The Neural Mechanisms by Which Facial Impressions Influence Memory

Grace Liu

Submitted in partial fulfillment of the
requirements for the degree of
Master of Arts in
Quantitative Methods in the Social Sciences
of the Graduate School of Arts and Sciences
Columbia University

May 9, 2025

Abstract

Just from a glance at another's face, we judge them on a variety of factors, often unconsciously. Social trait impressions — such as trustworthiness and dominance — are made rapidly and influence our behavior and decision-making. While prior research has shown that these impressions influence a variety of social outcomes, less is known about how they arise in the brain. Understanding the cognitive and neural mechanisms by which facial impressions influence memory can inform how facial biases form and affect our social interactions, relationships, and decisions, spanning diverse contexts from law to politics to the workplace. Currently, most studies focus on behavioral outcomes, and few explore the neural mechanisms of how face impressions influence memory. This study addresses that gap by investigating the neural mechanisms of facial impressions across race in the hippocampus and amygdala, and whether the differences in neural activity predict later memory performance. This is a secondary analysis where, in the original study, participants underwent fMRI scanning while completing a memory task involving faces that varied in trustworthiness and dominance ratings, paired with sentences, and subsequently followed by recall tests. In the present study, behavioral and univariate neuroimaging analyses were conducted, as well as brain-behavior correlations to link the trait-based neural processes to memory accuracy. Findings revealed that untrustworthy faces were better remembered than trustworthy faces, supporting existing evidence of threat prioritization. An interaction between trustworthiness and dominance was found, suggesting that trait impressions shape memory together rather than independently. While neural activation differences were observed in both the hippocampus and amygdala, brain-behavior correlations were non-significant. However, the findings indicate that there are possible race-based differences that affect how facial impressions influence memory performance.

Contents

Introduction

During our day-to-day activities, we are constantly interacting with others around us and forming judgments of them, without much thought. Face perception is a central pillar of human social interaction and we rely on it to assess others quickly in everyday interactions to guide our behaviors, decisions, and emotional responses. In just a brief glance, individuals can extract a wealth of socially meaningful information from faces, such as identity, emotion, and inferred traits. Two social judgments — trustworthiness and dominance — play a crucial role in shaping first impressions, influencing decisions about threat detection and social hierarchy (Oosterhof & Todorov, 2008). These trait impressions are formed automatically and have a powerful influence on decision making in various contexts and domains, spanning from politics (Levi & Stoker, 2000) to law (Porter et al., 2010) to interpersonal relationships (Rotter, 1980).

Two of the most robust dimensions of trait impressions are trustworthiness and dominance, which represent others' intentions and their ability to cause harm, respectively (Oosterhof & Todorov, 2008). These dimensions not only shape social judgments, but they also influence how we encode and remember faces. For example, people are more likely to remember faces they perceive as untrustworthy due to increased attention for potential threats (Rule et al., 2012).

Yet, while there is extensive research on the behavioral effects of facial impressions on memory, the connection between neural mechanisms and these processes has remained understudied. Previous research suggests that the hippocampus, which plays a crucial role in episodic memory and facial recognition memory, is sensitive to socially relevant trait impressions (Cao et al., 2021). It has also been determined that the amygdala is automatically responsive to facial trustworthiness even when faces are not consciously perceived (Freeman et

al., 2014). The amygdala, known for its role in emotions and threat detection, may also interact with the hippocampus to amplify encoding of socially meaningful or threatening faces (Phelps, 2004).

The present study builds upon this growing research by exploring how trait impressions — specifically trustworthiness and dominance — affect memory encoding, with a focus on hippocampal and amygdala neural activity. By linking individual differences in behavioral memory performance to neural activation, this study aims to grow our understanding of how socially driven face judgments influence memory and their implications for real-world decision making and cognitive bias.

Literature Review

I. Face Perception

Face perception is a fundamental component of human social interactions in every facet of our lives, shaping how we navigate and interpret the world. With just a brief glance at someone's face, individuals can quickly and efficiently extract social information about them that is necessary for navigating social environments, influencing communication, decision-making, and behavior. Faces are the main channel of social cognition, serving as powerful social stimuli that elicit a wealth of information such as social categories (e.g., sex, race, age), mental states, attractiveness, and identity (Stangor et al., 1992; Ekman, 1992).

II. Social Trait Inferences from Faces

In as little as 38 milliseconds, people can make rapid personality judgments of others based on their facial appearance, and these judgments can ultimately influence approach or

avoidance behaviors (Oosterhof & Todorov, 2008). This stems from an evolutionary perspective, where accurate assessments of threats are essential for survival (Bar et al., 2006). Principal component analysis of trait judgments and statistical modeling of facial features have found that these face impressions are made along the key dimensions of trustworthiness and dominance, but do not necessarily account for accurate judgments (Oosterhof & Todorov, 2008). These inferences of trustworthiness and dominance are crucial, as they can predict critical real-world consequences, such as hiring decisions, political elections, and criminal sentencing (Olivola et al., 2014; Wilson & Rule, 2015, 2016).

The idea that a person's facial features alone can reveal their personality has endured for centuries, grounded in evolutionary explanations of social and environmental adaptation (Zebrowitz, 2004). Historically, physiognomy — the practice of analyzing someone's personality based on their facial appearance — was widely accepted. In Chinese culture, physiognomy was used to analyze facial beauty (Wang, 2020), and in ancient Greece, it was used to label heroes and villains (Stavru, 2019), and Aristotle even wrote a treatise titled "Physiognomonica." Today, while it is shown not to accurately reflect actual personality traits, research has shown that trait judgments made from facial appearance are highly consistent across subjects and critically influence real-world decision-making in various domains (Hassin & Trope, 2000). One of the most consequential settings where facial judgments are made is in law sentencing decisions. In a study by Eberhardt et al. (2006), it was found that those with stereotypically Black facial features were more likely to receive the death penalty, especially when the victim was White. Similarly, in the electoral space, candidates who had more "competent" facial features were more likely to win elections, predicting electoral success (Todorov et al., 2005). Even in economic decisions, facial features, such as smiling, act as a socio-economic cue that increases cooperation in

one-shot trust games, where strangers risk monetary resources based on perceived trustworthiness (Scharlemann et al., 2001). Even those who are more attractive earn more in their occupation compared to those who are perceived as less attractive, reinforcing the "beauty premium" in the workplace (Hamermesh & Biddle, 1993) and illustrating the Halo Effect, where a few positive traits may lead others to assume an overall all positive impression of them (Thorndike, 1920).

These studies reveal just how impactful facial trait judgments are and how they are not simply arbitrary perceptions — they actively shape real-life decisions that make a lasting impact. However, such cognitive behavior should be considered with caution, as assumptions that individuals make about others purely from their facial impressions can reinforce social biases and disparities. Understanding how and why individuals subconsciously use facial features to make important social decisions is an important step to combating this.

Trustworthiness and Dominance

Trustworthiness and dominance represent the two core dimensions of trait inferences that people spontaneously infer from faces (Oosterhof & Todorov, 2008). Trustworthiness reflects a person's intent, as in whether they appear benevolent or threatening. Dominance reflects a person's ability to enforce their intent, as in whether they seem physically strong or weak. Trustworthy faces tend to have soft features such as a smiling mouth and wide eyes that look almost surprised (Oosterhof & Todorov, 2008; Todorov et al., 2007b). On the other hand, faces that are higher on the dominance dimension are often associated with small eyes, angular faces, lower brow positions, and large jawlines/chins (Toscano et al., 2014).

Oosterhof & Todorov (2008) found that when facial features became more exaggerated in the affectively positive, or happier looking, direction, participants classified these faces as

trustworthy, whereas when facial features became more exaggerated in the affectively negative, or angry looking, direction, participants classified these faces as untrustworthy or dominant. This suggests that judgments of trustworthiness and dominance are closely related to emotional facial expression cues (e.g., happiness, anger), which individuals often use to infer others' behavioral intentions (Fridlund, 1994). The overgeneralization theory (Zebrowitz et al., 2003) attributes this phenomenon to evolutionary adaptation responses: trait impressions arise due to survival-related responses, and these assumptions are erroneously applied to individuals who simply resemble those physical features. For instance, a happy expression suggests approachability, signaling that a person is safe to interact with, while an angry expression signals that the person should be avoided and may be a threat (Adams et al., 2006). Judgments of trustworthiness have been found to be negatively correlated with judgments of anger and positively correlated with judgments of happiness from emotionally neutral faces (Oosterhof & Todorov, 2009). Similarly, judgments of affiliation follow the same pattern (Montepare & Dobish, 2003). Jaeger & Jones (2022) found that resemblance to emotional expressions is the strongest predictor of face impressions, particularly for trustworthiness and dominance.

Perceptions of trustworthiness and dominance are not arbitrary — they are strongly related to facial expressions and emotions, and they influence how people perceive, categorize, and remember others. Emotion strongly influences various cognitive processes, including attention, perception, and memory, among others, as they are known to enhance encoding, consolidation, and retrieval of memory (Tyng et al., 2017). Since facial cues of trustworthiness and dominance often resemble emotional expressions, such as happiness and anger, they may act as emotionally salient signals that not only shape impressions but also memory retention (Zebrowitz et al., 2003; Todorov & Duchaine, 2008). For instance, a trustworthy face resembling

a smile may benefit from the memory-enhancing effects of the positive cue (Tsukiura & Cabeza, 2008), while a dominant-rated face engages memory due to its association with threat (Oosterhof & Todorov, 2008). Thus, it can be inferred that the social impressions made from facial appearance leverage the overlapping neural and cognitive pathways involved in emotional processing.

Dominance and Threat Perception

Social groups are categorized not only by trust but also by dominance, determining how individuals navigate threats. Dominance helps determine how individuals navigate power structures and potential dangers, influencing how they go about interpersonal interactions. Threatening faces are perceived as both untrustworthy and dominant (Oosterhof & Todorov, 2008), suggesting that dominance is closely linked to perceptions of physical and social threat. While trustworthiness judgments are focused on assessing intent, dominance judgments are focused on assessing someone's capacity to act on that intent.

Dominance plays a massive role in assessing threats and establishing social threat hierarchies. High-dominant faces, which include features such as strong jawlines, wider faces, and lower brows, are often perceived as more aggressive and authoritative (Carré et al., 2009). This perception influences how people evaluate threat potential and dominance in interpersonal relationships in a range of social contexts.

As we touched upon earlier, there are various real-world implications in making such dominance judgments. In the workplace, individuals with more dominant facial features are more likely to achieve leadership positions as they are more likely to be perceived as figures of authority (Rule & Ambady, 2008). In legal sentencing decisions, more dominant-looking

individuals (especially for Black men with strong facial features) are more likely to be perceived as a threat, contributing to the enormous racial disparities in law (Kleider-Offutt et al., 2017).

These findings establish a strong connection between dominance perception and threat evaluation. By shaping how individuals determine another's power and threat levels, dominance judgments can fuel biases that disproportionately affect certain groups. It is necessary to investigate how such trait-based perceptions influence memory and decision-making processes, particularly in the domains of facial recognition and social cognition.

III. Facial Impressions and Memory

Face impression judgments — particularly trustworthiness and dominance — help organize our social world. These judgments significantly influence how we encode and remember faces, shaping our everyday social interactions and decisions. For instance, in a trust game, it was found that people were more likely to make a decision to trust someone who was subjectively rated as more trustworthy (Van't Wout & Sanfey, 2008). Although we cannot purely rely on facial appearance to tell us about someone's intent or personality, we rapidly form impressions based on these social traits, impacting how effectively faces are remembered. In fact, people have better memory for information that violates their expectations based on facial appearance (Bell et al., 2015). These impressions influence whether we remember a face later on, often giving socially relevant (Cassidy & Gutchess, 2012) or emotionally charged (Cassidy, 2020) faces a memory advantage. More specifically, facial information contributes to recognition memory since faces are perceptually and ecologically salient (Zebrowitz, 2004). Therefore, understanding the relationship between facial impressions and memory is essential, providing insights into cognitive biases that affect real-world behaviors.

How Trustworthiness & Dominance Affect Memory

Previous research has determined that face recognition memory is influenced heavily by the perceived trustworthiness of faces. We make these trustworthiness judgments based on facial appearances automatically (Engell et al., 2007), and these judgments factor into our decisions concerning whom to trust (Van't Wout & Sanfey, 2008). This prioritization of trustworthiness is likely rooted in evolutionary processes for survival. Hou and Li (2019) conducted a series of experiments exploring the survival processing advantage in face recognition. They found that both trustworthy and untrustworthy faces were better recognized when processed in a survival-related context, indicating that facial trustworthiness provides important information for survival decisions. Along the spectrum of trustworthiness, faces that are perceived as more untrustworthy are more easily remembered than those perceived as trustworthy (Rule et al., 2012; Mattarozzi et al., 2015; Wendt et al., 2019). This preference for remembering untrustworthy faces likely functions as a protective mechanism against potential threats, making them highly salient since humans are naturally alert to cues of exploitation or deception (Cosmides & Tooby, 1989).

Going in another direction, further complexity arises when considering mismatches between facial impressions and behaviors. It was also found that memory was best for the pairing of trustworthy faces with untrustworthy behavior (Suzuki & Suga, 2010). This behavior is likely reflected by the protective mechanisms against exploitation by disguised cheaters, as it is more likely for humans to remember a wolf in sheep's clothing rather than a wolf in wolf's clothing. In other words, it is more likely to remember a trustworthy-looking cheater than an untrustworthy-looking cheater because we are more likely to trust the former, putting us in danger of manipulation. This unexpected pairing engages memory since it triggers humans'

inherent mechanisms against suspicious, disguised signs of trustworthiness or cheaters (Cosmides & Tooby, 1989).

However, the relationship between trustworthiness and memory is not always straightforward. Another possible explanation for why there is enhanced memory for trustworthy-looking cheaters is that trustworthy faces are more distinctive and, therefore, more easily remembered (Dewhurst et al., 2005). Alternatively, trustworthy faces actually reflect a memory advantage for happy-looking faces, since happy-looking faces are associated with being more trustworthy (Oosterhof & Todorov, 2008) — indicating that facial expressions modulate trustworthiness in memory.

Faces are tied to emotional context, and emotional expressions increase the likelihood that a face will be remembered later on (D'Argembeau & Van Der Linden, 2007). In an investigation by Mattarozi et al. (2015) on the effects of facial appearance — specifically perceived trustworthiness — and emotional context on face memory, it was established that trustworthy and untrustworthy faces were better remembered than neutral faces, and this was due to the emotional context that such faces are associated with. Whether the emotions perceived in the face were pleasant or unpleasant, it increased the likelihood of remembering semantic and even episodic details associated with faces, whereas neutral faces were simply recognized to just be familiar. Thus, established that perceived trustworthiness affects face memory. In line with previous findings, this study also confirmed that untrustworthy faces were more likely to lead to memory enhancement. Oosterhof & Todorov (2008) also demonstrated that changes in face trustworthiness correspond to subtle changes resembling expressions signaling approach or avoidance. This points toward the fact that the memory advantage for untrustworthy faces might be partially explained by their resemblance to negative emotional expressions.

While trustworthiness has been the primary focus in existing research, other research has shown that dominance is important for the social perception of faces, although its influence on memory encoding and recall is comparatively less understood. Dominant facial features — wide faces, strong jawlines, lowered brows — are essential in social interactions by communicating physical strength, authority, and threat (Carré et al., 2009). This stems from an evolutionary explanation as displays of dominance are carefully evaluated in decision-making when one is in a hostile or ambiguous situation (Oosterhof & Todorov, 2008). More dominant-looking faces often attract greater attentional resources, thereby increasing their salience and memorability. Additionally, research by Rule et al. (2010) demonstrated that dominant faces are more likely to be encoded deeply into memory when associated with contexts that highlight leadership or threat, reflecting evolutionary preparedness to prioritize social signals of power or potential harm.

Dominant faces are also associated with emotions, particularly angry emotions (Oosterhof & Todorov, 2008). Angry expressions often involve lowering the brow and raising the upper lip, which are linked to more dominant facial features. This increases the saliency of this trait, signaling an aggression judgment.

This enhanced memory encoding could also increase stereotype-driven memory errors, where more dominant faces are misremembered in contexts of aggression or hostility (Rule et al., 2010). Kleider-Offutt et al. (2017) revealed that dominant facial characteristics are more likely to trigger memory biases related to aggressive or threatening behavior, particularly for faces relating to minority group members. This has various implications in contexts such as eyewitness identification. These findings underscore the double-edged sword of dominance on memory:

while it can enhance recognition memory through salience and attention, it also risks encoding biases influenced by social stereotypes.

Neural Mechanisms of Trait-Based Memory Encoding

The aforementioned findings highlight that facial trait judgments of trustworthiness and dominance shape how faces are encoded and retrieved from memory. Consequently, it is essential to understand the underlying neural mechanisms by which these traits influence memory processes. Recent functional magnetic resonance imaging (fMRI) findings have begun to reveal neural activity patterns, particularly those associated with the hippocampus and amygdala, by which trustworthiness and dominance influence memory encoding and recall processes.

Hippocampus

Research consistently identifies the hippocampus as a critical neural structure involved in encoding facial impressions into memory. Paller and Wagner (2002) established that heightened hippocampal activation during initial exposure of the encoding phase reliably predicts later successful memory performance, and succeeding research has extended this paradigm for faces. To get more specific in the hippocampus's role in encoding socially relevant stimuli, fMRI studies have also shown that hippocampal activation is modulated by facial trustworthiness or dominance. Faces perceived as highly trustworthy and dominant evoke greater hippocampal activity, as these traits are socially relevant and influence memory encoding. Rule et al. (2012) found that faces judged as untrustworthy elicited stronger hippocampal responses than neutral faces, likely due to the evolutionary salience of detecting potential social threats. Along the same lines, Oosterhof & Todorov (2008) demonstrated that subtle variations in facial trustworthiness

and dominance affect memory encoding, with the hippocampus integrating these social signals into long-term face recognition. Expanding on this foundational knowledge and looking specifically looking into trustworthiness, a hippocampal fMRI study by Tsukiura & Cabeza (2008) found that not only were individuals more accurate in recalling smiling faces — which is associated with trustworthiness — than neutral faces, but the hippocampus showed successful encoding and retrieval activations that were also greater for smiling than neutral faces. This illustrates how social signals such as trustworthiness and smiling facial expressions can enhance face memory, exhibiting greater neural activity for such faces.

Amygdala

Beyond the hippocampus, the amygdala also plays a pivotal role in processing social signals, exhibiting activation patterns associated with facial impressions, especially those associated with threat detection or trustworthiness evaluations (Adolphs, 2010; Phelps & LeDoux, 2005). Existing work has demonstrated that the amygdala shows a range of emotional responses to the perceived valence of a face, and trustworthiness serves as a key indicator of that valence (Engell et al., 2007; Freeman et al., 2014). Amygdala activation was found for trustworthy faces (Whalen et al., 1998), and a study by Adolphs et al. (1998) further highlighted the amygdala's importance: individuals with bilateral amygdala damage were impaired in their ability to judge trustworthiness, perceiving untrustworthy-looking faces as trustworthy.

Empirical findings show that the amygdala is significantly more responsive to faces rated as untrustworthy compared to trustworthy, reflecting heightened awareness of potential social threats (Engell et al., 2007; Winston et al., 2013). Amygdala responses were particularly strong to faces perceived as untrustworthy or dominant, consistent with its established role in evaluating threats and emotional salience (Freeman et al., 2014). Todorov & Engell (2008) extended this by

investigating how the amygdala responds to neutral faces evaluated along various trait dimensions, including trustworthiness and dominance. They found that the amygdala's activation correlates more strongly with traits that have clear valence connotations, such as trustworthiness, compared to traits like dominance, which have less clear valence associations. Valence connotation here refers to whether a trait carries a positive or negative connotation. This automatic evaluation was less pronounced for dominance judgements, suggesting that the amygdala's role in processing dominance is less direct than for trustworthiness. Another study also reported a U-shaped, nonlinear amygdala response to face trustworthiness; the activation was stronger for faces at the extremes of trustworthiness rather than for faces in the middle of trustworthiness (Freeman et al., 2014). Together, these findings underscore the amygdala's central role in creating rapid, automatic social evaluations based on trait impressions from facial appearance.

Hippocampus and Amygdala Interactions

Interestingly, these studies suggest an interaction between the hippocampus and amygdala during the encoding of facial impressions. This ties into the amygdala's broader role in memory: the amygdala not only responds to emotional cues, but it enhances memory encoding and storage through interactions with the hippocampus, particularly for emotional experiences. More specifically, greater amygdala activation during encoding is associated with better memory later on (Phelps et al., 2004). It was found that amygdala and hippocampal neurons are responsible for spontaneous first impressions of faces, encoding a social trait space, specifically for trustworthiness and dominance (Cao et al., 2022). Building on this notion of a social trait space, more recent work has shown that such trait representations are not confined to memory regions alone. The findings from Chwe et al.'s (2024) study found that the two-dimensional

representation of faces' inferred traits occurs in the middle temporal gyrus (MTG), a region involved in the activation of new social concepts. The MTG interacts with both the hippocampus and amygdala: it strengthens its connection with the hippocampus to form new associations (Ren et al., 2020), and the right amygdala interacts with both the right hippocampus and MTG, integrating emotional processing with memory and conceptual processing (Iidaka et al., 2001). Supporting this MTG-amygdala network's role in socioemotional processing, Diano et al. (2017) found heightened amygdala activation connectivity with the MTG when participants viewed fearful or happy faces compared to neutral expressions.

Collectively, these neural findings portray the complexity of facial impressions on memory encoding processes. Trustworthiness and dominance do not simply guide surface-level social interactions, but they also affect neural activity within memory regions such as the hippocampus and amygdala. Understanding these neural mechanisms provides direction into the cognitive biases and memory distortions associated with social trait perception.

IV. Hippocampus

To fully appreciate how facial impressions such as trustworthiness and dominance influence memory encoding, it is essential to understand the neural foundations underlying these processes. The hippocampus, a brain structure central to episodic memory formation, consolidation, and retrieval, plays a critical role in determining what information is retained or forgotten (Squire et al., 2015). Due to its importance in integrating emotional, contextual and perceptual details into memories, exploring hippocampal activity can provide valuable insights into how facial impressions are encoded, stored, and recalled. Thus, the following section will

delve into the hippocampus's basic functions in memory, its involvement in face recognition, and the neural mechanisms of the hippocampus during encoding and retrieval.

Hippocampus and Episodic Memory

Episodic memory is what allows us to retain and later recall information about our experiences (Tulving, 1972), and the hippocampus is widely known to play a critical role in episodic memory after discoveries made in the notorious neuroscience case of patient H.M. (Scoville & Milner, 1957). In an attempt to control his seizures, H.M. underwent an experimental procedure that left him with severe memory impairment, although his epilepsy was controlled. The surgery resulted in anterograde amnesia, which is when an individual cannot form new memories after the occurrence of the injury. H.M.'s medial temporal lobe, including the hippocampus, was affected, demonstrating the importance of the hippocampus in memory formation (Squire & Wixted, 2011).

Once memory is encoded into the brain, the memory is stored for later retrieval, also known as consolidation, by the hippocampus. This is where memories are stabilized over long periods of time, and the hippocampus works again with other cortical regions. The memories are transitioned from their initially fragile, hippocampus-dependent state to more stable, cortically distributed memories (Eichenbaum et al., 1992). Lastly, during the retrieval phase, the hippocampus accesses and recalls the stored memory by piecing together the various aspects of the memory, including the time and place of the event (Tulving, 1972). This retrieval process also demonstrates the critical role of the hippocampus not only in initial encoding but in the reconstruction and recall of episodic memories.

Neural Mechanisms

The hippocampus represents a crucial neural substrate in the encoding, consolidation, and retrieval of episodic memories. Hippocampal cellular responses are strongly linked to intended movement and sensory input and often require highly environment-specific triggers (Eichenbaum et al., 1992). The hippocampus represents the configurations of various cues, including olfaction, spatial position, and temporal context, binding various elements to form coherent representations of memory.

At the neural level, the synaptic plasticity of hippocampal neurons also explains the strengthening of episodic memories over time. More specifically, long-term potentiation (LTP) is observed in the hippocampus during learning. LTP is the persistent strengthening of synapses, and it is induced by patterns of neural activity that occur during hippocampal learning (Eichenbaum et al., 1992). This means that there is increased neural activity during the encoding of memories, and it is associated with temporal cues. Additionally, Barnes et al. (1979) found a correlation between the duration of LTP and the retention of information, suggesting the hippocampus's involvement in consolidation and later retrieval. LTP is also regulated by theta oscillations, which time the precise firing of neurons. Theta oscillations are rhythmic patterns of neural activity, and they essentially represent the "online" state, so in the hippocampus, they are a primary indicator of activation. During encoding, theta power increases in the hippocampus, and it was found that there is greater theta activity when information is successfully retrieved than when information is not remembered (Joensen, 2023).

Hippocampal activity during encoding predicts later memory retrieval, which reinforces its role in memory consolidation (Eichenbaum et al., 1992). The hippocampus is not simply a passive repository; it is actively involved in retrieving stored information by reconstructing past

experiences. Wagner et al. (1998) conducted fMRI studies to explore how neural activation during word encoding predicts subsequent memory performance. Their findings revealed that the ability to later remember a verbal experience is predicted by the magnitude of activation in the left prefrontal and temporal cortices, which feed into the hippocampus, during that experience. This was echoed by Davachi & Wagner (2002), who also demonstrated that greater hippocampal activation during encoding was significantly associated with better subsequent memory performance. This supports the idea that heightened hippocampal activation facilitates stronger, more integrated memory traces, improving overall memory outcomes.

To further investigate the neural activity associated with the retrieval of information, Cansino et al. (2002) conducted a study where participants were exposed to novel images and then later asked to identify if the images were old or new during retrieval. The results showed that images that were correctly recognized during retrieval elicited greater activity in the hippocampus than those that were incorrectly recognized, specifically in the right hippocampal formation.

These findings display the importance of the hippocampus in episodic memory encoding, consolidation, and retrieval. It interacts with various sensory and cognitive systems to ensure that context-rich memories are formed, kept, and accessible in the brain. Understanding these mechanisms not only informs memory function but also provides insights into how the hippocampus shapes cognition, perception, and decision-making.

Hippocampus and Face Recognition Memory

The ability to recognize faces is a fundamental aspect of human social cognition, allowing people to differentiate between familiar and unfamiliar faces and facilitate interpersonal relationships. While face recognition is traditionally associated with the fusiform face area (FFA)

(Kanwisher et al., 1997, the hippocampus is also involved in face recognition, specifically the right hippocampus (Taylor et al., 2011), beyond its general role in episodic memory.

Face recognition depends on both perceptual and memory processes. While occipitotemporal regions are in charge of distinguishing and perceiving individual facial features, the hippocampus is responsible for encoding and recalling the contextual details related to faces (Prince et al., 2009). Unlike pure visual processing for immediate face perception, long-term face recognition requires the hippocampus to supplement the face with identifying details such as names, past interactions, and emotional significance (Yonelinas et al., 2010). The hippocampus is key for remembering where and when — the context or episodic details — a face was encountered.

Studies examining patients with hippocampal damage also inform the brain region's role in face recognition memory. Impairments in hippocampal function, such as in Alzheimer's disease, result in severe deficits in face recognition, particularly when contextual cues are necessary for retrieval (Ryan et al., 2001). Going back to the amnesia patient H.M., he was unable to learn new faces even when his basic perceptual abilities remained intact (Scoville & Milner, 1957).

Furthermore, the hippocampus is involved in memory for emotionally salient faces. The hippocampus has strong connections with the amygdala, which assigns social and emotional significance to faces, so it also plays a role in emotionally significant memories (Vuilleumier et al., 2004). Previous research has established that happy faces enhance memory performance of face-name associations better than neutral faces, consistent with evidence that reward can enhance episodic memory processes (Adcock et al., 2006). This suggests that a happy, more friendly expression may enhance episodic memory for new face-name associations, making it

advantageous to remember such faces for future social interactions. Faces associated with trustworthiness, dominance, or emotional expression are often better remembered, demonstrating that socially relevant facial impressions are more memorable than neutral faces (Rule et al., 2012). This interaction between the hippocampus and amygdala provides a potential explanation for why certain faces, especially those associated with threat and trustworthiness, may leave stronger memory traces.

The hippocampus plays a pivotal role in integrating faces with episodic memory. Unlike purely visual face-processing regions, the hippocampus allows for the long-term retention and contextualization of faces. The relationship between the hippocampus and face recognition provides the foundation for how memory and social cognition interact.

Neural Mechanisms

fMRI studies have shown that hippocampal activity is essential for recognizing faces, particularly in specific contexts, and for retrieving information about prior interactions with others (Yonelinas et al., 2010). Taylor et al. (2011) confirmed the hippocampus' role in face memory, starting from a young age. They explored the development of hippocampal and frontal lobe contributions to face recognition and found that there was consistent right hippocampal activation during face recognition tasks across various age groups. However, it was also discovered that other frontal regions of the brain become increasingly engaged in face recognition as individuals develop, suggesting that neural strategies for facial processing shift over time.

Using an fMRI dataset of face-name associative memory tasks, Sheng et al. (2022) found that the human brain had high-dimensional representations of faces specific to better episodic memory. More specifically, individuals with greater representational dimensionality (RD) in their

neural activity exhibited better subsequent episodic memory performance. RD relates to how effectively neural representations distinguish between different face-name associations. This suggests that faces are encoded more richly due to the high HD, resulting in better episodic memory recall for faces, and this occurs in the hippocampus and related cortical areas.

In collaboration with the amygdala, the hippocampus also plays a key role in encoding salient facial features (Cao et al., 2021). Specifically, hippocampal neurons selectively responded to facial features such as the eyes and mouth, suggesting that the hippocampus contributes to the encoding of facial identity and socially relevant information. It's also interesting to note that there is greater parahippocampal and amygdala activation during remembering negative faces than remembering neutral faces (Fenker et al., 2005). This implies that emotionally charged faces enhance hippocampal memory retrieval.

Together, prior research illustrates the hippocampus's pivotal role in face recognition memory, ensuring that faces are contextualized with emotional significance and social meaning. By interacting with the FFA, the amygdala, and other cortical regions, the hippocampus ensures that faces are encoded with rich details that allow for the retrieval of face memories and social trait information related to facial impressions(Cao et al., 2022).

V. Amygdala

The amygdala is the most vital brain region for processing emotion, crucial for fear processing and emotional learning, emotional modulation of memory (e.g., emotionally arousing events are remembered better than neutral ones), and also affects how we perceive emotionally salient stimuli (Phelps & LeDoux, 2005). Faces are highly salient social stimuli, and the amygdala is especially sensitive to emotionally relevant facial cues, such as anger, fear, or other indicators of threat (Adams et al., 2003). But, other researchers argue that the amygdala instead

assesses the valence of stimuli — whether something is positive or negative — and coordinates physiological and behavioral responses accordingly, rather than just focusing on threat.

Amygdala and Emotional/Threat Processing

The established, widely accepted view is that the amygdala plays a key role in the rapid, automatic evaluation of stimuli that signal potential threat or danger in the environment (Adolphs et al., 1999). Early fear conditioning studies have shown the amygdala's crucial role in acquiring and expressing learned fear (Phelps & LeDoux, 2005). The amygdala also responds strongly to ambiguous cues, such as fearful faces (Adams et al., 2003) or uncertain dangers (Fox et al., 2015), highlighting how it doesn't only process obvious threats.

However, more recent research suggests that the amygdala's role shifts away from being purely a "threat detector," but instead is more general in emotional processing, more so detecting relevance regardless of positive or negative valence (Sander et al., 2003). Previous fMRI studies have found that the amygdala is activated by both positive and negative stimuli: increased amygdala activation for fearful faces, and decreased activation for happy faces (Whalen et al., 1998). Meta-analyses further confirm that there is amygdala activation to both valences, although with a preference for emotionally salient faces (Sergerie et al., 2008). This includes distress (Blair et al., 1999), surprise (Kim et al., 2003), and social traits such as dominance (Watanabe & Yamamoto, 2015), illustrating the amygdala's broader processing role. Additionally, the amygdala can process emotional and social signals unconsciously, responding to fearful or salient facial cues even when they are presented outside of conscious awareness (such as during a masked paradigm) (Whalen et al., 1998). Moreover, amygdala responses to faces may contribute to biases in social evaluation, with greater activation toward either extremes of facial trustworthiness, following U-shaped activation patterns (Freeman et al., 2014), and potentially

influencing which faces are better encoded into memory (Phelps, 2004). This suggests that early emotional impressions shaped by amygdala activation can have lasting effects on social memory formation.

Neural Mechanisms

Research into the neural mechanisms underlying amygdala function highlights its broad response to emotionally salient stimuli. Fear still remains especially significant in amygdala activation, as a study by Phan et al. (2002), which explored brain responses to emotional stimuli (happiness, fear, anger, sadness, disgust) as well as valence, concluded that fear specifically activates the amygdala.

The amygdala also operates within a broader neural network that interacts with various other brain regions to regulate social- and memory-related processes. The prefrontal cortex (PFC) modulates amygdala output to inhibit maladaptive fear (Phelps & LeDoux, 2005), the MTG integrates emotional signals from the amygdala with the hippocampus for social judgments (Iidaka et al., 2001), and hippocampus-amygdala interactions support the encoding and consolidation of emotionally salient memories (Cao et al., 2021). These interactions are what allow trait impressions (such as trustworthiness and dominance) to influence memory encoding.

VI. The Present Study

Despite extensive prior research on face perception and memory, there remain various gaps in our understanding of how facial features, such as those resembling trustworthiness and dominance — the core dimensions of facial impressions — influence memory processes. While research has covered behavioral outcomes or general neural mechanisms, there has been little exploration of the specific neural activation that underlies the memory of faces along the

dimensions of trustworthiness and dominance. The present study aims to investigate how trustworthy versus dominant facial features influence memory, focusing on hippocampal and amygdala activation. While prior research has established that trustworthiness and dominance shape first impressions and decision-making (Oosterhof & Todorov, 2008), less is known about how these traits affect memory processes at the neural level. Specifically, this study will explore whether trustworthy or dominant facial impressions enhance memory performance and whether hippocampal activation during encoding predicts later memory recall.

VII. Hypotheses

H1: Along the trustworthiness dimension, untrustworthy faces will be better remembered than trustworthy faces.

Faces rated as more untrustworthy will be more accurately recalled than those rated as trustworthy due to evolutionary processes for survival. This will be reflected in greater hippocampal activation during encoding, predicting successful recall.

H2: Along the dominance dimension, dominant faces will be better remembered than submissive faces.

Dominant faces will be associated with increased memory accuracy as they increase perceptual salience and attention due to evolutionary processes for survival.

H3: Hippocampal and amygdala activation will predict memory performance.

Greater hippocampal/amygdala engagement during encoding will be associated with correct recall, while lower activation will predict memory errors. This will be assessed using univariate fMRI analyses comparing hippocampal/amygdala beta values across correct recall,

within-category errors (e.g., confusing a Black man for another Black man), and between-category errors (e.g., confusing a White man for a Black man), where category refers to the social categorization of race.

H4: Untrustworthy faces will elicit stronger amygdala activation than average trustworthy faces along the trustworthiness dimension.

Given the amygdala's role in evaluating threat and emotional salience, untrustworthy faces will evoke greater amygdala activation than average trustworthy faces. This aligns with prior findings where facial trustworthiness showed a U-shaped response pattern in amygdala activation.

H5: Differences in neural activation related to trustworthiness in the hippocampus and amygdala will be associated with memory performance.

Greater activation differences between trustworthy and untrustworthy faces during encoding are expected to predict greater recall accuracy.

By evaluating these hypotheses, this study will provide us with direction on how social trait impressions can influence memory at the neural level.

Methods

The present study is a secondary analysis using unidentified data collected from the Who-Said What (WSW) study at the Social Cognitive and Neural Sciences Lab at Columbia University.

Participants

The original WSW study included 32 participants (20 = female, 12 = male) recruited via RecruitMe, a subject recruiting platform, from the New York City area. Participants completed a standard screening form asking for demographic information (e.g., age, gender, race). Eight participants identified as Asian (25%), three identified as Black (9%), one identified as Indian (3%), eight identified as Hispanic/Latino (25%), and 12 identified as White (38%). All participants reported normal or corrected-to-normal vision, right-handedness, no use of psychoactive medications, and no history of neurological disease. All participants identified as native or near-native English speakers and were born and raised in the United States. One participant was excluded from the present study analysis due to technical issues during preprocessing that prevented the accurate generation of their brain mask.

Procedure

Materials

All face images used in the original WSW study were taken from the Chicago Face Database (CFD) (Ma et al., 2015). The CFD was developed at the University of Chicago and is intended for use in scientific research. It contains high-resolution, standardized photographs of male and female faces of varying ethnicity between the ages of 17 to 65. For the present study's analysis, the trustworthiness and dominance ratings were also provided by the CFD (Ma et al., 2015). The ratings were collected from a convenience sample of over 1,087 participants who viewed the standardized, neutral face images online. Raters assessed each face based on how trustworthy and dominant the face appeared on a 1-7 Likert scale (1 = Not at all, 7 = extremely).

Each rater rated a randomized subset of faces to prevent fatigue, and data was aggregated across raters to generate trustworthiness and dominance ratings for each face.

In the WSW study, 96 statements were used. 24 statements were created by Pietraszewski (2021), in which the statements were designed to convey interpersonal alliances, allowing researchers to test whether social categorization was based on race or coalition membership. The remaining 72 statements were generated by ChatGPT. The prompt given to ChatGPT instructed it to reference the original 24 sentences as a stylistic model, stating something along the lines of, "Here are 24 sentences from the original study. Create three new sets of sentences that match the style of these originals, but vary in content."

Category Localizer

Before the WSW study, participants underwent a category localizer task where they performed a 1-back memory task. Participants were presented with a sequence of stimuli and required to respond when the current stimulus matched the one presented immediately before it. Participants viewed face images consisting of 60 Black and 60 White male faces while undergoing fMRI scanning to record neural activity to identify the brain regions responsive to the face stimuli. They were required to remember and determine if the current face image was the same as the one presented right before and were instructed to press a button each time this occurred. For example, if a participant was shown three faces, but the third face was the same as the second, they would need to respond to that third face by pressing a button. Each face was displayed for 2 seconds with a 2-6 second interstimulus interval. The faces were shown across 6 runs (10 randomized Black male faces and 10 randomized White males per run), with each run lasting around 2 minutes. The 1-back task was implemented to ensure attention.

Who-Said-What

The WSW dataset is an fMRI dataset that was originally collected from 2023-2024 to explore how social categorization, memory, and neural representations interact in the context of race. There were four fMRI runs total and each run of the paradigm consisted of three parts: encoding, distractor, and retrieval. In each run, during the encoding phase (Fig. 1), participants viewed 8 male faces (4 Black and 4 White) where each face was shown alone for 3 seconds, followed by an interstimulus interval of 2–6 seconds. Then, each face was paired with three first-person behavioral sentences (e.g., “I was annoyed for the whole day after. But at the same time - I couldn't help but laugh at how absurd and funny the whole situation turned out to be.”), presented one at a time with the face for 7 seconds each. The function of the behavioral statement is to serve as unique pieces of information that participants must encode and subsequently retrieve, along with the faces, to assess memory accuracy and error patterns; the statements are fully randomized. The order of sentence presentation was structured using tertile ordering: the first 8 sentences seen during encoding were also the first 8 presented during retrieval, though randomized within the set, and this structure was maintained for the second and third tertiles.

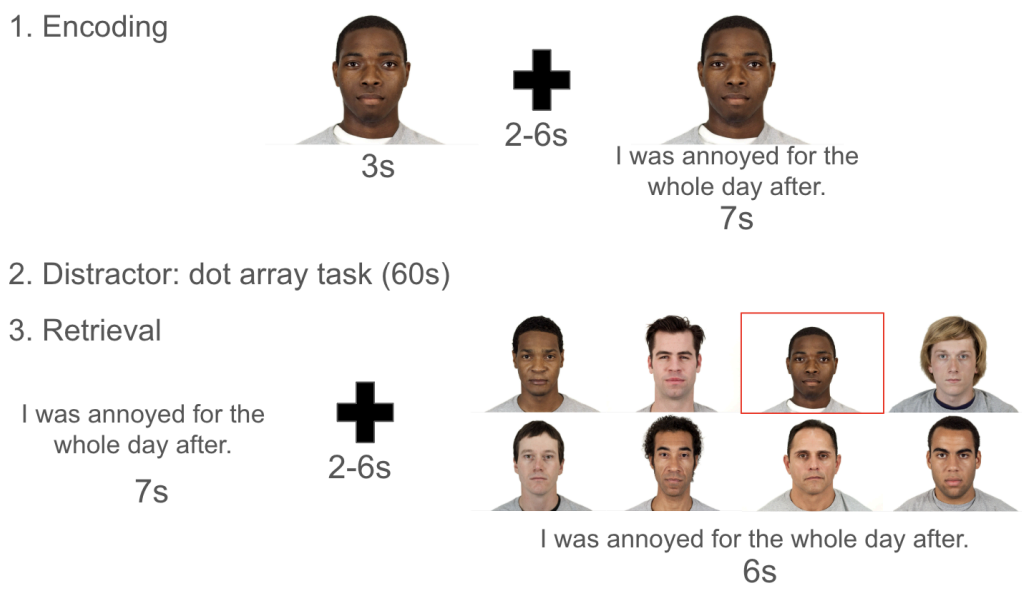

Figure 1. Who-Said-What paradigm

Following the encoding phase, participants underwent a 60-second distractor task. They viewed an array of dots, followed by a number, and indicated via button press whether the number of dots was greater or fewer than the number presented.

In each trial of the retrieval phase, participants were shown one of the previously encoded sentences for 7 seconds, followed by a 2–6 second interstimulus interval. Then, they were shown the same sentence along with a grid of all 8 previously seen faces from the encoding phase for 6 seconds. Participants were provided a red block to move so they could select the face they believed was originally paired with that sentence (Fig. 1). This entire process was repeated across four runs, each using different randomly selected face identities and sentence sets. Participants completed the task while undergoing fMRI scanning. For this current study, the WSW dataset was adapted to explore how facial social impressions, specifically trustworthiness and dominance, influence memory.

Measures

The present study investigated how trait impressions along the trustworthiness and dominance dimensions influenced memory accuracy and neural activation in the hippocampus and amygdala. Below, each key construct is described in terms of how it was operationalized, measured, and formatted in the dataset.

Dependent variable (DV)

Memory accuracy was the primary dependent variable in the behavioral and mixed-effects regression models. It was measured during the retrieval phase of the WSW study. In each trial, participants attempted to identify the face that had originally been paired with a specific behavioral sentence during the encoding phase. A binary accuracy score was assigned for each trial: responses were coded as 1 for correct face-statement matches and 0 for incorrect matches. Incorrect responses were further categorized based on the race of the selected face compared to the correct face. Within-category errors occurred when participants incorrectly selected a face of the same race as the correct one (e.g., choosing a different Black face when the correct face was Black), while between-category errors occurred when participants selected a face of the opposite race (e.g., choosing a Black face when the correct face was White). This allowed for additional insights into memory error patterns related to race-based social categorization.

Independent variables (IV)

Trustworthiness and dominance ratings were pre-rated scores from the CFD, rated on a 1-7 Likert scale. Each face in the category localizer and WSW tasks had a fixed trustworthiness

and dominance rating. In the present study's univariate fMRI analyses, trustworthiness was further split into "trustworthy" and "untrustworthy" groups and dominance was further split into "dominant" and "submissive" groups using an overall median split across all CFD ratings.

Each face in the category localizer and WSW tasks was labeled as either Black or White. This variable was used to subset the data in the present study for race-specific analyses and to account for race-based variance in memory and neural activation patterns.

Neural activation (mean beta values) was measured using beta values extracted from pre-defined anatomical regions of interest (ROIs) in the hippocampus and amygdala. These values were computed from subject-level general linear model (GLM) contrast maps during two different phases: 1) the encoding phase of the WSW task (to evaluate memory-related activation), and 2) the category localizer task (to evaluate trait-based encoding of face impressions). Mean beta values were calculated by averaging across all voxels within each ROI and were then used as dependent variables in the univariate and correlational neuroimaging analyses.

Quantitative Analysis Methods

This present study employed a combination of behavioral, univariate neuroimaging, and correlational analyses to examine how trait impressions, specifically trustworthiness and dominance, shape memory performance and modulate neural activity in the hippocampus and amygdala.

Behavioral Analysis

Behavioral data from the WSW memory task were analyzed to assess the difference in memory effect across participants and how impressions of trustworthiness and dominance influenced memory accuracy. Memory performance was categorized into three conditions:

1) Correct recall
2) Within-group errors (e.g., confusing a Black male face with another Black male face)
3) Between-group errors (e.g., confusing a Black male face with a White male face)

Descriptive statistics were first used to compare average trial counts across these conditions, followed by pairwise t-tests to test for differences in memory performance across conditions.

Then, to examine whether trustworthiness and dominance ratings could predict memory accuracy, generalized linear mixed models (glmer) were fit with a binary outcome variable (correct = 1, incorrect = 0) and trustworthiness and dominance ratings as predictors. The model accounted for subject-specific variance as well as race variance in Black and White faces. To explore whether trustworthiness and dominance interacted in predicting memory accuracy, a glmer was fit, again accounting for subject-specific and race variances. This allowed the investigation into whether the influence of one impression dimension on memory recall depended on the other. This was repeated with Black male faces only and White male faces only.

Univariate Neuroimaging Analysis

Functional MRI data were analyzed to explore the average activation in the two ROIs: the hippocampus and the amygdala, during the category localizer task and the encoding phase of the WSW memory task. The same analysis pipeline was applied to both ROIs.

Before getting started with any analysis, anatomical brain masks are used to define the hippocampus and amygdala ROIs. The initial association areas were sourced from Neurosynth, a

large-scale, meta-analytic database. Each ROI (Neurosynth search: “hippocampus” and “amygdala”) was thresholded at $z > 12$ to isolate the voxels most strongly associated with the hippocampus and the amygdala, ensuring that the selected ROIs corresponded to brain regions previously identified with these processes. These masks were then resampled to match the functional resolution of each subject’s individual contrast maps (Fig. 2a, 2b).

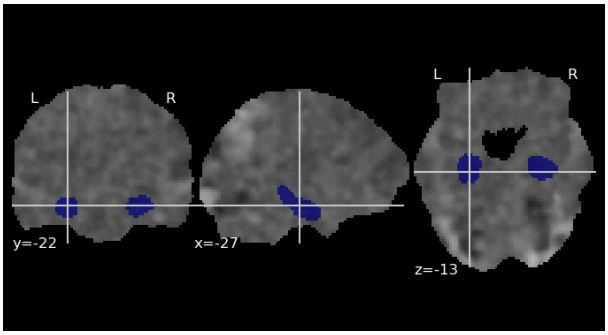

Figure 2a. Hippocampus ROI

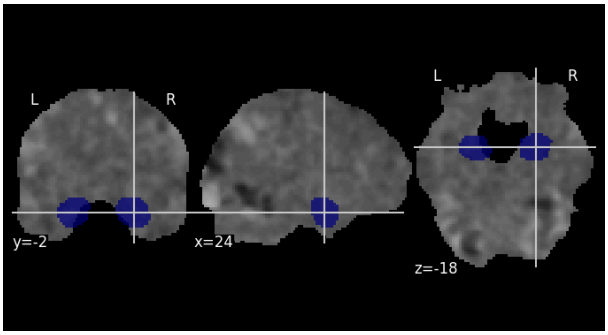

Figure 2b. Amygdala ROI

Univariate analysis in neuroimaging involves examining neural activity at each voxel independently to measure how much neural activation occurs within a specific brain region.

For each participant, beta values (reflecting fit with the hemodynamic response function) were extracted from the hippocampus and amygdala masks using subject-level general linear model (GLM) contrast maps. These beta values were then averaged across voxels within each ROI for each relevant condition (e.g., trustworthiness level, memory outcome).

To evaluate whether the hippocampus and amygdala showed differential activity based on memory performance, the neural activity during the encoding phase of the WSW task was analyzed. Mean beta values were computed for each of the three behavioral memory outcomes (correct recall, within-group error, between-group error). A repeated-measures ANOVA was

used to examine whether neural activity differed across memory outcomes, and pairwise t-tests were conducted to further investigate whether there were any significant differences in neural activation between the three memory conditions.

To determine how much neural activity was elicited when viewing faces of different trustworthiness and dominance levels, the data from the category localizer task was analyzed. The category localizer task was administered prior to the WSW memory task, and during this task, participants viewed a series of face images while undergoing fMRI scanning. The localizer task was used instead of the WSW task because it provided unconfounded neural data of faces, whereas the WSW task focused on memory.

Each face already had predetermined normed trustworthiness and dominance ratings on a 1-7 point scale provided by the CFD. The faces were then further categorized into four groups: trustworthy or untrustworthy, and dominant or submissive. This was done using an overall-level median split across all CFD ratings for all faces in the dataset. This ensured a consistent threshold across all participants and standard definitions of trustworthiness and dominance. However, this doesn't account for the variability in which faces each subject saw. An initial attempt to perform subject-specific median splits (based on only the faces each participant saw) yielded limited effects, so analysis proceeded with the overall-level median split for statistical power. For each ROI, the mean beta values (neural activation) were computed for each trait group. Analyses were repeated for:

- All faces
- Black male faces only
- White male faces only.

For each face group in both ROIs, pairwise t-tests were performed to compare average neural activation in response to faces along the trustworthiness dimension (i.e., trustworthy vs. untrustworthy) and the dominance dimension (i.e., dominant vs. non-dominant). These comparisons allowed for the examination of whether trait impressions elicited distinct neural activation responses within the hippocampus and amygdala, and whether these responses differed by race.

Brain-Behavior Correlations

Brain-behavior correlations were created to investigate if there were any relationships between neural activation of trustworthiness and the likelihood of recall. This was only done for trustworthiness, as near significance was found for only trustworthiness on recall accuracy in the behavioral analysis, and not dominance.

Pearson's correlations were calculated between 1) each participant's behavioral trustworthiness slope (the effect of trustworthiness on recall accuracy), and 2) their neural activation difference (mean beta values for trustworthy faces - untrustworthy faces) in each ROI. Correlations were computed separately for the hippocampus and amygdala, using all faces, Black male faces only, and White male faces only. These analyses aimed to determine whether individual differences in memory toward trustworthiness were reflected in differential encoding-related activity in the hippocampus and amygdala.

Results

I. Behavioral Results

Across all participants, a total of 3,072 trials were analyzed. First, participants' memory recall was categorized into three conditions: correct recall, within-category error, and between-category error. Before data analysis, the between-categories errors were multiplied by a constant, (n - 1)/n, to account for the greater number of possible confusions for between categories (i.e., all n members of the other category) than within categories (all members of the category minus the target: n - 1) since a target cannot be confused with themselves (Klauer & Wegener, 1998). For instance, during the retrieval phase of the WSW task, given that the correct choice is a Black man, there would be 3 other Black men remaining out of the grid of 8 faces — this constitutes a within-category error. The remaining 4 White men constitute a between-category error. Thus, there is 3/8 chance of making a within-category error, and 4/8 chance of making a between-category error, and so the between-category error was multiplied by 0.75 to correct for chance and allow for a fairer comparison between the two types of errors.

Participants correctly recalled 1,113 trials (36.2%), made 936 within-category errors (30.5%), and 767 between-category errors (25%). On average, participants correctly recalled approximately 35 out of 96 trials (M = 34.8, SD = 2.9) (Fig. 3), indicating strong performance in the WSW task. In contrast, participants made a moderate number of within-category errors (M = 29.2, SD = 1.4) and fewer between-category errors (M = 24.0, SD = 1.4) (Fig. 2).

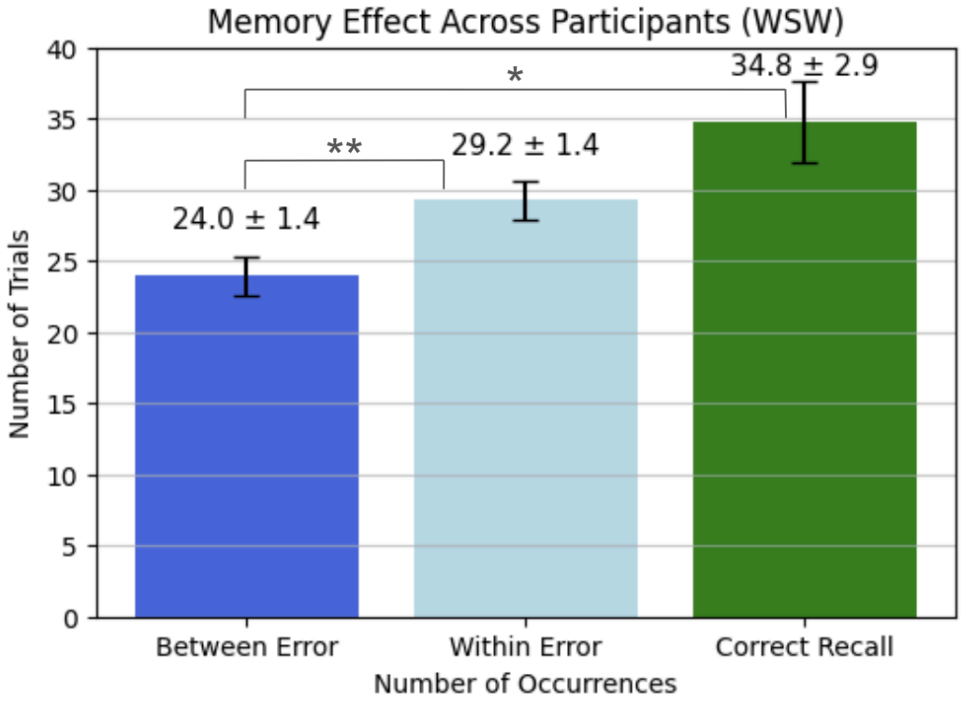

Figure 3. Memory performance across trial types.

Average number of trials for between-category errors (i.e., confusing a Black man for a White man), within-category errors (i.e., confusing one Black man for another Black man), and correct recall (left-right) in the WSW paradigm.

This pattern suggests that while memory for faces was generally reliable, as participants correctly recalled faces more than they made either within-group or between-group memory errors, they were also more likely to confuse individuals within the same race than across races. The standard errors indicate relatively low variability across participants. Pairwise t-tests revealed some significance between the three memory conditions:

- Correct versus Within: t = 1.337, p = 0.1909
- Correct versus Between: t = 2.577, p = 0.0149*
- Between versus Within: t = -4.632, p = 0.0001**

Participants made significantly more within-category errors than between-category errors, supporting the well-established finding that memory confusions are more likely to occur for individuals of the same social category. Additionally, correct recall significantly exceeded between-category errors, and also within-category errors, although the latter was insignificant. This suggests that people are more accurate when recalling faces from the same group, but they are also more likely to confuse those faces.

The glmers fit to examine how facial impressions may predict memory performance revealed some near significance. Results revealed a negative trend between trustworthiness rating and predicted probability of correct responses, where higher trustworthiness ratings are associated with lower memory recall accuracy and vice versa (Fig. 4a); this correlation was almost statistically significant (^β = -0.268, SE = 0.139, z = -1.93, p = 0.054).

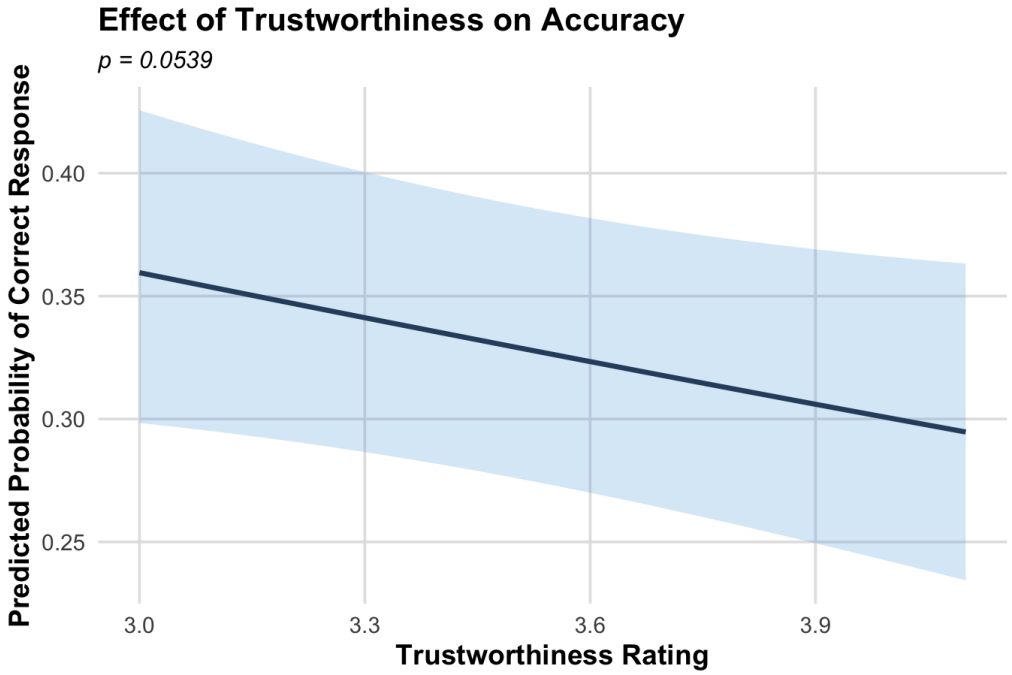

Figure 4a. Trustworthiness rating and predicted recall accuracy.

Predicted probability of correct recall responses as a function of a face's trustworthiness rating.

As a post hoc exploratory analyses, the data was subset by race to examine whether this relationship could vary by race (Fig. 4b). While the glmer models for both Black male faces only (^β = -0.173, SE = 0.216, z = -0.804, p = 0.422) and White male faces only (^β = -0.203, SE = 0.266, z = -0.760, p = 0.447) showed no significance, the correlations were a similar negative trend as the correlation with all faces (Fig. 3a). These results suggest that the inverse trustworthiness-accuracy relationship observed in the full dataset may not be robust within either race group independently.

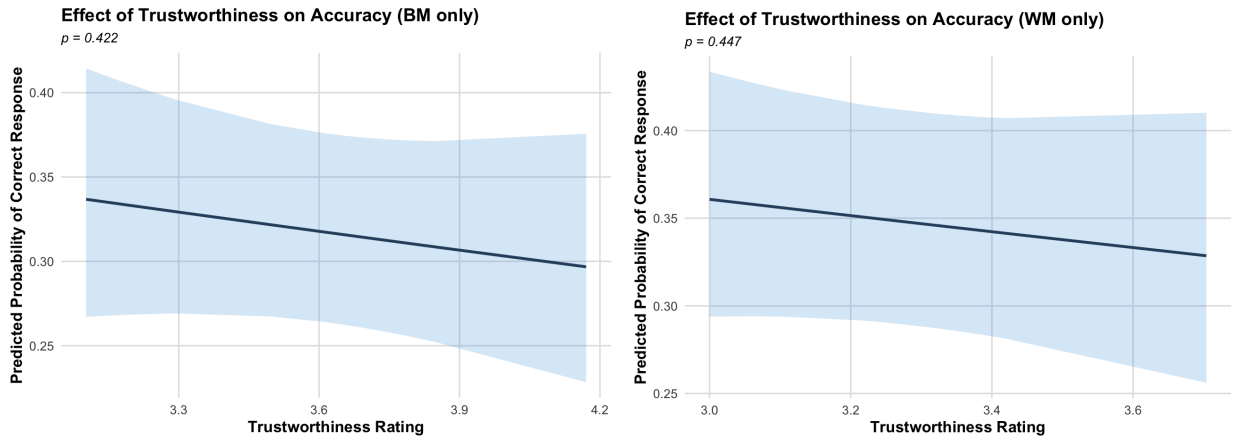

Figure 4b. Trustworthiness rating and predicted recall accuracy, subset for Black male faces only (left) and White male faces only (right).

Predicted probability of correct recall responses as a function of a face's trustworthiness rating for only Black male faces, and only White male faces.

Conversely, for dominance ratings, there was a weak, positive trend between dominance and predicted probability of correct response where higher dominance ratings were correlated with better memory recall accuracy, however, this was insignificant (^β = 0.053, SE = 0.066, z = 0.803, p = 0.422) (Fig. 5a). While dominance may contribute to how we perceive others, it is not as influential as trustworthiness when it comes to memory encoding.

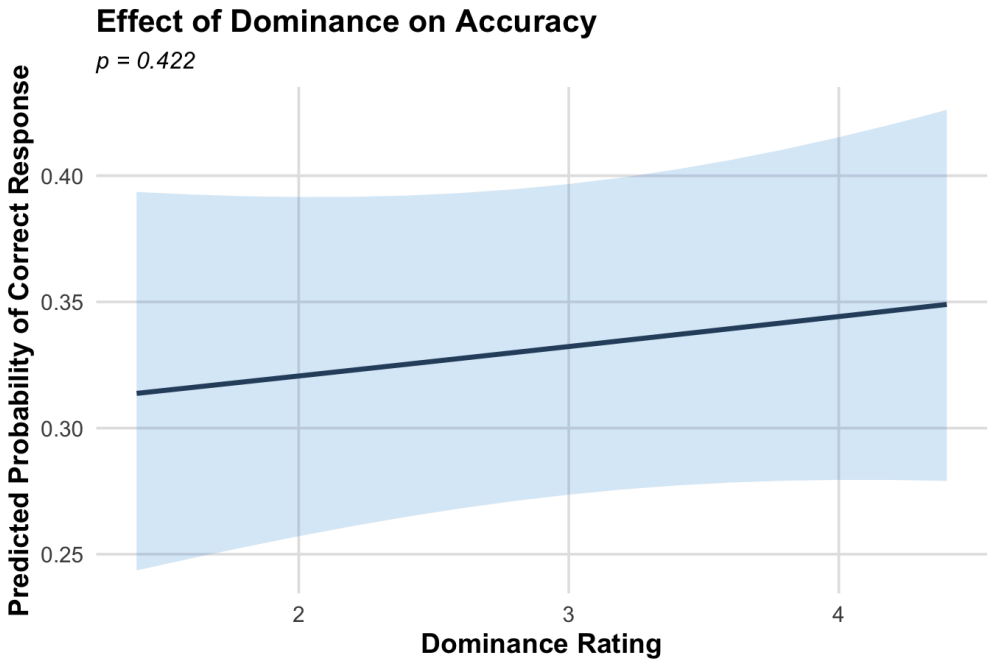

Figure 5a. Dominance rating and predicted recall accuracy.

Predicted probability of correct recall responses as a function of a face's dominance rating.

The data was subset again by race, and race-based differences were observed for dominance ratings' influence on recall accuracy (Fig. 5b). While Black males only had the same, weak positive association (^β = -0.170, SE = 0.119, z = -1.429, p = 0.153) as the all faces correlation (Fig. 5a), White males only revealed a stronger, significant pattern (^β = 0.166, SE = 0.079, z = 2.10, p = 0.036*). White faces rated higher on the dominance dimension were significantly

associated with greater recall accuracy, suggesting that the influence of facial dominance on memory is more pronounced for White male faces than Black male faces.

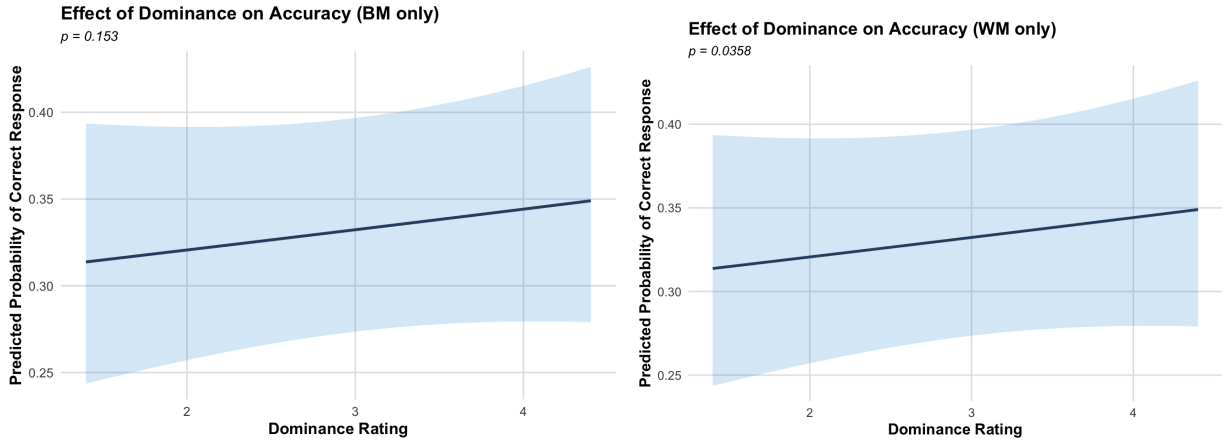

Figure 5b. Dominance rating and predicted recall accuracy, subset for Black male faces only (left) and White male faces only (right).

Predicted probability of correct recall responses as a function of a face's dominance rating for only Black male faces, and only White male faces.

The trustworthiness x dominance glmer results revealed a significant interaction between trustworthiness and dominance (^β = -0.650, SE = 0.214, z = -3.038, p = 0.002**). Fig. 6a shows low, average, and high dominance levels and their effects on trustworthiness. The X-axis represents how trustworthy a face appears, centered around the mean, so the main effect of dominance is intercepted at the average level of trustworthiness. The red line represents low dominance, exhibiting a positive trend: as trustworthiness increases while being paired with low dominance, the probability of a correct response increases. The green line represents average dominance, exhibiting no effect: trustworthiness does not have a meaningful effect when paired with average dominance. The blue line represents high dominance, exhibiting a negative trend:

as trustworthiness increases while being paired with high dominance, the probability of correct responses decreases (Fig. 6b). The lines intersected around the mean-centered trustworthiness value, suggesting that dominance-related differences in memory accuracy have the most effect at either extreme of trustworthiness, and the weakest effect for faces with average trustworthiness.

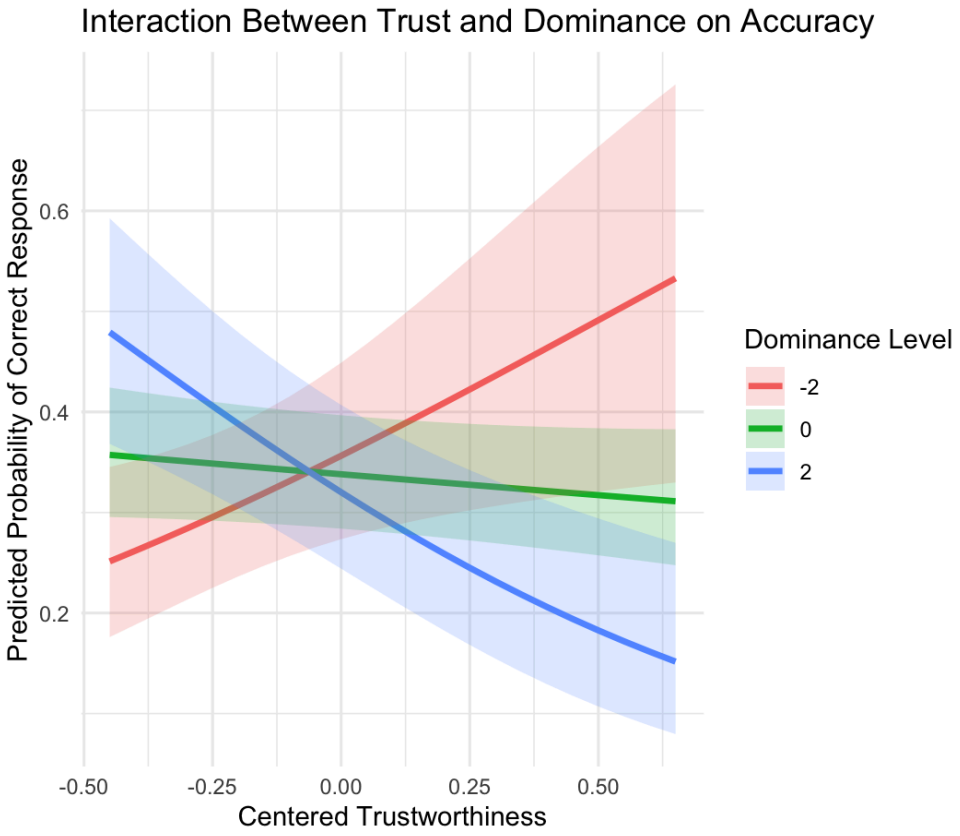

Figure 6a. Interaction between trustworthiness and dominance on recall accuracy of all faces.

Predicted probability of correct recall plotted as a function of centered trustworthiness scores, separated by dominance levels.

Dominance level	Trustworthiness effect
Low (-2)	Positive (accuracy increases with trust)
Average (0)	No effect
High (+2)	Negative (accuracy decreases with trust)

Figure 6b. Summary of interaction between dominance level and trustworthiness effect on memory accuracy.

The interaction was repeated with the dataset restricted to Black male faces only and White male faces only to determine if the relationship between trustworthiness, dominance, and memory accuracy varied by race (Fig. 6c).

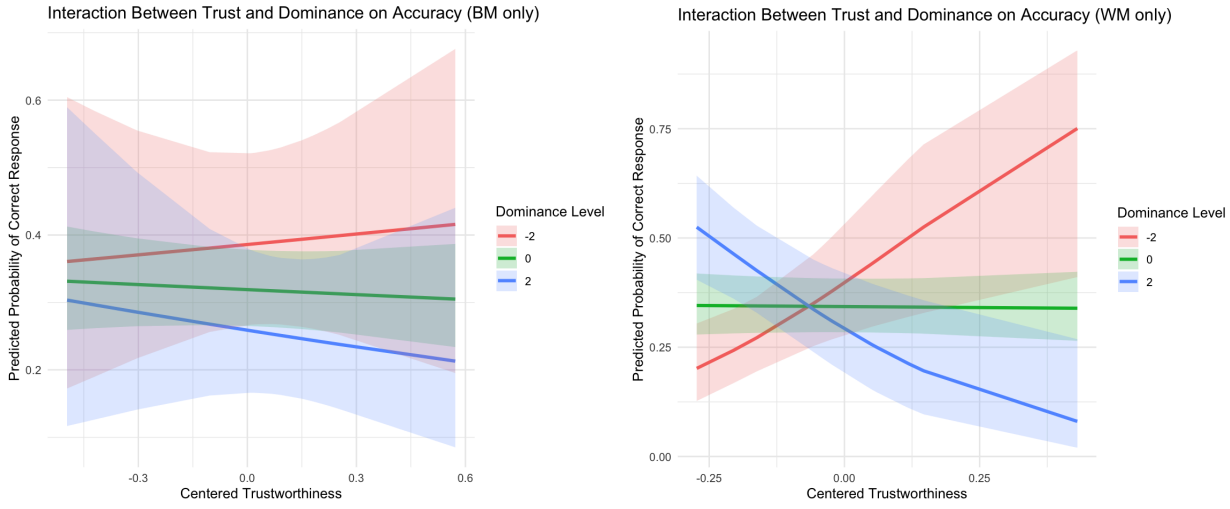

Figure 6c. Interactions between trustworthiness and dominance on recall accuracy, shown separately for Black male faces only (left) and White male faces only (right).

Predicted probability of correct recall plotted as a function of centered trustworthiness scores, separated by dominance levels, subset for Black male faces only, and White male faces only.

For Black male faces only, the trustworthiness x dominance glmer results revealed a non-significant interaction between trustworthiness and dominance (^β = -0.166, SE = 0.434, z = -0.382, p = 0.703), meaning that trustworthiness and dominance did not meaningfully interact to influence memory accuracy for Black male faces. Although the interaction patterns were similar to the overall pattern, the slopes for low dominance (red) and high dominance (blue) were less steep compared to the full dataset. The slightly positive trend for the low dominance level, paired with increasing trustworthiness, increases memory accuracy. The slightly negative trend for the high dominance level, paired with increasing trustworthiness, decreases memory accuracy. This suggests a weaker, but similar interaction pattern for Black male faces.

However, for White male faces only, the trustworthiness x dominance glmer results revealed a significant interaction between trustworthiness and dominance (^β = -1.779, SE = 0.600, z = -2.968, p = 0.003**). The slopes for all low dominance (red) and high dominance (blue) were much steeper compared to the Black male faces only interaction, and the interaction patterns replicate the overall pattern (Fig. 6a). The significant interaction suggests that among White male faces, trait impressions of trustworthiness and dominance strongly influenced memory accuracy: high trustworthiness, low dominance faces were better remembered, whereas high trustworthiness, high dominance faces were more likely to be misremembered.

II. Univariate Results

Memory Outcomes: Hippocampal Activity

To examine differences in hippocampal encoding across the three memory outcomes to examine whether neural activity during encoding could predict subsequent memory performance, average beta values were compared. The trial outcomes were already known as that is what participants answered correctly/incorrectly during the retrieval phase of the WSW task. As shown in Fig. 7, within-category errors were associated with higher hippocampal activation than other conditions, but a repeated-measures ANOVA confirmed there were no significant differences in hippocampal activation across memory conditions ($F = 0.269$, $p = 0.765$). The small effect size (0.006) means that 0.66% of the variance in beta values is due to the differences in memory conditions, so the conditions explain almost none of the variance.

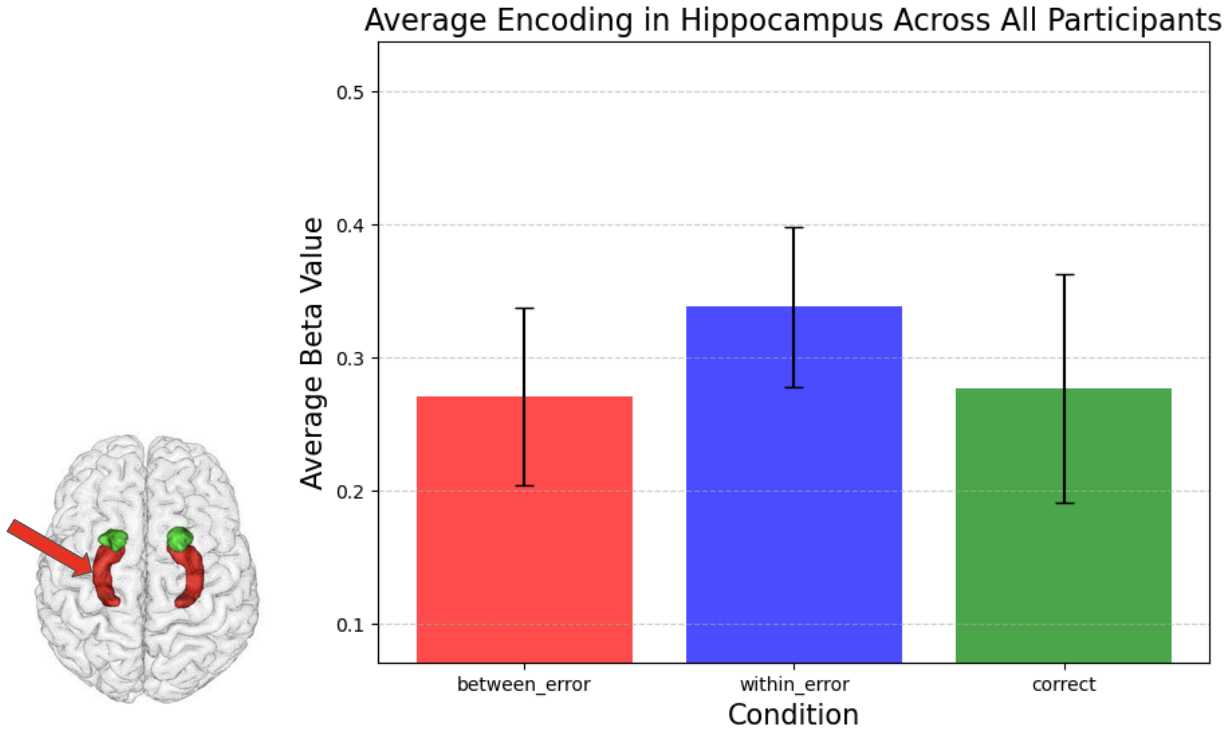

Figure 7. Hippocampal activation during encoding by memory outcome.

Mean beta values in the hippocampus during encoding trials that later resulted in between-category errors, within-category errors, and correct recall. Error bars represent standard error.

Follow-up pairwise t-tests revealed no significant differences between the three conditions:

- Correct versus Within: $t = -0.5840$, $p = 0.5617$
- Correct versus Between: $t = 0.053$, $p = 0.9581$
- Within versus Between: $t = 0.750$, $p = 0.4565$

Existing research by Eichenbaum et al. (1992) established that greater hippocampal activity during encoding is associated with better memory recall, so we would have expected the greatest activity for the correct memory outcome. However, while the results did not support this, they were also insignificant. Based on these results, there is no evidence that average hippocampal activation during encoding differentiates between whether a face would be later remembered or misremembered.

Memory Outcomes: Amygdala Activity

The same analysis was carried out in the amygdala (Fig. 8). Although it can be noted here that there is the greatest activation for the correct memory outcome, the repeated-measures ANOVA revealed that all results were insignificant ($F = 0.3049$, $p = 0.7379$).

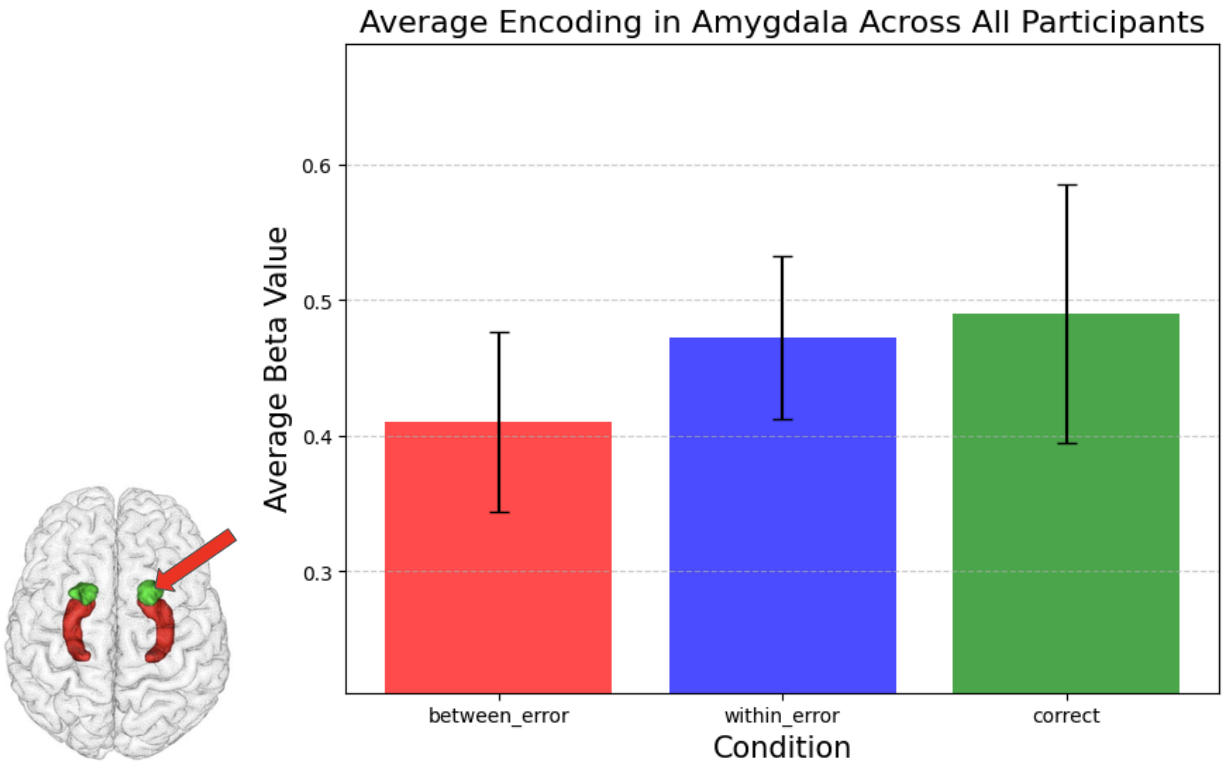

Figure 8. Amygdala activation during encoding by memory outcome.

Mean beta values in the amygdala during encoding trials that later resulted in between-category errors, within-category errors, and correct recall. Error bars represent standard error.

Follow-up pairwise t-tests revealed no significant differences between the three conditions:

- Correct versus Within: t = 0.696., p = 0.4888
- Correct versus Between: t = 0.681., p = 0.4988
- Within versus Between: t = 0.149., p = 0.8818

Neural Activity During Category Localizer Task

To investigate the hippocampal and amygdala neural activity when viewing faces rated along the trustworthiness dimension, as well as the dominant dimension, a median split was performed using the existing CFD trait ratings. The median split was an overall split based on ALL face ratings, instead of a subject-level split based only on faces that each participant saw, establishing the same threshold for everyone. This means that there was a consistent definition of trustworthy versus untrustworthy and dominant versus submissive across all participants. Initially, a subject-specific median split was performed; however, because it did not yield significant results, an overall median split was used instead, which produced significant findings.

To reiterate, this neural activity was collected during the category localizer task, which preceded the WSW task.

Hippocampus: Trustworthy versus Untrustworthy

The analysis was split by racial groups: All faces, Black male faces only, and White male faces only. This was done to examine if social group membership might influence how trait impressions are neurally encoded, as we observed significant differences that related to racial groups in the WSW behavioral analysis. There was no significance across all three face groups in their hippocampal activity by trustworthy levels (All faces: $p = 0.4361$, Black males only: $p = 0.2115$, White males only: $p = 0.8533$) (Fig. 9). However, it is important to point out that there is an inverse activation pattern between the Black males only group and the White males only group: there is greater activation for Black male faces rated higher on the trustworthiness dimension while the opposite is true for White male faces.

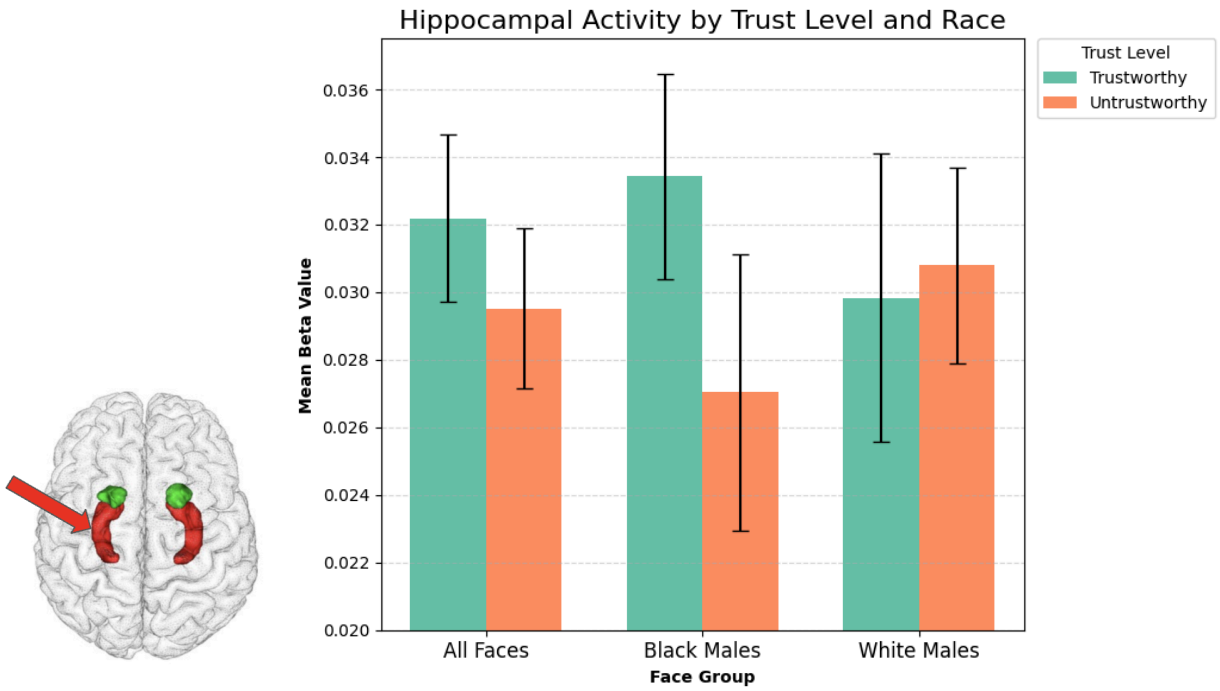

Figure 9. Hippocampal activation by trustworthiness level and race group.

Mean hippocampal activation during encoding of trustworthy versus untrustworthy faces across all faces, Black males, and White males. Error bars represent standard error.

Hippocampus: Dominant versus Submissive

When splitting the faces along the dominance dimension by dominant versus submissive, there was no significant difference in hippocampal activity across all three face groups (All faces: $p = 0.8531$, Black males only: $p = 0.7346$, White males only: $p = 0.5879$) (Fig. 10). Similar to the trustworthy versus untrustworthy level split in Fig. 7, the same inverse activation pattern was found between the Black males only group and the White males only group.

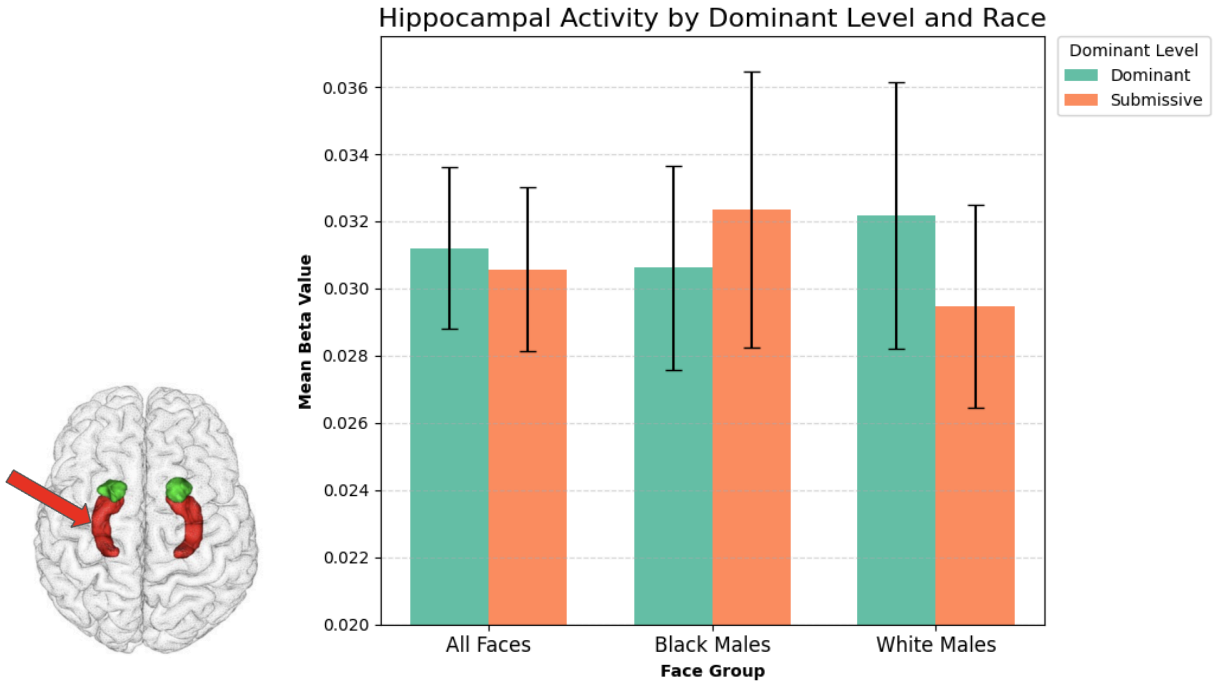

Figure 10. Hippocampal activation by dominant level and race group.

Mean hippocampal activation during encoding of dominant versus submissive faces across all faces, Black males, and White males. Error bars represent standard error.

Amygdala: Trustworthy versus Untrustworthy

In the amygdala, significant differences were observed when exploring neural activity by trustworthiness level and race (Fig. 11). There was a significant difference in activity level in the Black male only group ($p = 0.0324$*) where there was greater amygdala activity for higher trustworthy Black faces compared to faces rated as more untrustworthy. This suggests that Black faces on the higher side of the trustworthiness dimension evoke stronger emotional salience during encoding. Near significance was observed in the White male-only group ($p = 0.0697$), where there was greater amygdala activity for untrustworthy-rated White faces compared to higher trustworthy faces. No significance was found in the All faces group ($p = 0.7212$).

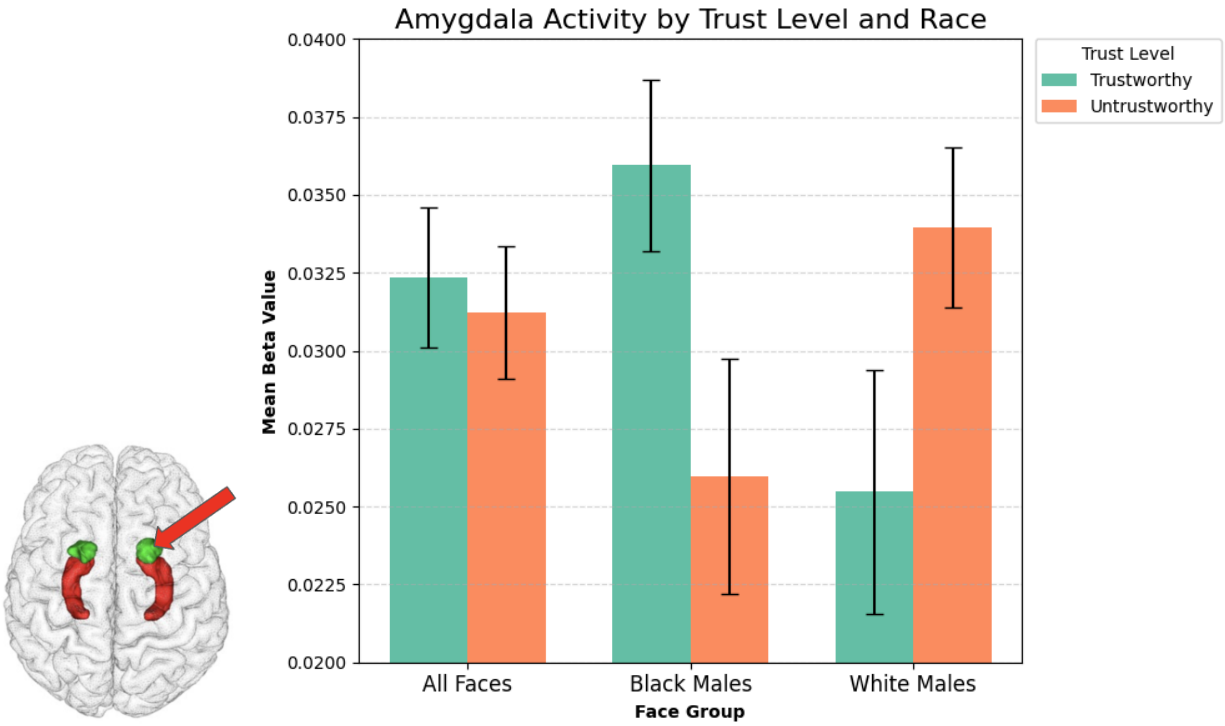

Figure 11. Amygdala activation by trustworthiness level and race group.

Mean amygdala activation during encoding of trustworthy versus untrustworthy faces across all faces, Black males, and White males. Error bars represent standard error.

Again, an inverse pattern of activity is found between the Black male faces only group and the White male faces only group.

Amygdala: Dominant versus Submissive

When split by dominant versus submissive levels along the dominance dimension, near significance was found in the White males face group ($p = 0.0894$) (Fig. 12). Greater amygdala activation was exhibited for dominant-rated White faces compared to submissive White faces. Although there was no significance found for the Black male-only group ($p = 0.7348$), there is

the same inverse activation pattern that was previously found in the previous plots. No significance was found for the All faces group as well.

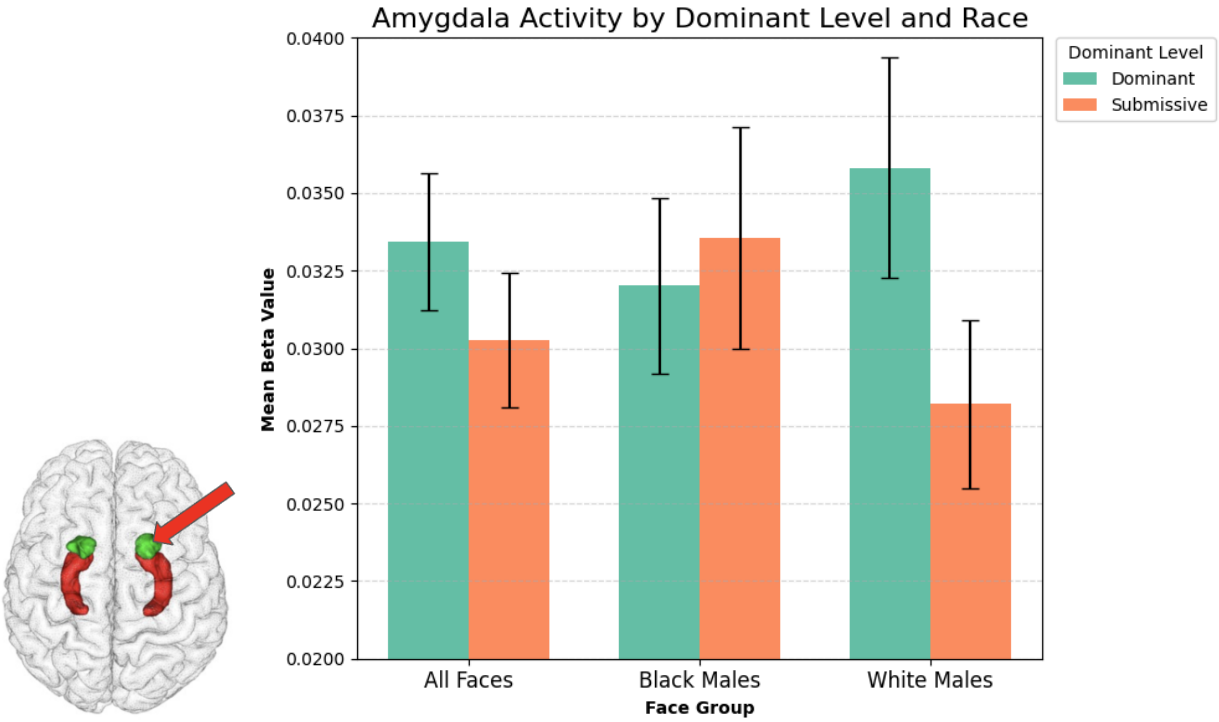

Figure 12. Amygdala activation by dominant level and race group.

Mean amygdala activation during encoding of dominant versus submissive faces across all faces, Black males, and White males. Error bars represent standard error.

III. Brain-Behavior Correlations

The brain-behavior correlations were only performed for trustworthiness since there was only a near-significant effect of trustworthiness on memory (Fig. 4a), and not dominance (Fig. 5b). The correlations were split up by race once again due to the univariate results in the univariate analysis of the category localizer task.

The brain-behavior correlation investigated the relationship between participants' trustworthiness slopes (X-axis), which were taken from the WSW behavioral results (Fig. 3), and the difference between mean trustworthy activation and untrustworthy activation (Y-axis), which were determined during the univariate analysis above. For the X-axis, a higher slope value means that the participant is more likely to remember faces that are rated higher along the trustworthiness dimension, while a lower slope value means that the participant is less likely to remember faces that are rated higher along the trustworthiness dimension. For the Y-axis, the difference of mean(trustworthy) - mean(untrustworthy) is used as the brain has different activation responses for trustworthy versus untrustworthy faces, and this difference score mitigates that difference in activation.

Hippocampus

When accounting for all faces, a very weak, flat, negative trend was observed (Fig. 13), and Pearson's correlation coefficient indicated a small, non-significant association ($r = -0.0863$, $p = 0.6502$). The extent to which trustworthiness influenced memory did not predict differential hippocampal activation between trustworthy-rated and untrustworthy-rated faces during encoding.

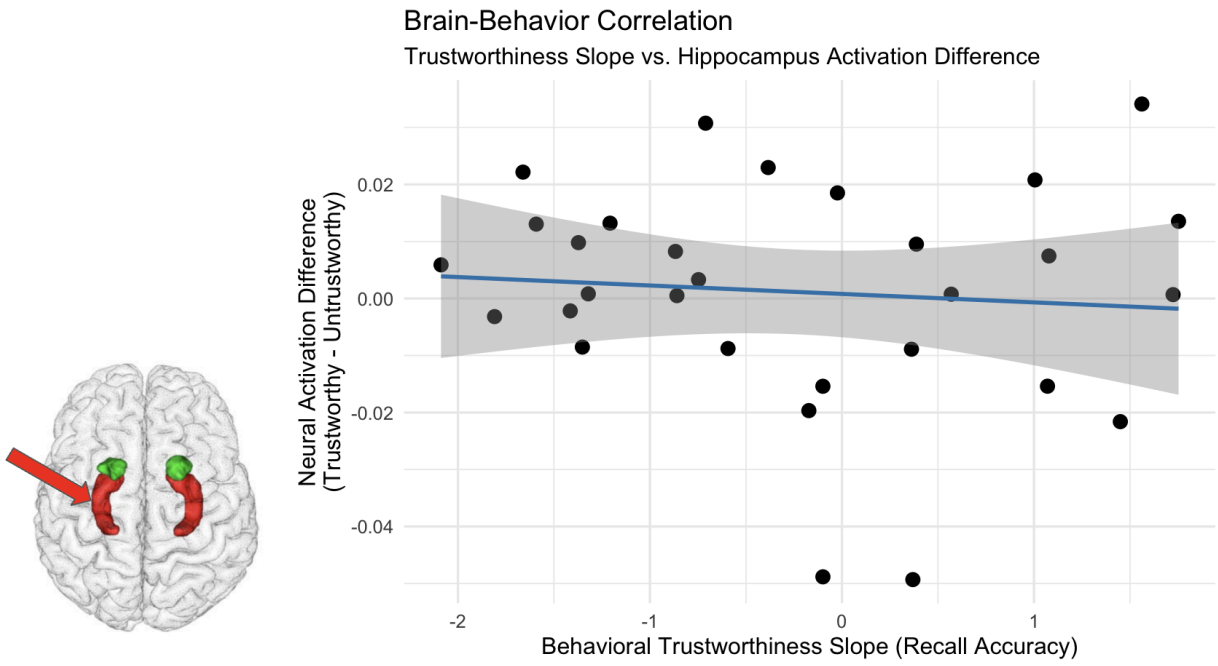

Figure 13. Hippocampus brain-behavior correlation (all faces).

The relationship between behavioral trustworthiness effects on memory and hippocampal activation differences mean(trustworthy) - mean(untrustworthy) across all faces.

When restricting analysis to Black male faces only, a non-significant, slightly positive trend was observed ($r = 0.2868$, $p = 0.1244$) (Fig.14). Participants who are better at recalling trustworthiness (higher slope value) tended to exhibit greater hippocampal activity. While this positive trend suggests the possibility of stronger encoding for socially positive traits in Black faces for some participants, it is non-significant and there is wide variability.

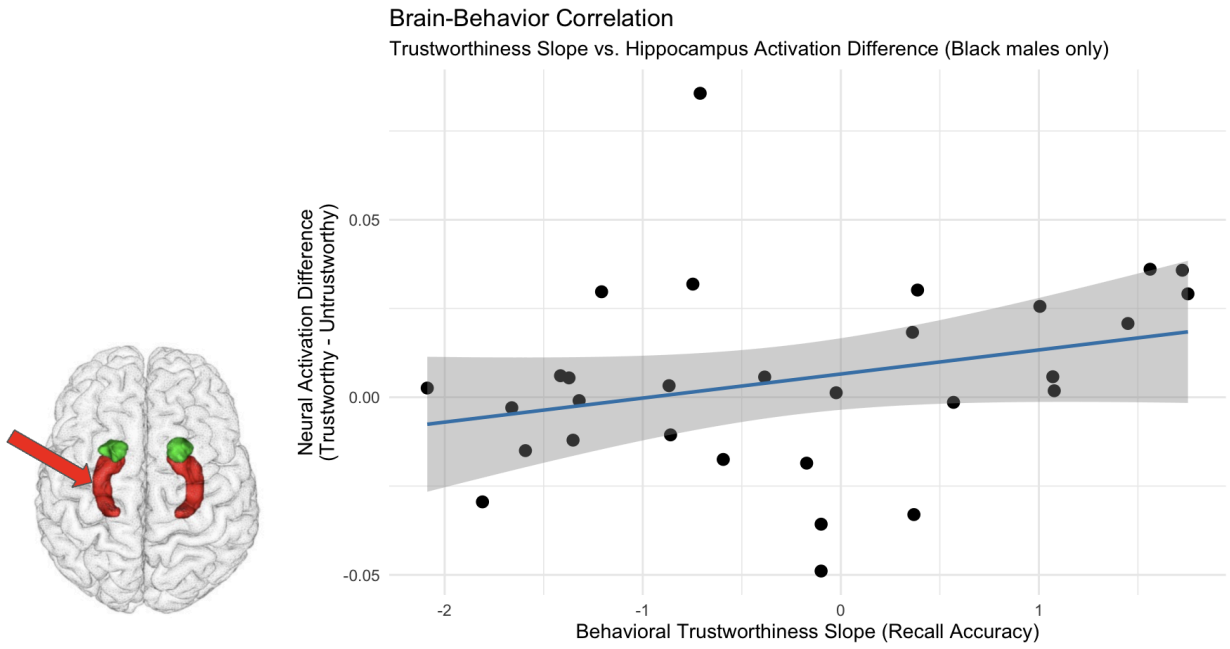

Figure 14. Hippocampus brain-behavior correlation (Black males only).

Scatterplot showing the relationship between behavioral trustworthiness effects on memory and hippocampal activation differences mean(trustworthy) - mean(untrustworthy) for Black males only.

When restricting analysis to White male faces only, a non-significant inverse trend was observed ($r = -0.2978$, $p = 0.11$) (Fig. 15). Participants who showed stronger memory for White male faces rated higher along the trustworthiness dimension were associated with less hippocampal activation. This pattern aligns with the possibility that, for White male faces, trustworthiness-related memory effects may not be correlated with hippocampal encoding mechanisms.

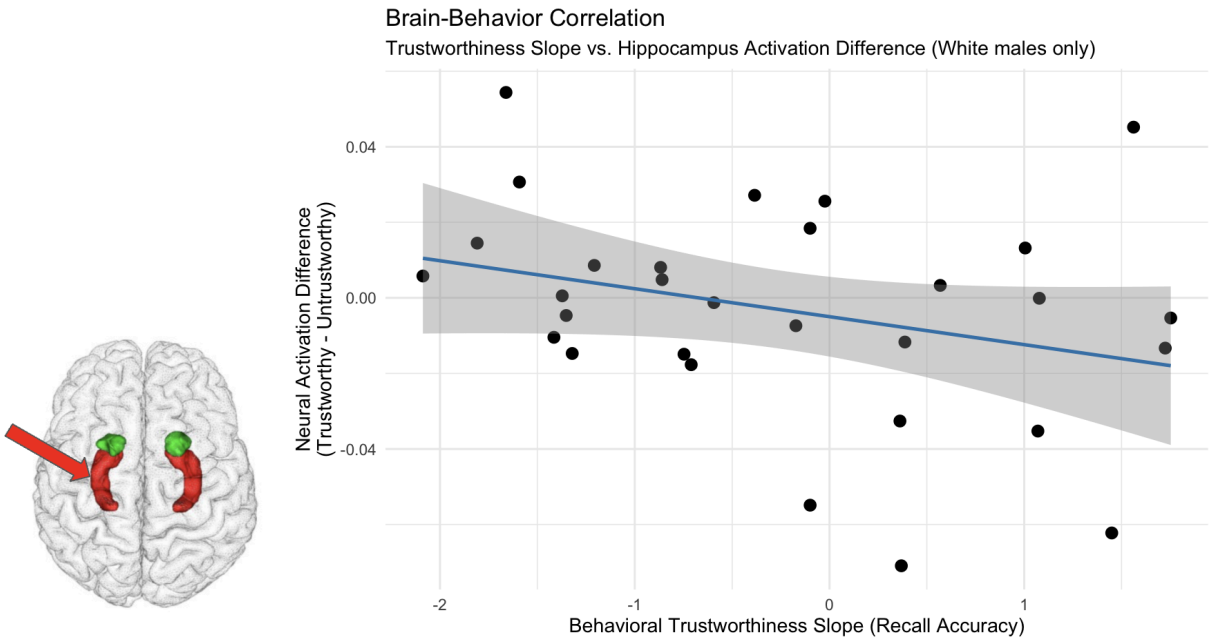

Figure 15. Hippocampus brain-behavior correlation (White males only).

Scatterplot showing the relationship between behavioral trustworthiness effects on memory and hippocampal activation differences mean(trustworthy) - mean(untrustworthy) for White males only.

This inverse pattern, compared to Black male faces only, could also suggest that different neural processes may underlie memory biases for different racial groups.

Amygdala

When exploring all faces, a very weak, flat, negative trend was observed in the scatterplot ($r = -0.1765$, $p = 0.3508$) (Fig. 16), similar to the hippocampus in Fig.13. Participants better at recalling trustworthiness (higher slope value) showed less amygdala activation differences between faces rated on either extreme of the trustworthiness dimension. This implies that higher

recall is associated with greater amygdala activation for untrustworthy-rated faces. Again, no meaningful link can be concluded between how strongly trustworthiness influenced memory and how differentially the amygdala responded to trustworthiness during encoding.

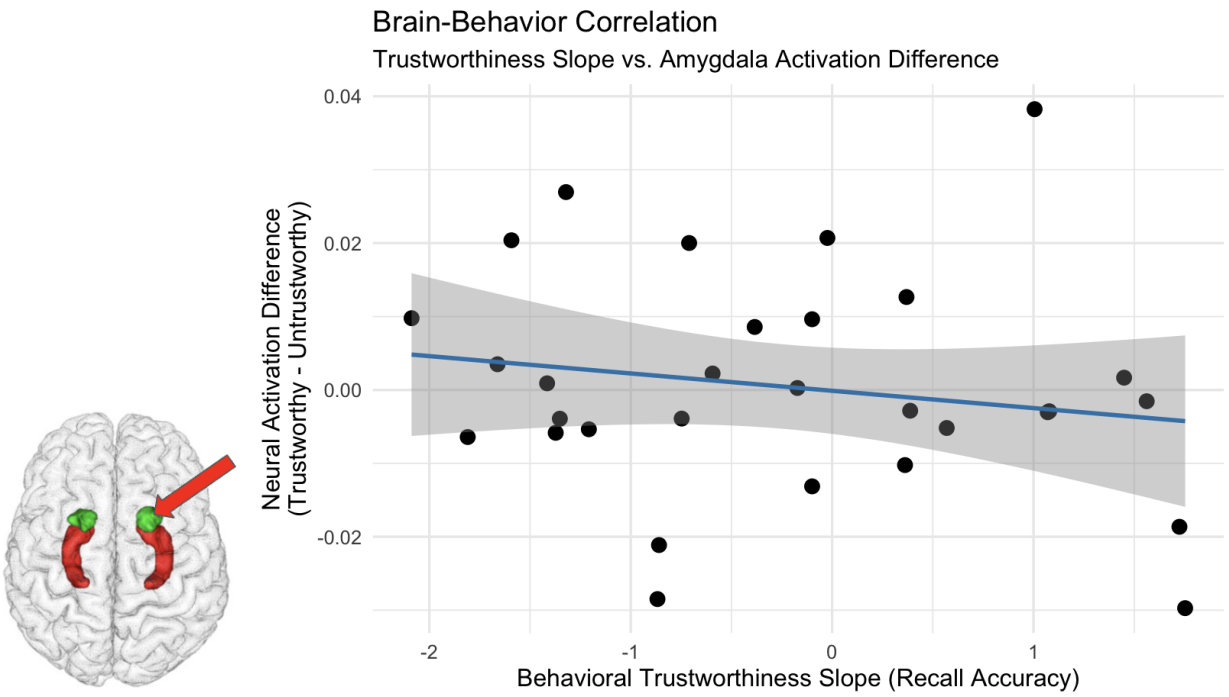

Figure 16. Amygdala brain-behavior correlation (all faces).

Scatterplot showing the relationship between behavioral trustworthiness effects on memory and amygdala activation differences mean(trustworthy) - mean(untrustworthy) across all faces.

When restricting analysis to Black male faces only, a non-significant, positive trend can be observed ($r = 0.1523$, $p = 0.4218$) (Fig. 17), similar to the hippocampal activity above in Fig. 14. Higher recall accuracy for trustworthy-rated Black faces is associated with slightly greater amygdala activation for those trustworthy-rated faces compared to untrustworthy-rated Black male faces. This aligns with the idea that higher trustworthy faces, particularly when they

challenge stereotypes, may trigger emotional salience that supports memory encoding (Sergerie et al., 2008). However, no conclusion can be made due to the weak and non-significant effect.

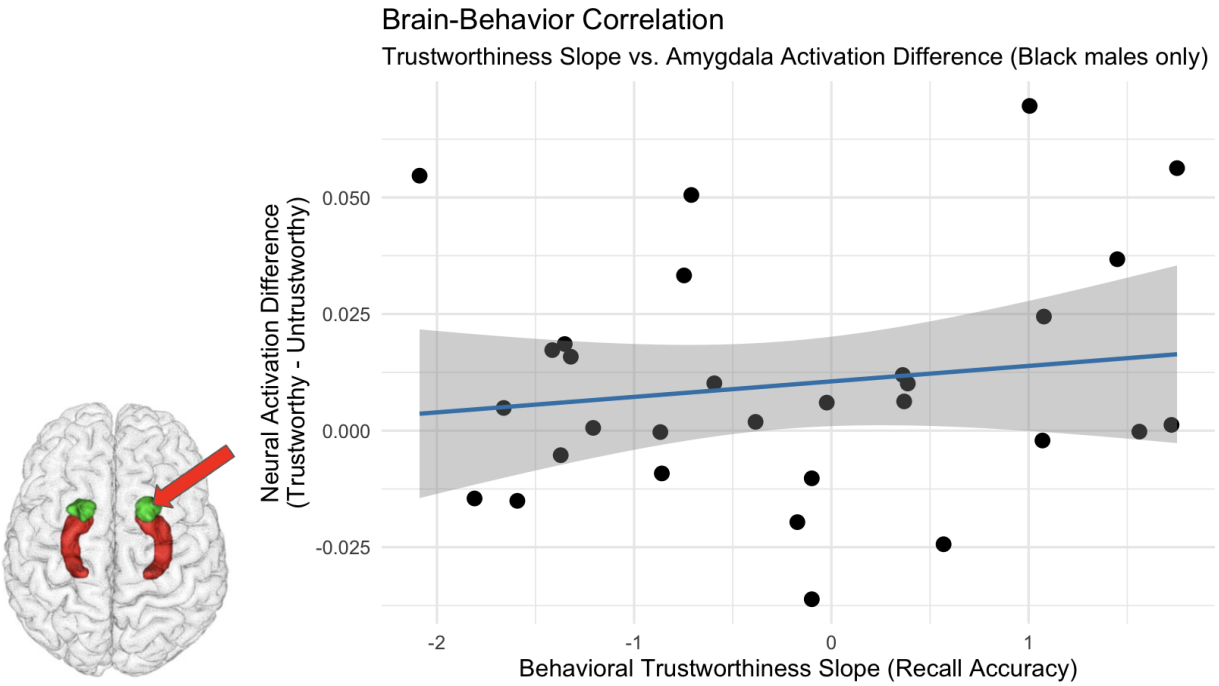

Figure 17. Amygdala brain-behavior correlation (Black males only).

Scatterplot showing the relationship between behavioral trustworthiness effects on memory and amygdala activation differences mean(trustworthy) - mean(untrustworthy) across Black males only.

When restricting analysis to White male faces only, a weak, negative trend was observed (r = -0.2960, p = 0.1122) (Fig. 18). Higher recall accuracy for trustworthy-rated White faces was associated with lower amygdala activation for those trustworthy-rated faces. This suggests that untrustworthy-rated White male faces stood out more, possibly due to violating social expectations and stronger emotional salience.

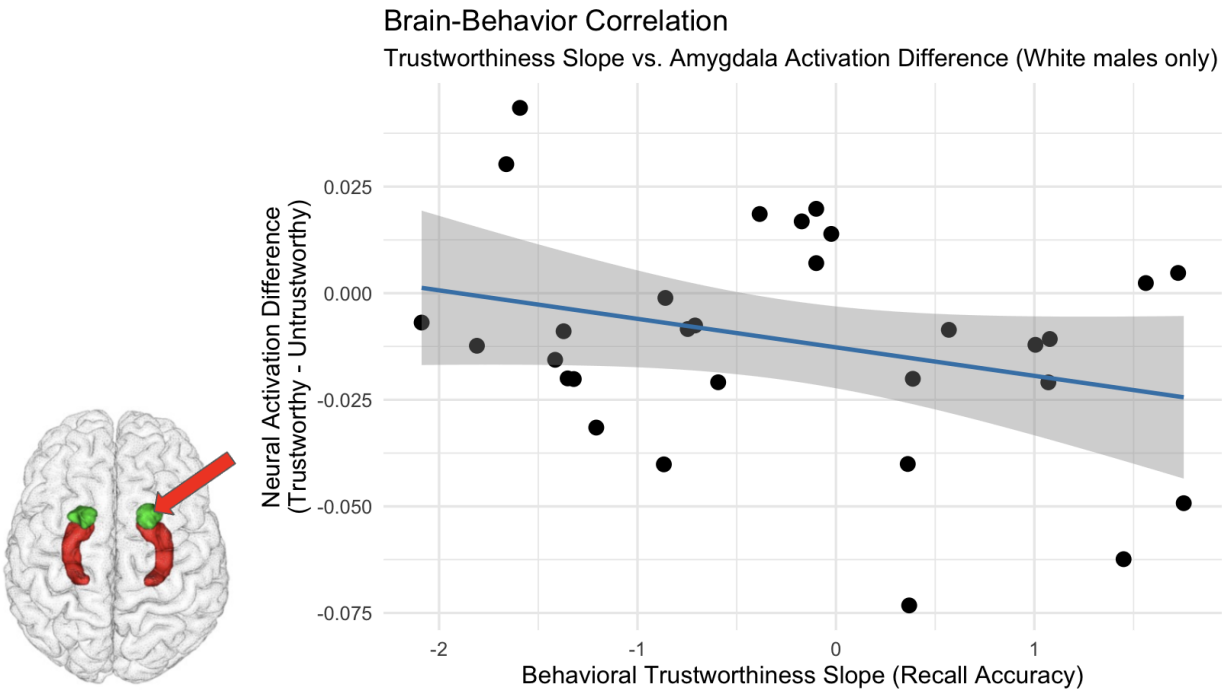

Figure 18. Amygdala brain-behavior correlation (White males only).

Scatterplot showing the relationship between behavioral trustworthiness effects on memory and amygdala activation differences mean(trustworthy) - mean(untrustworthy) across White males only.

Similarly to the hippocampus (Fig. 14, 15), the correlations between trustworthiness and amygdala activation are inverse for Black male faces only (Fig. 17) and White male faces only (Fig. 18). The directional difference may suggest potential race-based variations in how social impressions modulate amygdala activity during encoding but, again, there were no significant results.

Discussion

The present study investigates how trait impressions of trustworthiness and dominance influence face memory performance and neural activity in the hippocampus and amygdala. Overall, the results partially supported the hypotheses, providing nuanced insights into the complexity of social trait impressions, memory, and neural mechanisms.

The results provide support for H1, further establishing that untrustworthy-rated faces are remembered better than trustworthy-rated faces. This aligns with prior research that untrustworthy or threatening cues are prioritized in memory, likely due to evolutionary and social relevance (Rule et al., 2012). Although the negative slope between trustworthiness and memory accuracy was marginal ($p = 0.0539$), the trend suggests that individuals are more likely to remember faces that signal danger. However, post hoc exploratory analyses revealed that this effect did not hold when examining Black and White male faces separately, likely due to the small sample size that decreased in statistical power when subset again.

In contrast, dominance was not found to significantly predict memory accuracy and, therefore, H2 was not supported. Although the positive trend between dominance and memory accuracy supports prior research that dominant faces are more likely to be remembered (Rule et al., 2010), it was not significant ($p = 0.422$). However, exploratory race-specific analyses found a significant positive relationship among White male faces only ($p = 0.036$*) — more dominant-rated White male faces were remembered better. This is especially noteworthy given that dominance has been relatively underexplored compared to trustworthiness. One possible explanation is that higher dominant White male faces may activate culturally salient schemas about authority or threat, which could influence memory encoding, and these associations may not generalize to other racial groups due to differing social stereotypes. This aligns with previous

research that trait impressions interact with social categorization (Freeman & Chwe, 2024). It is again worth noting that these race-specific analyses were not initially planned and are exploratory.

However, the results did identify that dominance significantly moderates the salience of trustworthiness in memory. At low dominance levels, high trustworthiness improved memory; at high dominance levels, higher trustworthiness impaired memory. This interaction illustrates that trustworthiness and dominance work together to influence memory, indicating that social traits are not processed independently of one another but together. When the interaction was examined separately by Black males only and White males only, differences emerged. For Black male faces only, the interaction between trustworthiness and dominance on memory recall accuracy was not significant, and the slopes were flatter — trait impressions had less influence on memory for Black faces. On the other hand, for White male faces only, the interaction between trustworthiness and dominance on memory recall accuracy was significant and incredibly similar to that of the overall interaction patterns. This suggests that social trait impressions had a greater impact on memory for White faces. Together, the results point toward that the influence social trait impressions have on memory varies by race.

Hippocampal and amygdala activation was not found to predict memory performance, and so H3 was not supported. We should have been able to accept this hypothesis as Eichenbaum et al. (1992) found that greater hippocampal activity during encoding predicts later memory retrieval and Phelps (2004) found that greater amygdala activity during encoding is associated with better memory recall. While the greatest amygdala activity was found for correct recall, this was not true for hippocampal activity, albeit insignificant results for both the amygdala ($p =$

0.7379) and the hippocampus (p = 0.7651). This was likely due to noise, the limited sample size, or insufficient trials after splitting them by memory outcomes.

In terms of trait-based neural effects, no significant trustworthiness effects were found in hippocampus (p = 0.4361) or amygdala (p = 0.7212) activity. However, exploratory race-specific analyses partially support H4, which stated that untrustworthy-rated faces will elicit stronger amygdala activation. There was nearly significantly greater amygdala activation for untrustworthy-rated White faces compared to trustworthy-rated White faces (p = 0.0697). Again, if the sample size were larger, this would have reached significance. Interestingly, though, trustworthiness had opposite amygdala activation for Black faces: trustworthy-rated Black faces elicited significantly greater amygdala activation than untrustworthy-rated Black faces (p = 0.0324*). These inverse neural activity levels suggest that facial impressions have different neural signatures across racial groups in the amygdala, potentially reflecting differences in social schemas or cultural stereotypes. More specifically, when individuals view an untrustworthy-rated Black face, they might want to correct bias by labeling the face as trustworthy-rated so that they do not appear racially prejudiced or reinforce negative stereotypes, especially in a research setting where they may feel scrutinized.

Lastly, H5 cannot be supported, which states that differences in neural activation related to trustworthiness in the hippocampus and amygdala will be associated with memory performance. The brain-behavior correlations did not find any significant relationships between trustworthiness ratings and neural activation in both the hippocampus (p = 0.6502) and the amygdala (p = 0.3508). But, it might be possible that facial impressions engage the neural basis of memory encoding differently across racial groups. While exploratory analysis split by race did find possible race-based differences — a small, positive trend in the amygdala for Black faces

(Fig. 17) versus a small, negative trend in the amygdala for White faces (Fig. 18) — the results were insignificant. Overall, the results find that neural activity of varying degrees of trustworthiness during encoding was not predictive of memory.

Limitations

There are several limitations that need to be acknowledged. First, the original WSW paradigm was not designed to investigate face impressions — it was to explore the neural representations of racial social categorization and memory. The WSW dataset was adapted to the present study using existing CFD normative ratings to explore possibilities related to trustworthiness and dominance. Otherwise, stimuli that better represented trustworthiness and dominance would have been selected.

Furthermore, the CFD ratings are also a limitation concerning the interpretation of the race-based differences in neural activity. Although the findings saw the inverse neural activation pattern in both the hippocampus and amygdala for Black versus White male faces, this could have been due to biases in the CFD facial trait ratings used. The participants providing these ratings could have been influenced by social desirability concerns, adjusting their trustworthiness and dominance ratings to avoid appearing racially biased. Thus, the neural activation differences may not reflect genuine perceptual differences, but could be partially driven by social desirability and expectations.

Another major limitation was the small sample size, which limited statistical power for all analyses. It should also be noted that the race-specific analyses were exploratory and not initially included in the study, so these findings should be interpreted with a grain of salt. Finally, while univariate fMRI analyses provide useful insights, they do not capture more than where and how much neural activity is occurring.

Future Direction

While the present study provides insights into how social trait impressions, particularly trustworthiness and dominance, influence memory and neural activation, several future directions could extend and refine these findings.

First, the sample size could be increased for statistical power and generalizability. This study had only 31 participants, and many findings were near significance. A larger, more diverse participant pool would have pushed the results to significance and helped to validate the observed effects. Additionally, it would also better capture individual differences across demographic groups

Second, future work could move beyond univariate analysis. Multivariate analysis could provide a deeper, more comprehensive understanding of how trustworthiness and dominance are encoded in the brain. Multivoxel pattern analysis (MVPA) could determine whether patterns of activity in the hippocampus and amygdala can distinguish faces along the trustworthiness and dominance dimensions. Representational similarity analysis (RSA) could also explore how the neural representations of trustworthiness and dominance differ or overlap, revealing if they are encoded distinctly or along a shared dimension.

Beyond neuroimaging, future behavioral research could examine how trait-based memory biases influence real-world outcomes, such as eyewitness identification. For example, a face lineup memory study could assess whether individuals are more likely to misidentify or recall faces based on perceived trustworthiness or dominance, providing direct implications for the legal system.

And, with the increasing everyday relevance of virtual assistants and avatars, future studies could explore how the facial features of those artificial faces influence user trust and

engagement. Understanding how impressions of trustworthiness or dominance in faces affect memory and decision-making can provide direction for technology design, education, and human-computer interaction.

These future directions are promising avenues for delving into the behavioral and neural mechanisms underlying face impressions and memory, which can then inform new theoretical models as well as real-world applications.

Conclusion

In summary, this study investigated how social trait impressions — specifically trustworthiness and dominance — influence memory performance and neural activation during encoding in the hippocampus and amygdala. Behavioral results found that faces rated as untrustworthy along the trustworthiness dimension were more likely to be remembered than trustworthy faces, consistent with evolutionary perspectives of threat detection. An interaction was also found between trustworthiness and dominance, suggesting that trait impressions are considered holistically rather than independently. Neuroimaging analyses revealed that although there was no significant overall relationship between trait impressions and hippocampal or amygdala activation, race-specific activity differences were observed. Specifically, there were inverse neural responses to trustworthiness and dominance between Black male faces versus White male faces. The brain-behavior correlations, although non-significant, imply that the relationship between neural encoding of social impressions and memory performance is incredibly nuanced and likely influenced by other factors beyond trait impressions alone. These findings deepen our understanding of how automatic social judgments can influence memory, with consequential, real-life implications. Future work with larger, tailored datasets and

multivariate analysis could help further shed light on the neural mechanisms responsible for facial impressions and memory.

References

Adams Jr, R. B., Gordon, H. L., Baird, A. A., Ambady, N., & Kleck, R. E. (2003). Effects of gaze on amygdala sensitivity to anger and fear faces. Science, 300(5625), 1536-1536.

Adams RB, Ambady N, Macrae N, Kleck RE (2006) Emotional expressions forecastapproach-avoidance behavior. Motiv Emot 30:179 –188

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. (2006). Reward-motivated learning: mesolimbic activation precedes memory formation. Neuron, 50(3), 507-517.

Adolphs, R., Tranel, D., & Damasio, A. R. (1998). The human amygdala in social judgment. Nature, 393(6684), 470-474.

Adolphs, R., Tranel, D., Hamann, S., Young, A. W., Calder, A. J., Phelps, E. A., ... & Damasio, A. R. (1999). Recognition of facial emotion in nine individuals with bilateral amygdala damage. Neuropsychologia, 37(10), 1111-1117.

Adolphs, R. (2010). What does the amygdala contribute to social cognition? Year Cogn. Neurosci. 1191, 42–61. doi: 10.1111/j.1749-6632.2010.05445.x

Anderson, C., & Kilduff, G. J. (2009). Why do dominant personalities attain influence in face-to-face groups? The competence-signaling effects of trait dominance. Journal of personality and social psychology, 96(2), 491.

Bar, M., Neta, M., & Linz, H. (2006). Very first impressions. Emotion, 6(2), 269.

Barnes, C. A. (1979). Memory deficits associated with senescence: a neurophysiological and behavioral study in the rat. Journal of comparative and physiological psychology, 93(1), 74.

Bell, R., Mieth, L., & Buchner, A. (2015). Appearance-based first impressions and person memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 41(2), 456.

Blair, R. J. R., Morris, J. S., Frith, C. D., Perrett, D. I., & Dolan, R. J. (1999). Dissociable neural responses to facial expressions of sadness and anger. Brain, 122(5), 883-893.

Brewer, M. B. (1981). "Ethnocentrism and its role in interpersonal trust," in Scientific Inquiry and the Social Sciences, eds M. B. Brewer and B. E. Collins (San Francisco: Jossey-Bass), 345–359.

Cañadas, E., Rodríguez-Bailón, R., & Lupiáñez, J. (2015). The effect of social categorization on trust decisions in a trust game paradigm. Frontiers in psychology, 6, 1568.

Cansino, S., Maquet, P., Dolan, R. J., & Rugg, M. D. (2002). Brain activity underlying encoding and retrieval of source memory. Cerebral cortex, 12(10), 1048-1056.

Cao, R., Li, X., Brandmeir, N. J., & Wang, S. (2021). Encoding of facial features by single neurons in the human amygdala and hippocampus. Communications Biology, 4(1), 1394.

Cao, R., Lin, C., Hodge, J., Li, X., Todorov, A., Brandmeir, N. J., & Wang, S. (2022). A neuronal social trait space for first impressions in the human amygdala and hippocampus. Molecular Psychiatry, 27(8), 3501-3509.

Carré, J. M., McCormick, C. M., & Mondloch, C. J. (2009). Facial structure is a reliable cue of aggressive behavior. Psychological science, 20(10), 1194-1198.

Cassidy, B. S. (2020). Valenced appearance-behavior cues affect the extent of impression memory. Memory, 28(5), 642-654.

Cassidy, B. S., & Gutchess, A. H. (2012). Social relevance enhances memory for impressions in older adults. Memory, 20(4), 332-345.

Chwe, J. A. H., Vartiainen, H. I., & Freeman, J. B. (2024). A Multidimensional Neural Representation of Face Impressions. Journal of Neuroscience, 44(39).

Cosmides, L., & Tooby, J. (1989). Evolutionary psychology and the generation of culture, part II: Case study: A computational theory of social exchange. Ethology and sociobiology, 10(1-3), 51-97.

D'Argembeau, A., & Van der Linden, M. (2007). Facial expressions of emotion influence memory for facial identity in an automatic way. Emotion, 7, 507–515.

Davachi, L., & Wagner, A. D. (2002). Hippocampal contributions to episodic encoding: insights from relational and item-based learning. Journal of neurophysiology, 88(2), 982-990.

Devine, P. G. (1989). Stereotypes and prejudice: their automatic and controlled components. J. Pers. Soc. Psychol. 56, 5–18.

Dewhurst, S. A., Hay, D. C., & Wickham, L. H. (2005). Distinctiveness, typicality, and recollective experience in face recognition: A principal components analysis. Psychonomic Bulletin & Review, 12(6), 1032-1037.

Diano, M., Tamietto, M., Celeghin, A., Weiskrantz, L., Tatu, M. K., Bagnis, A., ... & Costa, T. (2017). Dynamic changes in amygdala psychophysiological connectivity reveal distinct neural networks for facial expressions of basic emotions. Scientific reports, 7(1), 45260.

Eberhardt JL, Davies PG, Purdie-Vaughns VJ, Johnson SL (2006) Looking deathworthy:Perceived stereotypicality of Black defendants predicts capital-sentencing outcomes.Psychol Sci 17:382–386.

Ekman, P. (1992). Are there basic emotions?.

Eichenbaum, H., Otto, T., & Cohen, N. J. (1992). The hippocampus—what does it do?. Behavioral and neural biology, 57(1), 2-36.

Engell, A. D., Haxby, J. V., & Todorov, A. (2007). Implicit trustworthiness decisions: automatic coding of face properties in the human amygdala. Journal of cognitive neuroscience, 19(9), 1508-1519.

Fenker, D. B., Schott, B. H., Richardson‐Klavehn, A., Heinze, H. J., & Düzel, E. (2005). Recapitulating emotional context: activity of amygdala, hippocampus and fusiform cortex during recollection and familiarity. European Journal of Neuroscience, 21(7), 1993-1999.

Fiske, S. T. (2003). "Five core social motives, plus or minus five," in Motivated Social Perception: The Ontario Symposium, Vol. 9, eds S. J. Spencer, S. Fein, M. Zanna, and J. Olson (Mahwah, NJ: Erlbaum), 233–246.

Fox, A. S., Oler, J. A., Tromp, D. P., Fudge, J. L., & Kalin, N. H. (2015). Extending the amygdala in theories of threat processing. Trends in neurosciences, 38(5), 319-329.

Freeman, J. B., & Chwe, J. A. (2024). Social Categorization: Looking Toward the Future.

Freeman, J. B., Stolier, R. M., Ingbretsen, Z. A., & Hehman, E. A. (2014). Amygdala responsivity to high-level social information from unseen faces. Journal of Neuroscience, 34(32), 10573-10581.

Fridlund AJ (1994) Human Facial Expression: An Evolutionary View (Academic, SanDiego).29.

Hamermesh, D. S., & Biddle, J. (1993). Beauty and the labor market.

Hassin, R., & Trope, Y. (2000). Facing faces: studies on the cognitive aspects of physiognomy. Journal of personality and social psychology, 78(5), 837.

Hou, C., & Liu, Z. (2019). The survival processing advantage of face: The memorization of the (un) trustworthy face contributes more to survival adaptation. Evolutionary Psychology, 17(2), 1474704919839726.

Iidaka, Tetsuya, Masao Omori, Tetsuhito Murata, Hirotaka Kosaka, Yoshiharu Yonekura, Tomohisa Okada, and Norihiro Sadato. "Neural interaction of the amygdala with the prefrontal and temporal cortices in the processing of facial expressions as revealed by fMRI." Journal of Cognitive Neuroscience 13, no. 8 (2001): 1035-1047.

Jaeger, B., & Jones, A. L. (2022). Which facial features are central in impression formation?. Social Psychological and Personality Science, 13(2), 553-561.

Joensen, B. H., Bush, D., Vivekananda, U., Horner, A. J., Bisby, J. A., Diehl, B., ... & Burgess, N. (2023). Hippocampal theta activity during encoding promotes subsequent associative memory in humans. Cerebral Cortex, 33(13), 8792-8802.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: a module in human extrastriate cortex specialized for face perception. Journal of neuroscience, 17(11), 4302-4311.

Kim, H., Somerville, L. H., Johnstone, T., Alexander, A. L., & Whalen, P. J. (2003). Inverse amygdala and medial prefrontal cortex responses to surprised faces. Neuroreport, 14(18), 2317-2322.

Klauer, K. C., & Wegener, I. (1998). Unraveling social categorization in the" who said what?" paradigm. Journal of Personality and Social Psychology, 75(5), 1155.

Kleider-Offutt, H. M., Bond, A. D., & Hegerty, S. E. (2017). Black stereotypical features: When a face type can get you in trouble. Current Directions in Psychological Science, 26(1), 28-33.

Levi, M., & Stoker, L. (2000). Political trust and trustworthiness. Annual review of political science, 3(1), 475-507.

Lin, C., Keles, U., & Adolphs, R. (2021). Four dimensions characterize attributions from faces using a representative set of English trait words. Nature Communications, 12(1), 5168.

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. Behavior research methods, 47, 1122-1135.

Mattarozzi, K., Todorov, A., & Codispoti, M. (2015). Memory for faces: the effect of facial appearance and the context in which the face is encountered. Psychological research, 79, 308-317.

Montepare, J. M., & Dobish, H. (2003). The contribution of emotion perceptions and their overgeneralizations to trait impressions. Journal of Nonverbal behavior, 27, 237-254.

Olivola, C. Y., Funk, F., & Todorov, A. (2014). Social attributions from faces bias human choices. Trends in cognitive sciences, 18(11), 566-570.

Olsson A. Ebert J.P. Banaji M.R. Phelps E.A. (2005). The role of social groups in the persistence of learned fear . Science , 309 , 785 – 7 .

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. Proceedings of the National Academy of Sciences, 105(32), 11087-11092.

Oosterhof, N. N., & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. Emotion, 9(1), 128.

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. Trends in cognitive sciences, 6(2), 93-102.

Phan, K. L., Wager, T., Taylor, S. F., & Liberzon, I. (2002). Functional neuroanatomy of emotion: a meta-analysis of emotion activation studies in PET and fMRI. Neuroimage, 16(2), 331-348.

Phelps, E. A. (2004). Human emotion and memory: interactions of the amygdala and hippocampal complex. Current opinion in neurobiology, 14(2), 198-202.

Phelps, E. A., & LeDoux, J. E. (2005). Contributions of the amygdala to emotion processing: from animal models to human behavior. Neuron, 48(2), 175-187.

Pietraszewski, D. (2021). The correct way to test the hypothesis that racial categorization is a byproduct of an evolved alliance-tracking capacity. Scientific reports, 11(1), 3404.

Porter, S., Ten Brinke, L., & Gustaw, C. (2010). Dangerous decisions: The impact of first impressions of trustworthiness on the evaluation of legal evidence and defendant culpability. Psychology, Crime & Law, 16(6), 477-491.

Prince, S. E., Dennis, N. A., & Cabeza, R. (2009). Encoding and retrieving faces and places: distinguishing process-and stimulus-specific differences in brain activity. Neuropsychologia, 47(11), 2282-2289.

Ren, J., Huang, F., Zhou, Y., Zhuang, L., Xu, J., Gao, C., ... & Luo, J. (2020). The function of the hippocampus and middle temporal gyrus in forming new associations and concepts during the processing of novelty and usefulness features in creative designs. Neuroimage, 214, 116751.

Richeson, J. A., & Shelton, J. N. (2003). When prejudice does not pay: Effects of interracial contact on executive function. Psychological science, 14(3), 287-290.

Rotter, J. B. (1980). Interpersonal trust, trustworthiness, and gullibility. American psychologist, 35(1), 1.

Rule, N. O., & Ambady, N. (2008). The face of success: Inferences from chief executive officers' appearance predict company profits. Psychological science, 19(2), 109-111.

Rule, N. O., Ambady, N., Adams Jr, R. B., Ozono, H., Nakashima, S., Yoshikawa, S., & Watabe, M. (2010). Polling the face: prediction and consensus across cultures. Journal of personality and social psychology, 98(1), 1.

Rule, N. O., Slepian, M. L., & Ambady, N. (2012). A memory advantage for untrustworthy faces. Cognition, 125(2), 207-218.

Ryan, L., Nadel, L., Keil, K., Putnam, K., Schnyer, D., Trouard, T., & Moscovitch, M. (2001). Hippocampal complex and retrieval of recent and very remote autobiographical memories: evidence from functional magnetic resonance imaging in neurologically intact people. Hippocampus, 11(6), 707-714.

Sander, D., Grafman, J., & Zalla, T. (2003). The human amygdala: an evolved system for relevance detection. Reviews in the Neurosciences, 14(4), 303-316.

Scharlemann, J. P., Eckel, C. C., Kacelnik, A., & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. Journal of Economic Psychology, 22(5), 617-640.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. Journal of neurology, neurosurgery, and psychiatry, 20(1), 11.

Sergerie, K., Chochol, C., & Armony, J. L. (2008). The role of the amygdala in emotional processing: a quantitative meta-analysis of functional neuroimaging studies. Neuroscience & Biobehavioral Reviews, 32(4), 811-830.

Sheng, J., Zhang, L., Liu, C., Liu, J., Feng, J., Zhou, Y., ... & Xue, G. (2022). Higher-dimensional neural representations predict better episodic memory. Science Advances, 8(16), eabm3829.

Squire, L. R., & Wixted, J. T. (2011). The cognitive neuroscience of human memory since HM. Annual review of neuroscience, 34(1), 259-288.

Squire, L. R., Genzel, L., Wixted, J. T., & Morris, R. G. (2015). Memory consolidation. Cold Spring Harbor perspectives in biology, 7(8), a021766.

Stangor, C., Lynch, L., Duan, C., & Glas, B. (1992). Categorization of individuals on the basis of multiple social features. Journal of Personality and social Psychology, 62(2), 207.

Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., and Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. Proc. Natl. Acad. Sci. U.S.A. 108, 7710–7715. doi: 10.1073/pnas.1014345108.

Stavru, A. (2019). Pathos, physiognomy and ekphrasis from Aristotle to the Second Sophistic. Visualizing the Invisible With the Human Body: Physiognomy and Ekphrasis in the Ancient World. Berlin: De Gruyter, 143-60.

Suzuki, A., & Suga, S. (2010). Enhanced memory for the wolf in sheep's clothing:: Facial trustworthiness modulates face-trait associative memory. Cognition, 117(2), 224-229.

Taylor, M. J., Mills, T., & Pang, E. W. (2011). The development of face recognition; hippocampal and frontal lobe contributions determined with MEG. Brain Topography, 24, 261-270.

Thorndike, E. L. (1920). A constant error in psychological ratings. Journal of applied psychology, 4(1), 25-29.

Todorov, A., & Engell, A. D. (2008). The role of the amygdala in implicit evaluation of emotionally neutral faces. Social cognitive and affective neuroscience, 3(4), 303-312.

Todorov A, Mandisodza AN, Goren A, Hall CC (2005) Inferences of competence fromfaces predict election outcomes. Science 308:1623–1626.

Todorov , A. , Pakrashi , M. , Loehr , V. R. and Oosterhof , N. 2007b . Evaluating faces on trustworthiness: Automatic assessment of face valence. , Manuscript submitted for publication.

Todorov, A., Baron, S. G., & Oosterhof, N. N. (2008). Evaluating face trustworthiness: a model based approach. Social cognitive and affective neuroscience, 3(2), 119-127.

Todorov, A., & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. Cognitive neuropsychology, 25(3), 395-410.

Todorov, A. (2012). The social perception of faces. The SAGE handbook of social cognition, 96-114.

Toscano, H., Schubert, T. W., & Sell, A. N. (2014). Judgments of dominance from the face track physical strength. Evolutionary Psychology, 12(1), 1-18.

Tsukiura, T., & Cabeza, R. (2008). Orbitofrontal and hippocampal contributions to memory for face–name associations: The rewarding power of a smile. Neuropsychologia, 46(9), 2310-2319.

Tulving, E. (1972). Episodic and semantic memory. Organization of memory, 1(381-403), 1.

Tyng, C. M., Amin, H. U., Saad, M. N., & Malik, A. S. (2017). The influences of emotion on learning and memory. Frontiers in psychology, 8, 235933.

Van't Wout, M., & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. Cognition, 108(3), 796-803.

Vuilleumier, P., Richardson, M. P., Armony, J. L., Driver, J., & Dolan, R. J. (2004). Distant influences of amygdala lesion on visual cortical activation during emotional face processing. Nature neuroscience, 7(11), 1271-1278.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., ... & Buckner, R. L. (1998). Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. Science, 281(5380), 1188-1191.

Wang, X. (2020). Physiognomy in ming China: Fortune and the body (Vol. 149). Brill.

Watanabe, N., & Yamamoto, M. (2015). Neural mechanisms of social dominance. Frontiers in neuroscience, 9, 154.

Wendt, J., Weymar, M., Junge, M., Hamm, A. O., & Lischke, A. (2019). Heartfelt memories: Cardiac vagal tone correlates with increased memory for untrustworthy faces. Emotion, 19(1), 178.

Weymar, M., Ventura-Bort, C., Wendt, J., & Lischke, A. (2019). Behavioral and neural evidence of enhanced long-term memory for untrustworthy faces. Scientific reports, 9(1), 19217.

Whalen PJ, Rauch SL, Etcoff NL, McInerney SC, Lee MB, Jenike MA (1998). Masked presentations of emotional facial expressions modulate amygdala activity without explicit knowledge. J Neurosci 18:411–418.

Wiggins JS, Philips N, Trapnell P (1989) Circular reasoning about interpersonal behav-ior: Evidence concerning some untested assumptions underlying diagnostic classifica-tion. J Pers Soc Psychol 56:296 –305.

Willis, J., & Todorov, A. (2006). First impressions: Making up your mind after a 100–ms exposure to a face. Psychological Science, 17, 592–598

Wilson, W., and Kayatani, N. (1968). Intergroup attitudes and strategies in game between opponents of the same or of a different race. J. Pers. Soc. Psychol. 9, 24–30. doi: 10.1037/h0025720

Wilson, J. P., & Rule, N. O. (2015). Facial trustworthiness predicts extreme criminal-sentencing outcomes. Psychological science, 26(8), 1325-1331.

Wilson, J. P., & Rule, N. O. (2016). Hypothetical sentencing decisions are associated with actual capital punishment outcomes: The role of facial trustworthiness. Social Psychological and Personality Science, 7(4), 331-338.

Winston, J. S., Strange, B. A., O'Doherty, J., & Dolan, R. J. (2013). Automatic and intentional brain responses during evaluation of trustworthiness of faces. In Social neuroscience (pp. 199-210). Psychology Press.

Yonelinas, A. P., Aly, M., Wang, W. C., & Koen, J. D. (2010). Recollection and familiarity: Examining controversial assumptions and new directions. Hippocampus, 20(11), 1178-1194.

Zebrowitz, L. A., Fellous, J. M., Mignault, A., & Andreoletti, C. (2003). Trait impressions as overgeneralized responses to adaptively significant facial qualities: Evidence from connectionist modeling. Personality and social psychology review, 7(3), 194-215.

Zebrowitz LA (2004) The origins of first impressions. J Cult Evol Psychol 2:93–10.

Zebrowitz, L. (2018). Reading faces: Window to the soul?. Routledge.