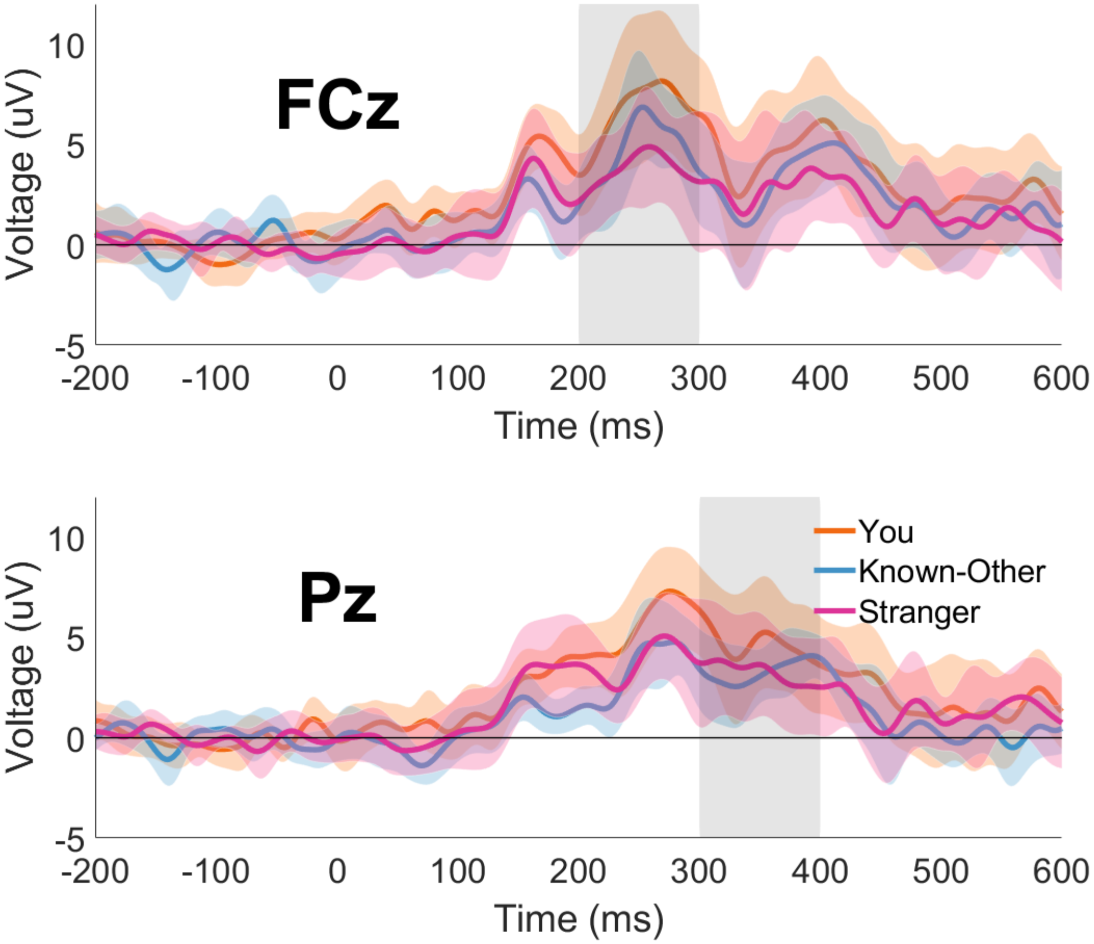

Figure 9. Feedback-locked difference waves (Win-Loss) for the three experimental conditions. The rectangles represent the ranges used to quantify the reward positivity and P300, and waveform shading reflects 95% within-subjects confidence intervals.

Pearson r correlations were computed between the target- and feedback-locked ERPs (Figure 10). Absolute significant correlations ranged from 0.53 to 0.76 representing medium to large strength relationships between ERP features. To examine how linear combinations of our ERP and EEG features predicted reward processing I used AIC-based stepwise linear multiple regression. The results of this analysis revealed models that predicted perceived reward positivity amplitude, $F(1, 43) = 5.00$, $p < .05$ ($r^2 = 0.08$, Table 3), but not P300 amplitude, $F(1, 43) = 2.89$, $p > .05$ ($r^2 = 0.04$, Table 3) .

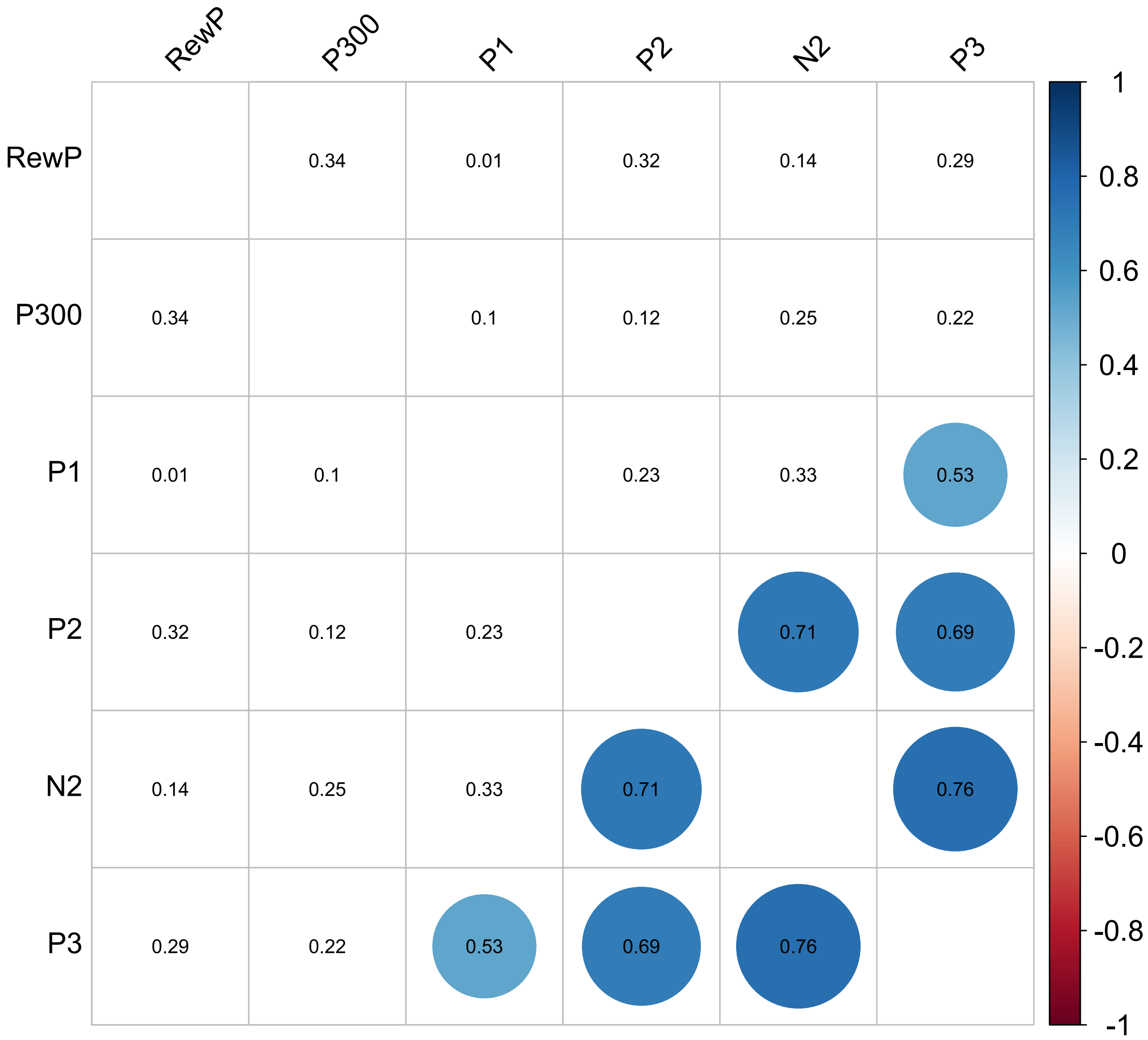

Figure 10. Correlation matrix for target- and feedback-locked ERP quantities. Correlations that did not reach significance (p<0.01) are presented without circles.

Table 3

Coefficients from stepwise linear regression models predicting reward component amplitudes.

Component	Predictor *F*(3, 41)	*B*	*B* (error)	*t*
Reward Positivity	Intercept	5.16	0.13	7.125***
	P2	0.28	0.12	2.23*
P300	Intercept	3.38	0.50	6.78***
	N2	0.12	0.07	1.70

Note. $p<.05$*, $p<.01$**, $p<.001$***

2. 4. Discussion

The current study aimed to investigate the attentional, perceptual, and reward-learning correlates of gambling on a scale of relatedness. Behaviourally, participants performed better when gambling for themselves and friends compared to others, evident from differences in the proportion of trials where the correct square was chosen. The amplitude of the P3 elicited by face cues demonstrated differences between one's self and strangers, as well as between their known-other and strangers. While reward positivity amplitude was only different between self and stranger gambles, the P300 elicited by feedback was larger for self and known-other gambles. Finally, the amplitudes of these reward components were successfully predicted by the P2 and N2 elicited by face cues, respectively, indicating reward signals are related to attentional and perceptual cues of relatedness. My results suggest that, during gambling, reward systems are sensitive to different non-self targets but still show bias towards the self. In summary, these results are in line with a large body of work on behavioural and neural indices of relatedness.

Previous work has demonstrated differences in attentional ERP components between self and other cues. Specifically, Turk et al. (2011) demonstrated reduced P1 for self-trials compared to other trials in an ownership task. Their results suggest an initial narrowing of attention (Handy et al., 2001) to upcoming stimuli that are related to the participant (Turk, van Bussel, Brebner, et al., 2011). Interestingly, I did not replicate this pattern in P1 amplitude. The paradigm in Turk & colleagues' (2011) second experiment introduced a probe before and after ownership cues, which was the stimuli P1 analysis was linked to, so that researchers could examine differences in spatial attention. Given that the current experiment lacked such probes, this is a likely reason I did not replicate their findings. One alternative explanation is the stimuli employed are more complex (i.e., processing a face and name instead of a coloured square), and the P1 can not capture that complexity. This is in line with other name-reading experiments where the P1 is either not present or not examined (Fan et al., 2013).

Additionally, previous work has identified ERP components are sensitive to self-relevance in attention and reward contexts. Specifically, P2 amplitude was larger for self-trials in reward-learning (Krigolson et al., 2013) and name-recognition tasks (Fan et al., 2013). In the former example, the P2 was elicited by cues representing who a subsequent reward would belong to (i.e., self or unknown other), indicating that the trial had reward value (Krigolson et al., 2013). Interestingly, while the RM ANOVA on P2 amplitude revealed an effect of condition, planned comparisons and post hoc analyses did not identify any differences between conditions. This result indicates that attention to the face cue in this experiment is similar regardless of who the participant is gambling for. A potential explanation is that in the current experiment each trial was important for learning which of the squares was rewarding, while in Krigolson and colleagues (2013) the task was unlearnable, and as such each trial warranted attention. Further work should consider how the context of learning may affect self-relevant attention, and as such attentional ERP components.

Later components yoked to target cues representative of perceptual processes were also sensitive to condition. Previously, a larger P3 in response to self-relevant cues has been attributed to top-down biasing of processing for stimuli with higher personal "emotional value" (Campanella, Gaspard, et al., 2002; Campanella, Quinet, et al., 2002; Carretié et al., 2004; Fan et al., 2013; Turk, van Bussel, Brebner, et al., 2011). Here, it seems the benefit of emotional value is not given to related others, as the amplitude of the P3 was similar in ‘Known-Other’ and ‘Stranger’ conditions. One potential explanation is that, in the context of gambling, the potential for self-rewards overrides processing for our known-others. Additionally, differences in the amplitude of the N2 were only apparent when comparing ‘You’ and ‘Stranger’ trials, supporting the idea that self-relevance has an ‘all-or-none’ (i.e., self or not-self) effect on perception. Taken together, my results suggest that, in the context of reward learning, perceptual components may only reflect the bias of self.

Present results further demonstrate reward learning systems are sensitive to self-relevant outcomes. Previously, the medial-frontal learning system that computes reward prediction errors (Holroyd et al., 2004a; C. C. Williams et al., 2017b) has been shown to selectively process self-related reward information (Hassall et al., 2016; Krigolson et al., 2013). My results replicate these findings, as the reward positivity exhibited increased amplitudes for self-gambles compared to stranger gambles. Importantly, ‘Stranger’ trials still elicited reward signals, whereas they did not in Krigolson et al. (2013). One potential explanation is that the present study introduced descriptive target information for the unknown other (i.e., A face), while Krigolson and colleagues (2013) used simple coloured borders to indicate trials were for the ‘Other’. This could lead to participants in Krigolson et al. (2013) not believing that the ‘Other’ was a real individual, and thus processing reward information for them is impractical. However, Hassall and colleagues (2016) used a similar paradigm and found that ‘Other’ trials still elicited a reward positivity, albeit a diminished one, suggesting the outcomes of other gambles are still processed. While the current results favour the latter example, there is no evidence to confirm that the delivery of the ‘Stranger’/’Known-Other’ information affects how those gambles are processed. Further research should explore different methods of presenting gambling targets to determine if this aspect of the paradigm could bias reward processing.

While there is evidence of a P300 elicited by reward feedback, there is no current work identifying this component in the context of self-relevance. Here, results showed the P300 was similar between ‘You’ and ‘Stranger’ trials, yet was largest for ‘Known-Other’ trials, which is inconsistent with existing accounts of how self-relevance affects reward processing (Hassall et al., 2016; Krigolson et al., 2013). One potential explanation is that this modified paradigm does not elicit a P300 in response to feedback, and the component as quantified here represents something else. While the timing of the P300 component identified with tsPCA analysis is consistent with previous examples, the topography was not as similar (Hammerstrom et al., 2021; Leng & Zhou, 2010; Wu & Zhou, 2009; Zhou et al., 2010). Nonetheless, while it is difficult to draw conclusions about the feedback-P300 from the current results, future work should still consider it in the context of reward and self-relevance.

A secondary goal of this experiment was to determine if differences in reward processing can be explained by how we perceive who we’re gambling for. Here, I have provided evidence that the reward positivity elicited by gambling on a scale of relatedness can be predicted by preceding attentional and perceptual components. As previously mentioned, ownership cues in a reward learning task elicited P2 components sensitive to the gambling condition (Krigolson et al., 2013). In line with this, our linear regression model predicting reward positivity amplitude isolated the P2 as a predictor, indicating it is the first signal of a cues reward value for the self. The output of this model suggests that as soon as determine the saliency of an upcoming gamble, is indexed by the P2, reward processing is primed for feedback evaluation accordingly (Fan et al., 2013; Krigolson et al., 2013).

Some have suggested that our sense of self is inherently rewarding, and as such self-relevance effects on cognition are a result of reward-seeking behaviour. Specifically, Northoff & Hayes (2011) suggested that self-relevance and reward information are processed in parallel in similar networks. Of particular interest is their suggestion that self-relevant processing involves a comparison of actual and anticipated self-relevance, similar to reward prediction errors. Given the reward positivity is thought to reflect a reward prediction error (Holroyd et al., 2004b; Holroyd & Coles, 2002; Holroyd & Krigolson, 2007; C. C. Williams et al., 2019), this model would suggest that its’ amplitude would also reflect the magnitude of difference in actual self-relevance (i.e. reflect that ‘Stranger’ trials are less self-relevant than

‘Known-Other’ trials). As the current results do not support this, its possible that this component of Northoff and Hayes's (2011) model is flawed. Importantly, the parallel processing model specifies that differences in self-relevant and reward processing are apparent on the time scale of the processing; attention to a self-related cue may occur before the computation of a reward prediction error which may occur before the processing of that reward’s self-relevance. While the current paradigm was not specifically designed to test this model, the temporal specificity involved in ERP experiments, and their widespread use for examining reward learning, indicate they can be employed to further examine the parallel processing theory.

In summary, it is clear that reward processing is sensitive to who we are gambling for. Specifically, in the present experiment, I demonstrated that the amplitude of the reward positivity and P300 are sensitive to different non-self targets. We are not completely selfish in the processing of reward, as known others still yield larger reward signals than their unknown counterparts. Further, it seems that this reward sensitivity is in part due to how we process visual information that cues the self-relevance of rewards. The implications of these findings are discussed further in *Chapter 3. 3. Implications for Future Research.*

Chapter 3. General Discussion

3. 1 Summary

The goal of this experiment was to determine whether reward learning is sensitive to a continuum of self-relevance. The amplitudes of the reward positivity and P300 ERP components were sensitive to the target of rewards, providing evidence in favour of this idea. Further, changes in the amplitude of the reward positivity were explained by the amplitudes of the P2, N2, and P3 elicited by the cues of gambling targets. In summary, this pattern of results indicates our reward learning systems are biased by self-relevance, and that bias is related to early cognitive processing of visual stimuli.

3. 2. Limitations

The current research can not be considered without a discussion of methodological considerations. In this paradigm, participants viewed photos of themselves and their known others that they submitted before the experiment. This diverges from similar studies, where photo stimuli were taken from videos (Keyes et al., 2010) or still pictures (Keyes & Dlugokencka, 2014) of the participant taken by researchers. Our participants were instructed to match the position, expression, and luminance of a sample photo from the database used for our 'Stranger' photos to the best of their ability, although this led to some deviations in characteristics (ex. Background was brighter in the 'You' photo than the 'Friend') that could affect ERP results. However, I argue that the face-name combinations in the current experiment acted more as repetitive categorical stimuli rather than the target of visual scrutiny, and the aforementioned differences in photos should not have affected results after repeated viewing throughout the experiment. This idea is supported by results from the ERPs yoked to these stimuli, which largely matched the patterns of results from experiments where categorical target stimuli (ex. Using the same-coloured border to indicate 'Self' trials throughout the experiment, Fan et al., 2013; Krigolson et al., 2013).

Another potential limitation of this experiment is the approach to ERP component analysis. Quantifying ERP components is a process ripe for bias and counterfeit effects based on the time range you quantify them within, and the method used to do so (Luck, 2014; Luck & Gaspelin, 2017). Thankfully, research from our laboratory used a large sample dataset to produce a rigorous method for quantifying the reward positivity in a two-armed bandit task, which was employed here (C. C. Williams et al., 2021). However, to the best of our knowledge, such research does not exist for any of the other ERP components of interest in this experiment. To circumvent this, I employed tsPCA to verify the existence of components that matched the timing and topography of the relevant attentional and perceptual components (Dien, 2012; Foti et al., 2009).

Analysis of ERP waveforms yoked to reward feedback did not elicit the expected P300 response. Typically, the P300 yoked to reward is maximal at channel Pz, 300ms or later after feedback presentation (Hammerstrom et al., 2021; Jarchi et al., 2011; Patel & Azzam, 2005; Polich, 2003, 2007). Here, channel Pz was maximal at 251 ms, suggesting that this is representative of a different ERP component. However, tsPCA did yield a component consistent with the P300.

A significant limitation of the current research is the sample size. Due to time constraints and logistical challenges (i.e., the COVID-19 pandemic), the current sample is 15 undergraduate student participants. In the context of ERP experiments examining reward processing, it has been suggested a sample size of at least 18 participants is needed to detect a reward positivity with a power of 0.95, let alone make within-groups comparisons (C. C. Williams et al., 2021). Previous work from my lab has adopted a policy of testing participants until 30 sets of sufficient data (i.e., EEG data with an artifact rejection rate of less than 40%, and matching behavioural data) are obtained, corresponding to a power of 0.99 (Hammerstrom et al., 2021).

Finally, the current investigation may not truly capture a scale of relatedness. Here, I had participants provide the name and likeness of someone close to them, namely a significant other, family member, or good friend. Consequently, results from our measure of relatedness, the IOtS scale (Aron et al., 1992), were heavily biased towards the higher end of scores. This highlights that a holistic view of relatedness requires including somewhat related individuals, who would be represented by lower IOtS scores than the aforementioned individuals, but would still be more related than strangers. As such, any conclusions drawn about relatedness from the current experiment may be limited.

3. 3. Implications for Future Research

Here, I showed that reward systems in the brain are not only sensitive to a dichotomy of self-relevant or not, but also to somewhat relevant stimuli. These findings are in line with previous work focused on other cognitive functions, namely attention and perception, that showed we differentially process information with a range of self-relevance (Fan et al., 2013; Keyes et al., 2010; Sui et al., 2006, 2012, 2014). The paradigm I implemented to explore this idea can be applied in the future to further understand how self-relevance affects reward processing. For example, how might reward processing change if more targets on the scale of self-relevance are used (ex. Someone somewhat known but less so than a friend)? Another interesting idea comes from the existing work on reward and self-relevance, where altering reward value has had varying effects on processing (Hassall et al., 2016; Sui & Humphreys, 2015a, 2015b). Future work can implement a nearly identical task as the one used here, but additionally, alter reward values based on who receives them.

The secondary goal of the current work represents a potential new line of self-reference research. While it has been determined that numerous cognitive functions are affected by self-relevance, the interactions between these functions are not well understood. Here, I demonstrated that attention and perception of relevance cues affect how we process that information via stepwise, multiple linear regression (see *Chapter 2.3.2.2. Feedback-Locked ERPs*). Future work could implement paradigms similar to the one used here, which allowed for measurement of attention and perception (ERPs yoked to target cues), as well as reward processing (ERPs yoked to feedback). This approach could help us better understand the associations between different cognitive functions affected by self-reference.

3. 4. Conclusions

The principal goal of this experiment was to determine if reward learning on a scale of relatedness could be explained by how we attend to and perceive relatedness information. The evidence presented here suggests that how we process rewards for ourselves and others is related to how we process the information that indicates who that reward is for. The conclusions drawn here stem from growing work on self-reference and reward (Hassall et al., 2016; Krigolson et al., 2013), attention (Humphreys & Sui, 2015; Turk, van Bussel, Brebner, et al., 2011), and perception (Sui et al., 2012, 2014). These findings can be applied to multiple facets of self-reference research, and encourage further work exploring interactions between cognitive processes as they pertain to ourselves and those around us.

References

Alexopoulos, T., Muller, D., Ric, F., & Marendaz, C. (2012). I, me, mine: Automatic attentional capture by self-related stimuli. *European Journal of Social Psychology*, *42*(6), 770–779. https://doi.org/10.1002/EJSP.1882

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the Structure of Interpersonal Closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Balota, D. A., Yap, M. J., Cortese, M. J., & Watson, J. M. (2008). Beyond mean response latency: Response time distributional analyses of semantic priming. *Journal of Memory and Language*, *59*(4), 495–523. https://doi.org/10.1016/J.JML.2007.10.004

Bargh, J. A. (1982). Attention and automaticity in the processing of self-relevant information. *Journal of Personality and Social Psychology*, *43*(3), 425–436. https://doi.org/10.1037/0022-3514.43.3.425

Bargh, J. A., & Pratto, F. (1986). Individual construct accessibility and perceptual selection. *Journal of Experimental Social Psychology*, *22*(4), 293–311. https://doi.org/10.1016/0022-1031(86)90016-8

Brainard, D. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. http://faculty.washington.edu/jpalmer/files/forMJ/NewPapersForMJ/brainard1997PsychToolbox.pdf

Brier, M. R., Maguire, M. J., Tillman, G. D., Hart, J., & Kraut, M. A. (2008). Event-related potentials in semantic memory retrieval. *Journal of the International Neuropsychological Society*, *14*(5), 815–822. https://doi.org/10.1017/S135561770808096X

Campanella, S., Gaspard, C., Debatisse, D., Bruyer, R., Crommelinck, M., & Guerit, J. M. (2002). Discrimination of emotional facial expressions in a visual oddball task: an ERP study. *Biological Psychology*, *59*(3), 171–186. https://doi.org/10.1016/S0301-0511(02)00005-4

Campanella, S., Quinet, P., Bruyer, R., Crommelinck, M., & Guerit, J. M. (2002). Categorical Perception of Happiness and Fear Facial Expressions: An ERP Study. *Journal of Cognitive Neuroscience*, *14*(2), 210–227. https://doi.org/10.1162/089892902317236858

Carretié, L., Hinojosa, J. A., Martín-Loeches, M., Mercado, F., & Tapia, M. (2004). Automatic attention to emotional stimuli: Neural correlates. *Human Brain Mapping*, *22*(4), 290–299. https://doi.org/10.1002/HBM.20037

Chiu, P. H., Holmes, A. J., & Pizzagalli, D. A. (2008). Dissociable recruitment of rostral anterior cingulate and inferior frontal cortex in emotional response inhibition. *NeuroImage*, *42*(2), 988–997. https://doi.org/10.1016/J.NEUROIMAGE.2008.04.248

Cloutier, J., & Neil Macrae, C. (2008). The feeling of choosing: Self-involvement and the cognitive status of things past. *Consciousness and Cognition*, *17*(1), 125–135. https://doi.org/10.1016/J.CONCOG.2007.05.010

Cockburn, D. (2001). Descartes: The Self and the World. *An Introduction to the Philosophy of Mind*, 1–12. https://doi.org/10.1057/9780230509108_1

Cohen, M. X. (2017). Where Does EEG Come From and What Does It Mean? *Trends in Neurosciences*, *40*(4), 208–218. https://doi.org/10.1016/J.TINS.2017.02.004

Colino, F. L., Heath, M., Hassall, C. D., & Krigolson, O. E. (2020). Electroencephalographic evidence for a reinforcement learning advantage during motor skill acquisition. *Biological Psychology*, *151*, 107849. https://doi.org/10.1016/J.BIOPSYCHO.2020.107849

Conway, M. A. (2005). Memory and the self. *Journal of Memory and Language*, *53*(4), 594–628. https://doi.org/10.1016/J.JML.2005.08.005

Conway, M. A., & Dewhurst, S. A. (1995). The self and recollective experience. *Applied Cognitive Psychology*, *9*(1), 1–19. https://doi.org/10.1002/ACP.2350090102

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

Cunningham, S. J., Brady-Van Den Bos, M., Gill, L., & Turk, D. J. (2013). Survival of the selfish: Contrasting self-referential and survival-based encoding. *Consciousness and Cognition*, *22*(1), 237–244. https://doi.org/10.1016/J.CONCOG.2012.12.005

Cunningham, S. J., & Turk, D. J. (2017). Editorial: A review of self-processing biases in cognition. *Quarterly Journal of Experimental Psychology*, *70*(6), 987–995. https://doi.org/10.1080/17470218.2016.1276609

Cunningham, S. J., Turk, D. J., Macdonald, L. M., & Neil Macrae, C. (2008). Yours or mine? Ownership and memory. *Consciousness and Cognition*, *17*(1), 312–318. https://doi.org/10.1016/J.CONCOG.2007.04.003

Cunningham, S. J., van den Bos, M. B., & Turk, D. J. (2011). Exploring the effects of ownership and choice on self-memory biases. *Memory*, *19*(5), 449–461. https://doi.org/10.1080/09658211.2011.584388

Daffner, K. R., Alperin, B. R., Mott, K. K., Tusch, E. S., & Holcomb, P. J. (2015). Age-related differences in early novelty processing: Using PCA to parse the overlapping anterior P2 and N2 components. *Biological Psychology*, *105*, 83–94. https://doi.org/10.1016/J.BIOPSYCHO.2015.01.002

Damasio, A. (1999). The Feeling of what Happens: Body and Emotion in the Making of Consciousness. In *Harcourt Inc*. Harcourt, Inc. https://books.google.ca/books?hl=en&lr=&id=RSOPDHP9QekC&oi=fnd&pg=PA1&dq=Damasio,+A.+(1999)+The+Feeling+of+What+Happens:+Body+and+Emotion+in+the+Making+of+Consciousness&ots=DRi93ejr_b&sig=rbNJo8DgkeV3qqYsODidZldY_5w#v=onepage&q=Damasio%2C%20A.%20(1999)%20The%20Feeling%20of%20What%20Happens%3A%20Body%20and%20Emotion%20in%20the%20Making%20of%20Consciousness&f=false

de Greck, M., Enzi, B., Prösch, U., Gantman, A., Tempelmann, C., & Northoff, G. (2010). Decreased neuronal activity in reward circuitry of pathological gamblers during processing of personal relevant stimuli. *Human Brain Mapping*, *31*(11), 1802. https://doi.org/10.1002/HBM.20981 de Greck, M., Rotte, M., Paus, R., Moritz, D., Thiemann, R., Proesch, U., Bruer, U., Moerth, S., Tempelmann, C., Bogerts, B., & Northoff, G. (2008). Is our self based on reward? Self-relatedness recruits neural activity in the reward system. *NeuroImage*, *39*(4), 2066–2075. https://doi.org/10.1016/J.NEUROIMAGE.2007.11.006

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. https://doi.org/10.1016/j.jneumeth.2003.10.009

Dennet, D. (1993). Consciousness explained. In *Penguin UK*. https://scholar.google.com/scholar_lookup?title=Consciousness%20Explained&publication_year=1999&author=D.%20Dennett

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

Dien, J. (2012). Applying Principal Components Analysis to Event-Related Potentials: A Tutorial. *Https://Doi.Org/10.1080/87565641.2012.697503*, *37*(6), 497–517. https://doi.org/10.1080/87565641.2012.697503

Dien, J., Spencer, K. M., & Donchin, E. (2003). Localization of the event-related potential novelty response as defined by principal components analysis. *Cognitive Brain Research*, *17*(3), 637–650. https://doi.org/10.1016/S0926-6410(03)00188-5

DiQuattro, N. E., & Geng, J. J. (2011). Contextual knowledge configures attentional control networks. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *31*(49), 18026–18035. https://doi.org/10.1523/JNEUROSCI.4040-11.2011

Donchin, E. (1981). Surprise!? Surprise? *Psychophysiology*, *18*(5), 493–513. https://doi.org/10.1111/j.1469-8986.1981.tb01815.x

Donchin, E., & Coles, M. G. H. (1988). Is the P300 component a manifestation of context updating? *Behavioral and Brain Sciences*, *11*(03), 357. https://doi.org/10.1017/S0140525X00058027

Duncan-Johnson, C. C. (1981). P300 Latency: A New Metric of Information Processing. *Psychophysiology*, *18*(3), 207–215. https://doi.org/10.1111/j.1469-8986.1981.tb03020.x

Duncan-Johnson, C. C., & Donchin, E. (1977). On Quantifying Surprise: The Variation of Event-Related Potentials With Subjective Probability. *Psychophysiology*, *14*(5), 456–467. https://doi.org/10.1111/j.1469-8986.1977.tb01312.x

Dupont, S., Bouilleret, V., Hasboun, D., Semah, F., & Baulac, M. (2003). Functional anatomy of the insula: New insights from imaging. *Surgical and Radiologic Anatomy*, *25*(2), 113–119. https://doi.org/10.1007/S00276-003-0103-4/FIGURES/3

Enzi, B., de Greck, M., Prosch, U., Tempelmann, C., & Northoff, G. (2009). Is Our Self Nothing but Reward? Neuronal Overlap and Distinction between Reward and Personal Relevance and Its Relation to Human Personality. *PLoS ONE*, *4*(12). https://go.gale.com/ps/i.do?id=GALE%7CA472753242&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=19326203&p=HRCA&sw=w&userGroupName=anon%7E93ddaed8

Ersner-Hershfield, H., Wimmer, G. E., & Knutson, B. (2009). Saving for the future self: Neural measures of future self-continuity predict temporal discounting. *Social Cognitive and Affective Neuroscience*, *4*(1), 85. https://doi.org/10.1093/SCAN/NSN042

Fan, W., Chen, J., Wang, X. Y., Cai, R., Tan, Q., Chen, Y., Yang, Q., Zhang, S., Wu, Y., Yang, Z., Wang, X. A., & Zhong, Y. (2013). Electrophysiological Correlation of the Degree of Self-Reference Effect. *PLOS ONE*, *8*(12), e80289. https://doi.org/10.1371/JOURNAL.PONE.0080289

Fichte, J. G. (1794). *Ueber dem Begriff der Wissenschaftslehre oder der sogenannten Philosophie ... - Johann Gottlieb Fichte - Google Books*. https://books.google.ca/books?hl=en&lr=&id=OONZAAAAcAAJ&oi=fnd&pg=PR1&dq=fichte+1794&ots=ryPsAtC4zz&sig=apLQ646O1-CjNbjzPgu9_vukodc#v=onepage&q=fichte%201794&f=false

Folstein, J. R., & van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. In *Psychophysiology* (Vol. 45, Issue 1, pp. 152–170). Blackwell Publishing Inc. https://doi.org/10.1111/j.1469-8986.2007.00602.x

Foti, D., Hajcak, G., & Dien, J. (2009). Differentiating neural responses to emotional pictures: evidence from temporal-spatial PCA. *Psychophysiology*, *46*(3), 521–530. https://doi.org/10.1111/J.1469-8986.2009.00796.X

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

Gallagher, S. (2000). Philosophical conceptions of the self: implications for cognitive science. *Trends in Cognitive Sciences*, *4*(1), 14–21. https://doi.org/10.1016/S1364-6613(99)01417-5

Gazzaniga, M. S. (1995). *Consciousness and the cerebral hemispheres.* (pp. 1391–1400). The Cognitive Neurosciences. https://psycnet.apa.org/record/1994-98810-092

Gill, C. (2017). The Self in Greek literature. *Oxford Research Encyclopedia of Classics*. https://doi.org/10.1093/ACREFORE/9780199381135.013.8109

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

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

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

Handy, T. C., Grafton, S. T., Shroff, N. M., Ketay, S., & Gazzaniga, M. S. (2003). Graspable objects grab attention when the potential for action is recognized. *Nature Neuroscience 2003 6:4*, *6*(4), 421–427. https://doi.org/10.1038/nn1031

Handy, T. C., Soltani, M., & Mangun, G. R. (2001). Perceptual load and visuocortical processing: Event-Related Potentials Reveal Sensory-Level Selection. *Psychological Science*, *12*(3), 213–218. https://doi.org/10.1111/1467-9280.00338

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

Hassall, C. D., Holland, K., & Krigolson, O. E. (2013). What do I do now? An electroencephalographic investigation of the explore/exploit dilemma. *Neuroscience*, *228*, 361–370. https://doi.org/10.1016/j.neuroscience.2012.10.040

Hassall, C. D., Mcdonald, C. G., & Krigolson, O. E. (2019). *Ready, set, explore! Event-related potentials reveal the time-course of exploratory decisions*. https://doi.org/10.1016/j.brainres.2019.05.039

Hassall, C. D., Silver, A., Turk, D. J., & Krigolson, O. E. (2016). We are more selfish than we think: The endowment effect and reward processing within the human medial-frontal cortex. *Quarterly Journal of Experimental Psychology*, *69*(9), 1676–1686. https://doi.org/10.1080/17470218.2015.1091849

Hayhoe, M. M. (2004). Advances in relating eye movements and cognition. *Infancy*, *6*(2), 267–274. https://doi.org/10.1207/S15327078IN0602_7

Heatherton, T. F., Wyland, C. L., Macrae, C. N., Demos, K. E., Denny, B. T., & Kelley, W. M. (2006). Medial prefrontal activity differentiates self from close others. *Social Cognitive and Affective Neuroscience*, *1*(1), 18–25. https://doi.org/10.1093/SCAN/NSL001

Heinzel, A., Walter, M., Schneider, F., Rotte, M., Matthiae, C., Tempelmann, C., Heinze, H. J., Bogerts, B., & Northoff, G. (2011). Self-related processing in the sexual domain: A parametric event-related fMRI study reveals neural activity in ventral cortical midline structures. *Social Neuroscience*, *1*(1), 41–51. https://doi.org/10.1080/17470910600663137

Holroyd, C. B. (2004). A note on the oddball N200 and the feedback ERN. *Errors, Conflicts, and the Brain*, 211–218. http://www.ece.uvic.ca/~bctill/papers/learning/Holroyd_2004.pdf

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

Holroyd, C. B., & Krigolson, O. E. (2007). Reward prediction error signals associated with a modified time estimation task. *Psychophysiology*, *44*, 913–917. https://doi.org/10.1111/j.1469-8986.2007.00561.x

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G. H., & Cohen, J. D. (2004a). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience 2004 7:5*, *7*(5), 497–498. https://doi.org/10.1038/nn1238

Holroyd, C. B., Nieuwenhuis, S., Yeung, N., Nystrom, L., Mars, R. B., Coles, M. G. H., & Cohen, J. D. (2004b). Dorsal anterior cingulate cortex shows fMRI response to internal and external error signals. *Nature Neuroscience*, *7*(5), 497–498. https://doi.org/10.1038/nn1238

Holroyd, C. B., Yeung, N., Coles, M. G. H., & Cohen, J. D. (2005). A mechanism for error detection in speeded response time tasks. *Journal of Experimental Psychology: General*, *134*(2), 163–191. https://doi.org/10.1037/0096-3445.134.2.163

Howse, A. D., Hassall, C. D., Williams, C. C., Hajcak, G., & Krigolson, O. E. (2018). Alcohol hangover impacts learning and reward processing within the medial-frontal cortex. *Psychophysiology*, *55*(8), e13081. https://doi.org/10.1111/PSYP.13081

Humphrey, D. G., Kramer, A. F., & Stanny, R. R. (1994). Influence of Extended Wakefulness on Automatic and Nonautomatic Processing. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *36*(4), 652–669. https://doi.org/10.1177/001872089403600407

Humphreys, G. W., & Sui, J. (2015). Attentional control and the self: The Self-Attention Network (SAN). *Https://Doi.Org/10.1080/17588928.2015.1044427*, *7*(1–4), 5–17. https://doi.org/10.1080/17588928.2015.1044427

Jarchi, D., Sanei, S., Mohseni, H. R., & Lorist, M. M. (2011). Coupled particle filtering: A new approach for P300-based analysis of mental fatigue. *Biomedical Signal Processing and Control*, *6*(2), 175–185. https://doi.org/10.1016/J.BSPC.2010.09.001

Jenkins, A. C., & Mitchell, J. P. (2011). Medial prefrontal cortex subserves diverse forms of self-reflection. *Social Neuroscience*, *6*(3), 211–218. https://doi.org/10.1080/17470919.2010.507948

Keyes, H., Brady, N., Reilly, R. B., & Foxe, J. J. (2010). My face or yours? Event-related potential correlates of self-face processing. *Brain and Cognition*, *72*(2), 244–254. https://doi.org/10.1016/J.BANDC.2009.09.006

Keyes, H., & Dlugokencka, A. (2014). Do I Have My Attention? Speed of Processing Advantages for the Self-Face Are Not Driven by Automatic Attention Capture. *PLOS ONE*, *9*(10), e110792. https://doi.org/10.1371/JOURNAL.PONE.0110792

Kircher, T. T. J., Senior, C., Phillips, M. L., Rabe-Hesketh, S., Benson, P. J., Bullmore, E. T., Brammer, M., Simmons, A., Bartels, M., & David, A. S. (2001). Recognizing one’s own face. *Cognition*, *78*(1), B1–B15. https://doi.org/10.1016/S0010-0277(00)00104-9

Krigolson, O. E. (2018). Event-related brain potentials and the study of reward processing: Methodological considerations. *International Journal of Psychophysiology*, *132*, 175–183. https://doi.org/10.1016/J.IJPSYCHO.2017.11.007

Krigolson, O. E., Hassall, C. D., Balcom, L., & Turk, D. (2013). Perceived ownership impacts reward evaluation within medial-frontal cortex. *Cognitive, Affective and Behavioral Neuroscience*, *13*(2), 262–269. https://doi.org/10.3758/S13415-012-0144-4/FIGURES/6

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. https://doi.org/10.1162/jocn_a_00509

Krigolson, O. E., Pierce, L. J., Holroyd, C. B., & Tanaka, J. W. (2009). Learning to become an expert: Reinforcement learning and the acquisition of perceptual expertise. *Journal of Cognitive Neuroscience, 21*(9), 1834–1841. https://doi.org/10.1162/jocn.2009.21128

Krigolson, O. E., Williams, C. C., Norton, A., Hassall, C. D., & Colino, F. L. (2017). Choosing MUSE: Validation of a low-cost, portable EEG system for ERP research. *Frontiers in Neuroscience*, *11*(MAR), 109. https://doi.org/10.3389/FNINS.2017.00109/BIBTEX

Lachter, J., Forster, K. I., & Ruthruff, E. (2004). Forty-five years after broadbent (1958): Still no identification without attention. *Psychological Review*, *111*(4), 880–913. https://doi.org/10.1037/0033-295X.111.4.880

Leng, Y., & Zhou, X. (2010). Modulation of the brain activity in outcome evaluation by interpersonal relationship: An ERP study. *Neuropsychologia*, *48*(2), 448–455. https://doi.org/10.1016/j.neuropsychologia.2009.10.002

Liu, X., Powell, D. K., Wang, H., Gold, B. T., Corbly, C. R., & Joseph, J. E. (2007). Functional Dissociation in Frontal and Striatal Areas for Processing of Positive and Negative Reward Information. *Journal of Neuroscience*, *27*(17), 4587–4597. https://doi.org/10.1523/JNEUROSCI.5227-06.2007

Lovero, K. L., Simmons, A. N., Aron, J. L., & Paulus, M. P. (2009). Anterior insular cortex anticipates impending stimulus significance. *NeuroImage*, *45*(3), 976–983. https://doi.org/10.1016/J.NEUROIMAGE.2008.12.070

Luck, S. J. (2014). *An Introduction to the Event-Related Potential Technique, second edition*. MIT Press Journals. https://books.google.ca/books?hl=en&lr=&id=y4-uAwAAQBAJ&oi=fnd&pg=PR5&dq=luck+2014&ots=jC00z6LNFH&sig=MK0BCjR9iu1A1mk_RANcoyrkh0Y#v=onepage&q=luck 2014&f=false

Luck, S. J., & Gaspelin, N. (2017). How to get statistically significant effects in any ERP experiment (and why you shouldn't). *Psychophysiology*, *54*(1), 146–157. https://doi.org/10.1111/PSYP.12639

MacNamara, A., Foti, D., & Hajcak, G. (2019). Tell me about it: Neural activity elicited by emotional pictures and preceding descriptions. *Emotion*, *9*(4), 531–543. https://psycnet.apa.org/record/2009-11528-012

Mill, J. S. (1863). *Utilitarianism*. https://scholar.google.ca/scholar?hl=en&as_sdt=0%2C5&q=mill+1863+utilitarianism&btnG=&oq=mill+1863+

Moray, N. (1959). Attention in dichotic listening: Affective cues and the influence of instructions. *Https://Doi.Org/10.1080/17470215908416289*, *11*(1), 56–60. https://doi.org/10.1080/17470215908416289

Morgan, H. M., Klein, C., Boehm, S. G., Shapiro, K. L., & Linden, D. E. J. (2008). Working Memory Load for Faces Modulates P300, N170, and N250r. *Journal of Cognitive Neuroscience*, *20*(6), 989–1002. https://doi.org/10.1162/JOCN.2008.20072

Mott, K. K., Alperin, B. R., Holcomb, P. J., & Daffner, K. R. (2014). Age-related decline in differentiated neural responses to rare target versus frequent standard stimuli. *Brain Research*, *1587*(1), 97–111. https://doi.org/10.1016/J.BRAINRES.2014.08.057

Neisser, U. (1988). Five kinds of self-knowledge. *Philosophical Psychology*, *1*(1), 35–59. https://doi.org/10.1080/09515088808572924

Neisser, U., & Fivush, R. (1994). *The remembering self: Construction and accuracy in the self-narrative*. https://books.google.com/books?hl=en&lr=&id=8wISpDE4K2IC&oi=fnd&pg=PR7&ots=jDmPoyRVxR&sig=4rxDZ3pd7bbV-T00k-IYDmL83UA

Northoff, G., & Hayes, D. J. (2011). Is Our Self Nothing but Reward? In *Biological Psychiatry* (Vol. 69, Issue 11, pp. 1019–1025). Elsevier. https://doi.org/10.1016/J.BIOPSYCH.2010.12.014

Northoff, G., Heinzel, A., de Greck, M., Bermpohl, F., Dobrowolny, H., & Panksepp, J. (2006). Self-referential processing in our brain—A meta-analysis of imaging studies on the self. *NeuroImage*, *31*(1), 440–457. https://doi.org/10.1016/J.NEUROIMAGE.2005.12.002

Patel, S. H., & Azzam, P. N. (2005). Characterization of N200 and P300: selected studies of the Event-Related Potential. *International Journal of Medical Sciences*, *2*(4), 147–154. http://www.ncbi.nlm.nih.gov/pubmed/16239953

Phan, K. L., Taylor, S. F., Welsh, R. C., Ho, S. H., Britton, J. C., & Liberzon, I. (2004). Neural correlates of individual ratings of emotional salience: a trial-related fMRI study. *NeuroImage*, *21*(2), 768–780. https://doi.org/10.1016/J.NEUROIMAGE.2003.09.072

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

Polich, J. (2007). Updating P300: An integrative theory of P3a and P3b. In *Clinical Neurophysiology* (Vol. 118, Issue 10, pp. 2128–2148). Elsevier. https://doi.org/10.1016/j.clinph.2007.04.019

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

Rogers, T. B., Kuiper, N. A., & Kirker, W. S. (1977). Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology*, *35*(9), 677–688. https://doi.org/10.1037/0022-3514.35.9.677

Rundell, J. F. (2004). *Fragments of a History of the Theory of Self-Consciousness from Kant to Kierkegaard*. Brill. https://brill.com/view/book/9789047406648/B9789047406648_s004.xml

Sambrook, T. D., & Goslin, J. (2015). A Neural reward prediction error revealed by a meta-analysis of ERPs using great grand averages. *Psychological Bulletin*, *141*(1), 213–235. https://doi.org/10.1037/BUL0000006

Sato, A., Yasuda, A., Ohira, H., & … K. M.-. (2005). Effects of value and reward magnitude on feedback negativity and P300. *Journals.Lww.Com*. https://journals.lww.com/neuroreport/fulltext/2005/03150/effects_of_value_and_reward_magnitude_on_feedback.20.aspx?casa_token=gs0jeAY9HKoAAAAA:1WbxYN0yFMRZx-I73ZH8ELCBKEGgYD4et8r0hj1gEpkP72X0sZ7W5g2EeeEDZs26wc0XyB2z7CZE2XKreLGy-yj7cg

Schomaker, J., & Meeter, M. (2014). Novelty detection is enhanced when attention is otherwise engaged: an event-related potential study. *Experimental Brain Research*, *232*(3), 995–1011. https://doi.org/10.1007/S00221-013-3811-Y/FIGURES/6

Seligman, M. (2020). Agency in Greco-Roman philosophy. *The Journal of Positive Psychology*, *16*(1), 1–10. https://doi.org/10.1080/17439760.2020.1832250

Shimada, S., Hiraki, K., & Oda, I. (2005). The parietal role in the sense of self-ownership with temporal discrepancy between visual and proprioceptive feedbacks. *NeuroImage*, *24*(4), 1225–1232. https://doi.org/10.1016/J.NEUROIMAGE.2004.10.039

Sui, J., He, X., & Humphreys, G. W. (2012). Perceptual effects of social salience: Evidence from self-prioritization effects on perceptual matching. *Journal of Experimental Psychology: Human Perception and Performance*, *38*(5), 1105–1117. https://doi.org/10.1037/A0029792

Sui, J., & Humphreys, G. W. (2013). The boundaries of self face perception: Response time distributions, perceptual categories, and decision weighting. *Visual Cognition*, *21*(4), 415–445. https://doi.org/10.1080/13506285.2013.800621

Sui, J., & Humphreys, G. W. (2015a). The interaction between self-bias and reward: Evidence for common and distinct processes. *Sage*, *68*(10), 1952–1964. https://doi.org/10.1080/17470218.2015.1023207

Sui, J., & Humphreys, G. W. (2015b). More of me! Distinguishing self and reward bias using redundancy gains. *Attention, Perception, and Psychophysics*, *77*(8), 2549–2561. https://doi.org/10.3758/S13414-015-0970-X/FIGURES/4

Sui, J., Liu, M., Mevorach, C., & Humphreys, G. W. (2015). The Salient Self: The Left Intraparietal Sulcus Responds to Social as Well as Perceptual-Salience After Self-Association. *Cerebral Cortex*, *25*(4), 1060–1068. https://doi.org/10.1093/CERCOR/BHT302

Sui, J., Rotshtein, P., & Humphreys, G. W. (2013). Coupling social attention to the self forms a network for personal significance. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(19), 7607–7612. https://doi.org/10.1073/PNAS.1221862110

Sui, J., Sun, Y., Peng, K., & Humphreys, G. W. (2014). The automatic and the expected self: Separating self- and familiarity biases effects by manipulating stimulus probability. *Attention, Perception, and Psychophysics*, *76*(4), 1176–1184. https://doi.org/10.3758/S13414-014-0631-5/FIGURES/2

Sui, J., Zhu, Y., & Han, S. (2006). Self-face recognition in attended and unattended conditions: An event-related brain potential study. *NeuroReport*, *17*(4), 423–427. https://doi.org/10.1097/01.WNR.0000203357.65190.61

Sutton, R. S., & Barto, A. C. (2018). *Reinforcement Learning, second edition: An Introduction*. MIT Press Journals. https://books.google.ca/books?hl=en&lr=&id=uWV0DwAAQBAJ&oi=fnd&pg=PR7&dq=Sutton,+R.S.,+Barto,+A.G.,&ots=misGu70_i1&sig=5cE3EkGhYGs0dRK2K0-w0z_AslM&redir_esc=y#v=onepage&q=Sutton%2C R.S.%2C Barto%2C A.G.%2C&f=false

Sutton, S., Braren, M., Zubin, J., & John, E. R. (1965). Evoked-potential correlates of stimulus uncertainty. *Science (New York, N.Y.)*, *150*(3700), 1187–1188. http://www.ncbi.nlm.nih.gov/pubmed/5852977

Sutton, S., Tueting, P., Zubin, J., & John, E. R. (1967). Information delivery and the sensory evoked potential. *Science (New York, N.Y.)*, *155*(3768), 1436–1439. http://www.ncbi.nlm.nih.gov/pubmed/6018511

Symons, C. S., & Johnson, B. T. (1997). The self-reference effect in memory: A meta-analysis. *Psychological Bulletin*, *121*(3), 371–394. https://psycnet.apa.org/record/1997-03609-003

Tenke, C. E., Kayser, J., Stewart, J. W., & Bruder, G. E. (2010). Novelty P3 reductions in depression: Characterization using principal components analysis (PCA) of current source density (CSD) waveforms. *Psychophysiology*, *47*(1), 133–146. https://doi.org/10.1111/J.1469-8986.2009.00880.X

Tong, F., & Nakayama, K. (1999). Robust representations for faces: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *25*(4), 1016–1035. https://doi.org/10.1037/0096-1523.25.4.1016

Treisman, A. M. (1960). Contextual cues in selective listening. *Http://Dx.Doi.Org/10.1080/17470216008416732*, *12*(4), 242–248. https://doi.org/10.1080/17470216008416732

Turk, D. J., Cunningham, S. J., & Macrae, C. N. (2008). Self-memory biases in explicit and incidental encoding of trait adjectives. *Consciousness and Cognition*, *17*(3), 1040–1045. https://doi.org/10.1016/J.CONCOG.2008.02.004

Turk, D. J., van Bussel, K., Brebner, J. L., Toma, A. S., Krigolson, O., & Handy, T. C. (2011). When "it" becomes "mine": Attentional biases triggered by object ownership. *Journal of Cognitive Neuroscience*, *23*(12), 3725–3733. https://doi.org/10.1162/JOCN_A_00101

Turk, D. J., van Bussel, K., Waiter, G. D., & Macrae, C. N. (2011). Mine and me: Exploring the neural basis of object ownership. *Journal of Cognitive Neuroscience*, *23*(11), 3657–3668. https://doi.org/10.1162/JOCN_A_00042

van den Bos, M., Cunningham, S. J., Conway, M. A., & Turk, D. J. (2010). Mine to remember: The impact of ownership on recollective experience. *Quarterly Journal of Experimental Psychology*, *63*(6), 1065–1071. https://doi.org/10.1080/17470211003770938

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

Wiggins, O. P. (1994). Commentary on "Self-Consciousness, Mental Agency, and the Clinical Psychopathology of Thought Insertion." *Philosophy, Psychiatry, & Psychology*, *1*(1). https://muse.jhu.edu/article/244996

Williams, B., & Long, A. A. (2008). *Shame and necessity*. 254.

Williams, C. C., Ferguson, T. D., Hassall, C. D., Abimbola, W., & Krigolson, O. E. (2021). The ERP, frequency, and time–frequency correlates of feedback processing: Insights from a large sample study. *Psychophysiology*, *58*(2), e13722. https://doi.org/10.1111/PSYP.13722

Williams, C. C., Hassall, C. D., Lindenbach, T., & Krigolson, O. E. (2019). Reward Prediction Errors Reflect an Underlying Learning Process That Parallels Behavioural Adaptations: A Trial-to-Trial Analysis. *Computational Brain and Behaviour*, *3*, 189–199. https://doi.org/10.1007/s42113-019-00069-4

Williams, C. C., Hassall, C. D., Trska, R., Holroyd, C. B., & Krigolson, O. E. (2017a). When theory and biology differ: The relationship between reward prediction errors and expectancy. *Biological Psychology*, *129*, 265–272. https://doi.org/10.1016/j.biopsycho.2017.09.007

Williams, C. C., Hassall, C. D., Trska, R., Holroyd, C. B., & Krigolson, O. E. (2017b). When theory and biology differ: The relationship between reward prediction errors and expectancy. *Biological Psychology*, *129*, 265–272. https://doi.org/10.1016/J.BIOPSYCHO.2017.09.007

Williams, C. C., Saffer, Y., Mcculloch, R. B., & Krigolson, O. E. (2016). The scarcity heuristic impacts reward processing within the medial-frontal cortex. *NeuroReport*, *27*, 522–526. https://doi.org/10.1097/WNR.0000000000000575

Williams, H. L., Conway, M. A., & Cohen, G. (2008). Autobiographical memory. In *Memory in the Real World* (pp. 21–90). Psychology Press. https://psycnet.apa.org/record/2007-03296-002

Winterer, G., Adams, C. M., Jones, D. W., & Knutson, B. (2002). Volition to Action—An Event-Related fMRI Study. *NeuroImage*, *17*(2), 851–858. https://doi.org/10.1006/NIMG.2002.1232

Wolford, G., & Morrison, F. (1980). Processing of unattended visual information. *Memory & Cognition*, *8*(6), 521–527. https://doi.org/10.3758/BF03213771

Woodman, G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics 2010 72:8*, *72*(8), 2031–2046. https://doi.org/10.3758/BF03196680

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

Yang, H., Wang, F., Gu, N., Gao, X., & Zhao, G. (2013). The cognitive advantage for one's own name is not simply familiarity: An eye-tracking study. *Psychonomic Bulletin and Review*, *20*(6), 1176–1180. https://doi.org/10.3758/S13423-013-0426-Z/TABLES/3

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

Zhang, W., & Luck, S. J. (2008). Feature-based attention modulates feedforward visual processing. *Nature Neuroscience 2008 12:1*, *12*(1), 24–25. https://doi.org/10.1038/nn.2223

Zhou, Z., Yu, R., & Zhou, X. (2010). To do or not to do? Action enlarges the FRN and P300 effects in outcome evaluation. *Neuropsychologia*, *48*(12), 3606–3613. https://doi.org/10.1016/j.neuropsychologia.2010.08.010

Zhu, Y., Zhang, L., Fan, J., & Han, S. (2007). Neural basis of cultural influence on self-representation. *NeuroImage*, *34*(3), 1310–1316. https://doi.org/10.1016/J.NEUROIMAGE.2006.08.047

Appendix A: Supplemental Data

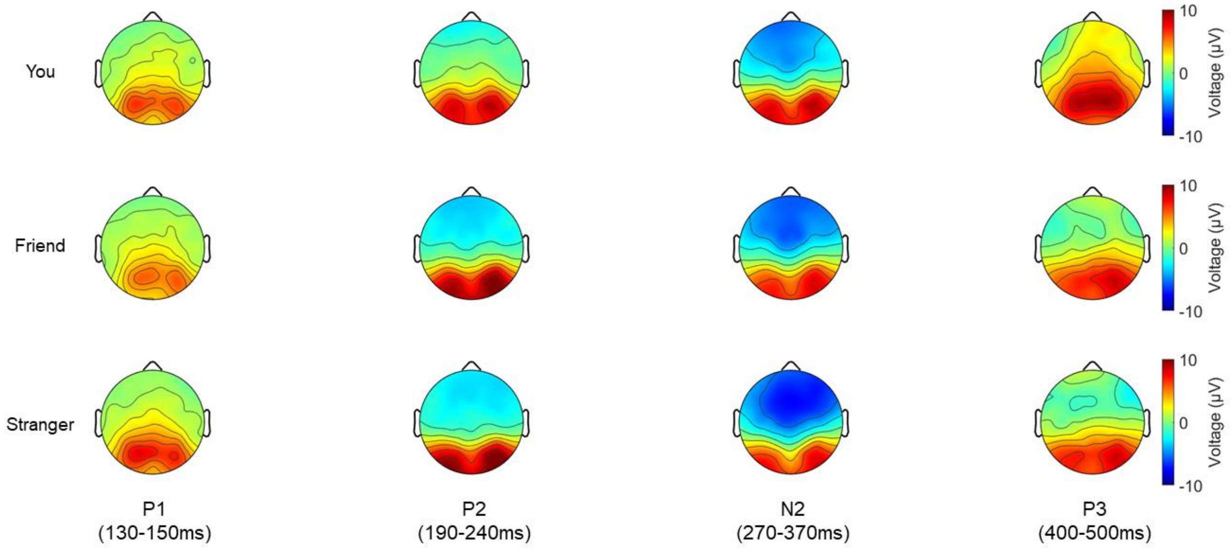

Figure A1. Grand average topographic maps on maximal points for the N1, P2, N2, and P3 components.

Appendix B: Supplemental Materials

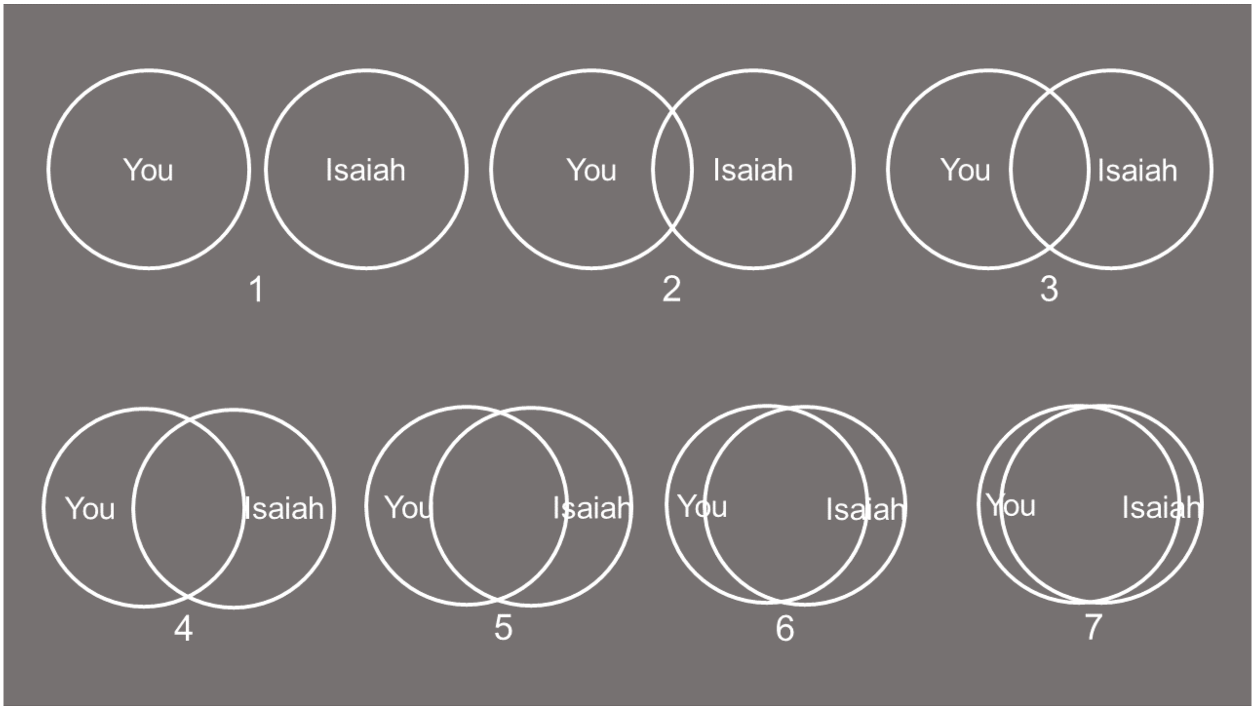

Figure B1. An example of the Inclusion of the Other in the Self (IOtS) scale presented to participants.