Neurocomputational understanding of decision-making in novel environments

Thesis by
Sanghyun Yi

In Partial Fulfillment of the Requirements for the
Degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2025
Defended September 16, 2024

Sanghyun Yi
ORCID: 0000-0003-1274-6523

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude to my academic advisor, John O'Doherty. From the bottom of my heart, I thank him for being a great supporter throughout my academic journey, even during times when I made mistakes or faced difficult situations that affected my entire Ph.D. journey. He consistently provided novel and insightful perspectives during our discussions and allowed me considerable freedom in my academic exploration. This thesis, and all the work I have completed during my Ph.D., would not have been possible without him. I also want to extend my heartfelt thanks to Ralph Adolphs, Kirby Nielsen, and Yisong Yue, who generously agreed to be on my thesis committee, as well as to Dean Mobbs, Antonio Rangel, and Colin Camerer for their guidance, teaching, and advice over the past six years.

I would also like to thank my incredible colleagues from the lab: Caroline Charpentier, Jeff Cockburn, Rani Gera, Cooper Grossman, Kiyohito Iigaya, Lisa Kluen, Vincent Man, Omar Perez, Tessa Rusch, Sneha Aenugu, Tomas Aquino, Jaron Colas, Logan Cross, Weilun Ding, Aniek Fransen, Qianying Wu, Amogh Johri, Sarah Oh, Julia Simon, and Sandy Tanwisuth. Their invaluable feedback and the time spent together outside of work were like a vitamin to my Ph.D. journey. I could not have completed this Ph.D. without them. The family-like, friendly culture of our lab exists thanks to all these wonderful people. I would like to express my gratitude to Mike Tyszka for his invaluable assistance with the MRI experiments. I am also deeply thankful to Mary Martin and Laurel Auchampaugh for their support and encouragement throughout my Ph.D. journey.

I also want to acknowledge my Social and Decision Neuroscience friends: Anastasia Buyalskaya, Xiaomin Li, Song Qi, Marcos Gallo, Brenden Eom, Wenning Deng, Thomas Henning, Seokyoung Min, Yi-Chuang Lin, and Noah Okada. Our weekly meetings filled with academic discussions provided me with great inspiration, and it was always a joy to spend time together outside of work.

Additionally, I would like to thank my friends from the Caltech community and beyond. They all gave me a safe space to relax and recharge after long, tiring days.

Lastly, I want to express my deepest gratitude to my beloved family—my father, mother, and sister—for their unwavering support, not just during my Ph.D. journey but throughout my entire life. I wouldn't be where I am today, nor would I have

completed my Ph.D., without them. Their support was truly essential.

And to my beloved wife, Hyeongyeong: Thank you for always being by my side, through both the highs and the lows. I couldn't have completed my Ph.D. without you. With all my love.

ABSTRACT

This thesis investigates the neural and computational mechanisms underlying human decision-making in unfamiliar environments through three interconnected studies. The first study demonstrates that aesthetic value computation for visual art can be systematically predicted from visual features, which are hierarchically represented along the brain's rostrocaudal axis, as revealed by combining deep neural networks with functional MRI data. The second study examines feature-based transfer learning, highlighting the importance of slow integration mechanisms, akin to glial cell functions, for effective knowledge transfer in humans. The third study explores how action affordance influences decision-making in novel environments, showing that action selection results from a competitive interaction between affordance-based and value-based systems, with meta-control exerted by the pre-supplementary motor area and anterior cingulate cortex. Taken together, these studies provide a comprehensive neuro-computational perspective for understanding how the brain navigates novel environments by doing feature-based value computation, transferring knowledge, and using affordance as a guide for action selection.

PUBLISHED CONTENT AND CONTRIBUTIONS

Kiyohito Iigaya, Sanghyun Yi, Iman A Wahle, Koranis Tanwisuth, and John P O'Doherty. Aesthetic preference for art can be predicted from a mixture of low- and high-level visual features. *Nature human behaviour*, 5(6):743–755, 2021. doi: https://doi.org/10.1038/s41562-021-01124-6. S.Y. performed experiments and analyzed the results with K.I.. S.Y. was particularly involved in the deep convolutional neural network training and related analyses. S.Y. wrote the deep convolutional neural network part of the manuscript.

Kiyohito Iigaya, Sanghyun Yi, Iman A Wahle, Sandy Tanwisuth, Logan Cross, and John P O'Doherty. Neural mechanisms underlying the hierarchical construction of perceived aesthetic value. *Nature Communications*, 14(1):127, 2023. doi: https://doi.org/10.1038/s41467-022-35654-y. S.Y. performed experiments and analyzed the results with K.I.. S.Y. was particularly involved in the deep convolutional neural network training and related analyses. S.Y. wrote the deep convolutional neural network part of the manuscript.

Sanghyun Yi and John P. O'Doherty. Computational and neural mechanisms underlying the influence of action affordances on value learning. *BioRxiv*, 2024. doi: https://doi.org/10.1101/2023.07.21.550102.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Chapter 1

GENERAL INTRODUCTION

1.1 Overview

Decision-making in novel environments, where an individual encounters a situation they have never experienced before, is challenging but occurs almost every day in our daily lives. For example, choosing what to eat in a restaurant you are visiting for the first time involves evaluating multiple factors, such as visual cues from menu photos or past experiences at similar restaurants. In such scenarios, decisions are made by integrating current environmental cues with knowledge from past experiences. This interplay between new information and prior knowledge enables individuals to make informed choices, even in unfamiliar situations.

The ability to navigate novel environments and make effective decisions is a fundamental aspect of human cognition, and understanding how the brain achieves this cognitive flexibility is crucial for advancing our knowledge of natural intelligence. It also has significant implications for the development of artificial intelligence systems that can replicate or assist in human decision-making.

This thesis explores the computational and neural mechanisms that enable decision-making in novel environments. First, feature-based value computation will be studied, as it is a key mechanism that supports value assessments in never-before-seen environments, facilitating reasonable decision-making. Imagine a situation where you make a food choice at an exotic restaurant solely based on pictures. You might focus on visual features, such as whether the dish has a lot of red, which might indicate the spiciness of the cuisine. Here, feature-based value computation will be discussed using one of the most complex visual stimuli we can imagine: visual arts. Computational modeling of the valuation process will be explored using feature-based computation and cutting-edge techniques from artificial intelligence which transforms high-dimensional pixel inputs are into a scalar liking value in a biologically plausible manner.

Next, we will explore transfer learning, which supports decision-making in novel environments by enabling humans to apply previously learned knowledge to the current situation. For example, when you play a new video game, you might rely on your experience with similar games to understand the basic mechanics, controls, and

objectives. This ability to transfer knowledge from one context to another allows for quicker adaptation and more effective decision-making, even in unfamiliar settings. The study of transfer learning in this thesis will focus on scenarios in which multiple environments share visual cues. The computational mechanisms behind learning the meaning of these cues and transferring this knowledge will be discussed using online and offline behavioral data.

Finally, the concept of action affordance will be examined as a critical factor in decision-making within real-world novel environments. Action affordance refers to the ability to perceive potential actions that an environment offers, based on both its physical characteristics and the individual's capabilities. For instance, when faced with an unfamiliar tool, you might infer its possible uses based on its shape and similarity to other tools you have used before. This aspect of decision-making is essential for quickly determining the most appropriate actions in a new environment, allowing for effective interaction and problem-solving. The neuro-computational mechanisms behind the effect of affordance in value-based decision-making will be explored.

Together, these studies will provide a comprehensive understanding of how the brain computes value, transfers knowledge, and perceives action opportunities in novel environments, thereby enabling adaptive and intelligent learning and decision-making.

1.2 Decision neuroscience

Decision neuroscience, also known as neuroeconomics, is a field that seeks to understand how the brain supports decision-making processes. It integrates principles and methods from neuroscience, psychology, economics, and computer science to study the neural basis of decision-making. Historically, decision-making was primarily studied within the realms of psychology and economics, with a focus on behavior and the outcomes of decisions rather than the underlying neural mechanisms. The foundation for neuroeconomics was established by combining economic theories with neurobiological data (Glimcher and Rustichini, 2004).

In classical economics, it was assumed that humans make logically optimal decisions, such as maximizing utilities, the outcomes of their decisions, based on a set of rational principles. This assumption is central to the concept of the "rational actor," a theoretical individual who always chooses the option that maximizes their utility. However, a series of behavioral data showed that classical economics does

not realistically describe actual human decision-making. For instance, real human decisions often violate the principle of transitivity. In classical economics, if a person prefers option A over option B and option B over option C, they should logically prefer option A over option C. However, empirical studies have shown that human preferences are not always consistent in this way (Tversky, 1969). Moreover, classical economics fails to explain phenomena such as loss aversion, framing effects, and overconfidence, where people deviate from rational behavior predicted by utility maximization (Kahneman and Tversky, 1984; Thaler, 1985).

The emergence of neoclassical economics was started by incorporating psychological insights into economic models. This transition was significantly influenced by the findings of psychologists like Daniel Kahneman and Amos Tversky (Kahneman and Tversky, 2013). Kahneman and Tversky proposed Prospect Theory, which became a cornerstone of behavioral economics and laid the groundwork for decision neuroscience. Prospect Theory introduced the concept that people value gains and losses differently, leading them to make decisions based on perceived potential losses rather than potential gains. This theory explains why people are often risk-averse when pursuing potential gains but become risk-seeking when trying to avoid losses. It also introduced the idea of loss aversion, where the pain of losing is psychologically more impactful than the pleasure of gaining (Tversky and Kahneman, 1992; Kahneman, 2011).

Prospect Theory challenged the traditional economic models by showing that human decision-making is influenced by biases, heuristics, and emotional responses, rather than purely rational utility maximization. This paradigm shift opened the door for further exploration into the cognitive and neural processes that support decision-making. Researchers began to investigate how different brain regions contribute to evaluating risks and rewards, processing uncertainty, value, and integrating past experiences with current information.

The field of decision neuroscience emerged from these interdisciplinary explorations, seeking to identify the specific brain circuits and mechanisms involved in decision-making (Glimcher and Rustichini, 2004). Significant progress in decision neuroscience has been driven by studies using model animals, such as macaques and rodents, where electrophysiological techniques have been employed to record neuronal activity with high temporal and spatial resolution. These animal studies have provided critical insights into the neural dynamics of decision-making processes, allowing researchers to link specific neural circuits to behavior (Schultz et al., 1997;

Shadlen and Newsome, 2001). In addition, advances in neuroimaging techniques, such as functional magnetic resonance imaging (fMRI), have allowed researchers to directly observe brain activity during decision-making tasks in humans as well. Together, these studies have revealed that decision-making is supported by a network of brain regions, including the prefrontal cortex (PFC), striatum, anterior cingulate cortex (ACC), and posterior parietal cortex (PPC) each playing a role in different aspects of the decision-making process (Gold and Shadlen, 2007).

The prefrontal cortex (PFC) is essential for higher-order cognitive functions and plays a pivotal role in decision-making. Within the PFC, the orbitofrontal cortex (OFC) is particularly important for encoding the value of different options, integrating sensory information with knowledge of expected outcomes to guide decision-making. One of the first evidence that OFC encodes economical value was shown in monkey experiments, which identified neural activities that encode the quantity of juice outcomes independent of the location to the choice options (Padoa-Schioppa and Assad, 2006). A series of fMRI studies also showed that OFC and ventromedial preforntal cortex (vmPFC) have activities correlated with subjective values independent of the type of outcome that resemble common currencies of value or the true meaning of utility (Chib et al., 2009; Levy and Glimcher, 2012; Plassmann et al., 2007). On top of that, recent studies showed that the subjective value is constructed in feature-based computation which feature representation is supported by lateral orbitofrontal cortex (lOFC). For example, it has shown that subjective value on food can be explained by the subjective assessment of the nutrient content of the food and those nutrient information are represented in lOFC (Suzuki et al., 2017b).

In addition, the neural correlates of decision-making have also been found in the posterior parietal cortex (PPC) and other effector-specific regions, such as the frontal eye field (FEF). The PPC, particularly the lateral intraparietal area (LIP), has been shown to play a crucial role in the representation of decision variables, particularly in tasks involving spatial attention and saccadic eye movements. For example, neural activities in the LIP were shown to be reflecting the outcomes of the choice options that invloves saccadic eye movments to making decisions (Platt and Glimcher, 1999; Dorris and Glimcher, 2004). Also, neurons in the LIP encode the accumulated evidence for making a decision about the direction of motion in a visual task, effectively reflecting the gradual formation of a decision over time (Gold and Shadlen, 2007). This area of the PPC is involved in integrating sensory evidence with motor plans, linking the perception of stimuli to the actions required to respond, which suggests

its role as a key node in the decision-making process (Andersen et al., 1997; Snyder et al., 1997; Andersen and Cui, 2009; **?**).

Moreover, the FEF, traditionally associated with the control of eye movements, has been implicated in decision-making processes related to visual attention and saccades. It is found that FEF neurons not only predict the direction of upcoming saccades but also encode information about the relative value of different eye movement choices, indicating that this region integrates decision-related signals with motor planning (Schall, 2001).

Furthermore, effector-specific regions outside the PPC and FEF, such as the dorsal premotor cortex (PMd), have been shown to encode decision variables related to the selection of motor actions. It is demonstrated that neurons in the PMd are involved in planning and deciding between competing motor actions, with neural activity representing the competition between potential movements before the final decision is made (Cisek and Kalaska, 2005, 2010; Cisek, 2007). This supports the idea that decision-making is not confined to prefrontal regions but involves a distributed network, including effector-specific areas that are directly related to the execution of chosen actions.

The anterior cingulate cortex (ACC) is implicated in monitoring conflicts and errors, assessing the costs associated with different choices (Rushworth et al., 2004; Shenhav et al., 2017). It helps the brain to adapt behavior in response to unexpected outcomes and is particularly active when decisions require weighing difficult trade-offs. For example, a series of studies using the Stroop task has shown that the ACC represents the cognitive control signal necessary to selectively process goal-related information when sensory information conflicts with task goals. In the Stroop task, where participants must name the color of the ink a word is printed in while ignoring the word itself (e.g., the word “red” printed in blue ink), the ACC is activated in response to the conflict between the automatic reading process and the task requirement to name the ink color. The ACC was found to be more active during high-conflict trials in the stroop task, supporting its role in conflict monitoring and signaling the need for increased cognitive control (Botvinick et al., 2001; Yeung et al., 2004b).

The striatum, particularly the ventral striatum, plays a crucial role in processing rewards and adapting choice behavior to maximize rewards. It receives dopaminergic signals from the midbrain that encode reward prediction errors—the difference between expected and actual outcomes (Schultz et al., 1997). This feedback mech-

anism allows the brain to learn from experience, updating expectations and refining future decisions. The dorsal striatum, on the other hand, is closely associated with habitual decision-making, where behaviors become automatic through repeated reinforcement. It has shown that the dorsal striatum is involved in the development of habits, as it integrates the history of actions and their outcomes, reinforcing behavior patterns that lead to consistent rewards (Yin and Knowlton, 2006; Tricomi et al., 2009b).

Studies on reinforcement learning (RL) have further demonstrated that the dorsal and ventral striatum support different aspects of the learning process. The dorsal striatum is primarily related to the prediction error when an action is involved, indicating its role in instrumental learning. In contrast, the ventral striatum is associated with prediction errors even in the absence of action, emphasizing its role in value-based learning where outcomes are anticipated without direct behavioral input (O'Doherty et al., 2004). This distinction underscores the specialized functions of these striatal regions in different forms of learning and decision-making.

Together, these neural circuits form an integrated network that supports the complex and dynamic process of decision-making. The interaction between these regions allows for the flexible evaluation of options, the integration of past experiences and the physical constraint, and the adjustment of behavior in response to new information and changing circumstances.

1.3 Artificial intelligence and reinforcement learning

Artificial intelligence (AI) has been widely accepted in modeling complex human behaviors, brain functions, and decision-making processes. One prominent example of AI's contribution to neuroscience is the use of Convolutional Neural Networks (CNNs) as models of the ventral visual stream (Yamins and DiCarlo, 2016; Kriegeskorte, 2015). The ventral visual stream, often referred to as the "what" pathway, is crucial for object recognition and visual perception. CNNs, inspired by the hierarchical organization of the visual cortex, have been widely used to model how the brain processes visual information. These networks consist of layers that mimic the stages of visual processing in the brain, from simple edge detection in early layers to complex object recognition in later layers that mimics the activities in the V4 and inferior temporal cortex (Yamins et al., 2014). The neural representations in CNNs are not limited to those learned through supervised training on image classification tasks; unsupervised contrastive training can also enable CNNs to mimic

the ventral visual stream(Zhuang et al., 2021a). Additionally, a study that trained a CNN using Q-learning, a reinforcement learning algorithm, demonstrated that the network could represent action-related state information, similar to the dorsal stream's function in the brain (Cross et al., 2021).

The other prominent sub-field in AI, reinforcement learning (RL), has been widely accepted as a biologically plausible account of value learning and decision-making in humans and animals. At its core, RL involves an agent learning to make decisions by interacting with an environment, receiving feedback in the form of rewards or punishments, and updating its behavior to maximize cumulative rewards. This process is mathematically formalized using concepts like the Bellman equation and Temporal Difference (TD) learning. The Bellman equation provides a recursive decomposition of the value of a state into immediate rewards plus the expected value of subsequent states, guiding the learning process. TD learning, in particular, captures the idea of prediction error—the difference between expected and actual outcomes—which has been closely linked to midbrain dopaminergic activities in the brain. This connection suggests that dopamine neurons encode prediction errors, providing a neurobiological substrate for RL-like learning mechanisms (Schultz et al., 1997; Sutton, 2018).

In reinforcement learning (RL), approaches are generally categorized into policy-based and value-based methods, with the value-based approach further divided into model-free and model-based RL. Model-free RL involves learning the value of actions directly from experience, without constructing an explicit model of the environment. This approach is often associated with habitual decision-making in humans and animals, where actions become automatic through repeated reinforcement. Conversely, model-based RL entails building an internal model of the environment (Tolman, 1948), enabling goal-directed behavior through planning and simulating future outcomes before making decisions. This distinction between model-free and model-based RL has been mapped onto human behavior and brain function, with model-free processes linked to the dorsolateral striatum and habitual behavior, while model-based processes are associated with the prefrontal cortex and goal-directed decision-making. Notably, studies using two-step Markov decision tasks have allowed researchers to behaviorally dissociate habitual and goal-directed behaviors, explaining them as a mixture of model-based and model-free RL processes (Daw et al., 2005; Lee et al., 2014b). The concept that human behavior results from the interaction of multiple specialized systems has been expanded to model learning

from observing others' behavior as well (O'Doherty et al., 2021a; Charpentier et al., 2020).

RL has also been integrated with neural networks, particularly in the form of deep reinforcement learning, to predict and model complex behaviors. By combining deep learning techniques with heuristics that stabilize the training process, deep RL models can handle high-dimensional inputs, such as visual data, allowing them to perform tasks like playing video games or controlling robotic systems with human-like proficiency (Mnih et al., 2015; Silver et al., 2016). These models learn to extract task-relevant features from raw sensory inputs and use them to make decisions, paralleling the way the brain processes information along the dorsal visual pathway (Cross et al., 2021). Recent advances in RL have introduced the use of multiple policies, modulated by an arbitrator based on each policy's performance on the current task, akin to the concept of a mixture of experts and arbitration among different systems (Badia et al., 2020; Fan et al., 2023). Utilizing multiple policies and reapplying them to new tasks has also been explored as a solution for transfer learning in the RL context (Fernández and Veloso, 2006; Fernández et al., 2010).

The flexibility of deep learning and deep RL has led to the development of a new class of behavior modeling with significantly enhanced predictive power. For instance, using recurrent neural networks (RNNs) for behavioral modeling has shown greater sensitivity in distinguishing the choice characteristics of mental health patients from those of the healthy population (Dezfouli et al., 2019b). Moreover, RNNs have demonstrated the ability to uncover underlying cognitive mechanisms, with trial-by-trial updating patterns of hidden activations that align with models that simulated the behaviors (Ji-An et al., 2023b; Miller et al., 2024). This advancement heralds a new era of cognitive modeling, capable of addressing more complex and naturalistic behaviors.

The flexibility and predictive power of deep learning and deep reinforcement learning models have not only enhanced our ability to simulate and predict human behavior and neural activities but have also opened new avenues for understanding the intricacies of mental health, cognitive disorders, and naturalistic decision-making. As AI continues to evolve, its applications in neuroscience will likely lead to even deeper insights into the workings of the human brain, providing a foundation for the development of more sophisticated and human-like AI systems.

1.4 Motivation for the Thesis

Chapter 2 will explore the computational mechanisms underlying value computation using naturalistic visual stimuli, specifically focusing on the domain of art. We aim to investigate how the brain transforms complex visual inputs into subjective aesthetic preferences by leveraging feature-based analysis and computational modeling.

The motivation for this project stems from the concept of feature-based value computation, which posits that preferences for complex stimuli can be constructed by integrating various features extracted from those stimuli (Palmer et al., 2013; Chatterjee, 2003; Pelletier and Fellows, 2019; Suzuki et al., 2017a). In the context of art, individual artworks possess a wide range of features—from low-level visual properties such as color, contrast, and texture to high-level semantic elements like meaning, emotion, and symbolism(Leder et al., 2004). Understanding how these features collectively shape subjective aesthetic judgments can reveal the computational principles underlying human preference formation (Ramachandran and Hirstein, 1999; Zeki, 2002).

Previous research has explored the psychological and neural bases of aesthetic judgment, highlighting the influence of multiple features and the role of PFC (Kawabata and Zeki, 2004; Cela-Conde et al., 2004; Leder et al., 2004). While models of aesthetic processing have suggested feature-based valuation as a mechanism, they rarely address how such features are extracted from complex, naturalistic images like works of art and used for constructing subjective preference.

Machine learning and computer vision research have advanced feature extraction techniques, with convolutional neural networks demonstrating success in modeling human-like visual recognition (Bishop, 2006; Yamins and DiCarlo, 2016; Dezfouli et al., 2019a; LeCun et al., 2015). However, few studies have applied these models to aesthetic valuation, where subjective preferences must be inferred from a high-dimensional feature space. Additionally, it remains unclear how feature representations extracted by computational models correspond to neural representations supporting human preferences. Another gap in the literature is the limited understanding of how various levels of features—ranging from low-level visual statistics to high-level semantic attributes—are integrated to form subjective value judgments.

To address these gaps, Chapter 2 will combine computational modeling with neuroimaging to investigate how human preferences for art emerge from a multi-level feature integration process. We will apply CNNs to extract low- and high-level features from visual art, and analyze brain activity to uncover neural mechanisms

supporting value computation. This approach will help advancing our understanding of the neural basis of art appreciation.

In chapter 3, I will explore human transfer learning, employing RL algorithms to model this complex cognitive process. The findings from this study are expected to provide valuable insights into addressing transfer learning challenges within the RL context, potentially offering novel solutions to this longstanding problem in AI (Parisi et al., 2019; Flesch et al., 2023).

Previous research in cognitive science has extensively explored transfer learning through declarative memory frameworks, and cognitive map theories (Tolman, 1948; Reber et al., 1996; Squire and Zola, 1996). While RL models have provided valuable insights into how humans adapt to new tasks, they typically assume task-specific learning and struggle to generalize across environments (Botvinick, 2012; Tessler et al., 2017). Declarative memory models emphasize flexible knowledge application but lack precise computational implementations of how such flexibility arises. Similarly, cognitive map theories have been useful for structural knowledge tasks but do not generalize well to feature-based transfer domains (Mark et al., 2020).

A notable gap in the literature is the limited exploration of how humans maintain long-term, transferable representations of feature-based information across tasks. Additionally, research on RNNs has demonstrated success in modeling sequential behavior, but RNNs typically lack biologically plausible mechanisms for retaining learned knowledge over extended periods (Ji-An et al., 2023b; Miller et al., 2024).

To address these limitations, we incorporate slow integration mechanisms inspired by the physiological properties of astrocyte glial cells, which presumably support long-term information retention (Mu et al., 2019; Perea et al., 2009a; Kofuji and Araque, 2021; Mederos et al., 2021; Wang et al., 2017). We hypothesize that such mechanisms enable the gradual accumulation of learned feature-based information, facilitating transfer learning across tasks. To test these models, we designed a feature-based multi-armed bandit task, requiring participants to learn action values based on shared visual features. We collected three independent behavioral datasets from both online and in-person experiments to ensure robust, generalizable results. Through model-driven analysis, we compare RL models with and without slow integration components, alongside RNN models, to identify the mechanisms best capturing human transfer learning behavior.

Chapter 4 will explore the computational and neural mechanisms underlying naturalistic decision-making, focusing on how action affordance shapes value-based learning in novel environments. Specifically, we aim to investigate how action affordances interact with value-driven decision-making processes (Cisek, 2007).

Previous research in cognitive neuroscience and psychology has extensively studied action affordance in terms of automatic action potentiation (Ellis and Tucker, 2000; Zhang et al., 2021). Studies have shown that affordances can prime actions compatible with an object's physical properties, facilitating action selection in visually guided tasks (Symes et al., 2007; Cisek and Pastor-Bernier, 2014). However, these accounts primarily focus on immediate action selection and do not explain how affordance-based processes might contribute to learning and adaptation in complex decision-making environments (Pastor-Bernier and Cisek, 2011).

One key gap in the literature is understanding how action affordances influence the learning process itself rather than merely biasing action selection. It remains unknown whether affordances act as a persistent bias, an initial prior to guide exploration, or as a fully independent controller that dynamically competes with value-driven policies.

To address these open questions, we designed a novel decision-making task that explicitly manipulates action affordance and reward contingencies. Using this task, we collected both behavioral and fMRI data to model the competing influences of affordance-based and value-based policies. Through computational modeling, we test whether these systems function as independent controllers governed by a meta-controller, and we examine the neural correlates of this arbitration process. By integrating insights from cognitive neuroscience and artificial intelligence, this work aims to advance our understanding of how action affordances shape adaptive behavior in complex and dynamic environments.

Chapter 2

AESTHETIC PREFERENCE FOR ART EMERGES FROM A WEIGHTED INTEGRATION OVER HIERARCHICALLY STRUCTURED VISUAL FEATURES IN THE BRAIN

The following chapter is adapted from Iigaya et al., 2021 and Iigaya et al., 2023 and modified according to Caltech Thesis format.

Kiyohito Iigaya, Sanghyun Yi, Iman A Wahle, Koranis Tanwisuth, and John P O'Doherty. Aesthetic preference for art can be predicted from a mixture of low- and high-level visual features. *Nature human behaviour*, 5(6):743–755, 2021. doi: https://doi.org/10.1038/s41562-021-01124-6.

Kiyohito Iigaya, Sanghyun Yi, Iman A Wahle, Sandy Tanwisuth, Logan Cross, and John P O'Doherty. Neural mechanisms underlying the hierarchical construction of perceived aesthetic value. *Nature Communications*, 14(1):127, 2023. doi: https://doi.org/10.1038/s41467-022-35654-y.

2.1 Abstract

It is an open question whether preferences for visual art can be lawfully predicted from the basic constituent elements of a visual image. Moreover, little is known about how such preferences are actually constructed in the brain. Here we developed and tested a computational framework to gain an understanding of how the human brain constructs aesthetic value. We show that it is possible to explain human preferences for a piece of art based on an analysis of features present in the image. This was achieved by analyzing the visual properties of drawings and photographs by multiple means, ranging from image statistics extracted by computer vision tools, subjective human ratings about attributes, to a deep convolutional neural network. Crucially, it is possible to predict subjective value ratings not only within but also across individuals, speaking to the possibility that much of the variance in human visual preference is shared across individuals. Neuroimaging data revealed that preference computations occur in the brain by means of a graded hierarchical representation of lower and higher level features in the visual system. These features are in turn integrated to compute an overall subjective preference in the parietal and prefrontal cortex. Our findings suggest that rather than being idiosyncratic, human

preferences for art can be explained at least in part as a product of a systematic neural integration over underlying visual features of an image. This work not only advances our understanding of the brain-wide computations underlying value construction but also brings new mechanistic insights to the study of visual aesthetics and art appreciation.

2.2 Introduction

From ancient cave paintings to digital pictures posted on Instagram, the expression and appreciation of visual art is at the core of human experience. As Kant famously pointed out, art is both subjective and universal (Kant (1987)). Each individual person may have his/her own taste, but a given piece of art can also appeal to a large number of people across cultures and history. This subjective universality raises a fundamental question: should artistic tastes be likened to the inscrutable, idiosyncratic, and irreducible, or is it possible to deduce lawful and generalizable principles by which humans form aesthetic opinions?

The nature of aesthetic judgment has long been subject to empirical investigation (Fechner (1876); Ramachandran and Hirstein (1999); Zeki (2002); Leder et al. (2004); Biederman and Vessel (2006); Chatterjee (2011); Shimamura and Palmer (2012); Palmer et al. (2013); Leder and Nadal (2014)). Some studies have focused on the visual and psychological aspects of art might influence aesthetics (e.g., see Ramachandran and Hirstein (1999); Chatterjee (2003); Leder et al. (2004); Bar and Neta (2006); Palmer et al. (2013); Van Paasschen et al. (2014)), while other work has highlighted the brain regions whose activity level correlates with aesthetic values (e.g., Cela-Conde et al. (2004); Kawabata and Zeki (2004)). However, attaining a mechanistic understanding of how humans compute aesthetic judgments in the first place from the raw visual input has thus far proved elusive.

A long-standing finding, which partly supports the idiosyncrasy of preference formation, is that prior experience with a specific stimulus can influence value judgment, such as the role of prior episodic memories involving the item, or prior associative history (Fechner (1876); Ramachandran and Hirstein (1999); Zeki (2002); Leder et al. (2004); Weber and Johnson (2006); Wimmer and Shohamy (2012); Barron et al. (2013)). However, while the influence of past experience on current preference is undeniable, humans can express preferences for completely novel stimuli, suggesting that value judgments can be actively and dynamically computed.

It is an open question how the brain can transform realistically complex stimuli into

a simple subjective value. The brain takes a massively high-dimensional input (e.g., a complex art image) and eventually reduces this input to a one-dimensional scalar output (e.g., how much do I like this?). Little is known about how dimensionality reduction can be performed at this scale, while generating reliable output (preference ratings) for all kinds of visual input.

In machine-learning, classification problems (e.g., dog vs. non-dog) are typically solved by projecting an input to a *feature space* (Bishop (2006)). Each feature is a useful attribute that guides the classification of the input. Features can be engineered by taking easily observable characteristics of an object (e.g., its height and weight), or in other cases can be implicitly generated in a more abstracted and less easily interpretable manner (e.g., the activation patterns of hidden layers in deep artificial neural networks).

While previous studies have hinted at the use of such a feature-based framework, in those prior studies the features involved were salient and obvious properties of a stimulus (e.g., multi-attribute artificial stimuli including the movement and the color of dots (Kahnt et al. (2011b); Mante et al. (2013); Pelletier and Fellows (2019)), or items that are suited to a functional decomposition such as food odor (Howard and Gottfried (2014)) or nutritive components (Suzuki et al. (2017a)); see also (Hare et al. (2009); Lim et al. (2013))). However, in the case of visual imagery, the sheer visual complexity of one art piece, as well as the enormous variation between pieces, renders the task of identifying the relevant features that underpin this process exceedingly challenging. In addition, it is not even clear if features are extracted and used for aesthetic judgment in the first place. Moreover, even if relevant features are identified, it is unknown to what extent people may idiosyncratically select the features they use to shape their preferences and how they weigh those features to generate value judgments. Finally, the manner by which a complex visual image gets transformed into relevant features and then into a subjective value, is unclear.

Here, we aimed to establish a general mechanism that could underpin the construction of aesthetic preference. We first extracted features of an art image that have been theorized to play a role in aesthetic valuation (Palmer et al. (2013); Li and Chen (2009); Chatterjee et al. (2010); Vaidya et al. (2017); Chatterjee (2003)). These features reflect subjective judgments about an image, and as such, we deemed them to be "high-level" features, as they required human judgment to determine their presence in an image. We augmented this with a bottom-up process that extracted visual features derived from each image's statistics and visual properties, a feature

set we labeled as "low-level." We then used ratings from human participants' across a large set of painting and photography images to ascertain the extent to which we could predict art preferences using our image feature set. Finally, we applied a deep convolutional neural network (DCNN) to establish the degree to which features for computing visual preference might emerge spontaneously while processing visual images in an (approximately) brain-like architecture. Finally, we applied both linear and DCNN models to functional magnetic resonance imaging (fMRI) data collected from human participants, which allowed us to identify the specific neural mechanisms underlying these feature representations during the evaluation of visual art, as well as to identify the mechanism by which such features are integrated to produce a value judgment.

2.3 Results

Linear feature summation (LFS) model predicts human valuation of visual art

Participants were asked to report how much they liked various pieces of art (images of paintings). The data were collected from both in-lab (N=7) and online participants using Amazon Mechanical-Turk (N=1359). In-lab participants were recruited from the local community. These participants visited our lab in person and performed the task in a standard laboratory setting, while online participants performed the task over the internet.

On each trial, participants were presented with an image of a painting on a computer screen and asked to report how much they liked it on a scale of 0 (not at all) to 3 (very much) (Figure 2.1A). Each of the in-lab participants rated all of the paintings without repetition (1001 different paintings), while online participants rated approximately 60 stimuli, each drawn randomly from the image set. The stimulus set consisted of paintings from a broad range of art genres (Figure 2.1B), and each online participant saw images that were taken with equal proportions from different genres to avoid systematic biases related to style and time-period.

Using this rating data, we tested our hypothesis that the subjective value of an individual painting can be constructed by integrating across features commonly shared across all paintings. For this, each image was decomposed into its fundamental visual and emotional features. These feature values are then integrated linearly, with each participant being assigned a unique set of features weights from which the model constructs a subjective preference (Figure 2.1C). This model embodies the notion that subjective values are computed in a feature space, whereby overall

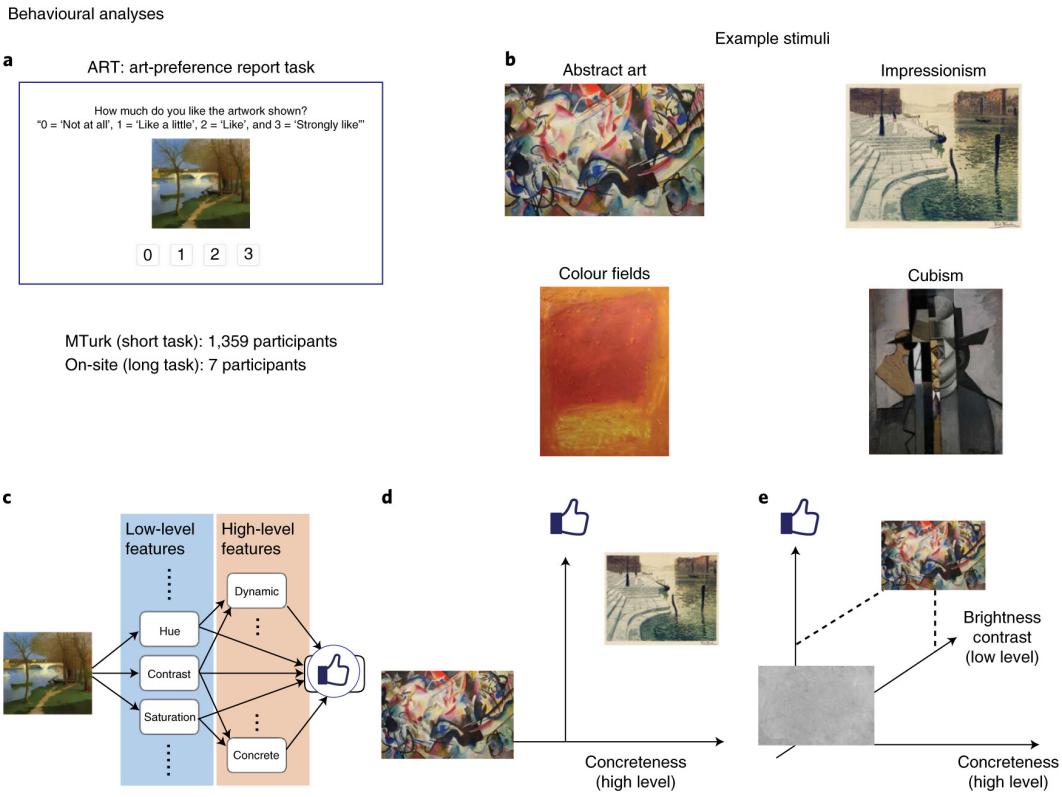


Figure 2.1: Testing the linear feature summation (LFS) model that constructs aesthetic value of visual stimuli. (**A**). The task (ART: art-liking rating task). Participants were asked to report how much they like a stimulus (a piece of artwork) shown on the screen using a four-point Likert rating ranging from 0 to 3. (**B**). Example stimuli. The images were taken from four categories from Wikiart.org.: Cubism, Impressionism, Abstract art and Color Fields, and supplemented with art stimuli previously used (Vaidya et al. (2017)). Each m-turk participant performed approximately 60 trials, while in-lab participants performed 1001 trials (one trial per image). (**C**). Schematic of the LFS model. A visual stimulus (e.g., artwork) is decomposed into various low-level visual features (e.g., mean hue, mean contrast), as well as high-level features (e.g., concreteness, dynamics). We hypothesized that high-level features are constructed from low-level features, and that subjective value is constructed from a linear combination of all low and high-level features. (**D**). How features can help construct subjective value. In this example, preference was separated by the concreteness feature. (**E**). In this example, the value over the concreteness axis was the same for four images; but another feature, in this case, the brightness contrast, could separate preferences over art. Due to copyright issues, some paintings presented here are not identical to what we actually used. Credit: History and Art Collection, ART Collection, Aleksandra Konoplya, Alamy Stock Photo, RISD Museum.

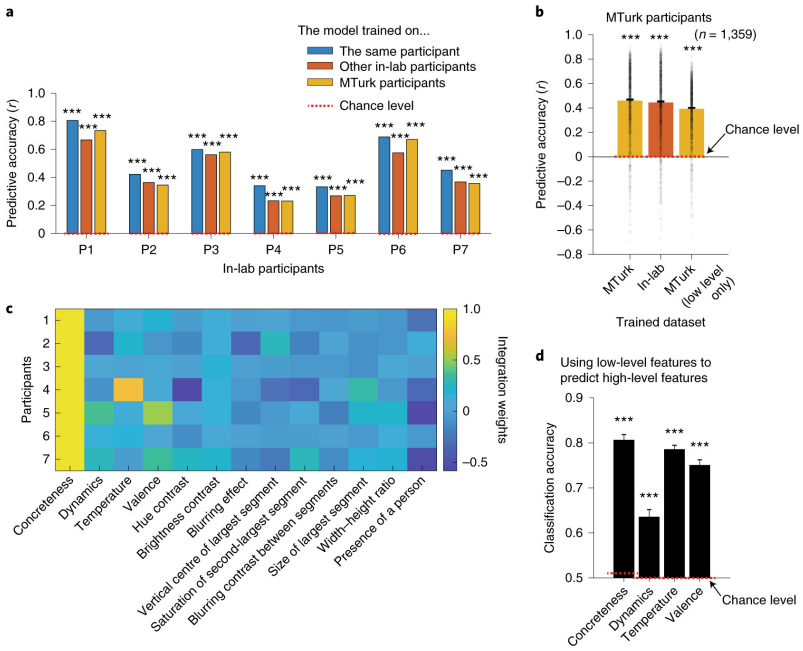


Figure 2.2: The LFS model successfully predicts the subjective value of paintings. (**A**). The LFS model with shared features captured in-lab participants' art liking ratings. The predictive score, defined by the Pearson correlation coefficient between the model's out-of-sample prediction and actual ratings, was significantly greater than chance for all subjects who performed the task in the lab. The model was trained on six participants and tested on the remaining participant (blue), trained and tested on the same participant (red), and trained on on-line participants and tested on in-lab participants (yellow). In-lab subjects performed a long task with 1001 trials. Statistical significance was tested against a null distribution of correlation scores constructed by the same analyses with permuted image labels. The chance level (the mean of the null distribution) is indicated by the dotted lines (at 0). The same set of features (shown in **C**) was used throughout the analysis. (**B**). Our model also successfully accounted for the on-line participants' liking of the art stimuli. We trained the model on all-but-one participants and tested on the remaining participants (left). We also fit the model separately to in-lab participants and tested it independently on all on-line participants (middle). The model predicted liking ratings significantly in all cases, even when we used low-level attributes alone (right). Each on-line participant performed approximately 60 trials. The error bars show the mean and the SEM over participants. The chance level (the mean of the null distribution constructed in the same manner as F) is indicated by the dotted line. (**C**). Weights on shared features that were estimated for in-lab participants. We estimated weights by fitting individual participants separately. (**D**). The low-level features can predict the variance of high-level features. Classification accuracy (high or low values, split by medians) are shown. Note that though the prediction is highly significant, there is still a small amount of variance remaining that is unique to high-level features. The chance level (the mean of the null distribution) is indicated by the dotted line. The error bars indicate the standard errors over cross-validation partitions. In all panels, three stars indicate $p < 0.001$ against permutation tests.

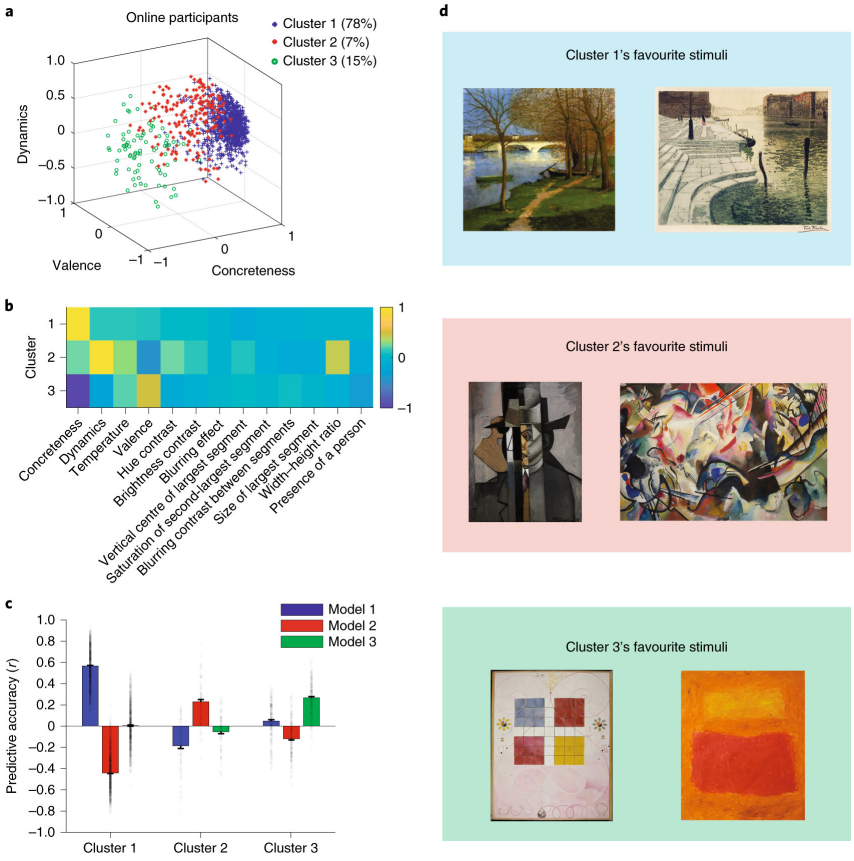


Figure 2.3: Cluster analysis in the feature space suggests the existence of distinct groups of individuals who vary in their preference computations across our online sample. (**A**). The estimated feature weights of all participants, colored by cluster membership. We fit the LFS model to each individual online participant. We then performed a clustering analysis on the estimated weights using a Gaussian mixture model. The number of Gaussians was optimized by comparing Bayes Information Criterion (BIC) scores. Estimated weights from three features are shown for illustration. (**B**). The estimated feature weights at the center of each cluster. Cluster 1 assigns a large positive value to concreteness, while cluster 3 assigns a large negative value to concreteness. Cluster 2 has a distinctively large weight on the dynamics. (**C**). Predictive accuracy of participants in each cluster, using a model with the mean of each gaussian as its parameters. The result suggests that Cluster 1 and 2 have conflicting preferences, while cluster 3 is rather distinct. (**D**). Example stimuli that were preferred by each cluster of participants. The stimuli preferred by participants in cluster 1 include realistic landscape paintings, some of which are from impressionism. The stimuli preferred by cluster 2 include abstract, complex paintings, e.g., in cubism. Cluster 3's favorite stimuli include simple paintings in color fields and abstract art. Art images are purchased from Alamy.com. Due to copyright, issues colour field paintings presented here are not identical to what we actually used. Credit: History and Art Collection, ART Collection, LatitudeStock, Volgi archive/Alamy Stock Photo, RISD Museum.

subjective value is computed as a weighted linear sum over feature content (Figure 2.1DE). We refer to this model as the Linear Feature Summation (LFS) model.

The LFS model extracts various low-level visual features from an input image using a combination of computer vision methods (e.g., Li and Chen (2009)). This approach computes numerical scores for different aspects of visual content in the image, such as the average hue and brightness of image segments, as well as the entirety of the image itself, as identified by machine learning techniques, e.g., Graph-Cuts Rother et al. (2004) (Details of this approach are described in the Methods section). Thus, we note that the LFS model performs a (weighted) linear sum over features, where features can be constructed non-linearly.

The LFS model also includes more abstract or "high-level" attributes that are likely to contribute to valuation. For this, we introduced three features based on previous studies: Chatterjee et al. (2010); Vaidya et al. (2017) the image is 'abstract or concrete,' 'dynamic or still,' 'hot or cold,' as well as a fourth high-level feature concerning whether the image had a positive or negative emotional valence. Note that "valence" is not necessarily synonymous with valuation: if a piece of art denotes content with a negative emotional tone (e.g., Edvard Munch's "The Scream"), it can still be judged to have a highly positive subjective value by the art appreciator. We hypothesized that these high-level features are constructed in downstream units using low-level features as input (Figure 2.1C). However, because we do not know the value of these high-level features a priori, following previous studies Chatterjee et al. (2010); Vaidya et al. (2017) we invited participants with familiarity and experience in art (n=13) to provide subjective judgments about the presence of each of these features in each of the images in our stimulus set (though we note that a previous study found that artistic experience did not affect feature annotations (Chatterjee et al., 2010)). We took the average score over these experts' ratings as the input into the model representing the content of each high-level attribute feature for each image.

The final output of the model is a linear combination of low- and high- level features. We assumed that weights over the features are fixed for each individual, which is a necessary requirement to derive generalizable conclusions about the features used to generate valuation across images. As our high-level features were annotated by humans, we treat low-level and high-level features equally, in a non-hierarchical manner, in order to determine the overall predictive power of our LFS model.

We first determined a minimal set of features that can reliably capture rating scores

across participants in order to gain insights into aesthetic preference universality. For this, we performed a group-level lasso regression on the data we collected in our in-depth in-lab study (n=7; each rated all 1001 images) using all of the low-level and high-level features that we constructed. By doing so, we removed from consideration those features that do not provide useful predictive information, ultimately selecting the features most uniquely predictive of subjective value, leaving 9 low-level and 4 high-level attribute features. The features include some low-level features computed from the entire images such as the 'mean hue contrast' and the 'blurring effect' as well as some low-level features computed using segmentation methods such as the 'position and the size of the largest segment,' in addition to high-level features (please see the Methods for more details). Note that while the integration weights can be tuned for individual participant(s), the feature values for each image remain consistent for all participants.

We then asked how a linear regression model with these features can predict an individual's liking for visual art. To our surprise, we found that we can predict subjective ratings in both a within-, and out-of-, participants manner; the model predicts subjective value not only when we trained the model's weights on the same participant (using a cross-validated procedure) but also when we trained the weights on other in-lab participants, and even when we trained the weights in an entirely independent sample of online participants (Figure 2.2A).

Similarly, we found that we could reliably predict value ratings for online participants (Figure 2.2B), not only when training the model on the online participants' data (using leave-one-out cross-validation) but also when the model had been trained using in-lab participants' data. We also tested the extent to which we can predict value from the low-level attributes alone. Removing the high-level features impaired predictive performance somewhat, but yielded highly significant prediction nonetheless (Figure 2.2B). These results suggest that a non-negligible proportion of the variance in participants' aesthetic ratings can be captured using simple visual features, and can be generalized across people and the art genres that we tested here.

Although we could predict each individual's ratings by training the model on the ratings of others, the degree to which each individual could be predicted from the pooled weights of other participants varied considerably. This suggests that while a common generic model of feature integration can predict individual liking ratings to a surprisingly high degree, there are also likely to be individual differences in how particular features are weighted, which reflects personal aesthetic tastes.

We therefore asked how the model's integration weights for each participant can be varied across participants if we fit the model to each participant separately. We found that the estimated model's feature weights varied across in-lab participants, though all seven of them assigned the largest positive value to the concreteness feature (Figures 2.2C and 2.18). This preference for concreteness generalized to many of the participants in our much larger scale dataset of online participants. However, we also found that a significant number of participants showed a small or even negative weight on the concreteness feature (Figure 2.3A).

To better understand the heterogeneity of aesthetic computation across our sample, we aimed to identify potential clusters of individuals that might use features similarly, using the large-scale online participants' data. We fit the LFS model weights to each individual participant, and then fit a Gaussian mixture model to the estimated weights over participants. By comparing the Bayes Information Criteria score between models with a different number of Gaussians, we identified three clusters in the data (Figure 2.3A,B). Clusters 1 and 2 show somewhat opposing preferences, while cluster 3 shows a distinct preference altogether (Figure 2.3C). Consistent with our in-lab dataset, the majority of individuals (78%) in our online dataset belonged to Cluster 1, showing a positive weight on the concreteness feature (Figure 2.3A,B), apparently preferring images of scenery and impressionism (Figure 2.3D). The remainder belong to one of two other groups: cluster 2 (7%) exhibited a strong preference for dynamic images (e.g., cubism), while cluster 3 (15%) had a large negative weight on concreteness and a positive weight on valence, exhibiting a preference for abstract art and color fields (Figure 2.3A,B,D). We also noticed that the difference between clusters was not well described by art categories.

One potential concern might be that the model's performance relies on the preferences for a particular art genre over the other (e.g., people may like impressionism over cubism); however, the same model trained on all images captures significant variations in preference *within* each art genre after taking out the effect of genre preferences (Figure 2.19), suggesting that the model captures variations in subjective preference both within and across genres. Indeed, we found in a representation dissimilarity analysis over visual stimuli, that low-level features seem to capture art genres, but high-level features go beyond the genres (Figure 2.20).

Although our model can capture significant variance in aesthetic liking judgements, it by no means captures everything. We compared our model's out-of-participant performance with the average correlations in ratings between participants, showing

that there is significant variance in ratings that the model fails to capture in all of the art genres (Figure 2.21). The latter provides an estimate of a noise ceiling, that is, the variance in ratings that can in principle be predicted from the data. The difference between the model's predictions and the average ratings shows that there is still significant remaining variance that the model fails to capture in all of the art genres.

The above results are based on a linear regression of low- and high-level features, but we also considered the possibility that high-level features are comprised of low-level features (illustrated in Figure 2.1C). To assess this, we probed the degree to which a linear combination of low-level features could predict the annotated ratings of high-level features. For this, we trained a linear support vector machine using all low-level features as input, and indeed we found that variance ascribed to high-level features could be predicted by low-level features (Figure 2.2D). This suggests that high-level features can be constructed using objective elements of the images, rather than subjective sensations, although the construction may well depend on additional nonlinear operations.

Finally, to test for the effects of the salience of features within the image on the behavioral prediction, we calculated a saliency map for each stimulus using the standard saliency toolbox Walther and Koch (2006). Then we re-calculated visual features (11 global features, 20 segmented features) with the saliency map, simply by filtering the features through the saliency map (please see Method for details). We added these saliency-weighted features to the original feature set, and performed linear regression analysis. We however found that the salience map filtered features did not improve the model's predictive accuracy (Figure 2.22).

The LFS model also predicts human valuation of photographs

One potential concern we had was that our ability to predict artwork rating scores using this linear model might be somehow idiosyncratic due to specific properties of the stimuli used in our stimulus-set. To address this, we investigated the extent to which our findings generalize to other kinds of visual images by using a new image database of 716 images Murray et al. (2012); Figure 2.4A), this time involving photographs (as opposed to paintings) of various objects and scenes, including landscapes, animals, flowers, and pictures of food. We obtained ratings for these 716 images in a new m-Turk sample of 382 participants. Using the low-level attributes alone (these images were not annotated with high-level features), the

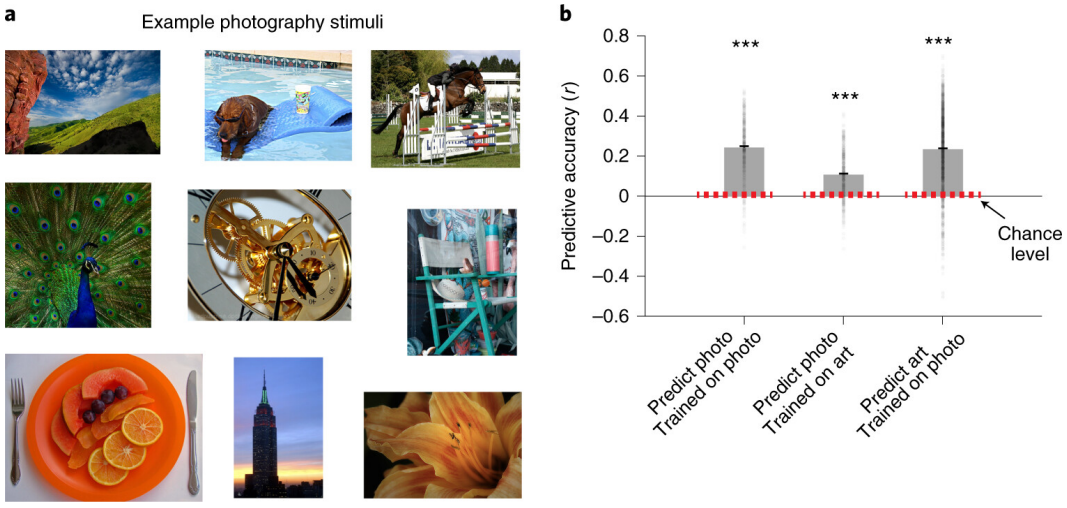


Figure 2.4: The LFS model also predicts subjective liking ratings for various kinds of photographs. (**A**). Example stimuli from the photography dataset. We took a wide range of images from the online photography (AVA) dataset Murray et al. (2012), and ran a further on-line experiment in new M-Turk participants (n = 382) to obtain value ratings for these images. (**B**). A linear model with low-level features alone captured liking ratings for photography. This model when trained on liking ratings for photography (data from the current experiment) also captured liking ratings for paintings (data from the previous experiment described in 2.1), and the model trained on liking ratings for paintings could also liking ratings for photography. We note that in all cases the model was trained and tested on completely separate sets of participants. Significance was tested against the null distribution constructed from the analysis with permuted image labels. The error bars indicate the mean and the s.e., while the dots indicate individual participants.

linear integration model could reliably predict photograph ratings (Figure 2.4B). The model performed well when trained and tested on the photograph database, but to our surprise, the same model (as trained on photographs) could also predict the ratings for paintings that we collected in our first experiment, and vice versa (a model trained on the painting ratings could predict photograph ratings). Of note, accuracy was reduced if trained on paintings and tested on photographs (though still highly above chance), suggesting that the photographs enabled improved generalization (possibly because the set of photographs were more diverse). We stress that here, in all cases the model was trained and tested on completely separate sets of participants.

We also tested whether the inclusion of high-level features can improve the model's predictive accuracy in the photograph dataset. Using the support vector machine that is trained on high-level features in the visual art dataset, we estimated binarized high

level features in the photograph dataset. We then tested how the model with both low- and high-level features predict ratings in photograph dataset. We found that the full model with both high- and low-level features performs significantly better than the prediction from average ratings, though the direct comparison between the model with low-level alone and the full model did not yield statistical significance (Figure 2.23). This indicates that abstract high-level features can contain information that is generalized across different images, which enables the model to go beyond the average ratings for each image.

A deep convolutional neural network (DCNN) model predicts human liking ratings for visual art

We now have shown that our LFS model can capture subjective preference for visual art; however, as we selected the model's features using a mixture of prior literature and bottom-up machine learning tools, we do not know if this strategy has any biological import. In particular, 1) because we handpicked the LFS model's features, it is not clear if and how a neural system learns to represent these features. It is unlikely that an actual neural system (e.g., human brain) is trained on the features explicitly. Rather, if the LFS model represents a biologically plausible computation, features should emerge out of training on value judgements without explicitly trained on features. Also, 2) it is not clear what kind of network architecture is sufficient to achieve the LFS model's computation. Specifically, it is unknown how a network architecture could end up representing low-level and high-level features hierarchically and integrating them to construct subjective value.

To address these issues, we utilized a standard deep convolutional neural network (DCNN; VGG 16 Simonyan and Zisserman (2014)), that had been pre-trained for object recognition with ImageNet Deng et al. (2009). This allows us to test if the computation of the LFS model can be realized in a standard feed-forward network. We used this network with fixed pre-trained weights in convolutional layers, but trained the weights for the last three fully-connected layers on averaged liking ratings. Mirroring the results of our LFS model, we found that our DCNN model can predict human participants' liking ratings across all participants (Figure 2.5A). This shows that it is indeed possible to predict preferences for visual art using a deep-learning approach without explicitly selecting stimulus features. In a supplementary analysis, we also opened the convolutional layers to training, but saw no improvement.

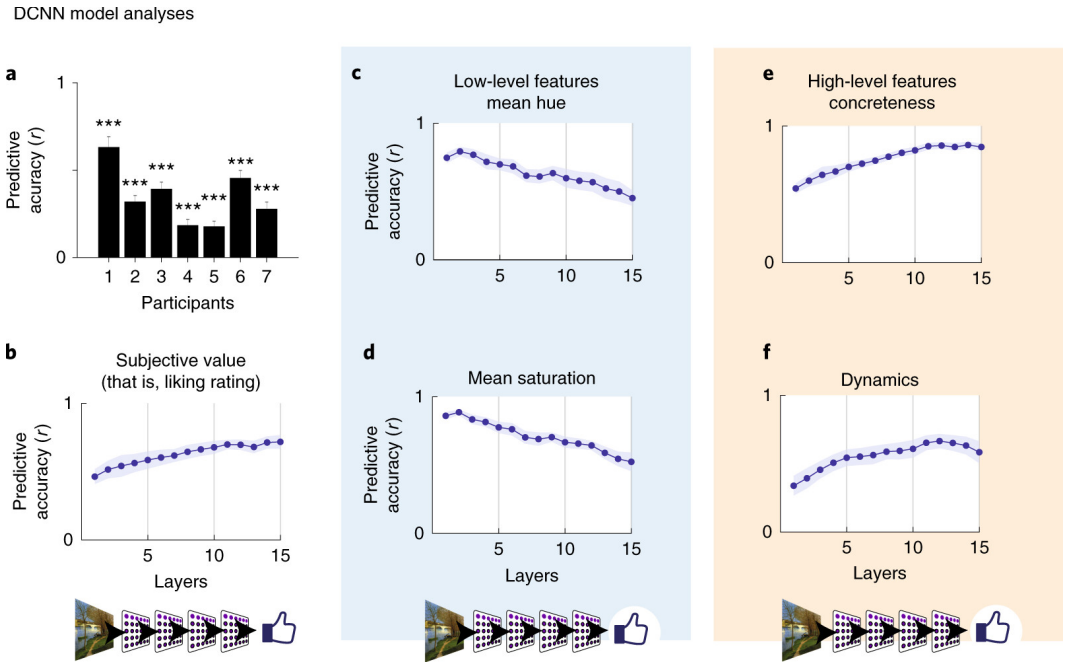


Figure 2.5: A deep convolutional neural network (DCNN) can predict subjective values (i.e., liking ratings) of art stimuli, and the features that we introduced to our LFS model spontaneously emerge in the hidden layers of the network. We utilized a standard convolutional neural network (VGG 16 Simonyan and Zisserman (2014)) that came pre-trained on object recognition with ImageNet Deng et al. (2009), consisting of 13 convolutional and three fully connected layers. We trained the last three fully connected layers of our network on average art liking scores without explicitly teaching the network about the LFS model's features. (**A**).The neural network could successfully predict human participants' liking ratings significantly greater than chance across all participants. The significance ($p < 0.001$, indicated by three stars) was tested by a permutation test. (**B**). We found that we can decode average liking ratings using activation patterns in each of the hidden layers. The predictive accuracy was defined by the Pearson correlation between (out-of-sample) model's predictions and the data. For this, we used a (ridge) linear regression to predict liking ratings from each hidden layer. We first reduced the dimensions of each layer with a PCA, taking top PCs that capture 80% of the variance in each layer. The accuracy gradually increases over layers despite the fact that most layers (layers 1-13) were not trained on liking ratings but on ImageNet classifications alone. (**C,D**). When performing the same analysis with the LFS model's features, we found some low-level visual features with significantly decreasing predictive accuracy over hidden layers (e.g., the mean hue and the mean saturation). We also found that a few computationally demanding low-level features showed the opposite trend (see the main text). (**E,F**). We found some high-level visual features with significantly increasing predictive accuracy over hidden layers (e.g., concreteness and dynamics). We also found that temperature, which we introduced as a putative high-level feature, actually shows the opposite trend, likely because it is a color-based feature that can be straightforwardly computed from pixel data. Credit: History and Art Collection/Alamy Stock Photo.

The LFS model's features emerge spontaneously in the DCNN model's hidden layers

We turn now to ask whether the features used in our LFS model are spontaneously encoded in the neural network. Mirroring our illustration of the LFS model in Figure 2.1C, we hypothesized that low-level visual features would be represented in early layers of the DCNN, while more abstract high-level attributes would be represented in later layers of the network. For our investigation, we performed decoding analyses to predict high- and low-level feature values using the activation patterns in each hidden layer. We first reduced the dimensions of each layer using principal component analysis (PCA), and using the top principal components (PCs) that capture 80% of the variance of each layer, we trained a regression model for a given variable that we aimed to predict (e.g., ratings or a feature) using the PCs of each layer.

We first tested to see if we can predict subjective liking ratings using the hidden layer activation patterns. We were able to decode subjective ratings across all layers, but noted that decodability gradually increased for layers deeper in the network (Figure 2.5B). This came as a surprise since all but the last three layers (layers 1 to 13, out of 16) were pre-trained not on the rating scales being decoded, but on image classifications alone using ImageNet, hinting at a tight relationship between value coding and visual recognition.

We then tested to see how the hidden layers related to the LFS model's features. This analysis showed that hidden layers could predict all 23 features included in the LFS model. Consistent with our hypothesis, six (of the 19) putative low-level features tested were represented more robustly in early layers, as shown by a significantly negative decoding slope across layers (Figure 2.5CD). We also found four (out of 19) low-level features had a decoding accuracy that increased as a function of the depth of the layer, suggesting those low-level features, in fact, may be better identified as high-level features. However, we note that the overall predictive accuracy of these features was low compared to those showing negative slopes. These positive slope features include: "the presence of a person," "the mass center for the largest segment," "mass variance of the largest segment," and "entropy in the 2nd largest segment," all of which require relatively complex computations (e.g., segmentation and the identification of the location of the segments) compared to the ones showing negative slopes (e.g., average saturation). We note that this result is consistent with a previous electrophysiological and computational modeling study in macaques

(Hong et al., 2016), which reported that the position of an object on the screen is more robustly represented in higher visual areas and deeper layers, as position identifications of a segment and an object likely involve similar computations. These object-related features were also referred to as 'low-level' features (Hong et al., 2016), in line with our original reference. The other 9 low-level features tested did not show either a strong positive or negative slope.

Similarly, for the putative high-level features, we found that two (of 4) features were more robustly represented in later layers (Figure 2.5EF). However, "temperature," which was labelled as a high-level feature, showed a significant negative slope. Given that this feature is based on color palettes in the image (i.e., whether the color palette is hot or cold), this feature's variance may already be well captured by low-level image statistics. The fourth putative high-level feature, valence, did not show either an increasing or decreasing trend in decoding across layers. Thus, the DCNN allows a more principled means to identify low and high-level features, enabling us to label 6 features as low-level (based on greater representation of those features in earlier layers of the network), and 6 as high-level features, indicated by representations that are present to a greater extent in later layers of the network. These LFS model-based analyses on the DCNN sheds light into what are often-considered-to-be "black box" computations in deep artificial neural networks, and may provide an empirical definition of computational complexity in feature extraction of visual as well as other sensory inputs.

Taken together, our DCNN analyses suggest that our conceptualized LFS model (Figure 2.1C) is, in fact, a natural consequence of training the neural network on object recognition and predicting subjective aesthetic value, without requiring any explicit feature engineering.

The LFS model's features emerge in the DCNN model's style features as well

In addition to our analysis of the decodability of low- and high-level features from the hidden layers, we also investigated whether these features could be decoded from the style representations of the convolutional layers. Specifically, we examined the Gram matrices of channel activations in each hidden layer, which capture the correlations between feature maps and are commonly associated with the style information of a painting genre (Gatys, 2015). Therefore, the style features could serve as a proper representation for the considered feature sets.

Similar to the previous analyses, the style features were first reduced to the dimen-

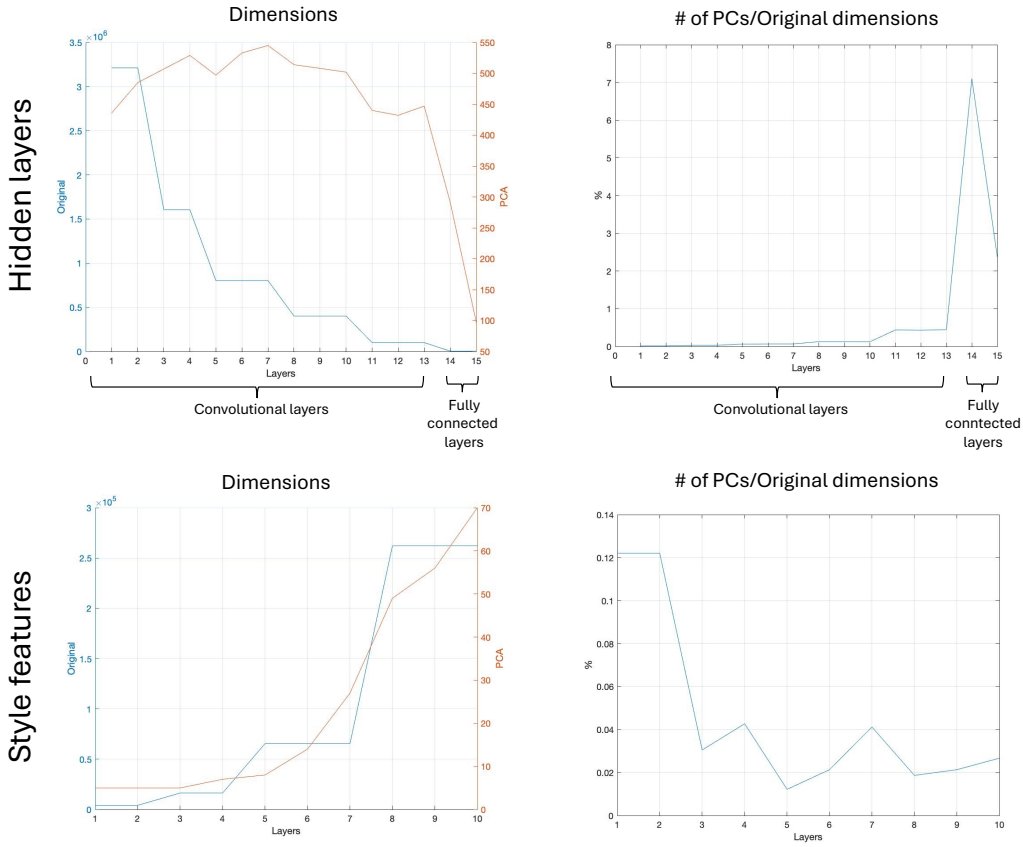


Figure 2.6: Comparison between the original and reduced dimensions after applying PCA. The upper row shows the hidden layer activations, while the lower row displays the style features of the hidden activations.

sions that explain 80% of the total variance using PCA. In particular, the number of principal components needed to explain 80% of the total variance was significantly lower when PCA was applied to the style features (Figure 2.6, left column). Additionally, the relative number of principal components required to explain the variance compared to the original dimensions was reduced by nearly a factor of 10 (Figure 2.6, right column).

We then conducted a decoding analysis on the style features analogous to the one performed on the hidden layers. Most of the subjective ratings and features included in the LFS model could be decoded from the style features, with the exceptions of mass skewness for the second largest segment and the vertical coordinate of the mass center for the largest segment (Figures 2.7 and 2.8). This is likely due to the style features losing spatial information as a result of the dot products between channels.

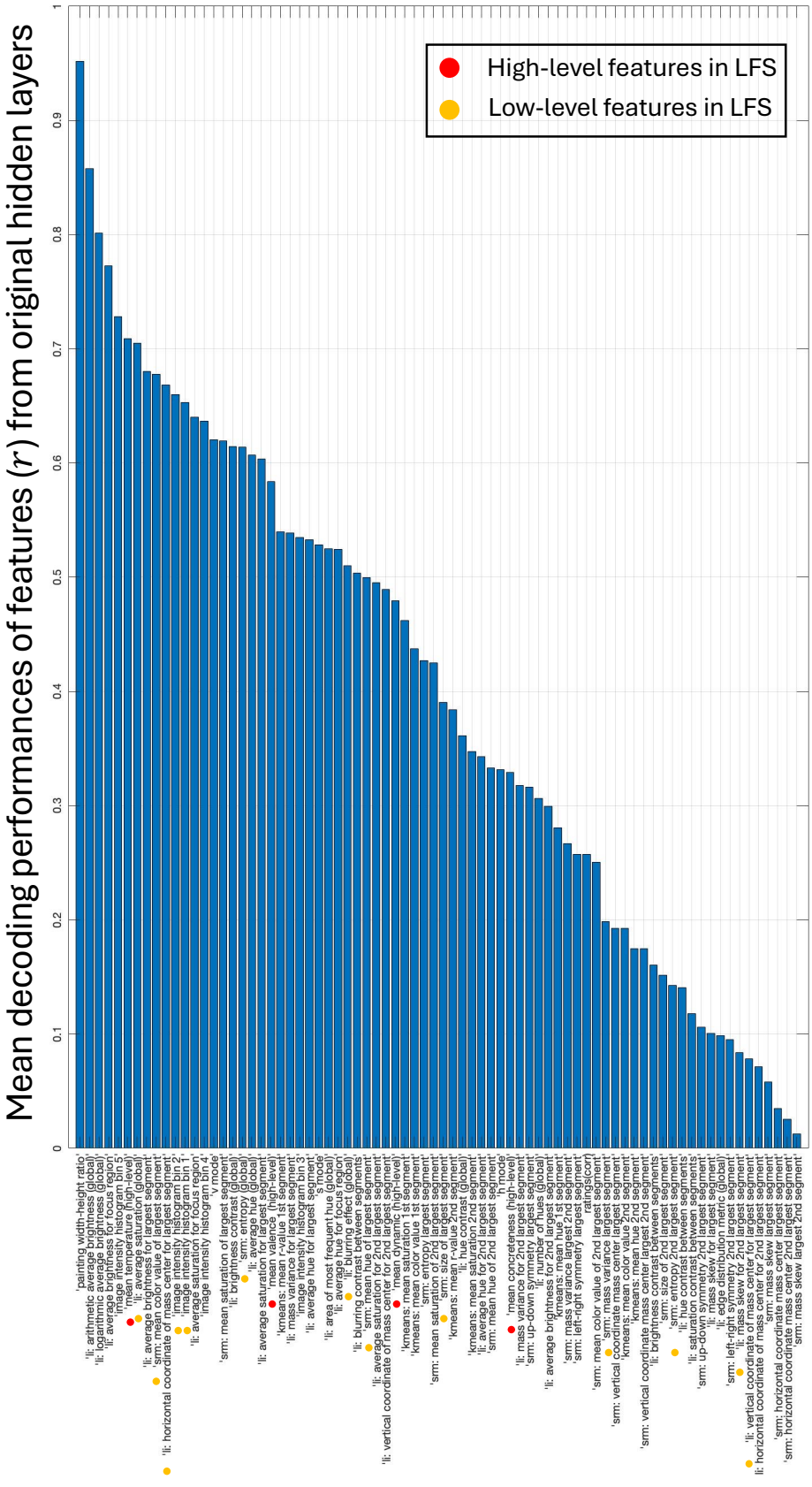


Figure 2.7: The predictive accuracies (Pearson correlation) for each feature averaged across hidden layers.

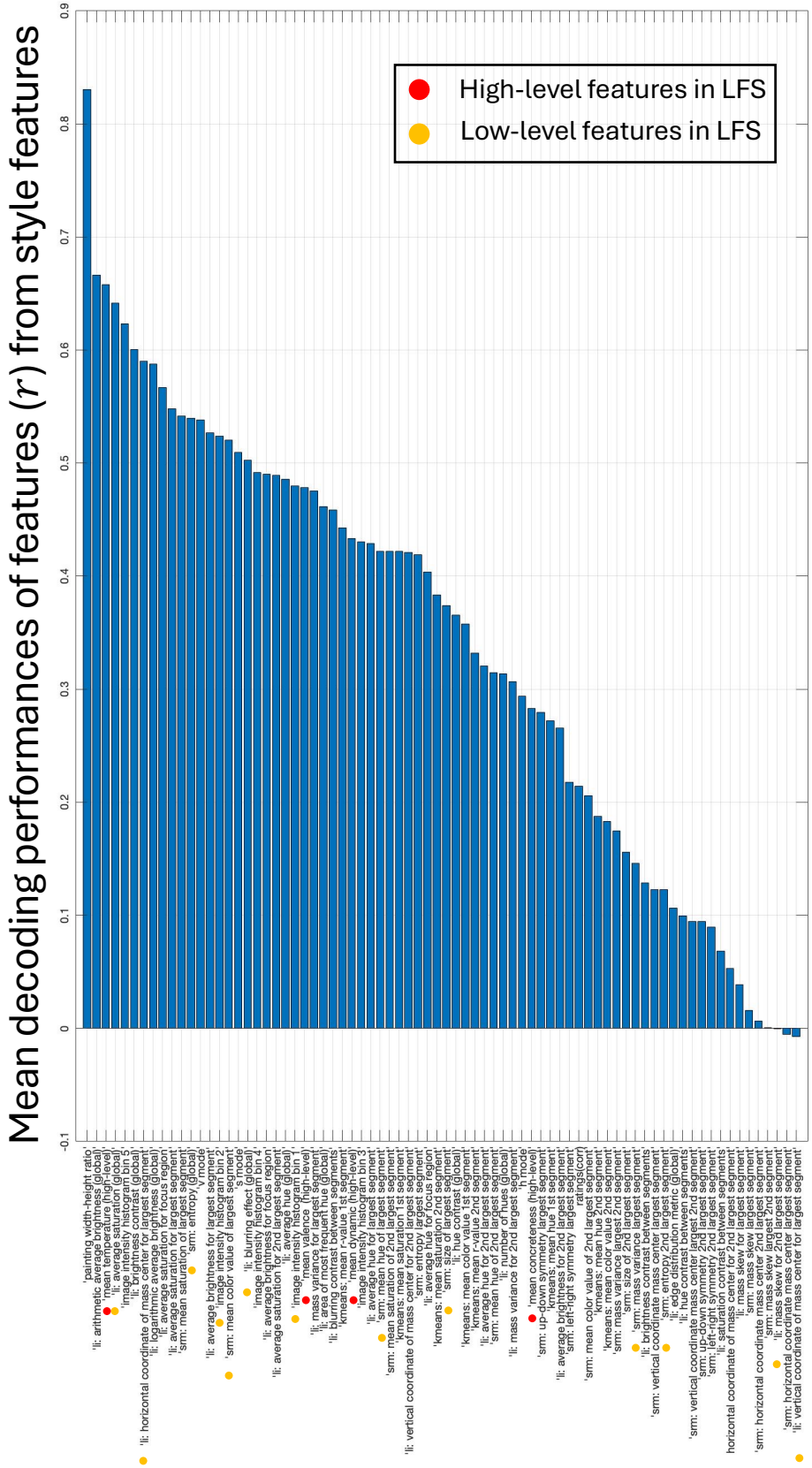


Figure 2.8: The predictive accuracies (Pearson correlation) for each feature averaged across style features from convolutional layers.

The decodability order of high-level features was identical to that observed in the hidden layer decoding analyses (strongest to weakest: temperature, valence, dynamics, and concreteness; see Figures 2.7 and 2.8). The order of decoding performance for low-level features from the style features was similar, though some discrepancies were noted compared to the original hidden layer decoding (Figures 2.7 and 2.8). For instance, the feature with the second highest decodability was the mean color value of the largest segment in the hidden layer decoding, but it was the horizontal coordinate of the mass center for the largest segment in the style feature decoding.

The decoding slope was mostly similar, but the style feature decoding showed a dip in the early layers, following the pattern of decoding from the original hidden layers in the higher convolutional layers (Figure 2.9). However, in general, the decoding accuracies were higher in the higher convolutional layers compared to the lower layers.

The subjective value of art is represented in the medial prefrontal cortex (mPFC)

We first tested for brain regions correlating with the subjective liking ratings of each individual stimulus at the time of stimulus onset. We expected to find evidence for subjective value signals in the medial prefrontal cortex (mPFC), given this is the main area found to correlate with value judgments for many different stimuli from an extensive prior literature, including for visual art (e.g., Cela-Conde et al. (2004); Kawabata and Zeki (2004); Padoa-Schioppa and Assad (2006); Kable and Glimcher (2007); Gläscher et al. (2008); Grabenhorst and Rolls (2011); Ishizu and Zeki (2013)). Consistent with our hypothesis, we found that voxels in the mPFC are positively correlated with subjective value across participants (Figure 2.12; See Figure 2.27 for the timecourse of the BOLD signals in the mPFC cluster). Consistent with previous studies, e.g., Hampton and O'doherty (2007); Serences (2008); Chatterjee et al. (2009); Stănişor et al. (2013); FitzGerald et al. (2013); Suzuki et al. (2017a); Bach et al. (2017), other regions are also correlated with liking value (Figure 2.29 and 2.30).

These subjective value signals could reflect other psychological processes such as attention. Therefore we performed a control analysis with the same GLM with additional regressors that can act as proxies for the effects of attention and memorability of stimuli, operationalized by reaction times, squared reaction times and

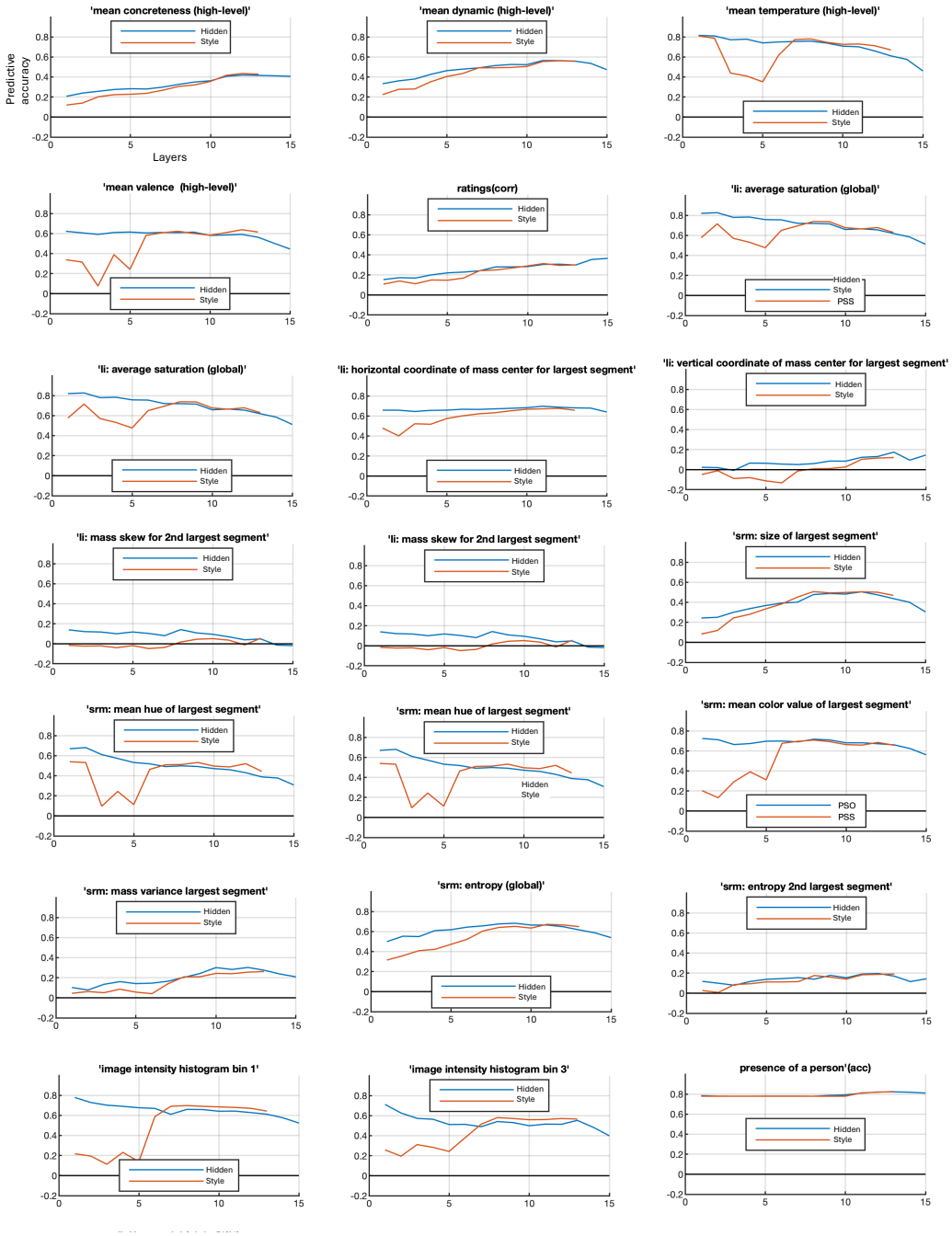


Figure 2.9: The predictive accuracies (Pearson correlation, or accuracy for the existence of human) for each feature for each layer.

the deviation from the mean rating (O'Doherty, 2014). We found that subjective value signals in all participants that we report in Figure 2.12c survived this control analysis (Figure 2.31).

Visual stream shows hierarchical, graded, representations of low-level and high-level features

As illustrated in Figure 2.10d, and reflecting our hypothesis regarding the encoding of low vs. high-level features across layers of the DCNN, we hypothesized that the brain would decompose visual input similarly, with early visual regions first representing low-level features, and with downstream regions representing high-level features. Specifically, we analyzed visual cortical regions in the ventral and dorsal visual stream (Wang et al., 2014) to test the degree to which low-level and high-level features are encoded in a graded, hierarchical manner. In pursuit of this, we constructed a GLM that included the shared feature time-locked to stimulus onset. We identified voxels that are significantly modulated by at least one low-level feature by performing an F-test over the low-level feature beta estimates, repeating the same analysis with high-level features. We then compared the proportion of voxels that were significantly correlated with low-level features vs. high-level features in each region of interest in both the ventral and dorsal visual streams. This method allowed us to compare results across regions while controlling for different signal to noise ratios in the BOLD signal across different brain regions (Barch et al., 2013). Regions of interest were independently identified by means of a detailed probabilistic visual topographical map (Wang et al., 2014). Consistent with our hypothesis, our findings suggest that low- and high-level features relevant for aesthetic valuation are indeed represented in the visual stream in a graded hierarchical manner. Namely, the relative encoding of high-level features with respect to low-level features dramatically increases across the visual ventral stream (Figure 2.13a). We found a similar, hierarchical organization in the dorsolateral visual stream (Figure 2.13b), albeit less clearly demarcated than in the ventral case. We also confirmed in a supplementary analysis that referring to feature levels (high or low) according to our DCNN analysis, i.e., by using the slopes of our decoding results (Iigaya et al., 2021), did not change the results of our fMRI analyses qualitatively and does not affect our conclusions (see Figure 2.32).

We also performed additional encoding analysis using cross validation at each voxel of each participant (Naselaris et al., 2011). Specifically, we performed a lasso regression at each voxel with the low- and high-level features that we considered in

our original analyses. Hyperparameters are optimized in 12-fold cross validation at each voxel across stimuli.

As a robustness check, we determined if our GLM results can be reproduced using the lasso regression analysis. We analyzed how low-level feature weights and high-level feature weights changed across ROIs. For this, we computed the sum of squares of low-level feature weights and the sum of squares of high-level feature weights at each voxel. Because these weights estimates include those that can be obtained by chance, we also computed the same quantities by performing the lasso regression with shuffled stimuli labels (labels were shuffled at every regression). The null distribution of feature magnitudes (the sum of squares) was estimated for low-level features and high-level features at each ROI. For each voxel, we asked if estimated low-level features and high-level features are significantly larger than what is expected from noise, by comparing the magnitude of weights against the weights from null distribution ($p < 0.001$). We then examined how encoding of low-level vs high-level features varied across ROIs, as we did in our original GLM analysis.

As seen in Figure 2.33, the original GLM analysis results were largely reproduced in the lasso regression. Namely, low-level features are more prominently encoded in early visual regions, while high-level features are more prominently encoded in higher visual regions. In this additional analysis, such effects were clearly seen across five out of six participants, while one participant (P1) showed less clear early vs late region-specific differentiation with regard to low vs high-level feature representation. We also note that the model's predictive accuracy in visual regions was lower for this participant (P1) than for the rest of the participants (Figure 2.34).

Non-linear feature representations

We found that features of the LFS model are represented across brain region and contribute to value computation. However, it is possible that nonlinear combinations of these features are also represented in the brain and that these may contribute to value computation. To explore this possibility, we constructed a new set of nonlinear features by multiplying pairs of the LFS model's features (interaction terms). We grouped these new features into three groups: interactions between pairs of low-level features (low-level x low-level), interactions between pairs of low-level and high-level features (low-level x high-level), and interactions between pairs of high-level features (high-level x high-level). To control the dimensionality of the new

feature groups, we performed principal component analysis within each of the three groups of non-linear features, and took the first five PCs to match the number of the high-level features specified in our original LFS model. We performed a LASSO regression analysis with these new features and the original features.

We found that in most participants, non-linear features created from pairs of high level features produced significant correlations with neural activity across multiple regions, while also showing similar evidence for a hierarchical organization from early to higher order regions, as found for the linear high level features (Figures 2.14 and 2.35). Though comparisons between separately optimized lasso regressions should be cautiously interpreted, the mean correlations of the model with both linear and nonlinear features across ROIs showed a slight improvement in predictive accuracy compared to the original LFS model with only linear features (Figure 2.34), while the DCNN model features out-performed both the original LFS model and the LFS model + nonlinear features.

Indeed, nonlinear features created from pairs of high-level features significantly contribute more to behavioral choice predictions than do other nonlinear features not built solely from high-level features (Figure 2.36). The first principal component of high level x high level features well captured three participants (3,5,6) behavior, while other participants show somewhat different weight profiles. However, we found that these newly added features only modestly improved the model's behavioral predictions (Figure 2.37).

DCNN model representations

We then tested whether activity patterns in these regions resemble the computations performed by the hidden layers of the DCNN model. We extracted the first three principal components from each layer of the DCNN, and included each as regressors in a GLM. Indeed, we found evidence that both the ventral and dorsal visual stream exhibits a similar hierarchical organization to that of the DCNN, such that lower visual areas correlated better with activity in the early hidden layers of the DCNN, while higher-order visual areas (in both visual streams) tend to correlate better with activity in deeper hidden layers of the DCNN (Figure 2.13cd).

We also performed additional analyses with LASSO regression using the DCNN features. To test if we can reproduce the DCNN results originally performed with the GLM approach (as shown in Figure 2.13), we first performed LASSO regression with the same 45 features from all hidden layers. Hyperparameters were optimized

by 12-fold cross-validation. The estimated weights were compared against the null distribution of each ROI constructed from the same analysis with shuffled stimuli labels. We then also performed the same analysis but with a larger set of features (150 features). In Figures 2.38 and 2.39, we show how the weights on features from different layers varied across different ROIs in the visual stream. We computed the sum of squared weights of hidden layer groups (layer 1–4, 5–9, 10–13, 14–15). Again, in order to discard weight estimates that can be obtained by chance, we computed a null distribution by repeating the same analysis with shuffled labels and took the weight estimates that are significantly larger than the null distribution (at $p < 0.001$) in each ROI. We again found that LASSO regression with within-subject cross validation reproduced our original GLM analysis results.

As a further control analysis, we asked whether similar results could be obtained from a DCNN model with random, untrained, weights (Kell et al., 2018). We repeated the same LASSO regression analysis as we did in our analysis with the trained DCNN model. We found that such a model does not reproduce the finding of a hierarchical representation of layers that we found across the visual stream and other cortical areas as in the analysis with trained DCNN weights (Figures 2.40 and 2.41).

PPC and PFC show mixed coding of low- and high-level features

We next probed these representations in downstream regions of association cortex (Baizer et al., 1991; Rao et al., 1997). We performed the same analysis with the same GLM as before in regions of interest that included the posterior parietal cortex (PPC), lateral prefrontal cortex (lPFC) and medial prefrontal cortex (mPFC). We found that both the LFS model features and the DCNN layers were represented in these regions in a mixed manner (Rigotti et al., 2013; Zhang et al., 2017). We found no clear evidence for a progression of the hierarchical organization that we had observed in the visual cortex; instead, each of these regions appeared to represent both low and high-level features to a similar degree (Figure 2.15a). Activity in these regions also correlated with hidden layers of the DCNN model (Figure 2.15b). We obtained similar results using a LASSO regression analysis with cross validation based on either the LFS model features (Figure 2.42) or the DCNN features (Figure 2.43 and 2.44). These findings suggest that, as we will see, these regions appear to play a primary role in feature integration as required for subjective value computations.

Features encoded in PPC and lPFC are strongly coupled to the subjective value of visual art in mPFC

Having established that both the engineered LFS model and the emergent DCNN model features are hierarchically represented in the brain, we asked if and how these features are ultimately integrated to compute the subjective value of visual art. First, we analyzed how aesthetic value is represented across cortical regions alongside the model features by adding the participant's subjective ratings to the GLM. We found that subjective values are, in general, more strongly represented in the PPC as well as in the lateral and medial PFC than in early and late visual areas (Figures 2.16a and Figure 2.45). Furthermore, value signals appeared to become more prominent in medial prefrontal cortex compared to the lateral parietal and prefrontal regions (consistent with a large prior literature, e.g., Cela-Conde et al. (2004); Padoa-Schioppa and Assad (2006); Kable and Glimcher (2007); Gläscher et al. (2008); Noonan et al. (2010); Grabenhorst and Rolls (2011); Ishizu and Zeki (2013)). This pattern was not altered when we control for reaction times and the distance of individual ratings from the mean ratings, proxy measures for the degree of attention paid to each image (Figure 2.46). In a further validation of our earlier feature encoding analyses, we found that the pattern of hierarchical feature representation in visual regions was unaltered by the inclusion of ratings in the GLM (Figure 2.47). We note that even when using the DCNN model to classify features as either high or low as opposed to relying on the a-priori assignment from the LFS model, this did not change the results of our fMRI analyses qualitatively and does not affect our conclusions (Figure 2.32).

These results suggest that rich feature representations in the PPC and lateral PFC could potentially be leveraged to construct subjective values in mPFC. However, it is also possible that features represented in visual areas are directly used to construct subjective value in mPFC. To test this, we examined which of the voxels representing the LFS model features across the brain are coupled with voxels that represent subjective value in mPFC at the time when participants make decisions about the stimuli. A strong coupling would support the possibility that such feature representations are integrated at the time of decision-making in order to support a subjective value computation.

To test for this, we first performed a psychological-physiological interaction (PPI) analysis, examining which voxels are coupled with regions that represent subjective value when participants made decisions (Figure 2.16b and Figure 2.48). We stress

that this is not a trivial signal correlation, as in our PPI analysis all the value and feature signals are regressed out. Therefore the coupling is due to noise correlations between voxels. Then we asked how much of the feature-encoding voxels overlap with these PPI voxels. Specifically, we tested for the fraction of feature-encoding voxels that are also correlated with the PPI regressor across each ROI. Finding overlap between feature encoding voxels and PPI connectivity effects would be consistent with a role for these feature encoding representations in value construction. We found that the overlap was most prominent in the PPC and lPFC, while there was virtually no overlap in the visual areas at all (Figure 2.16c), consistent with the idea that features in the PPC and lPFC, instead of visual areas, are involved in constructing subjective value representations in mPFC. A more detailed decomposition of the PFC ROI from the same analysis shows the contribution of individual sub-regions of lateral and medial PFC (Figure 2.49).

We also performed a control analysis to test the specificity of the coupling to an experimental epoch by constructing a similar PPI regressor locked to the epoch of inter-trial-intervals (ITIs). This analysis showed a dramatically altered coupling that did not involve the same PPC and PFC regions (Figure 2.50). These findings indicate that coupling between PPC and LPFC with mPFC value representations occurs specifically at the time that subjective value computations are being performed, suggesting that these regions are playing an integrative role of feature representations at the time of valuation. We however note that all of our analyses are based on correlations, which do not provide information about the direction of the coupling.

2.4 Discussion

Whether we can lawfully account for personal preferences in the aesthetic appreciation of art has long been an open question in the arts and sciences (Kant (1987); Fechner (1876); Zeki (2002); Chatterjee (2011)). Here, we addressed this question by engineering a hierarchical linear feature summation (LFS) model that generates subjective preference according to a weighted mixture of explicitly designed stimulus features. This model was verified with both in-depth lab-based small scale behavioural experiments and large-scale on-line behavioral experiments, and contrasted to a deep convolutional neural network (DCNN) model as well as in in-depth focused, within- subject, neuroimaging experiments. We found that it is indeed possible to predict subjective valuations of both paintings and photography using the same feature set, and we demonstrate hierarchical feature representations in a DCNN that predicts aesthetic valuations. Moreover, we demonstrate how the brain

transforms visual stimuli into subjective value, from the primary visual cortex to parietal and prefrontal cortices.

Our results indicate that linearly integrating a small set of visual features can explain human preferences for paintings and photography. Not only is it possible to predict an individual's ratings based on that particular individual's prior ratings for other images, but we also found that this strategy allowed us to predict one individual's preferences from the preferences of others, even for novel stimuli. This is achievable likely because the majority of participants shared substantial variance in their preferences, and the model efficiently extracted this, as shown by our clustering analysis whereby one dominant cluster was found to account for the majority of participants' liking ratings. Our results are consistent with a number of empirical aesthetics studies proposing that statistical properties of images can account for aesthetic values (e.g., Bar and Neta (2006); Mallon et al. (2014); Graham and Field (2008)).

We also found that the LFS model with the same visual feature set can predict subjective values for both visual art and for diverse photographic stimuli. This suggests that the features used for visual aesthetic judgement may not be domain-specific but universal, relying on a small set of visual features shared across visual stimuli. Our findings also hint that the extraction of these features might be a natural consequence of developing a visual system. We found that a DCNN model trained on object recognition and valuation represents those features throughout the hidden layers. Further studies could investigate whether such feature-extractions and feature-based value judgement are universal computations not only in visual processing but also other sensory domains such as in audition and olfaction.

The cluster analysis we ran on the large-scale online study showed that there is variation in preferences across individuals. A substantial component of that variation is whether or not participants liked concrete art or abstract art: the majority assigned large positive weights to concreteness, while the others assigned large negative weights. This indicates that concreteness alone is explaining a substantial part of the variance, and accounting for variation in preferences across groups of individuals. However, it should be noted that while concreteness does account for a substantial portion of variance in people's preferences, other high-level features also play an important role, including dynamics and valence.

We also note that, though such high-level features, including concreteness, can be used to predict preference, much if not most of the significant variance explained by such high-level features can also be explained directly as a linear combination

of a number of low-level visual features. This is consistent with the idea that many low-level sensory features (that are present in early layers in DCNN) are transformed into a smaller number of task-specific high-level features (that are present in deeper layers in DCNN), which are in turn used to predict subjective value.

It is also important to note there was in addition variance across individuals in preferences that the model did not capture well (Vessel et al. (2018); Vessel and Rubin (2010)). Thus, while art preferences share some commonalities across clusters of individuals, there is in addition some degree of individual variability. We found that the degree of commonality in subjective ratings in our study was in a similar range to previous studies (e.g., the average correlation between our M-turk samples was 0.45, which is similar to Vessel et al. (2018)). We nonetheless should stress that in our study there are two limitations. One is that most of our in-lab and online participants are not art experts and it thus remains possible that artistically experienced people might judge artworks differently. The second is that we only covered a relatively narrow subset of art genres, leaving open the possibility that there may be some art genres for which the model may not perform well. That said, we also validated our models using a wide range of photographs, indicating the potential generalizability of our findings even beyond drawings. In fact, previous studies (e.g., Graham and Field (2008)) suggest that artworks and natural scenes share some statistical regularities, of which our model might be able to take advantage.

It should also be noted that the predictive power of our model varied across participants. One possibility is that some participants were more reliable/consistent in reporting their ratings. Unfortunately, we did not present the same stimuli multiple times to each participant, making it difficult to directly test the consistency of participants' choices. However we did present the same set of stimuli to each participant. We thus directly tested how ratings of each painting were similar across participants. To test this, we computed the average ratings over n-1 participants of each painting and computed correlation between the average ratings and the ratings of the remaining participant. We performed this for each participant, and found that the correlation systematically co-varied with the within participant predictions of the model. This suggests that variability in predictability is largely due to noise in participant's preferences.

Here, utilizing a set of interpretable visual and emotional features, we showed that these features are employed by individuals to make value judgments for art. We note that this is by no means a complete enumeration of the features used by humans.

For instance, the semantic meaning of a painting, its historical importance, as well as memories of past experiences elicited by the painting, are also likely to play important roles (see, e.g., Leder et al. (2004); Palmer et al. (2013)). Thus, rather than offering a feature catalogue, our findings shed light on the general principles by which feature integration yields to aesthetic valuation. However, the features that we identified are likely to be important, particularly as we utilized a reasonably large set of potential features in our initial feature set which was subsequently narrowed down to a set of only the most relevant features.

We found evidence using a DCNN model that the features engineered for use in the LFS model spontaneously emerge in a neurally plausible manner. Our deep network model was not explicitly trained on any of the LFS model's features, nor were the convolutional hidden layers of the network trained on liking ratings (only the later fully connected layers were trained using rating scores). Nevertheless, we were still able to identify LFS features from the hidden layers of the network, which suggests that the features used for aesthetic valuation likely emerge spontaneously through more basic and generalizable aspects of visual development. Further, those features may well be utilized for a wide range of visual tasks, including object classification, prediction, and identification. Thus, we speculate that these findings suggesting a common feature space shared across different tasks may provide insights into transfer learning (Bengio (2012)) in machine learning.

One important consideration is whether linear feature operations are sufficient to describe the computations underlying aesthetic valuation. Notably, the highly non-linear deep network did not substantively outperform the simple linear model. However, in the LFS model, the feature extraction process itself is not necessarily linear (e.g., segmentation). As such, our results do not rule out the possibility of non-linearity in feature extraction processes in the brain, but they do suggest that the final feature value integration for computing subjective art valuation can be approximated by a linear operation. This computational scheme resonates with a widely-used machine-learning technique referred to as the kernel method, whereby inputs are transformed into a high-dimensional feature space in which categories are linearly separable (Leshno et al. (1993); Hofmann et al. (2008)), as well as with high-dimensional task-related variables represented in the brain (Rigotti et al. (2013)).

Focusing first on the visual system, we found that low-level features that predict visual art preferences are represented more robustly in early visual cortical areas,

while high-level features that predict preferences are increasingly represented in higher-order visual areas. These results support a hierarchical representation of the features required for valuation of visual imagery, and further support a model whereby lower-level features extracted by early visual regions are integrated to produce higher-level features in the higher visual system Chatterjee (2003). While the notion of hierarchical representations in the visual system is well established in the domain of object recognition Van Essen and Maunsell (1983); Felleman and Van (1991); Hochstein and Ahissar (2002); Konen and Kastner (2008), our results substantially extend these findings by showing that features relevant to a very different behavioral task — forming value judgments, are also represented robustly in a similar hierarchical fashion.

We then showed that the process through which feature representations are mapped into a singular subjective value dimension in a network of brain regions, including the posterior parietal cortex (PPC), lateral and medial prefrontal cortices (lPFC and mPFC). While previous studies have hinted at the use of such a feature-based framework in the prefrontal cortex (PFC), especially in orbitofrontal cortex (OFC), in those previous studies the features were more explicit properties of a stimulus (e.g., the movement and the color of dots (Kahnt et al., 2011b; Mante et al., 2013; Pelletier and Fellows, 2019), or items that are suited to a functional decomposition such as food odor (Howard and Gottfried, 2014) or nutritive components of food (Suzuki et al., 2017a); see also Hare et al. (2009); Lim et al. (2013)). Here we show that features relevant for computing subjective value of visual stimuli are widely represented in lPFC and PPC, whereas subjective value signals are more robustly represented in parietal and frontal regions, with the strongest representation in mPFC.

Further, we showed that PFC and PPC regions encoding low- and high-level features enhanced their coupling with the mPFC region encoding subjective value at the time of image presentation. While further experiments are needed to infer the directionality of the connectivity effects, our findings are compatible with a framework in which low and high-level feature representations in lPFC and PPC are utilized to construct value representations in mPFC, as we hypothesized in the LFS model.

Going beyond our original LFS model, we also found that in most participants, nonlinear features created from pairs of high level features specified in the original model produced significant correlations with neural activity across multiple regions, while largely showing similar evidence for a hierarchical organization from early to higher

order regions, as found for the linear high level features. These findings indicate that the brain encodes a much richer set of features than our original proposed set of low-level and high-level features as specified in the original LFS model. It will be interesting to see if the nonlinear features that we introduced here, especially the ones that were constructed from pairs of high-level features, can also be used to support behavioral judgments beyond the simple value judgments studied here, such as object recognition and other more complex judgements (Durkin et al., 2020). We also note that there are other ways to construct nonlinear features. Further studies with richer set of features, e.g, other forms of interactions, may improve behavioral and neural predictions

Accumulating evidence has suggested that value signals can be found widely across the brain including even in sensory regions (e.g., Hampton and O'doherty (2007); Serences (2008); Chatterjee et al. (2009); Stănişor et al. (2013); FitzGerald et al. (2013); Suzuki et al. (2017a); Bach et al. (2017)), posing a question about the differential contribution of different brain regions if value representations are so ubiquitous. While we also saw multiple brain regions that appeared to correlate with value signals during aesthetic valuation, our results suggest an alternative account for the widespread prevalence of value signals, which is that some components of the value signals especially in sensory cortex might reflect features that are ultimately used to construct value in later stages of information processing, instead of the value itself. Because neural correlates of features have not been probed previously, our results suggest that it may be possible to reinterpret at least some apparent value representations as reflecting the encoding of precursor features instead of value per se. In the present case even after taking into account feature representations, value signals were still detectable in the medial prefrontal cortex and elsewhere, supporting the notion that some brain regions are especially involved in value coding more than others. In future work it may be possible to even more clearly dissociate value from its sensory precursors by manipulating the context in which stimuli are presented, wherein features remain invariant across contexts, while the value changes. In doing so, further studies can illuminate finer dissociations between features and value signals (O'Doherty et al., 2021c).

While previous studies have suggested similarities between representations of units in DCNN models for object recognition and neural activity in the visual cortex (e.g., Cadieu et al. (2014); Khaligh-Razavi and Kriegeskorte (2014); Güçlü and van Gerven (2015); Hong et al. (2016)), here we show that the DCNN model can

also be useful to inform how visual features are utilized for value computation across a broader expanse of the brain. Specifically, we found evidence to support the hierarchical construction of subjective value, where the early layers of DCNN correlate early areas of the visual system, and the deeper layers of DCNN correlate higher areas of the visual system. All of the DCNN layers' information was equally represented in the PPC and PFC.

These findings are consistent with the suggestion that the hierarchical features which emerge in the visual system are projected into the PPC and PFC to form a rich feature space to construct subjective value. Further studies using neural network models with recurrent connections (Kar and DiCarlo, 2020) may illuminate more detail, such as the temporal dynamics, of value construction in such a feature space across brain regions.

Although the deep neural network approach has been successfully applied to a wide range of machine learning problems (e.g., Esteva et al. (2017); LeCun et al. (2015)), the underlying computational mechanisms that deep neural networks leverage in order to attain high performance across domains are opaque and often poorly understood. Here, we provide evidence that the hidden layers in the deep network encode low-level and high-level features relevant for computing aesthetic visual preferences in a hierarchical manner, which are utilized to produce coherent behavioral outputs. Thus our study provides a clear link from distributed neuronal computations to interpretable, explicit, hierarchical feature representations. Our study thus highlights the merits of a model-based analysis of artificial neural networks in order to better understand the nature of the computations implemented therein. We however caution that we do not claim that the DCNN model necessarily provides a plausible account of actual neural computations going on in the brain. Unlike a DCNN which is exclusively feedforward in its connections between layers, the brain is heavily recurrent, and thus is likely to be better approximated by networks with recurrent architecture.

One open question is how the brain has come to be equipped with a feature-based value construction architecture. We showed that a DCNN model trained solely on object recognition tasks represents the LFS's low- and high-level features in the hidden layers in a hierarchical manner, suggesting the possibility that such features could naturally emerge over development (Iigaya et al., 2021). While the similarity between the DCNN and the LFS model correlations with fMRI responses in adult participants provides a promising link between these models and the brain, further

investigations applying these models to studies with children or other species has the potential to inform understanding of the origin of feature-based value construction across development and across species.

Following the typical approach utilized in non-human primate and other animal neurophysiology as well as in human visual neuroimaging, we performed in-depth scanning (20 sessions) in a relatively small number of participants (six) in order to address our neural hypotheses. Because we were able to obtain a sufficient amount of fMRI data in individual participants, we were able to reliably perform single-subject inference in each participant and evaluate the results across participants side-by-side. This approach contrasts with a classic group-based neuroimaging study in which results are obtained from the group average of many participants, where each participant performs short sessions, thus providing data with low signal to noise. One advantage of our approach over the group averaging approach is that we can treat each participant as a replication unit, meaning that we can obtain multiple replications (Smith and Little, 2018) from one study instead of just one group result. If every participant shows similar patterns, then it is unlikely that those results are spurious, and much more likely they reflect a true property of human brain function. We indeed found that all participants similarly performed our-hypothesized feature-based value construction across the brain. Another advantage of our methodological approach concerns possible heterogeneity across participants. Not all brains are the same, and there is known to be considerable variation in the location and morphology of different brain areas across individuals (Llera et al., 2019). Thus, it is unlikely that all brains actually represent the same variable at the same MNI coordinates. The individual subject-based approach to fMRI analyses used here takes individual neuroanatomical variation into account, allowing for generalization that goes beyond a spatially smoothed average that does not represent any real brain. We note that one important limitation of this in-depth fMRI method is that it is not ideal for studying and characterizing differences across individuals. To gain a comprehensive account of such variability across individuals. it would be necessary to collect data from a much larger cohort of participants. As it is not feasible to scale the in-depth approach to such large cohorts due to experimenter time and resource constraints, such individual difference studies would necessarily require adopting more standard group-level scanning approaches and analyses.

While we found that results from the visual cortex were largely consistent across participants, the proportion of features represented in PCC and PFC, as well as

the features that were used, were quite different across participants. Understanding such individual differences will be important in future work. For instance, there is evidence that art experts tend to evaluate art differently from people with no artistic training (Hekkert and van Wieringen, 1996; Chatterjee and Vartanian, 2014). It would be interesting to study if feature representations may differ between experts and non-experts, while probing whether the computational motif that we found here (hierarchical visual feature representation in visual areas, value construction in PPC and PFC) might be conserved across different levels of expertise. We should also note that the model's predictive accuracy about liking ratings varied across participants. It is likely that some participants used features that our model did not consider, such as personal experience associated with stimuli. Brain regions such as the hippocampus may potentially be involved in such additional feature computations. Further, behavior and fMRI signals can be inherently noisy in that there will be a portion of data that cannot be predicted (i.e., a noise ceiling). Characterizing the contribution of these noise components will require further experiments with repeated measurements of decisions about the same stimuli.

The present findings offer a mechanism through which artistic preferences can be predicted. It is of course important to note that aesthetic experience more broadly defined goes beyond the simple one-dimensional liking rating (a proxy of valuation) that we study here (e.g., Zeki (2002); Chatterjee (2011); Palmer et al. (2013)), and that judgments can be context-dependent (Brieber et al. (2015)). Art is likely to be perceived along many dimensions, of which valuation is but one, with some dimensions relying more on idiosyncratic experience than others. Nevertheless, we speculate that just as it is possible to explain aesthetic valuation in terms of underlying features, many other aspects of the experience of art can also likely be decomposed into more atomic feature-based computations, with different dimensions employing different weights over those features. Indeed, subjective value can itself be considered to be a "feature" in a feature space, albeit a high-level one, alongside other judgments that might be made about a piece of art. Further, although we did not find evidence for this in the present study, it is undoubtedly the case that various psychological processes such as attention are likely to dynamically modulate the relative weights over features that construct subjective value, as well as modulating underlying neural activity (Lim et al. (2013)).

Taken together, these findings are consistent with the existence of a large-scale processing hierarchy in the brain that extends from early visual cortex to medial pre-

frontal cortex, whereby visual inputs are transformed into various features through the visual stream. These features are then projected to PPC and lPFC, and subsequently integrated into subjective value judgment in mPFC. Crucially, the flexibility afforded by such a feature-based mechanism of value construction ensures that value judgments can be formed even for stimuli that have never before been seen, or in circumstances where the goal of valuation varies (e.g., selecting a piece of art as a gift). Therefore, our study proposes a brain-wide computational mechanism that does not limit to aesthetics, but can be generalized to value constrictions of a wide range of visual and other sensory stimuli.

2.5 Methods

Participants

All participants provided informed consent for their participation in the study, which was approved by the Caltech IRB.

Online study: A total of 1936 volunteers (female: 883 (45.6%). age 18-24 yr: 285 (14.8%); 25-34 yr: 823 (42.8%); 35-44 yr: 435 (22.6%); 45 yr and above: 382 (19.8%)) participated in our on-line studies in the Amazon Mechanical Turk (M-turk). 1545 of them participated in the ART task, and 391 of them participated in the AVA photo task. Among these, participants who missed trials and failed to complete 50 trials were excluded from our analyses, leaving us with online 1359 participants in the ART task data and 382 participants in the AVA photo task data.

In-lab study: Seven volunteers (female: 3. age 18-24 yr: 5; 25-34 yr: 5; 35-44 yr: 3. 4 Asian, 3 Caucasian) were recruited to our in-lab study The in-lab participants did not include lab members but were instead recruited from the local community in Pasadena. Seven participants completed master's degree or higher. None of the participants possessed an art degree. Six of the participants reported that they visit art museums less than once a month, while one participant reported visiting art museums at least once but less than four times a month.

fMRI study: Six volunteers (female: 6; age 18-24 yr: 4; 25-34 yr: 1; 35-44 yr: 1. 4 White, 2 Asian) were recruited into our fMRI study. 1 participants completed master's degree or higher, 4 participants earned a college degree as the highest level, and 1 participant had a high-school degree as the highest degree. None of the participants possessed an art degree. All of the participants reported that they visit art museums less than once a month.

Additionally, thirteen art-experienced participants [reported in our previous behav-

ioral paper (Iigaya et al., 2021)] (female: 6; ages 18-24 yr: 3; 25-34 yr: 9; 35-44 yr: 1) were invited to evaluate the high-level feature values (outside the scanner). These participants for annotation were primarily recruited from the ArtCenter College of Design community.

Stimuli

The same stimuli as our recent behavioral study (Iigaya et al., 2021) were used in the current fMRI study. Painting stimuli were taken from the visual art encyclopedia www.wikiart.org. Using a script that randomly selects images in a given category of art, we downloaded 206 or 207 images from four categories of art (825 in total). The categories were 'Abstract Art,' 'Impressionism,' 'Color Fields,' and 'Cubism'. We randomly downloaded images with each tag using our custom code in order to avoid subjective bias. We supplemented this database with an additional 176 paintings that were used in a previous study (Vaidya et al., 2017). For the fMRI study reported here, one image was excluded from the full set of 1001 images to have an equal number of trials per run (50 images/run $\times$ 20 runs = 1000 images).

Picture images were taken from the Aesthetic Visual Analysis (AVA) dataset. This dataset consists of images from multiple online photo contests. We took images from the following categories (about 90 images from each): 'Animals,' 'Floral,' 'Nature,' 'Sky,' 'Still Life,' 'Advertisement,' 'Sky,' and 'Abstract Pictures'. In a total of 716 images were used.

Tasks

Behavioral task

Liking rating task: On each trial, participants were presented with an image of the artwork (in the Art-liking Rating Task: ART) or a picture image (in the AVA photo task) on the computer screen. Participants reported within 6 seconds how much they like the artwork (or the picture image), by pressing buttons corresponding to a scale that ranged from 0, 1, 2, 3, where 0 = not like at all, 1= like a little, 2 = like, and 3 = strongly like, presented at the bottom of the image. Each of the on-line ART participants performed on average 57 trials of the rating task, followed by a familiarity task in which they reported if they could recognize the name of the artist who painted the artwork for the same images that they reported their liking ratings. The images for each online participant were drawn to balance different art genres. On-site ART participants performed 1001 trials of rating tasks. On-site participants had a chance to take a short break approximately every 100 trials. Each of the

on-line AVA photo task participants performed on average 115 trials of the rating task.

Feature annotation: The four high-level features were annotated in a manner following Chatterjee et al. (2010); Vaidya et al. (2017). On each trial, participants were asked about the feature value of a given stimulus, ranged from -2, -1, 0, 1, 2. Following Chatterjee et al. (2010), example figures showing extreme feature values are always shown on the screen as a reference (please see Figure 2.24). Each participants completed four separate tasks (for four features) in a random order, where each task consists of 1001 trials (with 1001 images).

fMRI task

On each trial, participants were presented with an image of the artwork on the computer screen for three seconds. Participants were then presented with a scale from 0, 1, 2, 3 in which they had to indicate how much they liked the artwork. The location of each numerical score was randomized across trials. Participants had to press a button of a button box that they hold with both hands to indicate their rating within three seconds, where each of four buttons corresponded to a particular location on the screen from left to right. The left (right) two buttons were instructed to be pressed by their left (right) thumb. After a brief feedback period showing their chosen rating (0.5 sec), a center cross was shown for inter-trial intervals (jittered between 2 to 9 seconds). Each run consists of 50 trials. Participants were invited to the study over four days to complete twenty runs, where participants completed on average five runs on each day.

fMRI data acquisition

fMRI data were acquired on a Siemens Prisma 3T scanner at the Caltech Brain Imaging Center (Pasadena, CA). With a 32-channel radiofrequency coil, a multi-band echo-planar imaging (EPI) sequence was employed with the following parameters: 72 axial slices (whole-brain), A-P phase encoding, –30 degrees slice tilt with respect to AC-PC line, echo time (TE) of 30ms, multi-band acceleration of 4, repetition time (TR) of 1.12s, 54-degree flip angle, 2mm isotropic resolution, echo spacing of 0.56ms. 192mm x 192mm field of view, in-plane acceleration factor 2, multi-band slice acceleration factor 4.

Positive and negative polarity EPI-based field maps were collected before each run with very similar factors as the functional sequence described above (same acquisition box, number of slices, resolution, echo spacing, bandwidth and EPI

factor), single band, TE of 50ms, TR of 5.13s, 90-degree flip angle.

T1-weighted and T2-weighted structural images were also acquired once for each participant with 0.9mm isotropic resolution. T1's parameters were: repetition time (TR) 2.4 s: echo time (TE), 0.00232 s; inversion time (TI) 0.8 s; flip angle, 10 degrees; , in-plane acceleration factor 2. T2's parameters were: TR 3.2 s; TE 0.564s; flip angle, 120 degrees; in-plane acceleration factor 2.

fMRI data processing

Results included in this manuscript come from preprocessing performed using *fMRIPrep* 1.3.2 (Esteban et al. (2018a); RRID:SCR_016216), which is based on *Nipype* 1.1.9 (Gorgolewski et al. (2018); RRID:SCR_002502).

Anatomical data preprocessing

The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with N4Bias Field Correction Tustison et al. (2010), distributed with ANTs 2.2.0 (Avants et al., 2008, RRID:SCR_004757), and used as T1w-reference throughout the workflow. The T1w-reference was then skull-stripped with a *Nipype* implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as target template. Spatial normalization to the *ICBM 152 Nonlinear Asymmetrical template version 2009c* was performed through nonlinear registration with `antsRegistration` (ANTs 2.2.0), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using `fast`.

Functional data preprocessing

For each of the 20 BOLD runs found per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. A deformation field to correct for susceptibility distortions was estimated based on two echo-planar imaging (EPI) references with opposing phase-encoding directions, using `3dQwarp`(AFNI 20160207). Based on the estimated susceptibility distortion, an unwarped BOLD reference was calculated for a more accurate co-registration with the anatomical reference.

The BOLD reference was then co-registered to the T1w reference using `flirt` with the boundary-based registration cost-function. Co-registration was configured with nine degrees of freedom to account for distortions remaining in the BOLD reference. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcflirt` The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD time-series were resampled to MNI152NLin2009cAsym standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD and DVARS are calculated for each functional run, both using their implementations in *Nipype*.

The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction. Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). Six tCompCor components are then calculated from the top 5% variable voxels within a mask covering the subcortical regions. This subcortical mask is obtained by heavily eroding the brain mask, which ensures it does not include cortical GM regions. For aCompCor, six components are calculated within the intersection of the aforementioned mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run (using the inverse BOLD-to-T1w transformation).

The head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. All resamplings can be performed with *a single interpolation step* by composing all the pertinent transformations (i.e., head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and template spaces). Gridded (volumetric) resamplings were performed using ants Apply Transforms (ANTs), configured with Lanczos

interpolation to minimize the smoothing effects of other kernels.

Computational models

The computational methods and behavioral modeling reported in this manuscript overlap with that reported in our recent article focusing exclusively on behavior Iigaya et al. (2021). For completeness, we reproduce some of the descriptions of these methods as first described in Iigaya et al. (2021).

Linear feature summation model (LFS model)

We hypothesized that subjective preferences for visual stimuli are constructed by the influence of visual and emotional features of the stimuli. As its simplest, we assumed that the subjective value of the i-th stimulus v_i is computed by a weighted sum of feature values $f_{i,j}$:

$$v_i = \sum_{j=0}^{n_f} w_j f_{i,j} \tag{2.1}$$

where w_j is a weight of the j-th feature, $f_{i,j}$ is the value of the j-th feature for stimulus i, and n_f is the number of features. The 0-th feature is a constant $f_{i,0} = 1$ for all i's.

Importantly, w_j is not a function of a particular stimulus but shared across all visual stimuli, reflecting the *taste* of a participant. The same taste (w_j's) can also be shared across different participants, as we showed in our behavioral analysis. The features $f_{i,j}$ were computed using visual stimuli; we used the same feature values to predict liking ratings across participants. We used the simple linear model Eq.(2.1) to predict liking ratings in our behavioral analysis (see below for how we determined features and weights).

As we schematically showed in Figure 2.10, we hypothesized that the input stimulus is first broke down into low-level features and then transformed into high-level features, and indeed we found that a significant variance of high-level features can be predicted by a set of low-level features. This hierarchical structure of the LFS model was further tested in our DCNN and fMRI analysis.

Features

Because we did not know a priori what features would best describe human aesthetic values for visual art, we constructed a large feature set using previously published

methods from computer vision augmented with additional features that we ourselves identified using additional existing machine learning methods.

Visual low-level features introduced in Li and Chen (2009)

We employed 40 visual features introduced in Li and Chen (2009). We do not repeat descriptions of the features here; but briefly, the feature sets consist of 12 global features that are computed from the entire image that include color distributions, brightness effects, blurring effects, and edge detection, and 28 local features that are computed for separate segments of the image (the first, the second and the third largest segments). Most features are computed straightforwardly in either HSL (hue, saturation, lightness) or HSV (hue, saturation, value) space (e.g., average hue value).

One feature that deserves description is a blurring effect. Following Ke et al. (2006); Li and Chen (2009), we assumed that the image I was generated from a hypothetical sharp image with a Gaussian smoothing filter with an unknown variance σ. Assuming that the frequency distribution for the hypothetical image is approximately the same as the blurred, actual image, the parameter σ represents the degree to which the image was blurred. The σ was estimated by the Fourier transform of the original image by the highest frequency, whose power is greater than a certain threshold.

$$f_{\text{blur}} = max\left(k_x, k_y\right) \propto \frac{1}{\sigma} \tag{2.2}$$

where $k_x = 2(x - n_x/2)/n_x$ and $k_y = 2(y - n_y/2)/n_y$ with (x, y) and (n_x, n_y) are the coordinates of the pixel and the total number of pixel values, respectively. The above max was taken within the components whose power is larger than four (Li and Chen, 2009).

The segmentation for this feature set was computed by a technique called kernel GraphCut (Rother et al., 2004; Salah et al., 2010). Following Li and Chen (2009), we generated a total of at least six segments for each image using a C^{++} and Matlab package for kernel graph cut segmentation (Salah et al., 2010). The regularization parameter that weighs the cost of cut against smoothness was adjusted for each image in order to obtain about six segments. See Salah et al. (2010); Li and Chen (2009) for the full description of this method and examples.

Of these 40 features, we included all of them in our initial feature set except for local features for the third-largest segment, which were highly correlated with features for

the first and second-largest segments and were thus deemed unlikely to add unique variance to the feature prediction stage.

Additional Low-Level Features

We assembled the following low-level features to supplement the set by Li and Chen (2009). These include both global features and local features. Local features were calculated on segments determined by two methods. The first method was statistical region merging (SRM) as implemented by Nock and Nielsen (2004), where the segmentation parameter was incremented until at least three segments were calculated. The second method converted paintings into LAB color space and used k-means clustering of the A and B components. While the first method reliably identified distinct shapes in the paintings, the second method reliably identified distinct color motifs in the paintings.

The segmentation method for each feature is indicated in the following descriptions. Each local feature was calculated on the first and second-largest segments.

Local Features:

- Segment Size (SRM): Segment size for segment i was calculated as the area of segment i over the area of the entire image:

$$f_{\text{segment size}} = \frac{\text{area segment } i}{\text{total area}} \tag{2.3}$$

- HSV Mean (SRM): To calculate mean hue, saturation, and color value for each segment, segments were converted from RGB to HSV color space.

$$f_{\text{mean hue}} = \text{mean(hue values in segment } i) \tag{2.4}$$

$$f_{\text{mean saturation}} = \text{mean(saturation values in segment } i) \tag{2.5}$$

$$f_{\text{mean color value}} = \text{mean(color values in segment } i) \tag{2.6}$$

- Segment Moments (SRM):

$$f_{\text{CoM X coordinate}} = \frac{\sum_{k \in \text{segment } i} x_k}{\text{area segment } i} \tag{2.7}$$

$$f_{\text{CoM Y coordinate}} = \frac{\sum_{k \in \text{segment } i} y_k}{\text{area segment } i} \tag{2.8}$$

$$f_{\text{Variance}} = \frac{\sum_{k \in \text{segment } i} (x_k - \bar{x})^2 + (y_k - \bar{y})^2}{\text{area segment } i} \tag{2.9}$$

$$f_{\text{Skew}} = \frac{\sum_{k \in \text{segment } i} (x_k - \bar{x})^3 + (y_k - \bar{y})^3}{\text{area segment } i} \tag{2.10}$$

where $(\bar{x}, \bar{y})$ is the center of mass coordinates of the corresponding segment.

- Entropy (SRM):

$$f_{\text{entropy}} = -\sum_j (p_j * \log_2(p_j)) \tag{2.11}$$

where p equals the normalized intensity histogram counts of segment i.

- Symmetry (SRM): For each segment, the painting was cropped to maximum dimensions of the segment. The horizontal and vertical mirror images of the rectangle were taken, and the mean squared error of each was calculated from the original.

$$f_{\text{horizontal symmetry}} = \frac{\sum_{x,y \in \text{segment}} (\text{segment}_{x,y} - \text{horizontal\_flip}(\text{segment})_{x,y})^2}{\text{\# pixels in segment}} \tag{2.12}$$

$$f_{\text{vertical symmetry}} = \frac{\sum_{x,y \in \text{segment}} (\text{segment}_{x,y} - \text{vertical\_flip}(\text{segment})_{x,y})^2}{\text{\# pixels in segment}} \tag{2.13}$$

- R-Value Mean (K-Means): Originally, we took the mean of R, G, and B values for each segment, but found these values to be highly correlated, so we reduced these three features down to just one feature for mean R value.

$$f_{\text{R-value}} = \text{mean(R-values in segment)} \tag{2.14}$$

- HSV Mean (K-Means): As with SRM generated segments, we took the hue, saturation, and color value means of segments generated by K-means segmentation as described in equations 2-4.

Global Features:

- Image Intensity: Paintings were converted from RGB to grayscale from 0 to 255 to yield a measure of intensity. The 0-255 scale was divided into five equally-sized bins. Each bin count accounted for one feature.

$$f_{\text{intensity count bin } i \in \{1,4\}} = \frac{\text{\# pixels with intensity} \in \{\frac{255(i-1)}{5}, \frac{255i}{5}\}}{\text{total area}} \tag{2.15}$$

- HSV Modes: Paintings were converted to HSV space, and the modes of the hue, saturation, and color value across the entire painting were calculated. While we took mean HSV values over segments in an effort to calculate overall-segment statistics, we took the mode HSV values across the entire image in an effort to extract dominating trends across the painting as a whole.

$$f_{\text{mode hue}} = \text{mode(hue values in segment } i) \tag{2.16}$$

$$f_{\text{mode saturation}} = \text{mode(saturation values in segment } i) \tag{2.17}$$

$$f_{\text{mode color value}} = \text{mode(color values in segment } i) \tag{2.18}$$

- Aspect (width-height) Ratio:

$$f_{\text{aspect ratio}} = \frac{\text{image width}}{\text{image height}} \tag{2.19}$$

- Entropy: Entropy over the entire painting was calculated according to equation 9.

High-Level Feature Set (Chatterjee et al., 2010; Vaidya et al., 2017)

We also introduced features that are more abstract and not easily computed by a simple algorithm. Chatterjee et al. (2010) pioneered this by introducing 12 features (color temperature, depth, abstract, realism, balance, accuracy, stroke, animacy, emotion, color saturation, complexity) that were annotated by human participants for 24 paintings, in which the authors have found that annotations were consistent across participants, regardless of their artistic experience. Vaidya et al. (2017) further collected annotations of these feature sets from artistically experienced participants for an additional 175 paintings and performed a principal component analysis, finding three major components that summarize the variance of the original 12 features. Inspired by the three principal components, we introduced three high-level features: concreteness, dynamics, and temperature. Also, we introduced valence as an additional high-level feature. The four high-level features were annotated in a similar manner to the previous studies (Chatterjee et al., 2010; Vaidya et al., 2017). We took the mean annotations of all 13 participants for each image as feature values. In addition, we also annotated our image set with whether or not each image included a person. This was done by manual annotation, but it can also be done with a human detection algorithm (e.g., see Zhu et al. (2006)). We included this presence-of-a-person feature in the low-level feature set originally (Iigaya et al.,

2020b), though we found in our DCNN analysis that the feature shows a signature of a high-level feature (Iigaya et al., 2020b). Therefore in this current study, we included this presence of a person to the high-level feature set. As we showed in the main text, classifying this feature as a low-level feature or as a high-level feature does not change our results.

Identifying the shared feature set that predicts aesthetic preferences

The above method allowed us to have a set of 83 features in total that are possibly used to predict human aesthetic valuation. These features are likely redundant because some of them are highly correlated, and many may not contribute to decisions at all. We thus sought to identify a minimal subset of features that are commonly used by participants. In Iigaya et al. (2020b), we performed this analysis using Matlab Sparse Gradient Descent Library [1]. For this, we first orthogonalized features by sparse PCA (Hein and Bühler, 2010). Then we performed a regression with a LASSO penalty at the group level using participants' behavioral data with a function $group - lasso - problem$. We used Fast Iterative Soft Thresholding Algorithm (FISTA) with cross-validation. After eliminating PC's that were not shared by more than one participant, we transformed the PC's back to the original space. We then eliminated one of the two features that were most highly correlated ($r^2 > 0.5$) to obtain the final set of shared features.

To identify relevant features for use in the current fMRI analysis, we utilized behavioral data from both our previous in-lab behavioral study (Iigaya et al., 2020b) and the fMRI participants included in the current study (13 participants in total). Because the goal of the fMRI analysis is to highlight the hierarchical nature in neural coding between low and high-level features, we first repeated the above procedure with low-level features alone (79 features in total) and then we added high-level features (the concreteness, the dynamics, the temperature, and the valance) to the obtained shared low-level features.

The identified shared features are the following: the concreteness, the dynamics, the temperature, the valence, the global average saturation from Li and Chen (2009), the global blurring effect from Li and Chen (2009), the horizontal coordinate of mass center for the largest segment using the Graph-cut from Li and Chen (2009), the vertical coordinate of mass center for the largest segment using the Graph-cut from Li and Chen (2009), the mass skewness for the second largest segment using

[1] https://github.com/hiroyuki-kasai/SparseGDLibrary

the Graph-cut from Li and Chen (2009), the size of the largest segment using SRM, the mean hue of the largest segment using SRM, the mean color value of largest segment using SRM, the mass variance of the largest segment using SRM, global entropy, the entropy of the second-largest segment using SRM, the image intensity in bin 1, the image intensity in bin 2, and the presence of a person.

Nonlinear interaction features

We constructed additional feature sets by multiplying pairs of LFS features. We grouped the resulting features into three groups. 1) features created from interactions between high level features 2) features created from interactions between low level features and 3) features created from interactions between a high-level and a low-level feature. In order to determine the contribution of these three groups of features, we performed PCA on each group so that we can take the same number of components from each group. In our analysis, we took five PCs from each group to match with the number of features of original high-level features.

Behavioral Model fitting

We tested how our shared-feature model can predict human liking ratings using out-of-sample tests. All models were cross-validated in twenty folds, and we used ridge regression unless otherwise stated. Hyperparameters were tuned by cross-validation. We calculated the Pearson correlation between model predictions (pooled predictions from all cross-validation sets) and actual data, and defined it as the predictive accuracy.

We estimated individual participant's feature weights by fitting a linear regression model with the shared feature set to each participant. For illustrative purposes, the weights were normalized for each participant by the maximum feature value (concreteness) in Figures 2.10g, 2.25 and 2.36.

The significance of the above analyses was measured by generating a null distribution constructed by the same analyses but with permuted image labels. The null distribution was construed by 10000 permutations. The chance level was determined by the mean of the null distribution.

Deep Convolutional Neural Network (DCNN) analysis

Network architecture

The deep convolutional neural network (DCNN) we used consists of two parts. An input image feeds into convolutional layers from the standard VGG-16 network that is pre-trained on ImageNet. The output of the convolutional layers then projects to fully connected layers. This architecture follows the current state-of-the-art model on aesthetic evaluation Murray et al. (2012); Murray and Gordo (2017).

The details of the convolutional layers from the VGG network can be found in Simonyan and Zisserman (2014); but briefly, it consists of 13 convolutional layers and 5 intervening max pooling layers. Each convolutional layer is followed by a rectified linear unit (ReLU). The output of the final convolutional layer is flattened to a 25088-dimensional vector so that it can be fed into the fully connected layer.

The fully connected part has two hidden layers, where each layer has 4096 dimensions. The fully connected layers are also followed by a ReLU layer. During training, a dropout layer was added with a drop out probability 0.5 after every ReLU layer for regularization. Following the current state of the art model Murray and Gordo (2017), the output of the fully connected network is a 10-dimensional vector that is normalized by a softmax. The output vector was weighted averaged to produce a scalar value Murray and Gordo (2017) that ranges from 0 to 3.

Network training

We trained our model on our behavioral data set by tuning weights in the fully connected layers. We employed 10-fold cross-validation to benchmark the art rating prediction. The model was optimized using a Huber loss metric, which is robust to outliers Huber (1964).

We used stochastic gradient descent (SGD) with momentum to train the model. We used a batch size of 100, a learning rate of 10^{-4}, the momentum of 0.9, and weight decay of 5×10^{-4}. The learning rate decayed by a factor of 0.1 every 30 epochs.

To handle various sizes of images, we used the zero-padding method. Because our model could only have a 224×224 sized input, we first scaled the input images to have the longer edges be 224 pixels long. Then we filled the remaining space with 0 valued pixels (black).

We used Python 3.7, Pytorch 0.4.1.post2, and CUDA 9.0 throughout the analysis.

Retraining DCNN to extract hidden layer activations

We also trained our network on single fold ART data in order to obtain a single set of hidden layer activations. We prevented over-fitting by stopping our training when the model performance (Pearson correlation between the model's prediction and data) reached the mean correlation from the 10-folds cross-validation.

Decoding features from the deep neural network

We decoded the LFS model features from hidden layers by using linear (for continuous features) and logistic (for categorical features) regression models, as we described in Iigaya et al. (2020b). We considered the activations of outputs of ReLU layers (total of 15 layers). First, we split the data into ten folds for the 10-fold cross-validation. In each iteration of the cross-validation, because dimensions of the hidden layers are much larger ($64 \times 224 \times 224 = 3211264$) than the actual data size, we first performed PCA on the activation of each hidden layer from the training set. The number of principal components was chosen to account for 80% of the total variance. By doing so, each layer's dimension was reduced to less than 536. Then the hidden layers' activations from the test set were projected onto the principal component space by using the fitted PCA transformation matrices. The hyperparameter of the ridge regression was tuned by doing a grid search, and the best performing coefficient for each layer and feature was chosen based on the scores from the 10-folds cross-validation. We tested for a total of 19 features, including all 18 features that we used for our fMRI analysis, as well as the simplest feature that was not included into our fMRI analysis (as a result of our group-level feature selection) but that was also of interest here: the average hue value. For the continuous features (e.g., rating, mean hue), Pearson correlation between the model's predication and data were used as the metric for goodness of fit, while for the categorical features (e.g., presence of person), we calculated accuracy, area under curve (AUC), and F1 scores. The sign of slopes of decoding plots from these metrics were identical. In a supplementary analysis, we also explored whether adding 'style matrices' of hidden layers Gatys et al. (2016) to the PCA-transformed hidden layer's activations can improve the decoding accuracy; however, we found the style matrices do not improve the decoding accuracy. Sklearn 0.19.2 on Python 3.7 was used.

Reclassifying features according to the slopes of the decoding accuracy across hidden layers

In our LFS model, we classified putative low-level and high-level features simply by whether a feature is computed by a computer algorithm vs annotated by humans, respectively. In reality, however, some putative low-level features are more complex in terms of how they could be constructed than other lower level features, while some putative high-level features could in fact be computed straightforwardly from raw pixel inputs. Using the decoding results of the features from hidden layers in the DCNN, we identified DCNN-defined low-level and high-level features. For this, we fit a linear slope to the estimated decoding accuracy vs hidden layers. We permuted layer labels 10,000 times and performed the same analysis to construct null distribution as described earlier. We classified a feature as high-level if the slope was significantly positive at $p < 0.001$, and we classified a feature as a low-level feature if the slope was signifcantly negative at $p < 0.001$.

The features showing negative slopes were: the average hue, the average saturation, the average hue of the largest segment using GraphCut, the average color value of the largest segment using GraphCut, the image intensity in bin 1, the image intensity in bin 3, and the temperature.

The features showing positive slopes were: the concreteness, the dynamics, the presence of a person, the vertical coordinate of the mass center for the largest segment using the Graph Cut, the mass variance of the largest segment using the SRM, the entropy in the 2nd largest segment using SRM. All of these require relatively complex computations, such as localization of segments or image identification. This is consistent with a previous study showing that object-related local features showed a similar increased decodability at a deeper layer (Hong et al., 2016).

fMRI analysis

Standard GLM analysis

We conducted a standard GLM analysis on the fMRI data with SPM 12. The SPM feature for asymmetrically orthogonalizing parametric regressors was disabled throughout. We collected enough data from each individual participant (four days of scanning) so that we can analyze and interpret each participant's results separately. The following regressors were obtained from the fmriprep preprocessing pipeline and added to all analysis as nuisance regressors: framewise displacement, compcor, non-steady, trans, rot. The onsets of Stimulus, Decision, and Action were also

controlled by stick regressors in all GLMs described below. In addition, we added the onset of the Decision period, the onset of feedback to all GLM as nuisance regressors, because we focused on the stimulus presentation period.

Identifying subjective value coding (GLM 1)

In order to gain insight into how the subjective value of art was represented in the brain, we performed a simple GLM analysis with a parametric regressor at the onset of Stimulus (GLM 1). The parameter was linearly modulated by participant's liking ratings on each trial. The results were cluster FWE collected with a height threshold of $p < 0.001$.

Identifying feature coding (GLM 2,3, 2,' 3')

In order to gain insight into how features were represented in the brain, we performed another GLM analysis with a parametric regressor at the onset of Stimulus (GLM 2,3). In GLM 2, there are in total 18 feature-modulated regressors; each representing the value of one of the shared features for the fMRI analysis. We then performed F-tests on high-level features and low-level features (a diagonal contrast matrix with an entry set to one for each feature of interest was constructed in SPM) in order to test whether a voxel is significantly modulated by any of the high and/or low-level features. We then counted the number of voxels that are significantly correlated ($p < 0.001$) in each ROI (note that the F-value for significance is different for high and low features due to the difference in the number of consisting features). We then displayed the proportions of two numbers in a given ROI.

We performed a similar analysis using the DCNN hidden layers (GLM 3). We took the first three principal components of each convolutional and fully connected layers (three PCs times 15 layers = 45 parametric regressors). We then performed F-tests on PCs from layers 1 to 4, layers 5 to 9, layers 10 to 13, and fully connected layers (layers 14 and 15). The proportions of the survived voxels were computed for each ROI.

In addition, we also performed the same analyses with GLMs to which we added liking ratings for each stimulus. We call these analyses GLM 2' and GLM 3,' respectively.

We note that, because in our LFS model the liking rating is a linear integration of features, adding liking rating regressor means to identify neural correlates of the

liking ratings that are outside of the LFS model's prediction.

Region of Interests (ROI)

We constructed ROIs for visual topographic areas using a previously published probabilistic map (Wang et al., 2014). We constructed 17 masks based on the 17 probabilistic maps taken from Wang et al. (2014), consisting of 8 ventral–temporal (V1v, V2v, V3v, hV4, VO1, VO2, PHC1, and PHC2) and 9 dorsal–lateral (V1d, V2d, V3d, V3A, V3B, LO1, LO2, hMT, and MST) masks. In this, ventral and dorsal regions for early visual areas V1, V2, V3 are separately defined. Each mask was constructed by thresholding the probability map at $p > 0.01$. We defined V_{123} as V1v +V2v+ V3v+ V1d + V2d + V3d + V3A + V3B, and V_{high} as hV4 + VO1 + VO2 + PHC1 + PHC2 + LO1 + LO2 + hMT + MST. (hV4: human V4, VO: ventral occipital cortex, PHC: posterior parahippocampal cortex, LO: lateral occipital cortex, hMT: human middle temporal area, MST: medial superior temporal area.)

We also constructed ROIs for parietal and prefrontal cortices using the AAL database. Posterior parietal cortex (PPC) was defined by bilateral MNI-Parietal-Inf + MNI-Parietal-Sup. lateral orbitofrontal cortex (lOFC) was defined by bilateral MNI-Frontal-Mid-Orb + MNI-Frontal-Inf-Orb + MNI-Frontal-Sup-Orb, and medial OFC (mOFC) was defined by bilateral MNI-Frontal-Med-Orb + bilateral MNI-Rectus. Dorsomedial PFC (dmPFC) was defined by bilateral MNI-Frontal-Sup-Medial + MNI-Cingulum-Ant, and dorsolateral PFC (dlPFC) was defined by bilateral MNI-Frontal-Mid + MNI-Frontal-Sup. Ventrolateral PFC (vlPFC) was defined by bilateral MNI-Frontal-Inf-Oper + MNI-Frontal-Inf-Tri.

We also constructed lateral PFC (LPFC) as vlPFC + dlPFC +lOFC, and medial PFC (MPFC) as mOFC + dmPFC.

PPI analysis (GLM 4, 4')

We conducted a psychobiological-physiological interaction analysis. We took a seed from the GLM 1 identified cluster showing subjective value in MPFC (Figure 2.48), and a psychological regressor as a box function, which is set to one during the stimulus epoch and 0 otherwise. We added the time course of the seed, the PPI regressor, to a variant of GLM 2' (the parametric regressors in which feature values and liking values were constructed using a boxcar function at stimulus periods, instead of its onsets) and determined which voxels were correlated with the PPI

regressor (GLM 4). Following Suzuki et al. (2017a), boxcar functions were used because feature integration can take place throughout the duration of each stimulus presentation.

We also conducted a control PPI analysis. For this we took the same seed, but now the psychological regressor was a box function which is one during ITI, and 0 otherwise. We added the time course of the seed and the PPI regressor, the box function for ITI, and the PPI regressor to the same variant of GLM 2' (the parametric regressors with feature and liking values were constructed using boxcar function at Stimulus periods, instead of its onsets). We refer to this as GLM 4'.

Feature integration analysis

We conducted an F-test using GLM 2, to test whether any of the shared features were significantly correlated with a given voxel (a diagonal with one at all features in SPM). The resulting F-map is thresholded at $p < 0.05$ cFWE at the whole-brain with height threshold at $p < 0.001$. We then asked within the survived voxels, which of them were also significantly positively correlated with PPI regressor in GLM 4, using at value thresholded at $p < 0.001$ uncorrected. We then counted the fraction of voxels that survived this test in a given ROI.

Regression analysis with cross validation

In addition to the SPM GLM analysis, we also performed regression analyses with cross validation within each participant (Naselaris et al., 2011). We first extracted beta estimates at stimulus presentation time on each trial from a GLM with regressors at each stimulus onset, where the GLM also included other nuisance regressors, including framewise displacement, comp-cor, non-steady, trans, rot, the onsets of Decision, Action and feedback. We then used these beta estimates at the stimulus presentation time as dependent variables in our regression analysis. In all fMRI analyses, we used a Lasso penalty unless otherwise stated. The hyperparameters were optimized using 12-fold cross validation. The Matlab lasso function was used. We note that each stimulus was presented only once in our experiment in a given participant.

We performed a feature coding analysis analogous to what we performed using SPM. We first estimated the weights of the LFS model features using lasso regression at each voxel. We then computed a sum of squared weights for low-level features and high-level features separately. In order to discard weight estimates that can

be obtained by chance, we also performed the same lasso regression analysis using shuffled stimuli labels. We then constructed a null distribution with a sum of squared weights at each ROI using the weight estimates from this analysis. If the sum of squared weights of low (or high) -level features obtained from correct stimuli labels at a given voxel is significantly larger than the null distribution of low (or high) -level features in the ROI ($p < 0.001$), we identified the voxel as encoding low-level (or high-level) features.

We also ran a similar analysis with the LFS model's features where we also included 'nonlinear features' that are constructed by multiplying pairs of the LFS model's features. As described above, we grouped the nonlinear features into three groups. 1) features created from interactions between high level features 2) features created from interactions between low level features, and 3) features created from interactions between high-level and low-level features. We took five PCs from each group to match with the number of original high-level features from the model.

When comparing predictive accuracy across different models, we calculated Pearson correlations between the data and each model's predictions, where the model's predictions were pooled over predictions from testing sets across cross-validations.

We performed a similar analysis using the DCNN's features, where the DCNN was trained to predict behavioral data. Using the obtained results, we computed the sum of squared features from layers one to four, layers five to nine, layers ten to thirteen, and layers fourteen to fifteen. Again, estimates that are significantly greater than the ones obtained by chance (at $p < 0.001$) were included in our results, using the same regression analysis with shuffled labeled data. We performed analyses with 45 features (3 PCs from each layer) and 150 features (10 PCs from each layer).

We also performed the same DCNN analysis using untrained, random, weights.

Data availability

The data that support the findings of this study are available at https://github.com/kiigaya/Art or from the corresponding author upon request.

Code availability

The code that support the findings of this study are available at https://github.com/kiigaya/Art or from the corresponding author upon request.

Acknowledgements

We thank Peter Dayan, Shin Shimojo, Pietro Perona, Lesley Fellows, Avinash Vaidya, Jeff Cockburn for discussions and suggestions. We also thank Ronan O'Doherty for drawing the bird and fruits, Seiji Iigaya and Erica Iigaya for drawing the color field paintings presented in this manuscript. This work was supported by NIDA grant R01DA040011 and the Caltech Conte Center for Social Decision Making (P50MH094258) to JOD, the Japan Society for Promotion of Science, the Swartz Foundation and the Suntory Foundation to KI, and the William H. and Helen Lang SURF Fellowship to IW.

Author Contributions

K.I. and J.P.O. conceived and designed the project. K.I., S.Y., I.A.W., S.T., performed experiments and K.I., S.Y., I.A.W., L.C., J.P.O. analyzed and discussed results. K.I., S.Y., I.A.W., J.P.O. wrote the manuscript.

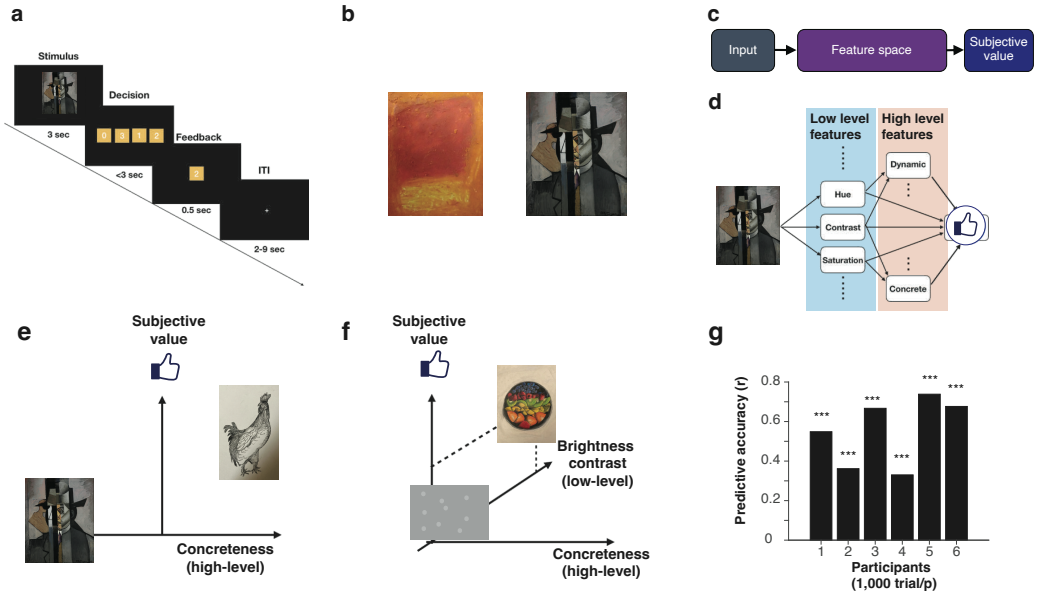


Figure 2.10: Neuroimaging experiments and the model of value construction. (**a**). Neuroimaging experiments. We administered our task (ART: art rating task) to human participants in an fMRI experiment. Each participant completed 20 scan sessions spread over four separate days (1,000 trials in total with no repetition of the same stimuli). On each trial, a participant was presented with a visual art stimulus (paintings) for 3 sec. The art stimuli were the same as in our previous behavioral study (Iigaya et al., 2021). After the stimulus presentation, a participant was presented with a set of possible ratings (0,1,2,3), where they had to choose one option within 3 seconds, followed by brief feedback with their selected rating (0.5 sec). The positions of the numbers were randomized across trials, and the order of presented stimuli was randomized across participants. (**b**). Example stimuli. The images were taken from four categories from Wikiart.org.: Cubism, Impressionism, Abstract art and Color Fields, and supplemented with art stimuli previously used (Vaidya et al., 2017). (**c**). The idea of value construction. An input is projected into a feature space, in which the subjective value judgment is performed. Importantly, the feature space is shared across stimuli, enabling this mechanism to generalize across a range of stimuli, including novel ones. (**d**). Schematic of the LFS model (Iigaya et al., 2021). A visual stimulus (e.g., artwork) is decomposed into various low-level visual features (e.g., mean hue, mean contrast), as well as high-level features (e.g., concreteness, dynamics). We hypothesized that in the brain high-level features are constructed from low-level features, and that subjective value is constructed from a linear combination of all low and high-level features. (**e**). How features can help construct subjective value. In this example, preference was separated by the concreteness feature.Reproduced from Iigaya et al. (2021). (**f**). In this example, the value over the concreteness axis was the same for four images; but another feature, in this case, the brightness contrast, could separate preferences over art. Reproduced from Iigaya et al. (2021). (**g**). The LFS model successfully predicts participants' liking ratings for the art stimuli. The model was fit to each participant (cross-validated). Statistical significance was determined by a permutation test (one-sided). Three stars indicate $p < 0.001$. Due to copyright considerations, some paintings presented here are not identical to that used in our studies. Credit: Jean Metzinger, Portrait of Albert Gleizes (public domain; RISD Museum).

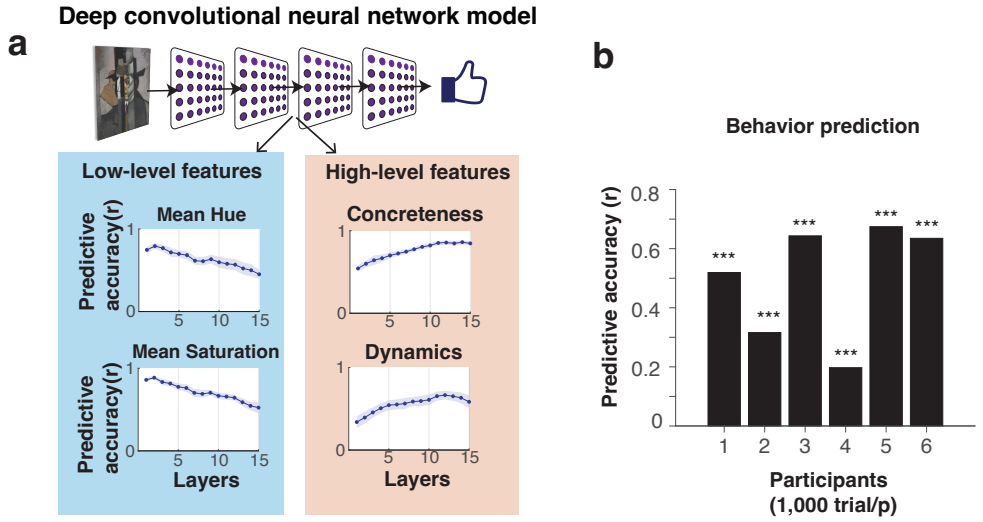


Figure 2.11: The deep convolutional neural network (DCNN) model naturally encodes low-level and high-level features and predict participants' choice behavior. (**a**). Schematic of the deep convolutional neural network (DCNN) model and the results of decoding analysis (Iigaya et al., 2021). The DCNN model was first trained on ImageNet object classifications, and then the average ratings of art stimuli. We computed correlations between each of the LFS model features and activity patterns in each of the hidden layers of the DCNN model. We found that some low-level visual features exhibit significantly decreasing predictive accuracy over hidden layers (e.g., the mean hue and the mean saturation). We also found that a few computationally demanding low-level features showed the opposite trend (see the main text). We further found that some high-level visual features exhibit significantly increasing predictive accuracy over hidden layers (e.g., concreteness and dynamics). Results reproduced from Iigaya et al. (2021).(**b**).The DCNN model could successfully predict human participants' liking ratings significantly greater than chance across all participants. Statistical significance ($p < 0.001$, indicated by three stars) was determined by a permutation test (one-sided). Credit: Jean Metzinger, Portrait of Albert Gleizes (public domain; RISD Museum).

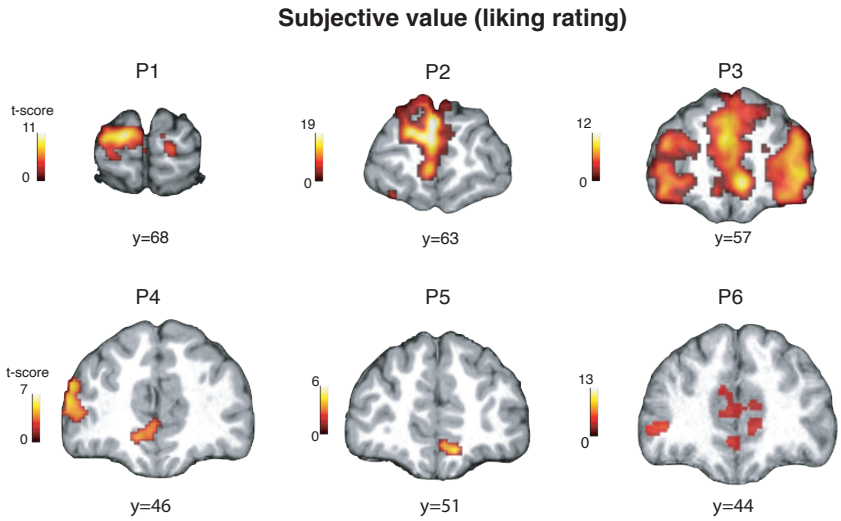


Figure 2.12: Subjective value (i.e., liking rating). Subjective value for art stimuli at the time of stimulus onset was found in the medial prefrontal cortex in all six fMRI participants (One-sided t-test. An adjustment was made for multiple comparisons: whole-brain cFWE $p < 0.05$ with height threshold at $p < 0.001$).

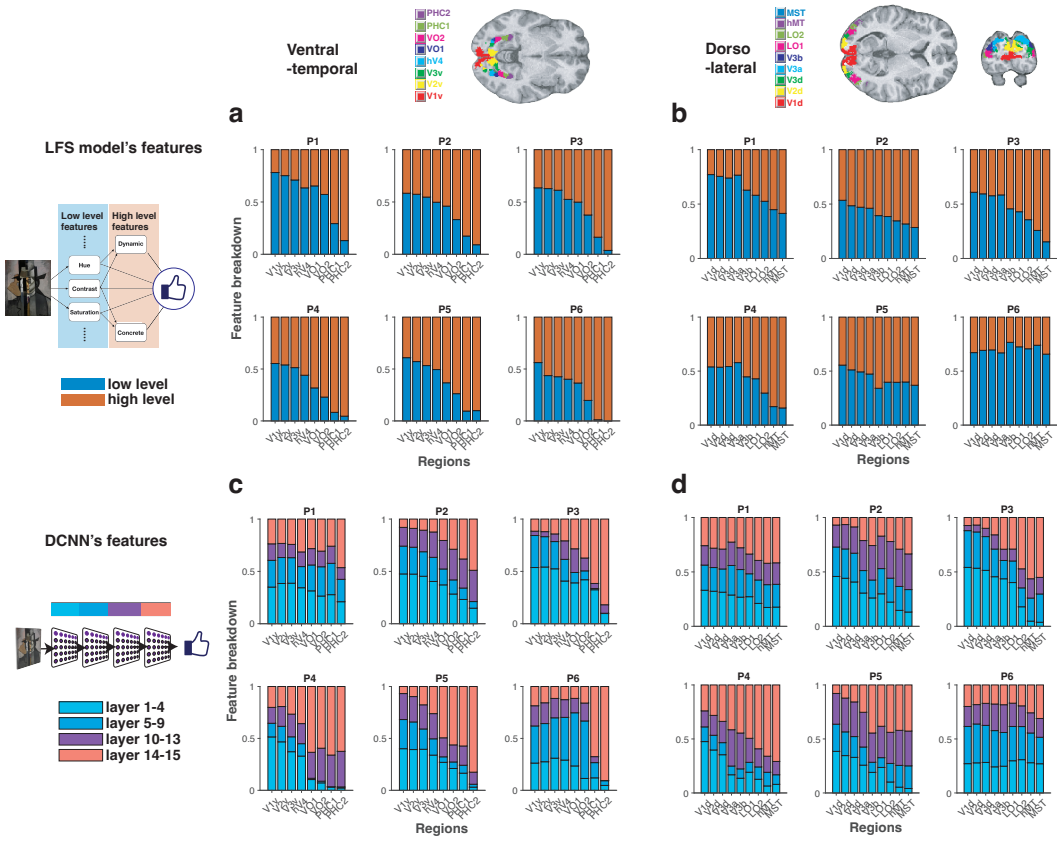


Figure 2.13: fMRI signals in visual cortical regions show similarity to our LFS model and DCNN model. (**a**). Encoding of low and high-level features in the visual ventral-temporal stream in a graded hierarchical manner. In general, the relative encoding of high-level features with respect to low-level features increases dramatically across the ventral-temporal stream. The maximum probabilistic map (Wang et al., 2014) is shown color-coded on the structural MR image at the top to illustrate the anatomical location of each ROI. The proportion of voxels that significantly correlated with low-level features (blue; one-sided F-test $p < 0.001$) against high-level features (red; one-sided F-test $p < 0.001$) are shown for each ROI. See the Methods section for detail. (**b**). Encoding low and high-level features in the dorsolateral visual stream. The anatomical location of each ROI (Wang et al., 2014) is color-coded on the structural MR image. (**c**). Encoding of DCNN features (hidden layers' activation patterns) in the ventral-temporal stream. The top three principal components (PCs) from each layer of the DCNN were used as features in this analysis. In general, early regions more heavily encode representations found in early layers of the DCNN, while higher-order regions encode representations found in deeper CNN layers. The proportion of voxels that significantly correlated with PCs of convolutional layers 1 to 4 (light blue), convolutional layers 5 to 9 (blue), convolutional layers 10 to 13 (purple), fully connected layers 14-15 (pink) are shown for each ROI. The significance was set at $p < 0.001$ by one-sided F-test. (**d**). Encoding of DCNN features in the dorsolateral visual stream. Credit: Jean Metzinger, Portrait of Albert Gleizes (public domain; RISD Museum).

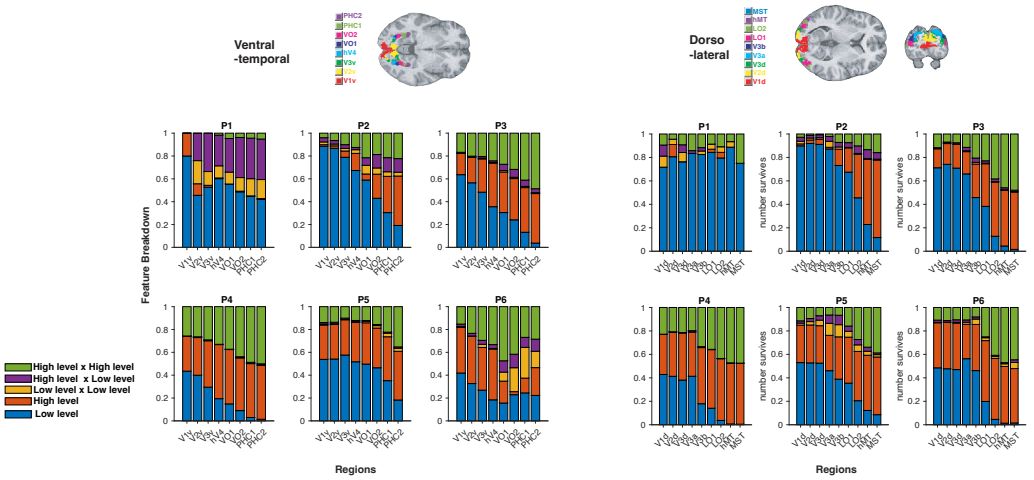


Figure 2.14: Encoding of nonlinear feature representations. We performed encoding analysis of low-level, high-level, and interaction term features (low x low, high x high, low x high), using lasso regression with cross validation within subject. The results of ROIs in the ventral-temporal and dorso-lateral visual streams are shown.

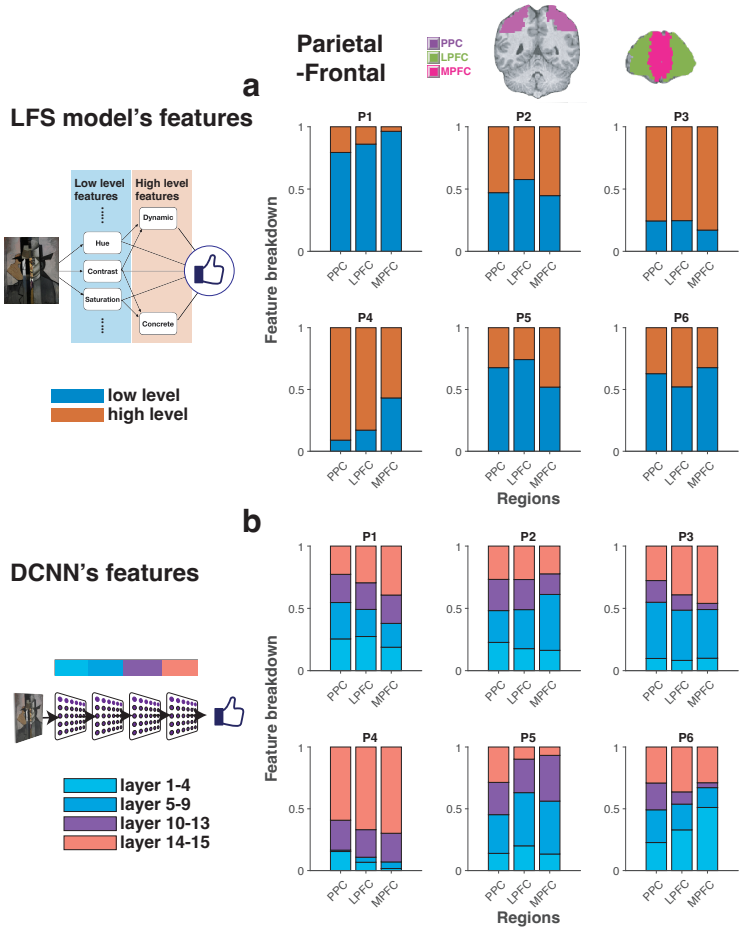


Figure 2.15: Parietal and prefrontal cortex encode features in a mixed manner. (**a**). Encoding of low- and high-level features from the LFS model in posterior parietal cortex (PPC), lateral prefrontal cortex (lPFC) and medial prefrontal cortex (mPFC). The ROIs used in this analysis are indicated by colors shown in a structural MR image at the top. (**b**). Encoding of the DCNN features (activation patterns in the hidden layers) in PPC and PFC. The same analysis method as Figure 2.13 was used. Credit: Jean Metzinger, Portrait of Albert Gleizes (public domain; RISD Museum).

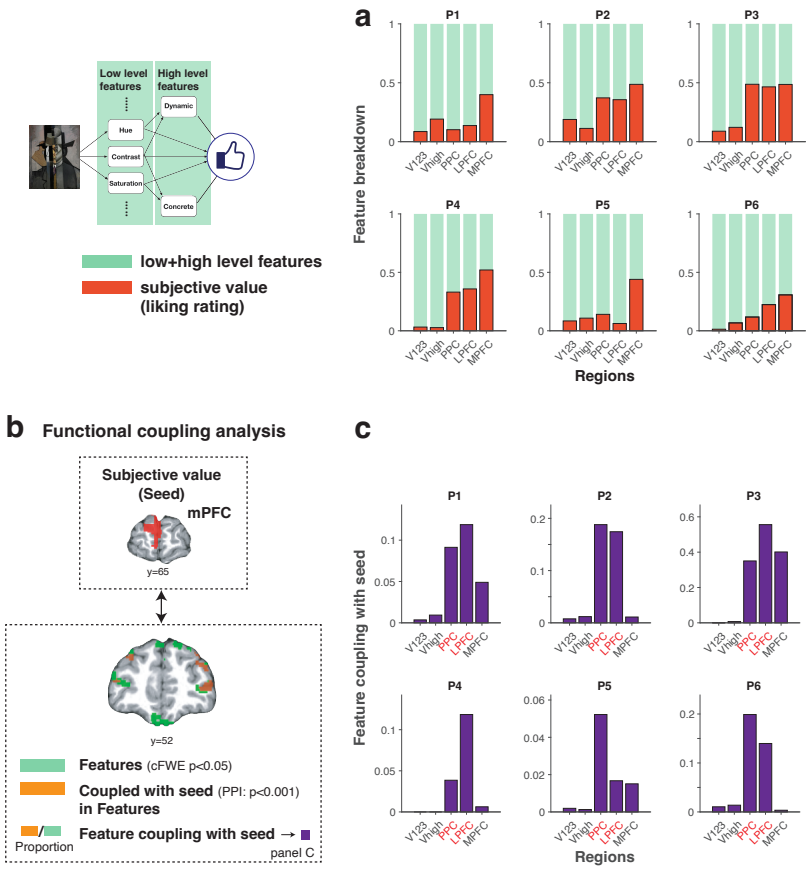


Figure 2.16: Features are integrated from PPC and lateral PFC to medial PFC when constructing the subjective value of visual art. (**a**). Encoding of low- and high-level features (green) and liking ratings (red) across brain regions. Note that the ROIs for the visual areas are now grouped as V1-2-3 (V1, V2 and V3) and V-high (Visual areas higher than V3). See the Methods section for detail. (**b**). The schematics of functional coupling analysis to test how feature representations are coupled with subjective value. We identified regions that encode features (green), by performing a one-sided F-test ($p < 0.05$ whole-brain cFWE with the height threshold $p < 0.001$). We also performed a psychophysiological interaction (PPI) analysis (orange: $p < 0.001$ uncorrected) to determine the regions that are coupled to the seed regions in mPFC that encode subjective value (i.e., liking rating) during stimulus presentation (red: seed, see Figure 2.48). We then tested for the proportion of overlap between voxels identified in these analyses in a given ROI. (**c**). The results of the functional coupling analysis show that features represented in the PPC and lPFC are coupled with the region in mPFC encoding subjective value. This result dramatically contrasts with a control analysis focusing on ITI instead of stimulus presentations (Figure 2.50). Credit: Jean Metzinger, Portrait of Albert Gleizes (public domain; RISD Museum).

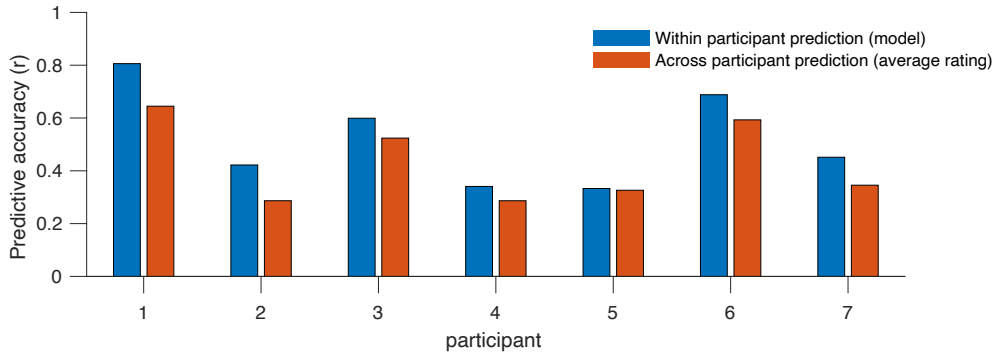


Figure 2.17: The predictive accuracy of subjective ratings of art in the in-lab participants. Blue: within-participant prediction using our computational model. Red: across-participant prediction using the average rating of each stimulus.

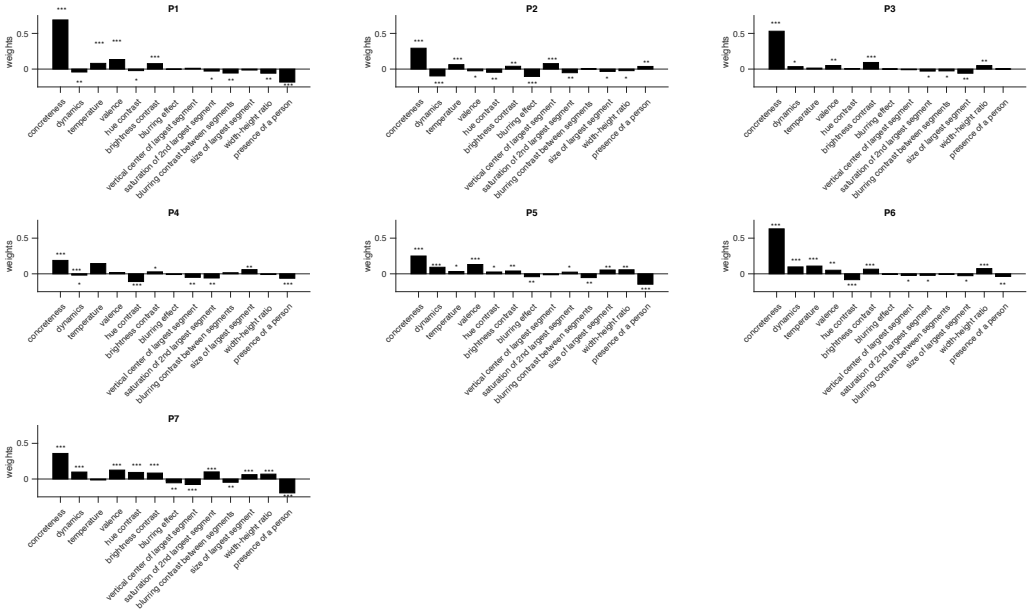


Figure 2.18: The estimated feature weights of in-lab participants. The significance was estimated against the null distribution of weighs constructed by model fittings to permuted data. One star, two starts, three stars, indicates $p < 0.05$, $p < 0.01$, $p < 0.001$, respectively.

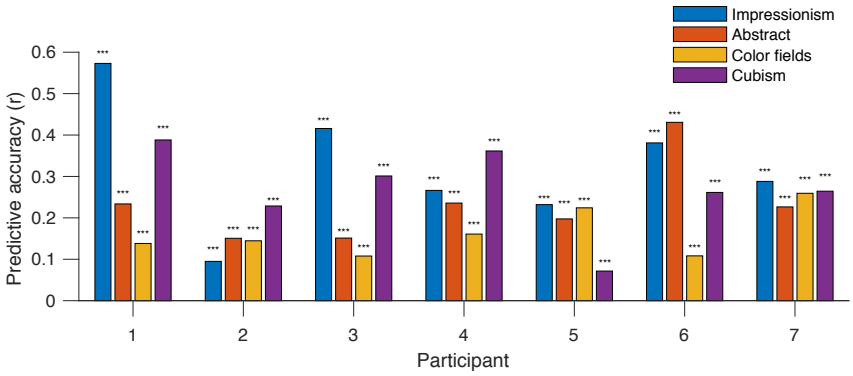


Figure 2.19: Predictive accuracy of the LFS model within different art genres. The model was trained on all images using 20 fold cross validation in each participant. Predictions for images in each art genre were compared with the actual data. The predictive accuracy was measured by Pearson correlation. This figure shows that our overall predictive accuracy is not merely an artifact of the fact that people like different genres differently, i.e., that the LFS model is sensitive only to differences between images as a result of genre and that this alone enables it to have success. Here, even within specific genres, the model can still succeed in predicting liking ratings just as it can across genres. Note that correlation values are smaller than the overall value presented in Figure 1. This is because between-genre correlation is indeed present in Figure 1.

.

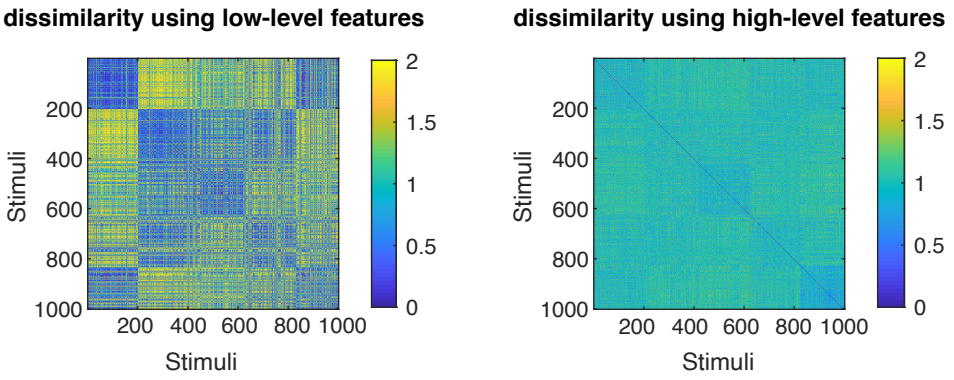


Figure 2.20: Representation dissimilarity matrix using low-level and high-level features. Image index; 1-204: Impressionism. 205-417: Abstract art. 418-621: Color fields. 622-826: Cubism. 827-1000: Pictures from the stimulus set of Vaidya et al. (2017) .

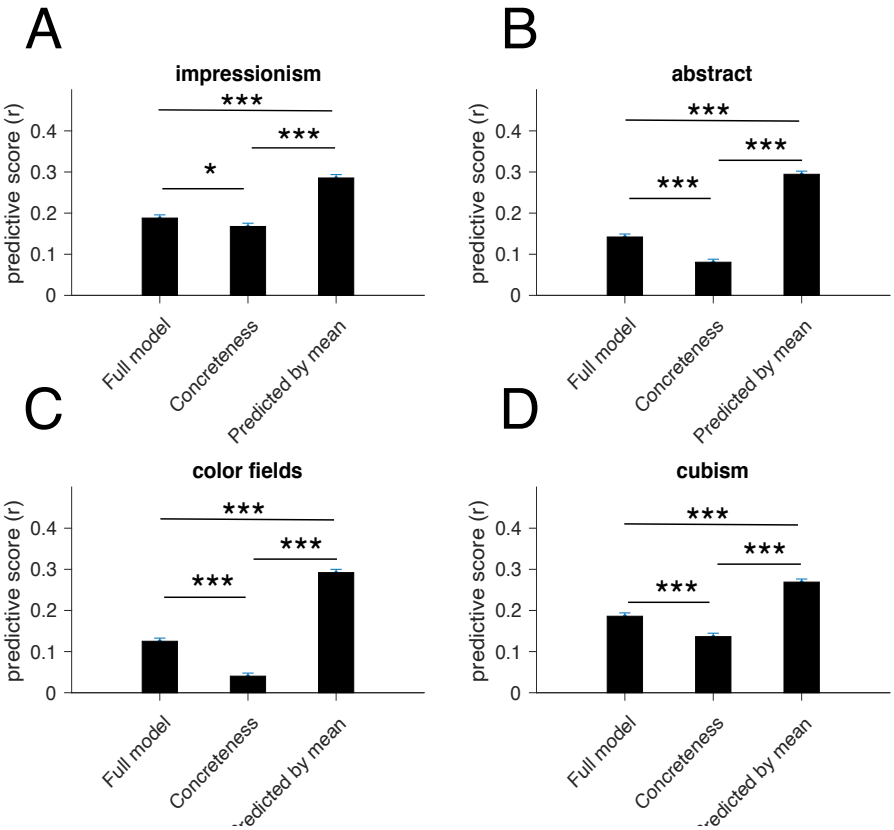


Figure 2.21: The model's predictive accuracy when using full features or the concreteness feature alone, tested in the large scale online dataset. The full model significantly outperforms the model with concreteness feature alone, but shows room to improve when compared against the performance of an average rating model. The error bars indicate the mean and SEM.

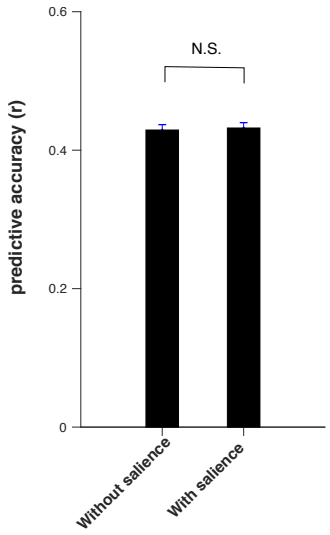


Figure 2.22: The inclusion of salience-weighted features does not improve our model's predictive accuracy in M-Turk participants. The error bars indicate SEM. N.S. indicates not significant in permutation test.

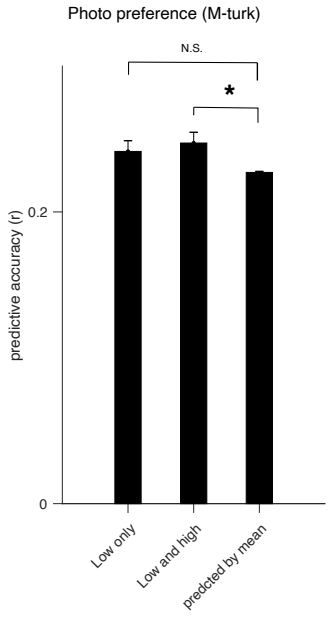


Figure 2.23: The predictive accuracy of model on photograph ratings. Left: the original model with low-level features. Middle: the model with low-level and high-level features, where the binary high level features are approximated by a nonlinear support vector machine trained on visual art set using low-level features. Right: correlations with the average ratings for each image. The one star indicates $p < 0.05$ in permutation test across participants. The error bars indicate SEM.

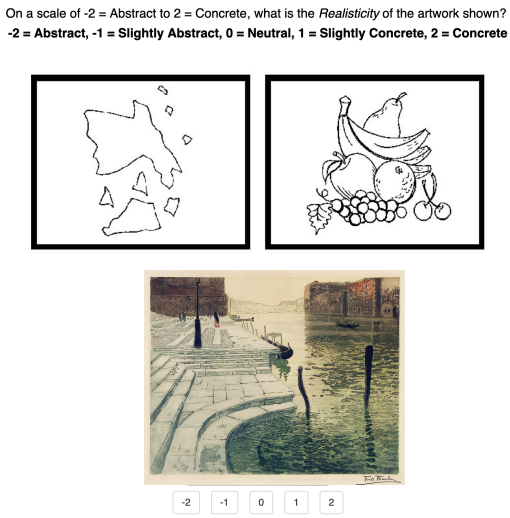


Figure 2.24: An example trial of feature annotation. Annotators were asked to evaluate high-level feature values (from –2 to 2). Frits Thaulow-Marmortrappen credit: ART Collection / Alamy Stock Photo.

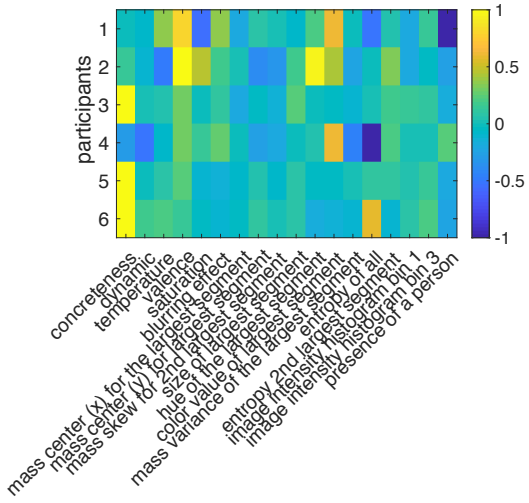


Figure 2.25: Feature weights from each fMRI participant, determined by fitting the LFS model to the liking ratings from each participant.

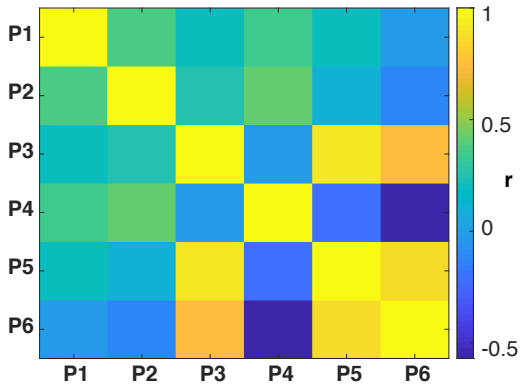


Figure 2.26: Correlations between feature weights across fMRI participants.

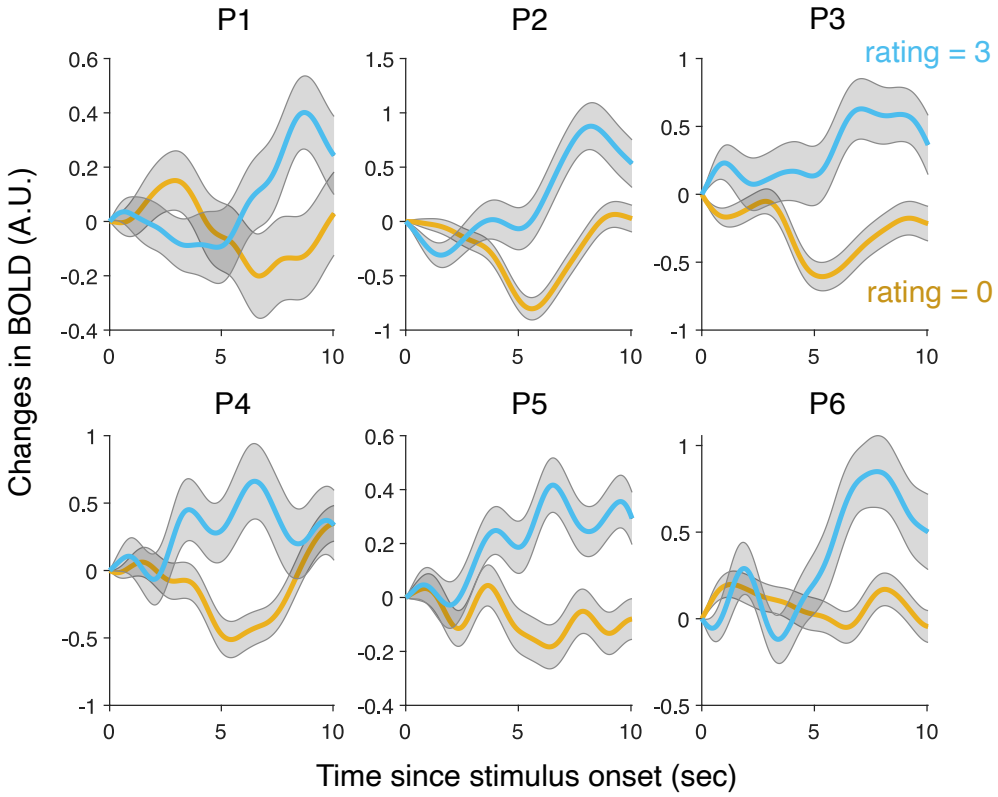


Figure 2.27: The time course of BOLD signals in mPFC. BOLD signals are extracted from the cluster in mPFC correlating with liking ratings from each participant depicted in Figure 2.12 in the main paper. The signals are up-sampled (Iigaya et al., 2020a) and shown for trials in which participants gave the highest (liking rating =3; blue) and the lowest (liking rating =0; orange). The error bars indicate the mean ± SEM over trials.

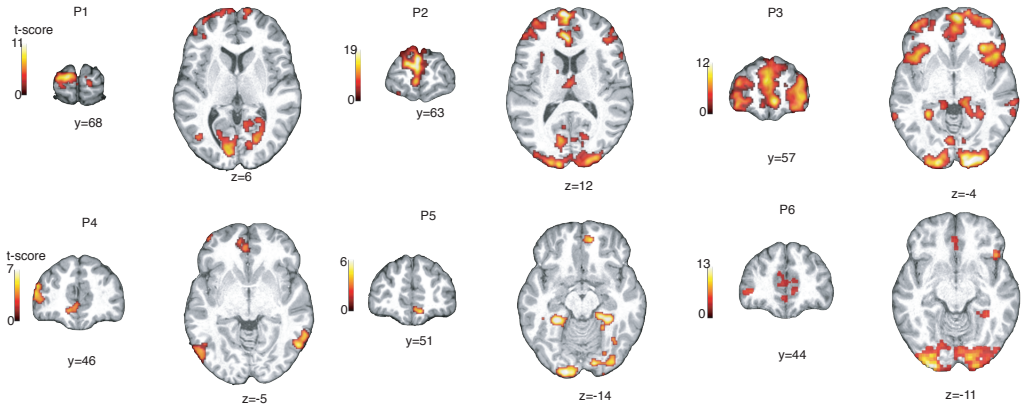


Figure 2.28: Neural correlates of subjective value. One-sided t-test. An adjustment was made for multiple comparison correction: clusters at whole-brain cFWE $p < 0.05$ with height threshold at $p < 0.001$ are shown.

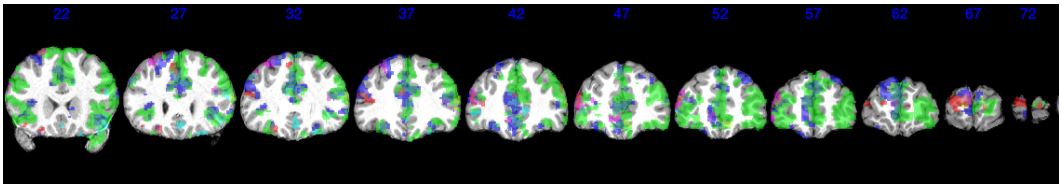


Figure 2.29: Neural correlates of subjective value. One-sided t-test. An adjustment was made for multiple comparisons: clusters at whole-brain cFWE $p < 0.05$ with height threshold at $p < 0.001$ are shown for slices ranging from $y = 22$ and $y = 72$ for participant 1 (red), participant 2 (blue), participant 3 (green), participant 4 (violet), participant 5 (yellow), participant 6 (cyan).

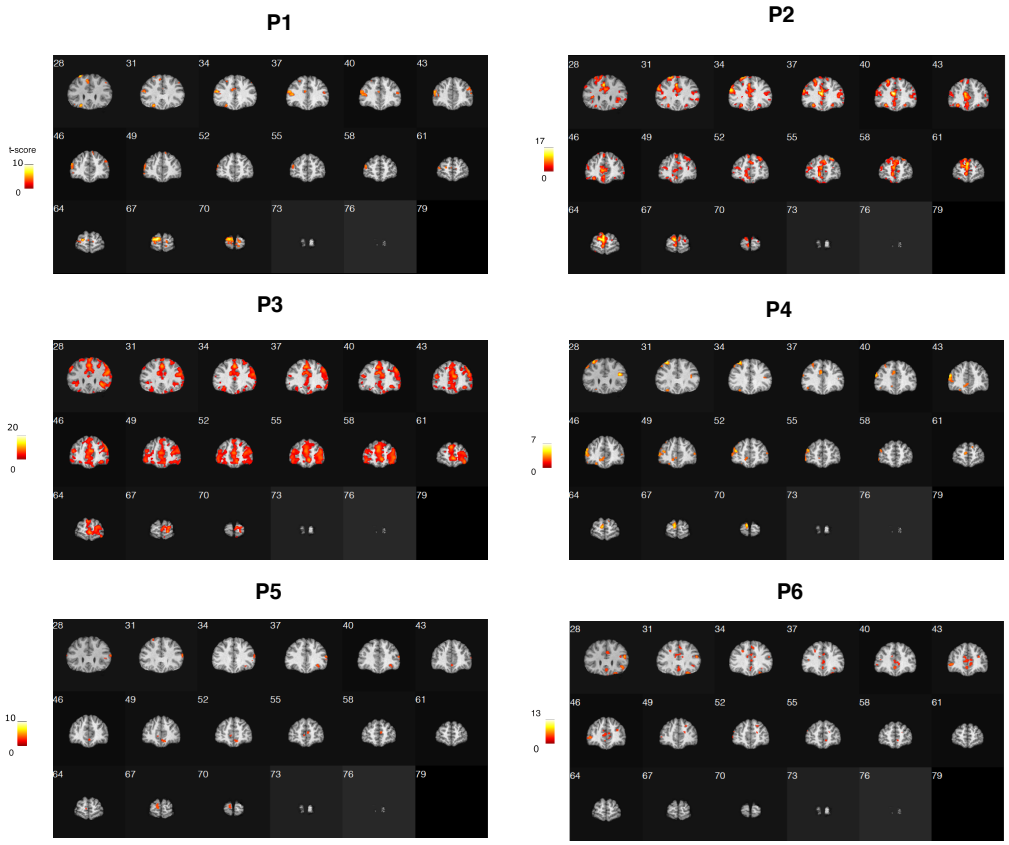


Figure 2.30: Neural correlates of subjective value. Voxels at $p < 0.001$ uncorrected are shown for slices between $y = 28$ and $y = 79$. One-sided t-test.

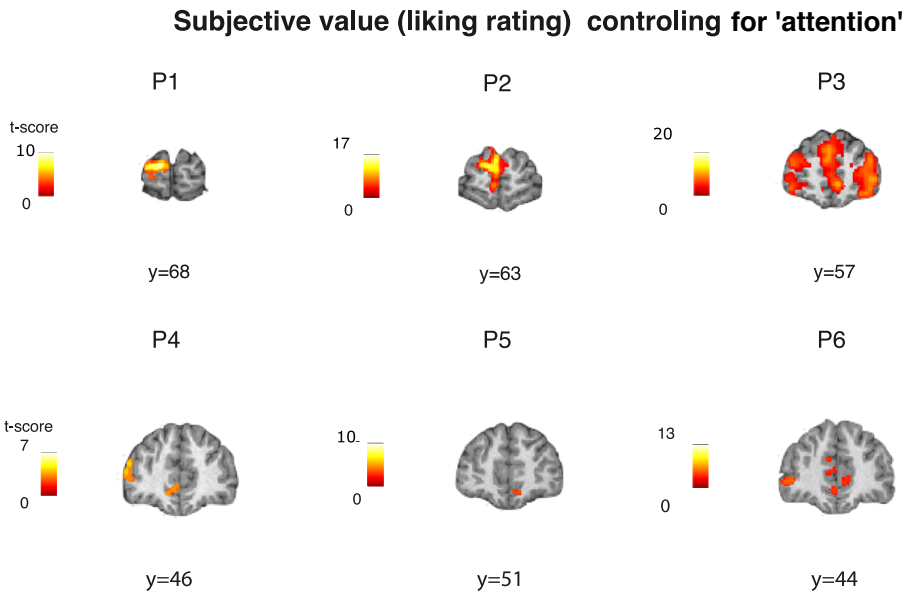


Figure 2.31: Neural correlates of subjective value when controlling for the effects of attention (reaction time, squared reaction time, distance from the mean rating). One-sided t-test. An adjustment was made for multiple comparisons: clusters significant at whole-brain cFWE $p < 0.05$ with height threshold at $p < 0.001$ are shown.

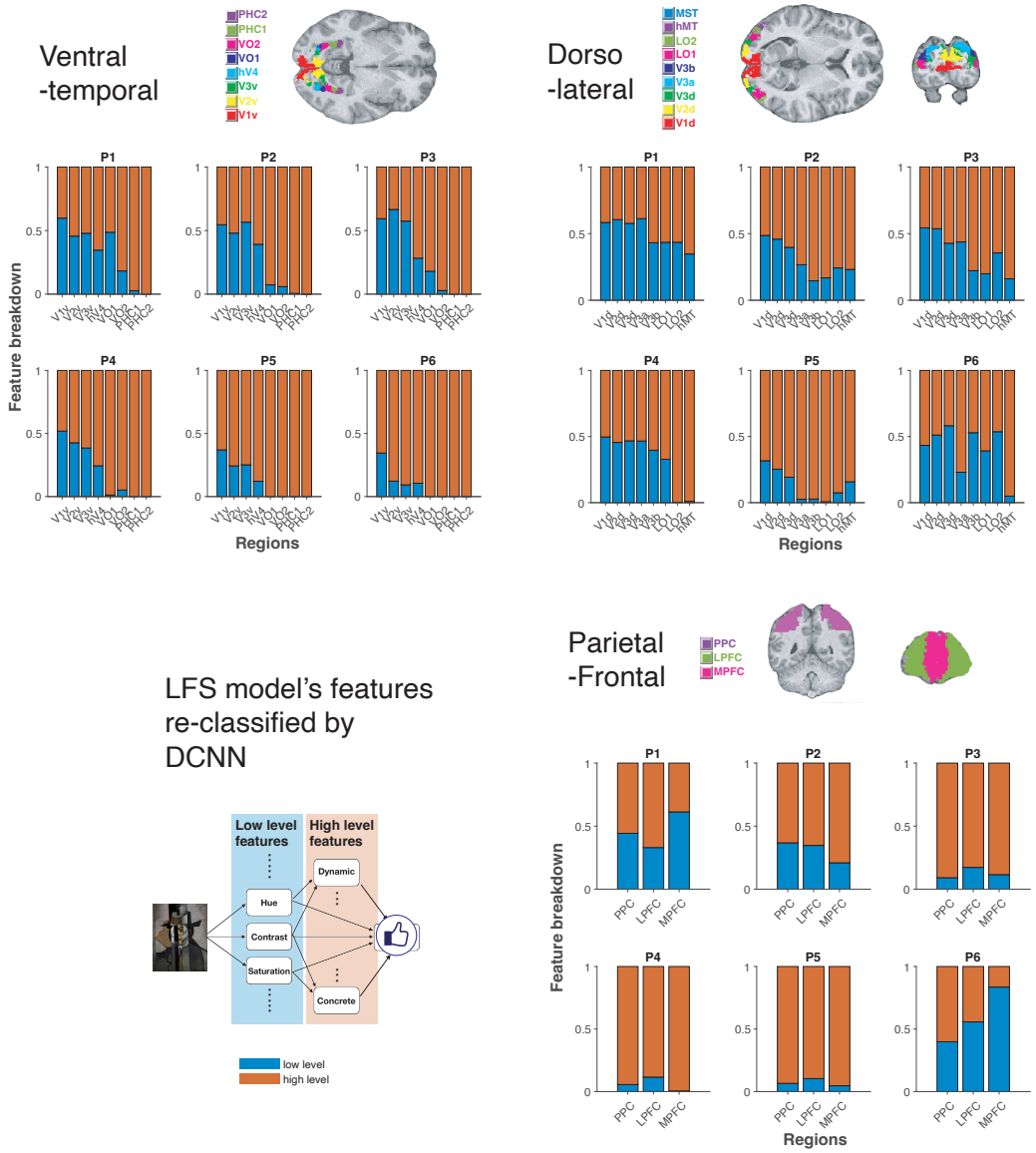


Figure 2.32: Results of fMRI encoding analysis of low- and high-level features, using the features that are reclassified according to the DCNN results. Among the features that were originally considered, the features showing significantly positive slopes across layers in the DCNN were defined as high-level features, while the features showing significantly negative slopes across layers in the DCNN were defined as low-level features. The results did not qualitatively change from our original analysis with the original definition of low- and high- level features. Credit: Jean Metzinger, Portrait of Albert Gleizes (public domain).

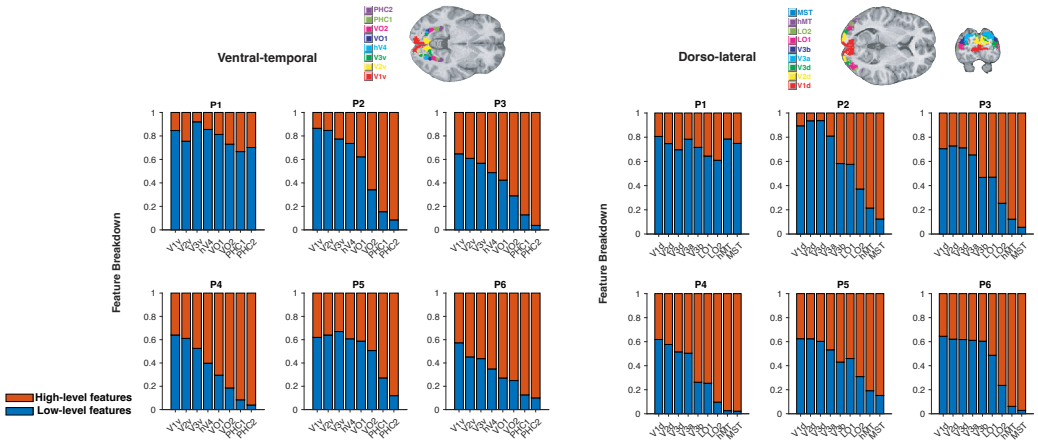


Figure 2.33: Encoding analysis of low-level, high-level, features using lasso regression with cross validation within subject. The results of ROIs in the ventral-temporal visual stream are shown.

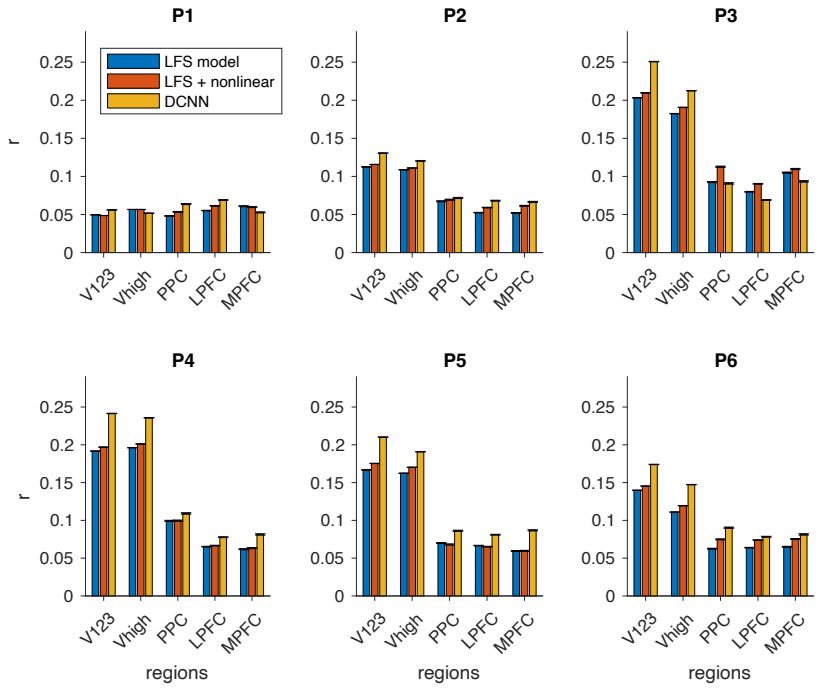


Figure 2.34: Contrasting different model predictive accuracies, measured by Pearson correlations, across ROIs. Blue is the model with original low and high level features. Red is the model with original features and nonlinear features that are constructed by interactions between pairs of original features. Yellow is the DCNN hidden layers (150 PCs in total). The correlation was computed for each voxel for each participant. The colored bars indicate the mean and the error bars indicate standard errors across voxels in each ROI.The number of voxels varied across ROIs.

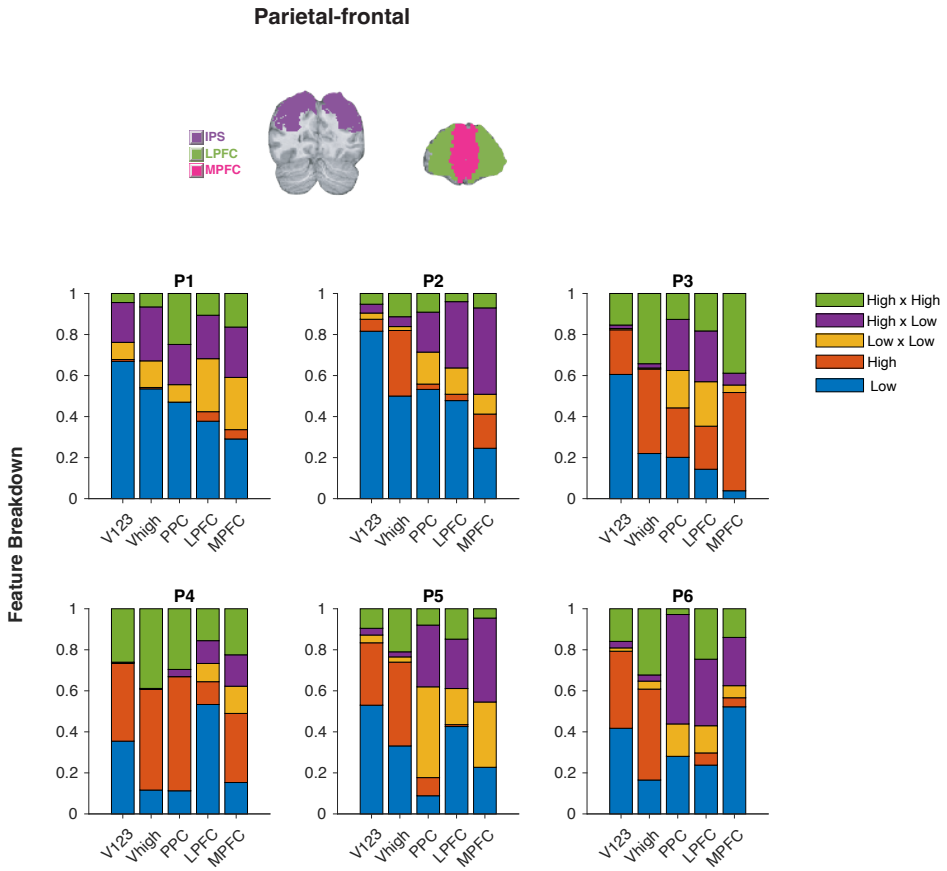


Figure 2.35: Encoding analysis of low-level, high-level, and interaction term features (low x low, high x high, low x high), using lasso regression with cross validation within subject. The results of ROIs in visual areas, PPC, PFC are shown.

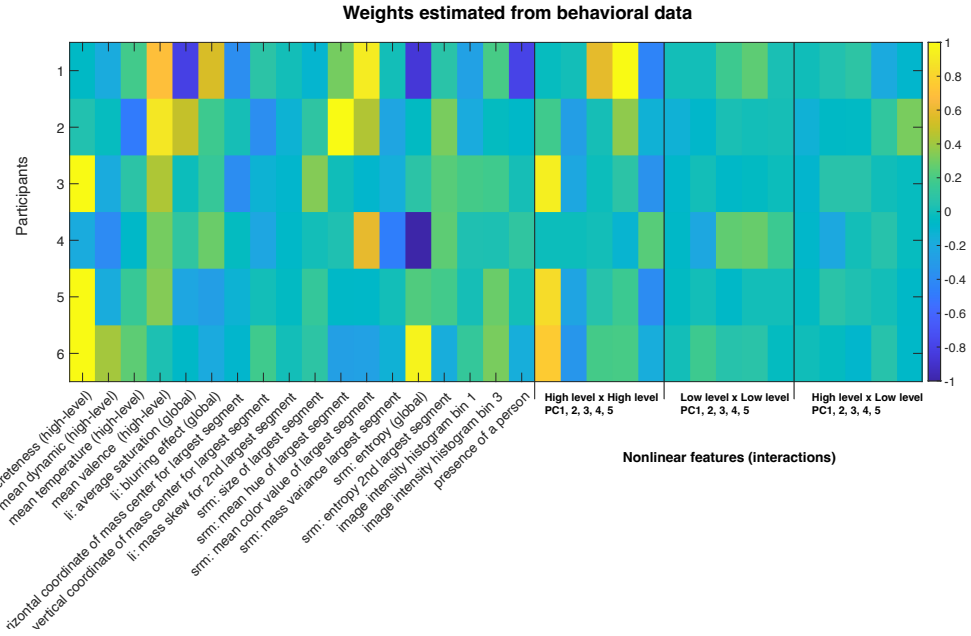


Figure 2.36: The weights estimation of original and nonlinear features across particiapnts.

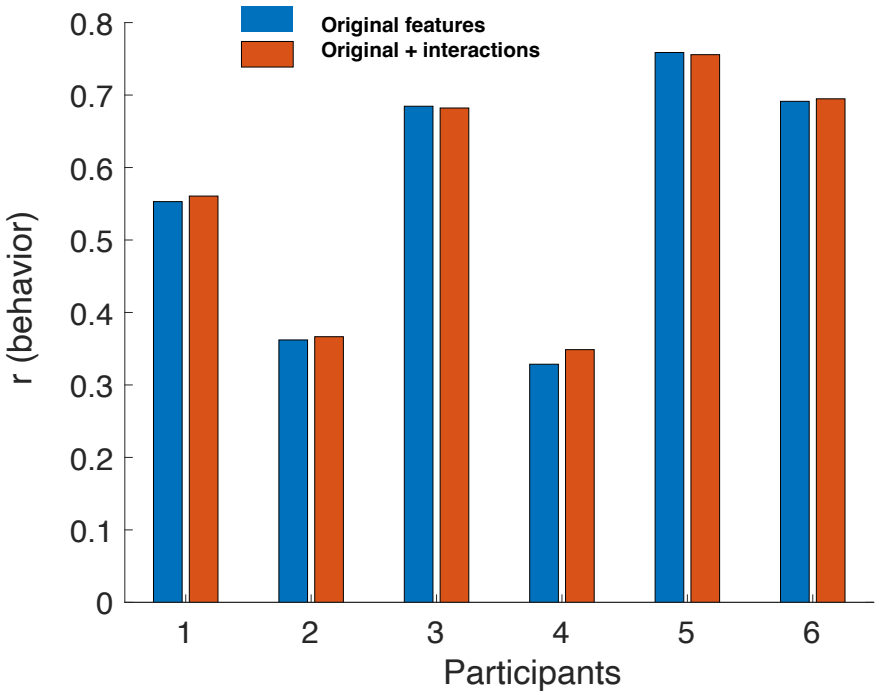


Figure 2.37: Behavioral prediction with original low and high level features (blue), and with these features plus nonlinear interaction features (red).

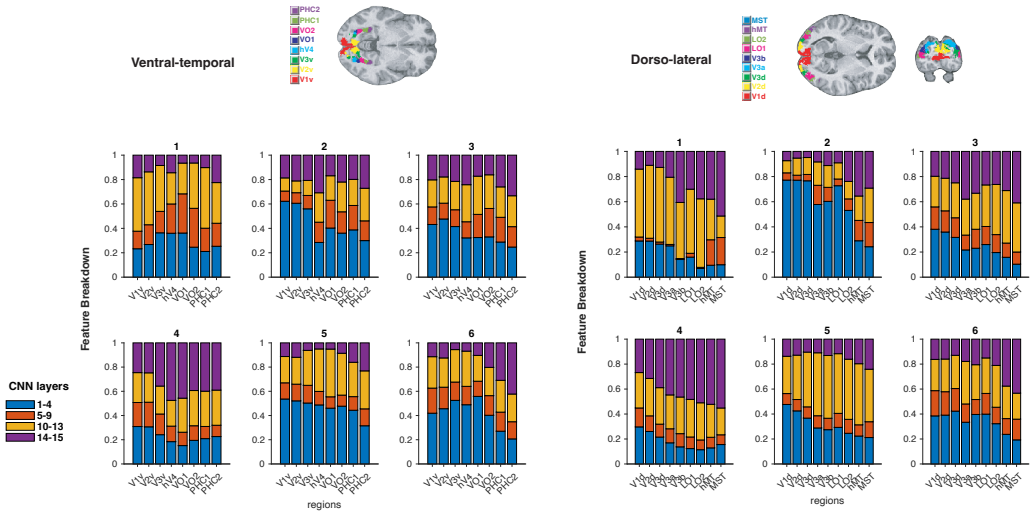


Figure 2.38: Encoding analysis of DCNN features (45 features in total, 3 features per layer) using lasso regression with cross validation within subject. The results of ROIs in the ventral-temporal and dorso-lateral visual stream are shown.

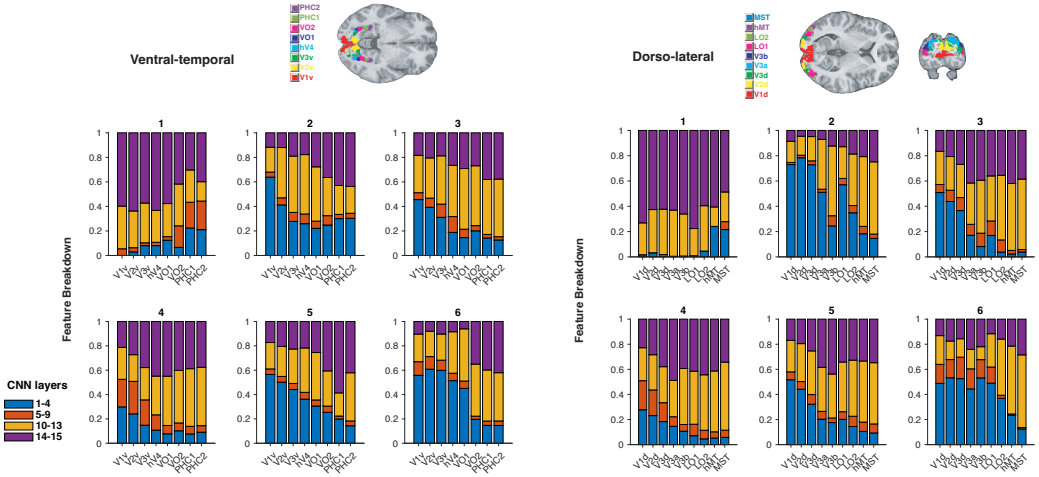


Figure 2.39: Encoding analysis of DCNN features (150 features in total, 10 features per layer) using lasso regression with cross validation within subject. The results of ROIs in the ventral-temporal and dorso-lateral visual stream are shown.

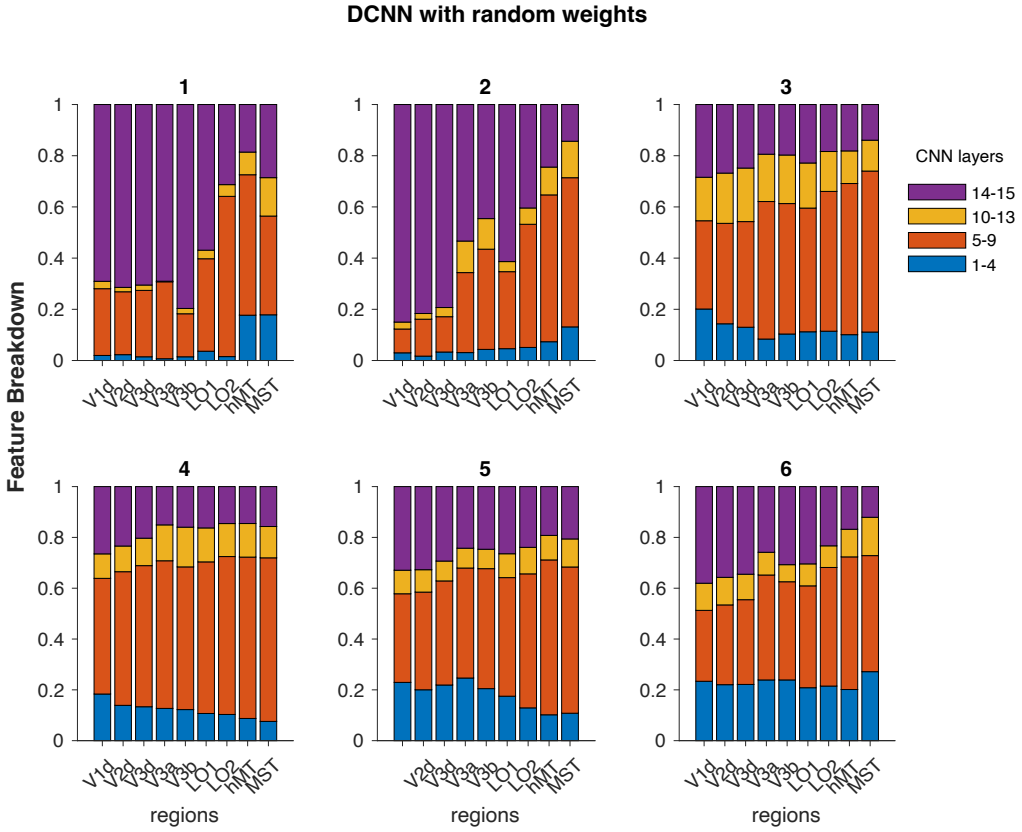


Figure 2.40: Encoding analysis of the DCNN model features, where the model weights were set to random, using lasso regression with cross validation within subject. The results of ROIs in the ventral-temporal visual stream are shown.

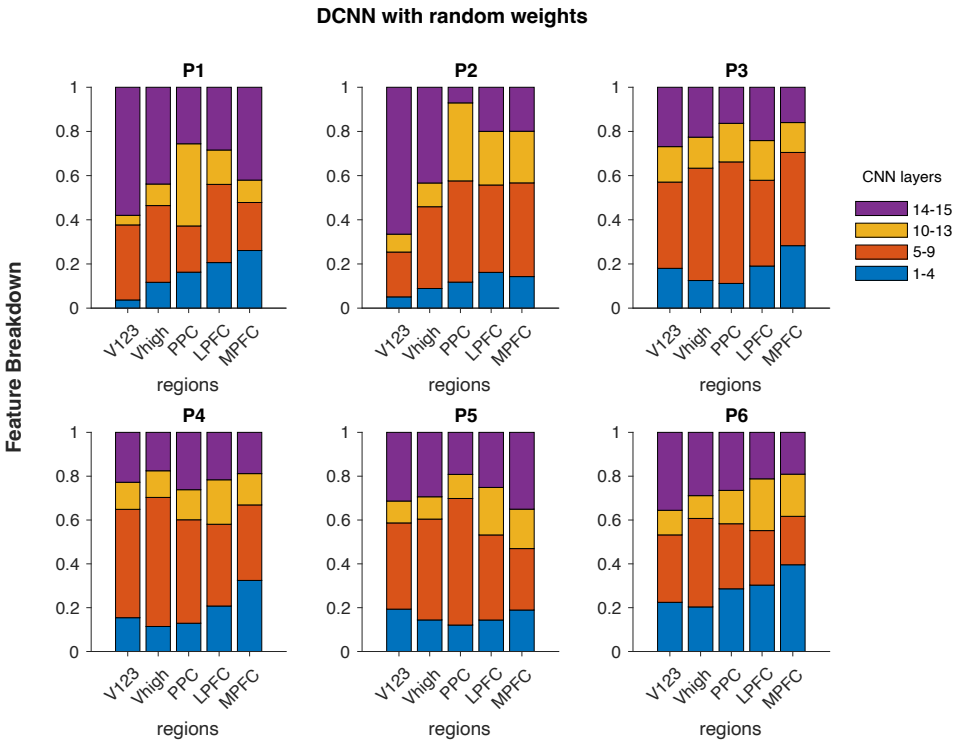


Figure 2.41: Encoding analysis of DCNN model features, where the model weights were set to random, with lasso regression with cross validation within subject. The results of ROIs in visual areas, PPC, PFC are shown.

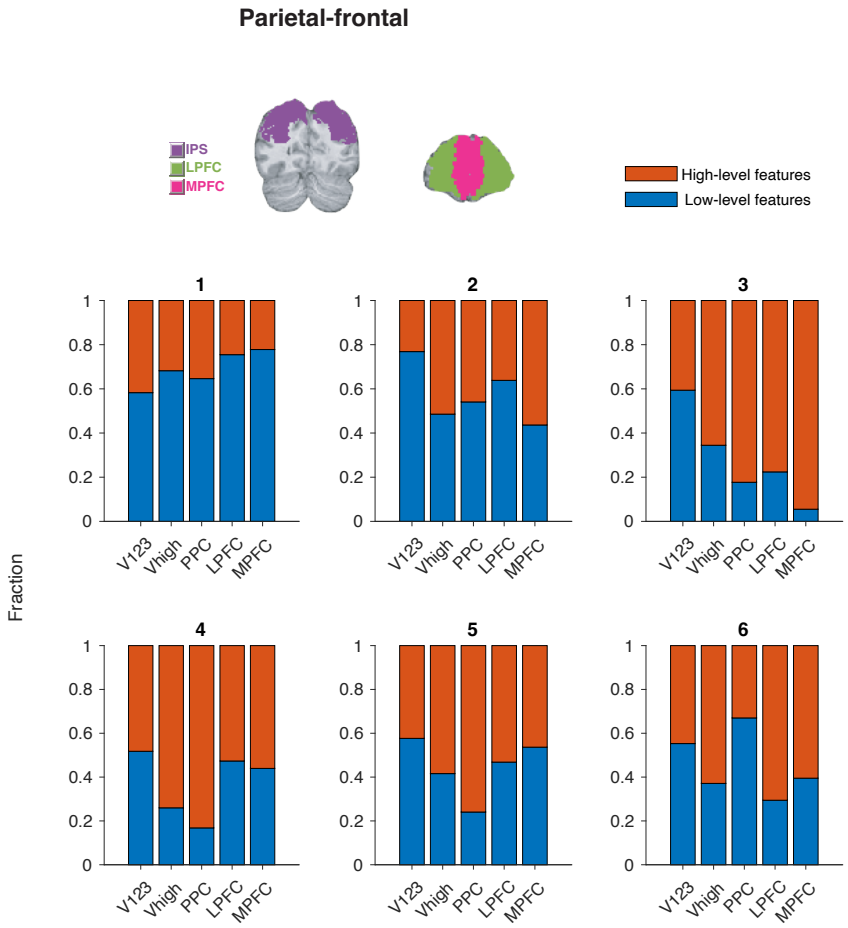


Figure 2.42: Encoding analysis of low-level, high-level features using lasso regression with cross validation within subject. The results of ROIs in visual areas, PPC and PFC are shown.

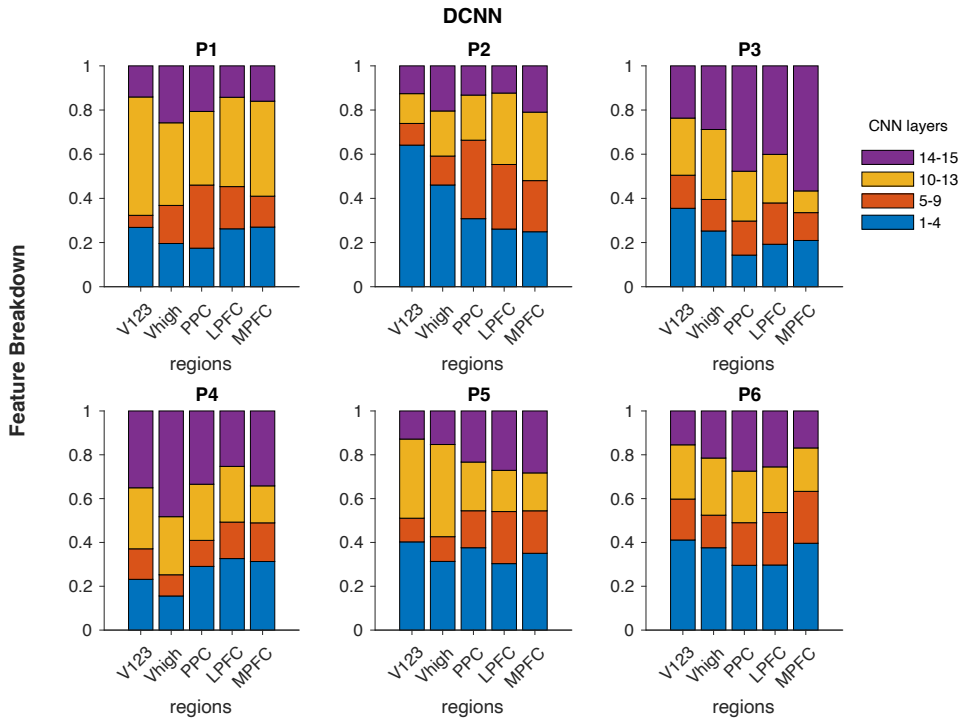


Figure 2.43: Encoding analysis of DCNN features using lasso regression (45 features in total, 3 features per layer) using cross validation within subject. The results of ROIs in visual areas, PPC, PFC are shown.

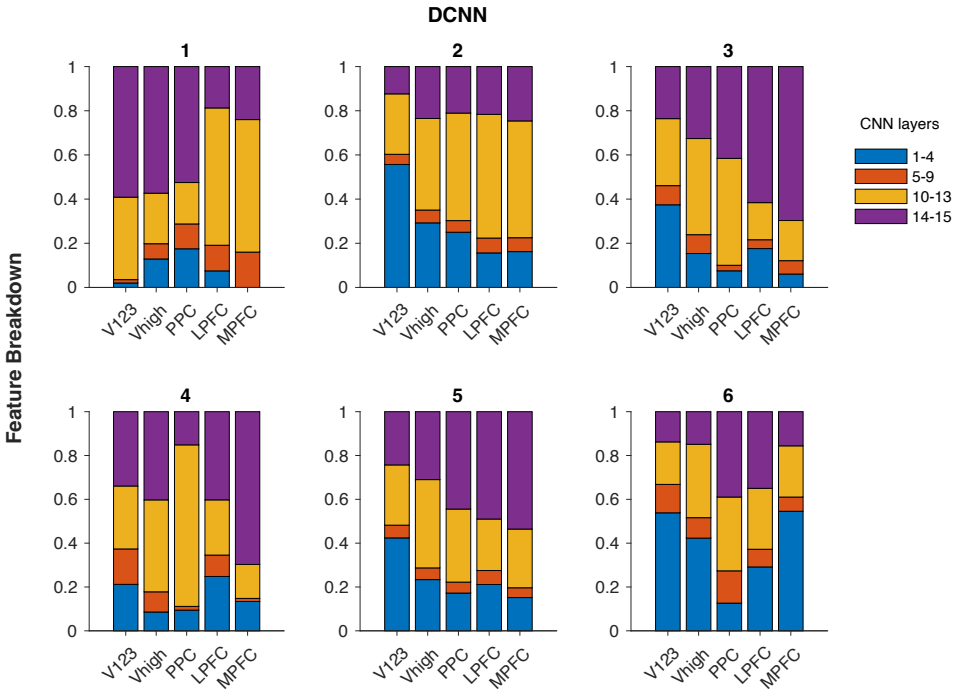


Figure 2.44: Encoding analysis of DCNN features (150 features in total, 10 features per layer) using lasso regression using cross validation within subject. The results of ROIs in visual areas, PPC, PFC are shown.

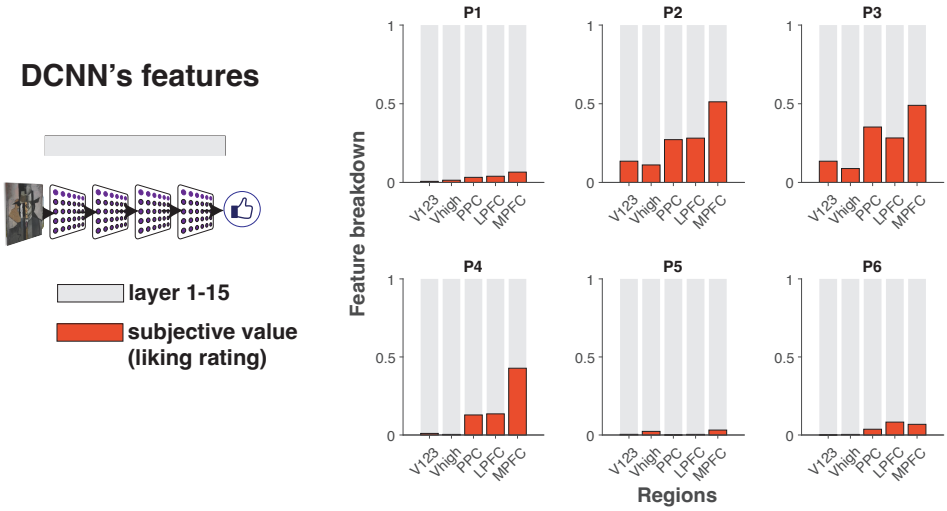


Figure 2.45: The same analysis as Figure 2.16A but now with the DCNN model features. Credit: Jean Metzinger, Portrait of Albert Gleizes (public domain).

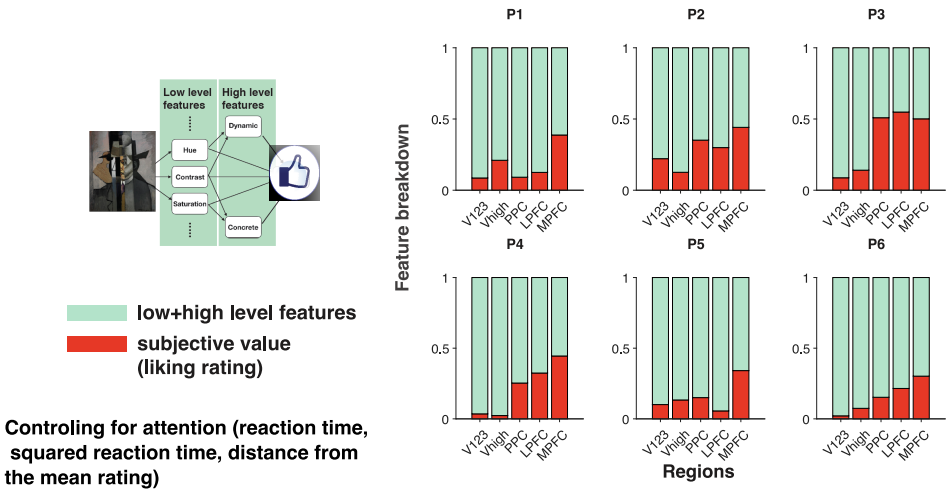


Figure 2.46: The breakdown of feature representations and value representations across cortical regions when controlling for reaction time, squared reaction time, and distance from the mean rating. Credit: Jean Metzinger, Portrait of Albert Gleizes (public domain).

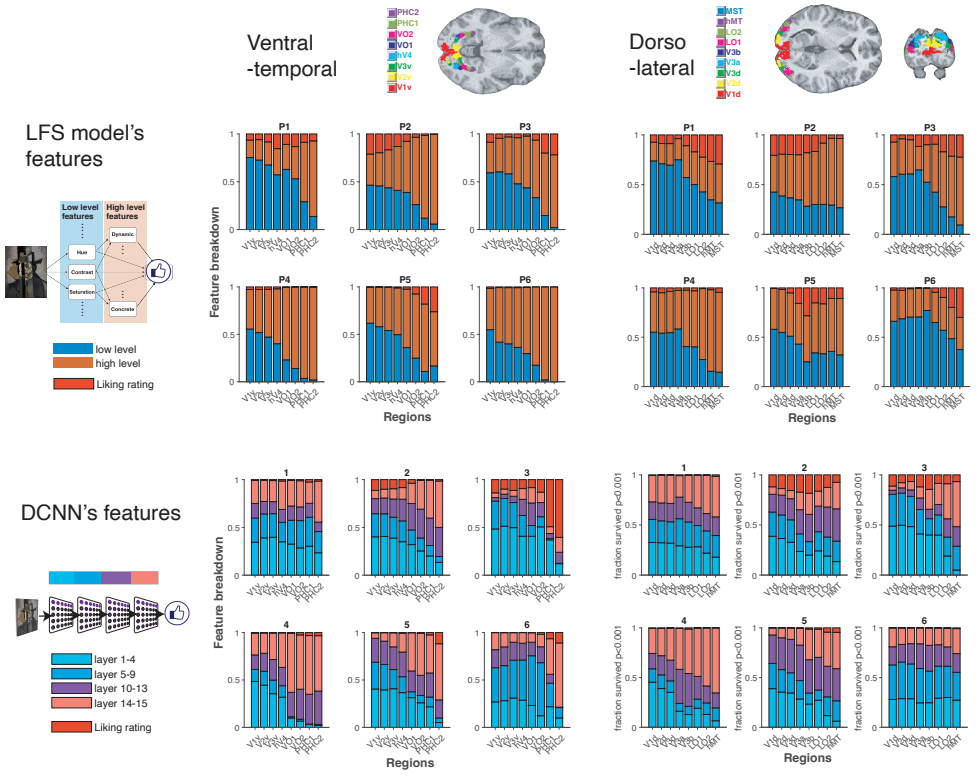


Figure 2.47: Encoding analysis of low- and high-level features when subjective liking ratings are also included into the same GLM. The results of ROIs in the ventral-temporal and dorso-lateral visual streams are shown. Credit: Jean Metzinger, Portrait of Albert Gleizes (public domain).

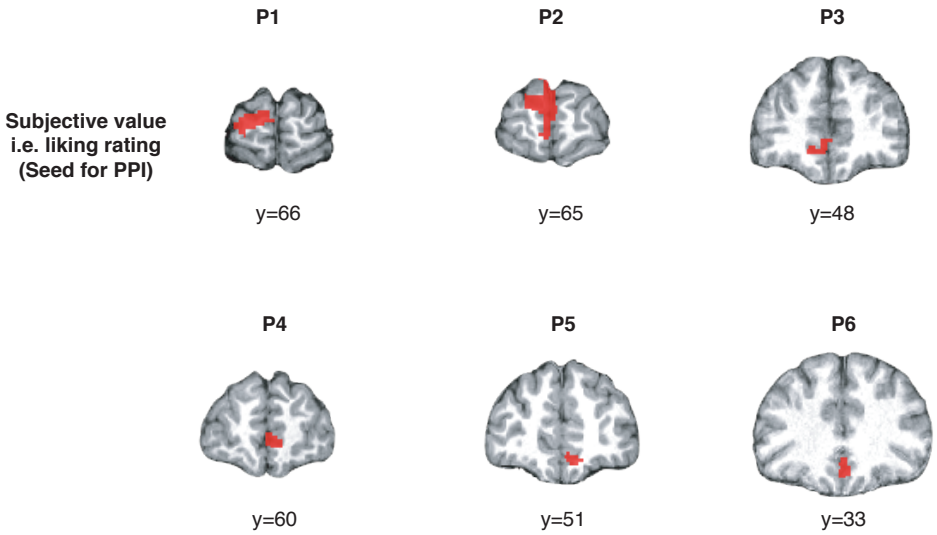


Figure 2.48: The seeds of the PPI analysis. The seeds used for the PPI analysis correspond to a medial PFC cluster drawn from each participant that was found to show significant correlation with subjective value. The cluster used for each participant is shown here (one-sided t-test. An adjustment was made for multiple comparisons: $p < 0.05$ cFWE at whole-brain with a height threshold of $p < 0.001$).

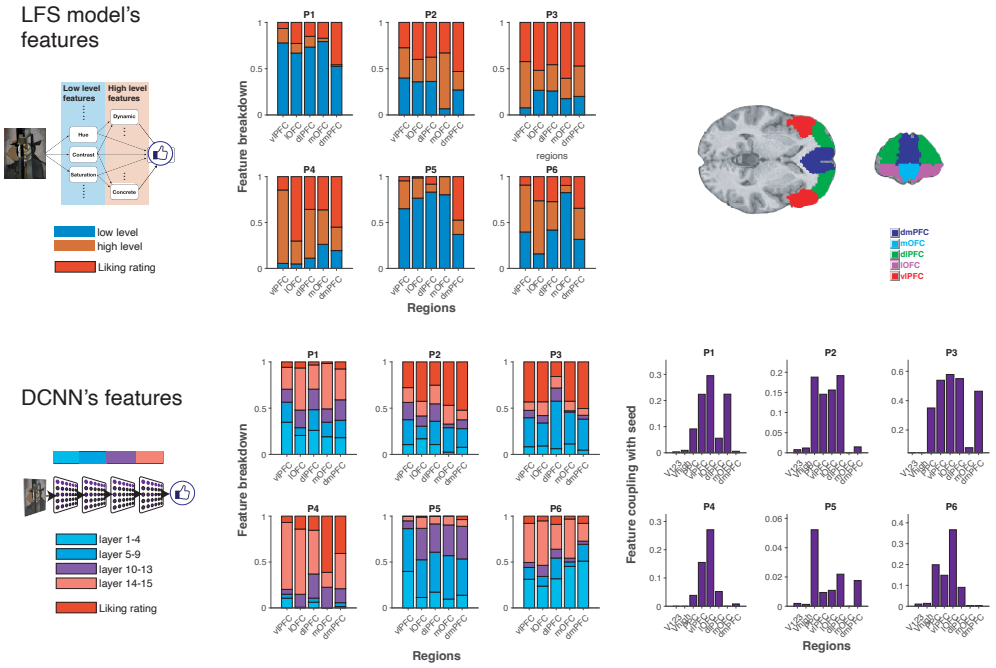


Figure 2.49: The same results as in Figures 2.16 and 2.45, but now broken down to show separate results for each sub-region of the PFC. Credit: Jean Metzinger, Portrait of Albert Gleizes (public domain).

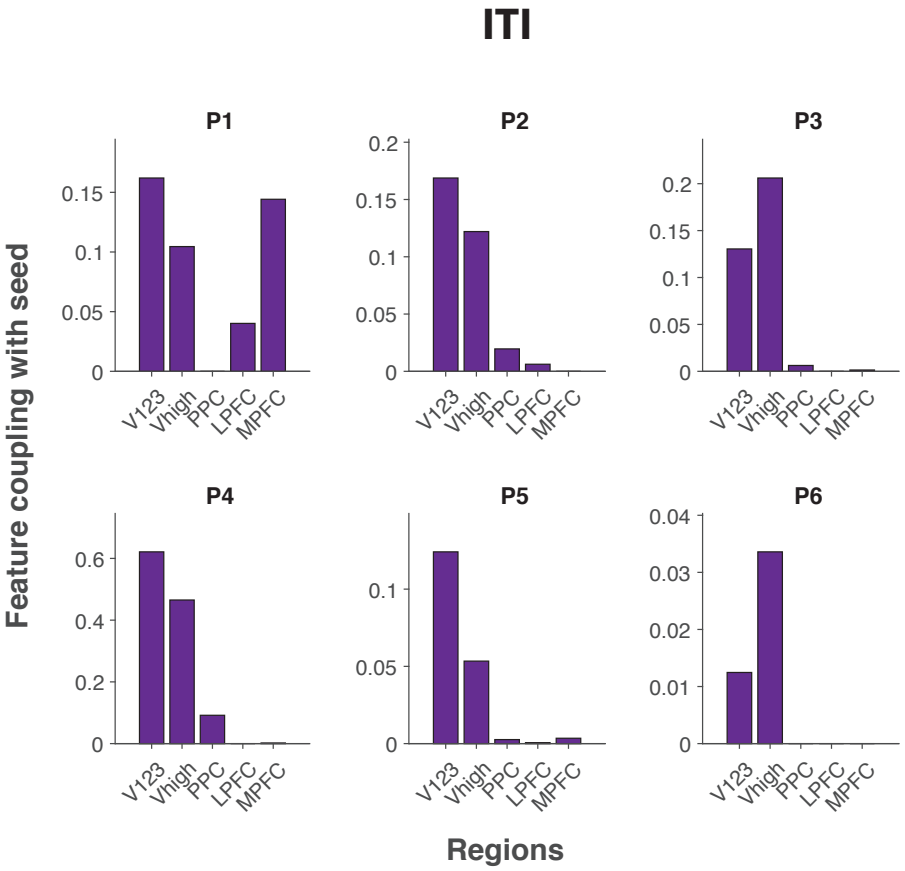


Figure 2.50: The same analysis as in Figure 2.16C, except here the epoch of the ITIs are taken as the psychological regressor, as opposed to the epoch of presentation of the visual stimuli. In this situation, we did not observe robust coupling between mPFC value areas and lateral PFC and PPC, thereby supporting the possibility that increased coupling between lPFC, PPC and mPFC occurs specifically at the time of stimulus evaluation.

Chapter 3

HUMAN TRANSFER LEARNING AND FEATURE LEARNING ACROSS DIFFERENT ENVIRONMENTS

3.1 Abstract

Human cognitive flexibility, particularly the ability to transfer knowledge across different environments, is crucial for continual life-long learning and decision-making. This study investigates the computational mechanisms underlying feature-based transfer learning in humans. We examined how individuals apply previously acquired knowledge to new environments with shared features by designing a feature-based multiarmed bandit task, collecting behavioral data, and employing various computational models. Our results demonstrate that participants successfully transferred knowledge across tasks, and reinforcement learning (RL) models incorporating glia-like slow integration components explain the behavior characteristics in transfer learning the best. These components, modeled after the physiological properties of astrocyte-glial cells, enabled the retention of learned information over time, thus supporting knowledge transfer. While recurrent neural network (RNN) models showed strong predictive accuracy in participant choices, they failed to capture the temporal dynamics observed in human transfer learning. This research highlights the complexity of human transfer learning and suggests that incorporating slow integration mechanisms is essential to achieve the cognitive flexibility necessary for effective transfer learning.

3.2 Introduction

Transfer learning is a critical aspect of human cognitive capacity, allowing individuals to apply knowledge acquired in previously experienced environments to novel contexts without forgetting it (Zhuang et al., 2021b; Flesch et al., 2023). In cognitive science, the mechanisms supporting this ability to transfer knowledge across different tasks have been studied from various perspectives. One line of research focuses on declarative memory, which is characterized by its flexibility and ability to be applied in new contexts (Keresztes et al., 2018; Reber et al., 1996; Poldrack et al., 2001; Squire and Zola, 1996). Another is the concept of cognitive maps, which suggests that spatial information is stored within the brain and can be leveraged to solve new problems (Mark et al., 2020; Tolman, 1948). Recently, shared value

representation across tasks that can handle different context information has been suggested to support flexible knowledge transfer (Tomov et al., 2021). However, these concepts do not fully explain scenarios where different environments share features, and recognizing these features can be advantageous for achieving goals in novel environments.

For instance, an individual's previous experience with red-colored, spicy cuisine from their home country might guide their food choices when confronted with unfamiliar options in an exotic location. This phenomenon, referred to as feature-based transfer learning (Zhuang et al., 2021b), illustrates how feature-based value computation can facilitate decision-making in novel environments (Farashahi et al., 2017; Niv et al., 2015; Leong et al., 2017; Suzuki et al., 2017b; Iigaya et al., 2021, 2023). Nevertheless, the exact computational mechanisms underlying how feature learning occurs, how this information is maintained, and in what ways previously acquired knowledge influences decision-making in new tasks remain unclear.

One of the hypotheses in understanding human transfer learning is that the brain operates with computational components that operate on different time scales. This temporal differentiation allows slower processes to retain information longer without forgetting, enabling effective transfer learning (Parisi et al., 2019). Hierarchical reinforcement learning models have been proposed to embody such properties, retaining context-specific sub-policies or skills that operate on shorter time scales and are dynamically employed in solving new environments when the skills are applicable (Tessler et al., 2017; Botvinick, 2012; Eckstein and Collins, 2020).

Another promising avenue for understanding the biological underpinnings of transfer learning lies in astrocyte glial cells. These cells exhibit properties that facilitate long-term information tracking and serve as computational resources alongside neurons in the brain by controlling the efficacy of neuronal synapses and facilitating learning (Kofuji and Araque, 2021; Mu et al., 2019; Perea et al., 2009a). Integrating properties of glial cells into artificial neural networks has also been shown to enhance performance in classification tasks and perform computations analogous to transformers, the backbone of current advances in large language models (Porto-Pazos et al., 2011; Kozachkov et al., 2023; Vaswani, 2017).

Based on these insights, our aim is to unveil the computational mechanisms behind transfer learning when the transferred information comprises values of features in environments. To address these research questions, we designed a novel feature-based multiarmed bandit task (or "contextual bandit") consisting of multiple two-

armed bandit problems and collected three independent human choice datasets from both online and in-person experiments. Using a computational model-driven analysis, we explored the possible mechanisms that facilitate human transfer learning across environments that share visual features. First, we tested various reinforcement learning (RL) models, with and without a slow integration component designed to mimic the physiological properties of glial cells. Furthermore, we explored the efficacy of recurrent neural network (RNN) models with gated memory units to explain human transfer learning, as these models have proven to be powerful tools in modeling behavioral patterns (Dezfouli et al., 2019b; Ji-An et al., 2023b; Miller et al., 2024).

From quantitative and qualitative model comparisons, we suggest that reinforcement learning models with separate components that gradually track the learned action values can explain human transfer learning. However, the complexity of this cognitive process necessitates further investigation to unravel the full extent of the computational and biological substrates involved. This research represents a step towards understanding how the brain maintains and utilizes feature-based information, contributing to the broader quest to replicate human-like learning in artificial systems.

3.3 Task design

In the behavioral task, participants engaged in multiple two-armed bandit tasks sequentially, which were framed as a Casino tour. Each "casino" had two slot machines, one of which had a higher reward probability. Participants interact with the slot machines for an average of 16 trials in the online tasks or 8 trials in the onsite experiment. The participants were instructed to maximize the collected reward. Consequently, participants had to interact with the slot machines, using trial and error to identify the more rewarding one (see Fig. 3.1).

A key characteristic of the task is the presence of shared visual cues across casinos, each with a different level of contingency to the winning slot machines. (see Fig. 3.2). The contingencies between each cue and the winning slot machine were varied to create blocks with different levels of information for identifying the better slot machine. This contingency was manipulated by varying the frequency with which each cue appeared on winning slot machines. For example, a cue named *H1* was associated with the winning slot machine in 18 out of 24 casinos in the online tasks (and in 24 out of 36 casinos in the onsite task).

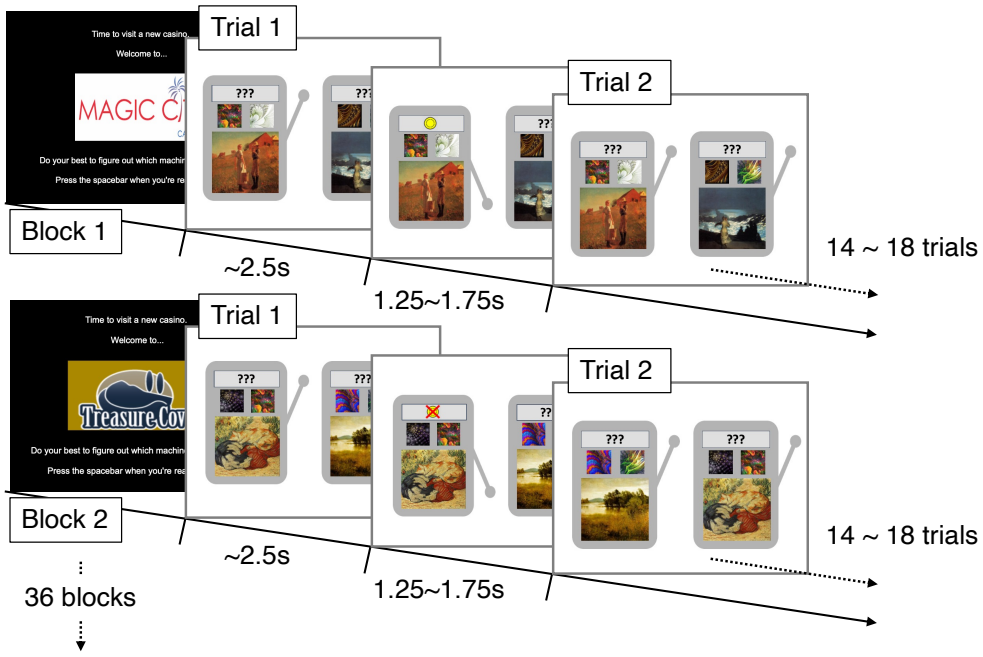
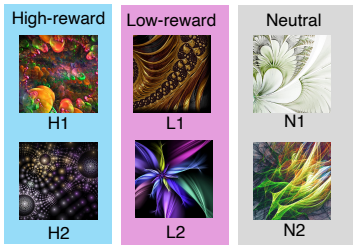


Figure 3.1: The task consists of multiple blocks or casinos that participants visited. In each casino, participants played with two slot machines and determined the more rewarding one through trial and error. The two cues displayed on top of the slot machines served as hints to help participants find the better slot machine faster, while the paintings at the bottom indicated that participants were visiting a novel, unique environment. The onsite tasks consist of 72 blocks, with trial lengths ranging from 7 to 9 trials, instead of 36 blocks with 14 to 18 trials.

Based on these contingencies, the 6 different cues were categorized into into 3 groups: high-reward, low-reward and neutral (see Fig. 3.3). In online tasks, for example, casinos where two high-reward cues were on one slot machine and two low-reward cues on the other were labeled as strongly-informative blocks. This is because, among the 12 casinos with these cue pairs, the slot machines with two high-reward cues were the better choice in 10 casinos. Similarly, other cues were associated with the better slot machine at different frequencies. Participants could learn these contingencies since high-reward cues were more frequently associated with the better slot machine compared to the other cue groups and low-reward cues were more frequently associated with the worse slot machine (See Figs. 3.2 and 3.3 for details on the task design and cue contingencies)

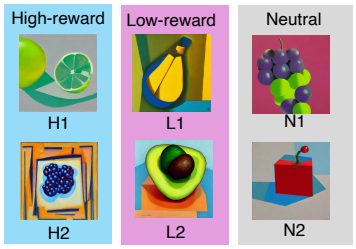
It is important to note that the onsite task comprise 72 blocks instead of 36, with each block consisting of an average of 8 trials, while the online tasks had an average

Online task



Block type	Slot1's cue	Slot2's cue	Number of blocks where slot1 is MORE rewarding	Number of blocks where slot1 is LESS rewarding
Strong informative	H1H2	L1L2	10	2
Weak informative	H1N1	L1N2	8	4
Noninformative	H2L2	N1N2	6	6

Onsite task



Block type	Slot1's cue	Slot2's cue	Number of blocks where slot1 is MORE rewarding	Number of blocks where slot1 is LESS rewarding
Strong informative	H1N1	N2L2	10	2
	H2N2	N1L1	10	2
Weak informative	H2N2	H1L2	8	4
	H1N1	H2L1	8	4
Noninformative	H1L2	H2L1	6	6
	N1L1	N2L2	6	6

Figure 3.2: Based on the contingency between the more rewarding slot machine and the cues, the task featured three different block types. The task was designed so that participants could identify the better slot machine more quickly in the informative blocks. In the onsite task, artistic fruit images were used as cues to reduce the cognitive demand of memorizing their meanings.

of 16 trials per block. The onsite task design was chosen to ensure that participants experienced not only the 3 different cue combination pairs (H1H2 vs L1L2, H1N1 vs L1N2, and H2L2 vs N1N2) but also more diverse cue combination pairs. This approach prevented the displayed cue combination pairs from directly indicating the block type, encouraging participants to learn the values of the cues rather then relaying on a 1-to-1 mapping between cue combination pairs and block types.

Using behavioral tasks, two sets of online data and one onsite dataset were collected. Amzaon Mechanical Turk (MTurk) and Prolific were utilized to collect the online data set (for more details, please refer to the Methods section)

3.4 Results

The behavioral experiment tested three hypotheses regarding transfer learning. The first hypothesis posits that participants' choices are influenced by knowledge gained from previous casinos, allowing them to more quickly identify the better slot machine, as evidenced by a bias in exploration. The second hypothesis suggests that this exploration bias and the improved performance from knowledge transfer will

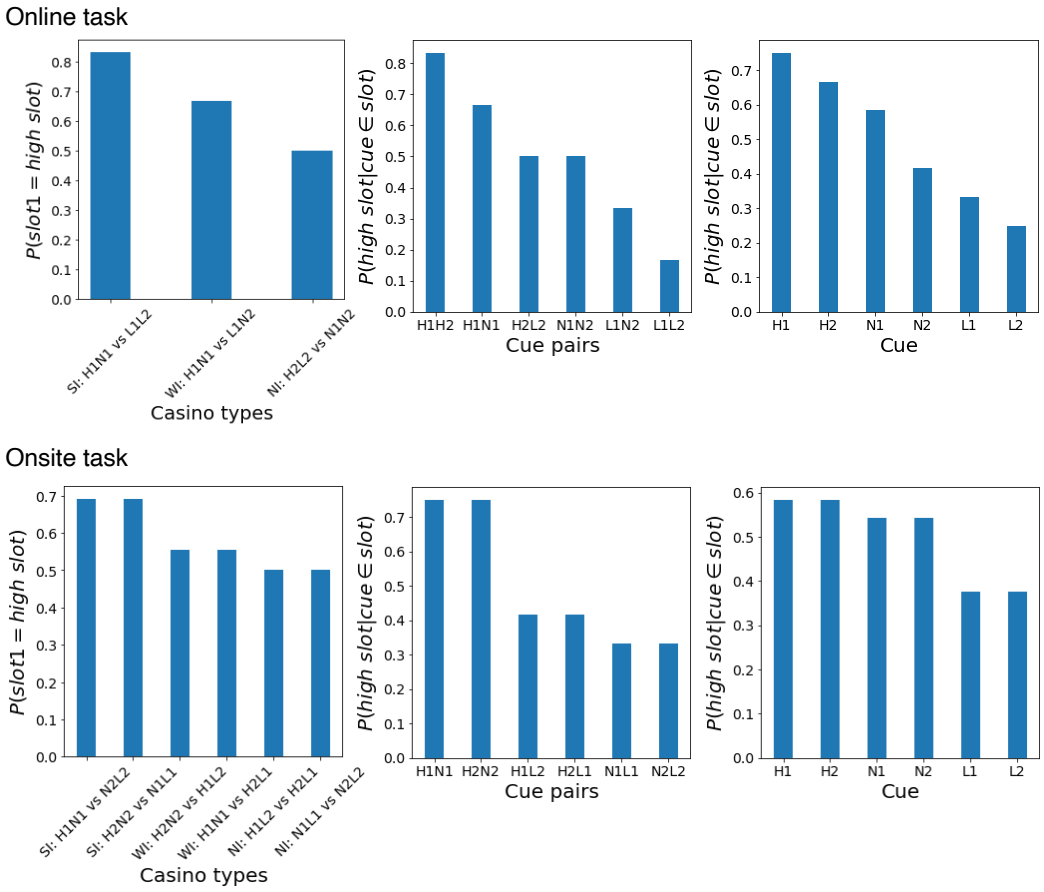


Figure 3.3: Contingency between cues and the most rewarding slot machine. SI: strongly informative, WI: Weakly informative, NI: Non-informative.

become more pronounced in the later phase of the task due to learning effect. The last hypothesis is that there will be a correlation between exploration bias and task performance.

To test the first hypothesis, we analyzed the initial choices made in informative blocks and the learning curves by block type. The initial choices in informative blocks depend entirely on knowledge transfer from previous experiences in other blocks, which serves as a behavioral indicator of transfer learning. As shown in the second column of Fig. 3.4, the initial choices in informative blocks (exploration bias) were significantly biased towards slot machines with high cues (t-tests t(88)=8.73, t(77)=6.12, t(29)=2.84, $p = 1.5 \times 10^{-}13, 3.6 \times 10^{-}8$ and 0.008 each for Mturk, Prolific, and Onsite data, respectively). This initial exploration bias enabled participants to identify the better slot machine more quickly in informative

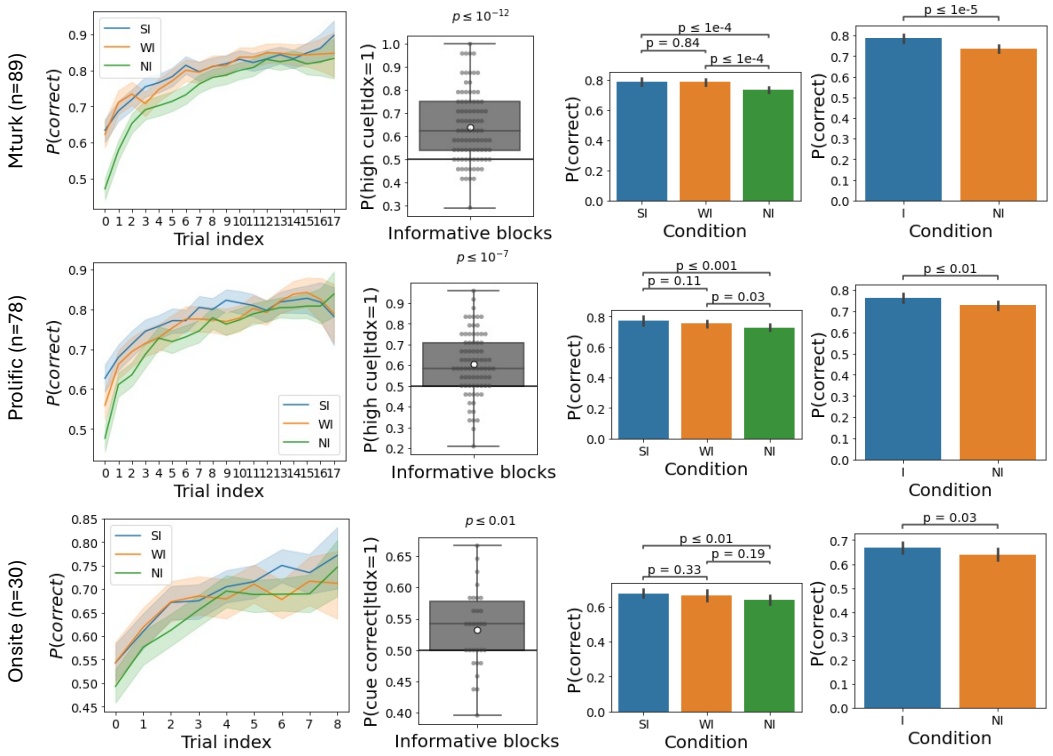


Figure 3.4: The cue information is transferred and facilitates learning. In all three datasets, The learning curve started higher from the initial trials in informative blocks compared to noninformative blocks. The probability of choosing the slot machine with high cues were significantly higher than the chance level 50% (t-tests t(88)=8.73, t(77)=6.12, t(29)=2.84, $p = 1.5 * 10^{-13}, 3.6 * 10^{-8}, 0.008$ each for Mturk, Prolific, and Onsite data, respectively). Moreover, participants' performance in choosing the more rewarding slot machine was higher in the informative blocks but the difference between strongly informative and weakly informative blocks were marginal. (t-test paired, SI vs NI: t(88)=4.53, t(77)=3.46, t(29)=2.85, $p = 1.8 * 10^{-5}, 8.8 * 10^{-4}, 0.008$ each for Mturk, Prolific, and Onsite data, respectively; WI vs NI: t(88)=4.47, t(77)=2.23, t(29)=1.36, $p = 2.3 * 10^{-5}, 0.028, 0.185$ each for Mturk, Prolific, and Onsite data, respectively; SI vs WI: t(88)=0.20, t(77)=1.60, t(29)=1.00, $p = 0.841, 0.114, 0.325$ each for Mturk, Prolific, and Onsite data , respectively; I vs NI: t(88)=5.10, t(77)=3.39, t(29)=2.34, $p = 1.9 * 10^{-6}, 3.6 * 10^{-13}, 0.027$ each for Mturk, Prolific, and Onsite data, respectively)

blocks compared to non-informative blocks. This pattern is also evident in the learning curve plots, where choices in non-informative blocks and choices in informative blocks, where slot machines with high cues were indeed the better choice (congruent blocks), are displayed.

The effect of transfer learning in the exploration phase also contributed to overall

task performance. As observed in the third and fourth columns of Fig. 4, the frequency of choosing the more rewarding slot machine was higher in strongly and weakly informative blocks compared to non-informative blocks (Paired t-tests, SI vs NI: t(88)=4.53, t(77)=3.46, t(29)=2.85, $p = 1.8 * 10^{-}5, 8.8 * 10^{-}4, 0.008$ for MTurk, Prolific, and Onsite data, respectively; WI vs NI: t(88)=4.47, t(77)=2.23, t(29)=1.36, $p = 2.3 * 10^{-}5, 0.028, 0.185$ for MTurk, Prolific, and Onsite data, respectively; I vs NI: t(88)=5.10, t(77)=3.39, t(29)=2.34, $p = 1.9 * 10^{-}6, 3.6 * 10^{-}13, 0.027$ for Mturk, Prolific, and Onsite data, respectively). Interestingly, there was no significant performance difference between the strongly and weakly informative blocks in all three datasets (Paired t-tests, SI vs WI: t(88)=0.20, t(77)=1.60, t(29)=1.00, p=0.841, 0.114, 0.325 for MTurk, Prolific, and Onsite data, respectively).

The second hypothesis was tested by analyzing the exploration bias and performance over the course of the task. As shown in Fig. 3.5, the exploration bias increased in later blocks. Specifically, the exploration bias in the last block was significantly larger than in the first block, especially in the MTurk dataset (Paired t-tests, t(88)=3.15, t(77)=1.16, t(29)=1.11, p=0.002, 0.246, and 0.271 for MTurk, Prolific, and Onsite data, respectively). When focusing on trials in strongly informative blocks, the Prolific dataset also showed a significantly larger exploration bias in the last block compared to the first block (Paired t-tests, t(88)=2.10, t(77)=2.04, t(29)=1.08, p=0.039, 0.045, and 0.283 each for MTurk, Prolific, and Onsite data, respectively).

The performance pattern showed a similar trend. For example, performance in the first strongly informative block was lower than in the last strongly informative block (Paired t-tests, t(88)=1.28, t(77)=2.85, t(29)=2.88, p=0.204, 0.006, 0.007 for MTurk, Prolific, and Onsite data, respectively, see Fig. 3.6). When pooling choices in strongly and weakly informative blocks together, a performance increase from the first to the last block was still observable (Paired t-tests, t(88)=0.18, t(77)=1.70, t(29)=1.77, p=0.854, 0.093, 0.087 for MTurk, Prolific, and onsite data, respectively, see Fig. 6). An ANOVA analysis of the performance increase showed similar results, indicating that being in informative blocks affects task performance increase (ANOVA on first vs last block & SI vs NI: F(1)=1.15, 7.39, 3.50, p=0.284, 0.007, 0.064 for Mturk, Prolific, and onsite data respectively; ANOVA on first vs last block & I vs NI: F(1)=0.13, 4.71, 1.24, p=0.716, 0.031, 0.268 for Mturk, Prolific, and Onsite data, respectively). From these analyses, it can be concluded that participants' choices are influenced by knowledge gained from previous casinos,

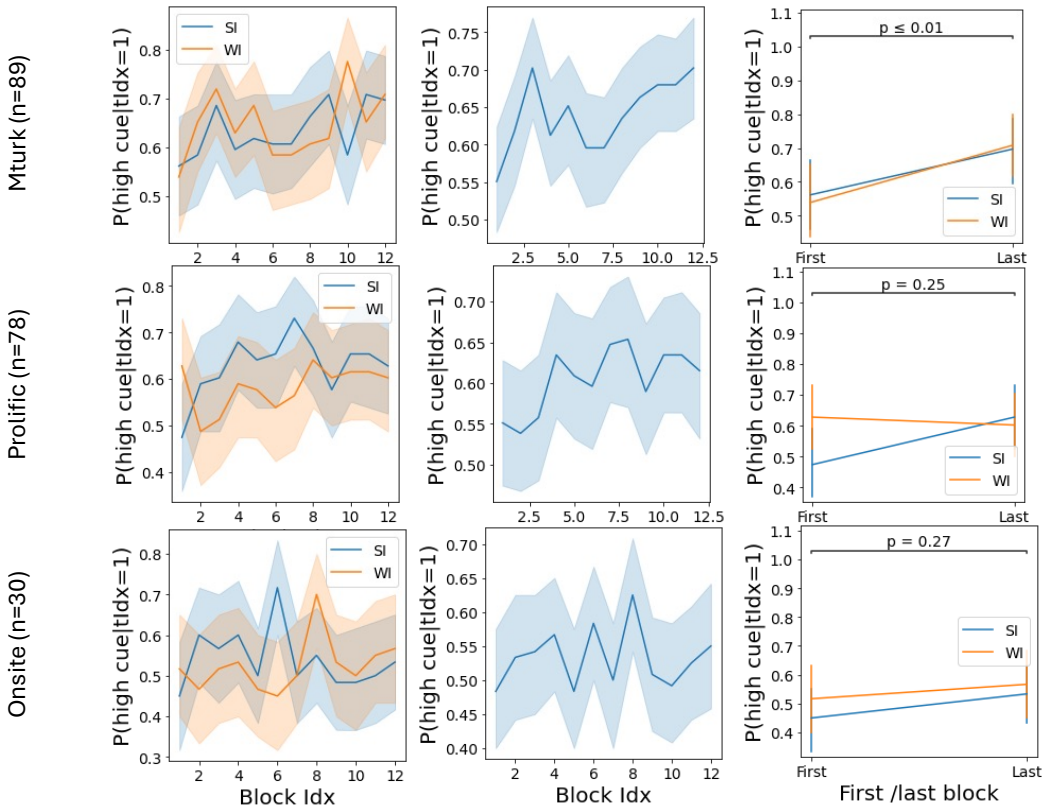


Figure 3.5: Exploration bias getting stronger in later blocks. The plots showing the exploration bias, or the frequency of choosing the slot machine with better cues at the initial trials, by the block index and the block type. The average exploration bias of informative blocks across participants was stronger in the very last block compared to the first block but only significant in the Mturk data (t-test paired, t(88)=3.15, t(77)=1.16, t(29)=1.11, $p = 0.002, 0.246, 0.271$ for Mturk, Prolific, and Onsite data, respectively) The exploration bias in the strongly informative blocks were also stronger in the very last block but was not significant in the onsite data (t-test paired, t(88)=2.10, t(77)=2.04, t(29)=1.08, $p = 0.039, 0.045, 0.283$ for Mturk, Prolific, and Onsite data, respectively).

with choice characteristics from knowledge transfer becoming more prominent in the later phases of the task due to learning effects.

Additionally, weak correlations between exploration bias and task performance was observed in all three datasets (Pearson correlation, $r = 0.18, 0.35, 0.15$, $p = 0.100, 0.002, 0.437$ for Mturk, Prolific, and Onsite data, respectively). Although only the Prolific data showed statistical significance, the general tendency for the two metrics to be correlated was evident in the other two datasets as well (see Fig. 3.7)

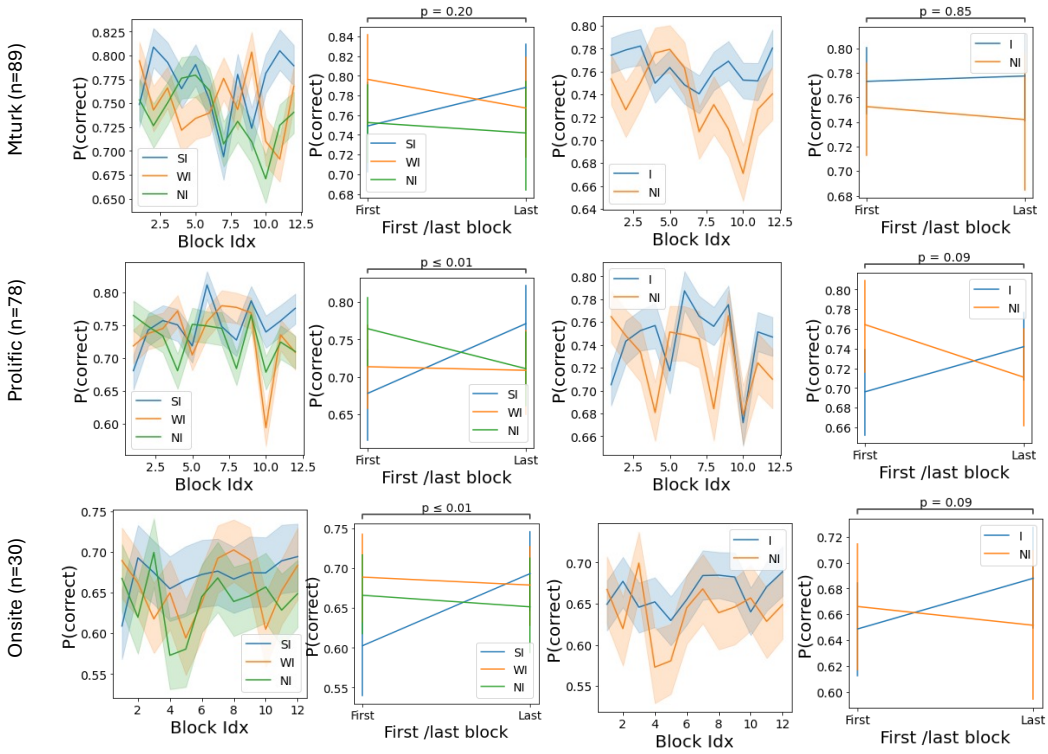


Figure 3.6: Task performance by block index. The plots showing the task performance, or the frequency of choosing the more rewarding slot machines, by the block index and the block type. Collectively, the data suggest that the performance improved in the later block and the effect was stronger in strongly informative blocks. The statistics test results in the second column show the performance difference between the first and the last strongly informative blocks (t-test paired, t(88)=1.28, t(77)=2.85, t(29)=2.88, $p = 0.204, 0.006, 0.007$ for Mturk, Prolific, and Onsite data, respectively). The test results in the fourth column show the performance difference between the first and the last informative blocks (t test paired, t(88)=0.18, t(77)=1.70, t(29)=1.77, $p = 0.854, 0.093, 0.087$ for Mturk, Prolific, and Onsite data, respectively). ANOVA analysis on the effect of condition (SI vs NI or I vs NI) in the performance increase also shows that SI cues had stronger effect in the performance improvement (ANOVA on SI vs NI: F(1)=1.15, 7.39, 3.50, $p = 0.284, 0.007, 0.064$ for Mturk, Prolific, and Onsite data, respectively; ANOVA on I vs NI: F(1)=0.13, 4.71, 1.24, $p = 0.716, 0.031, 0.268$ for Mturk, Prolific, and Onsite data, respectively).

3.5 Glia-inspired computational models

Given the observed behavioral patterns in transfer learning, we fit various reinforcement learning (RL) model variants to the choice datasets and conducted Bayesian model comparisons. Our initial hypothesis posited that transfer learning is supported

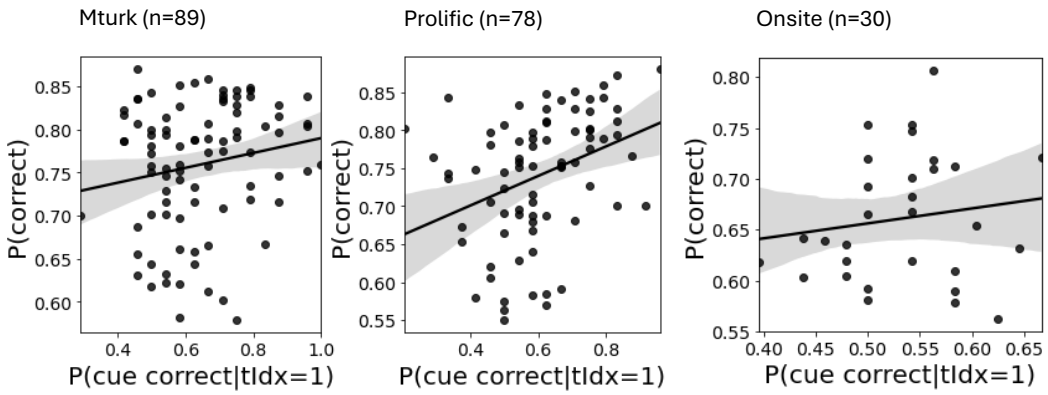


Figure 3.7: Correlations between the exploration bias and task performance in informative blocks. The task performance and the frequency of choosing the slot machine with higher cues in the initial trials is correlated but only significant in the Prolific data (Pearson correlation, $r = 0.18, 0.35, 0.15$, $p = 0.100, 0.002, 0.437$ for Mturk, Prolific, and Onsite data, respectively).

by components that slowly track learned values, allowing them to be maintained over longer timescales without forgetting. To test this hypothesis, we implemented five different RL models (see Fig. 3.8 and Methods for details). The first model was a Naive RL model, which tracks the values of slot machines in each block independently using the delta rule (Sutton, 2018) . The second model was a feature RL (fRL) model, which computes the slot machine value as a linear integration of the values of its features, such as the two cues and the painting. A forgetting mechanism was also tested by applying a decay term to the learned values at each time step. As the component that tracks learned value and maintains the information for longer timescale than the value tracking unit, we implemented non linear units for each value learning units. The exact implementation of the slower nonlinear units were inspired from physiological properties of astrocyte glia cells, which are believed to support information accumulation and event tracking (Mu et al., 2019; Porto-Pazos et al., 2011). These nonlinear units enhanced the efficacy of learned values in calculating the slot machine value when the corresponding value learning unit maintained high values for a sufficient number of time steps (see Methods for details).

To identify the underlying computational mechanisms that explain the transfer learning effect, we fit the free parameters of each model to the choice data and conducted Bayesian model comparisons (see Methods). As shown in Fig. 3.9, the onsite data

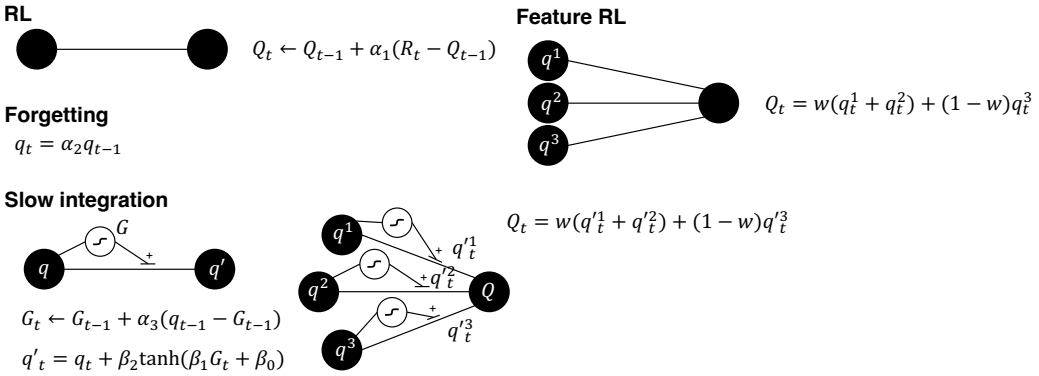


Figure 3.8: Conceptual diagram of the computational models. RL model learns the slot machine values, agnostic about the features while Feature RL model learns the values of each feature (cues and the painting) separately and linearly integrates those values to calculate the slot machine value. The models with the forgetting component forget the learned values every trial with a fixed ratio (α_2). The models with slow integration component implemented the glia-like computation by including the component G that tracks the activation of value neuron q and non-linearly affect the efficacy of neuron output using the tanh function.

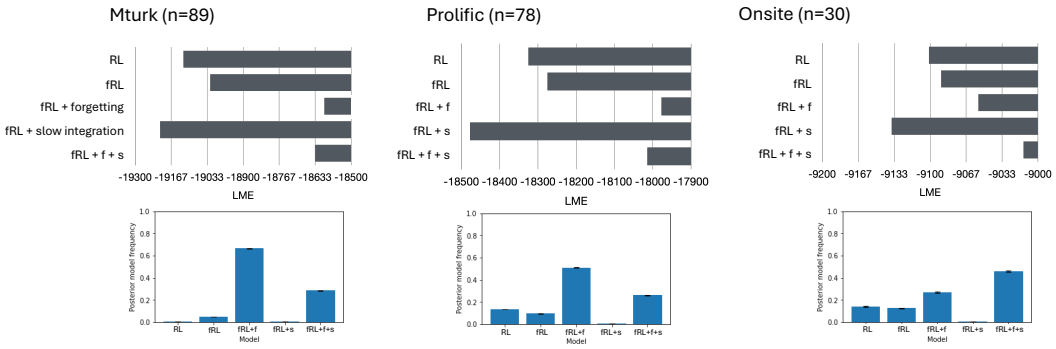


Figure 3.9: Model comparison among conventional cognitive models. The feature RL model with forgetting and glia-like slow integration component all together explained the onsite data best but it was the second best model in explaining online datasets.

was best explained by the feature RL model with forgetting and slow integration units, while the online datasets were better explained by the feature RL model with forgetting alone. However, the model frequency distribution revealed that nearly one-third of the online participants were best explained by the model with the glia-like slow integration component.

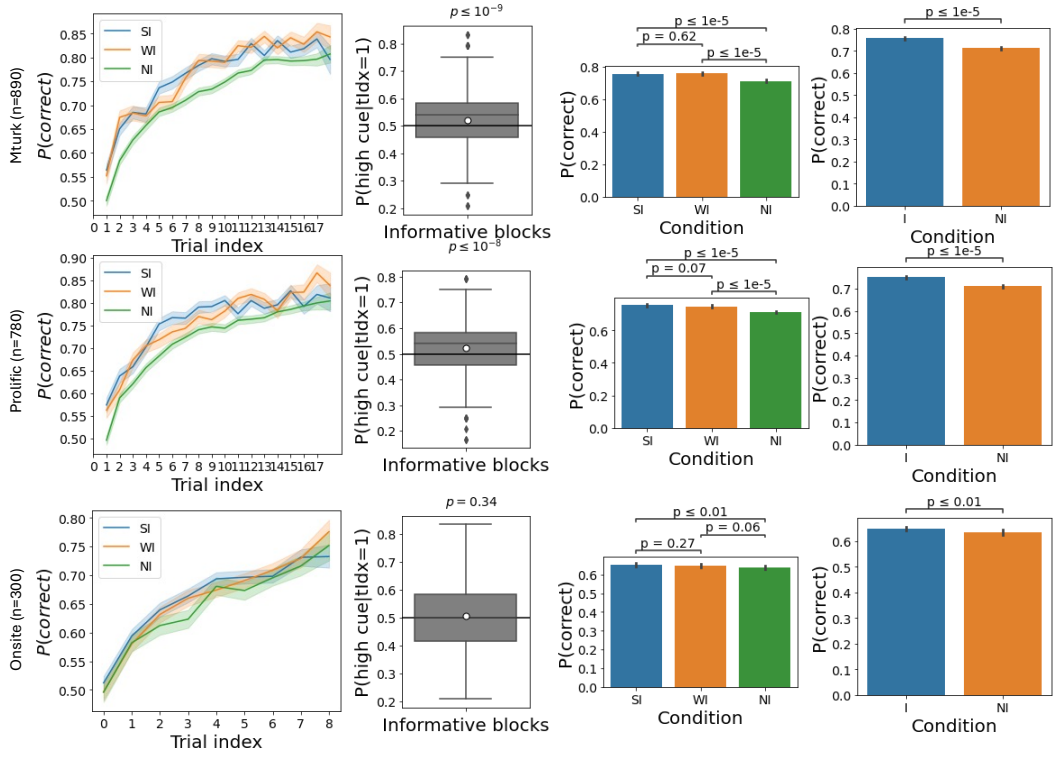


Figure 3.10: Posterior predictive check of the transfer learning effect in exploration using the feature RL with forgetting and slow integration. The model could reproduce the actual behavior in Fig 3.4. in online data but the initial exploration bias was not significant in the onsite data. Model with fit parameters were simulated 10 times for each participants. (Initial exploration bias t-test: t(889)=6.32, t(779)=5.80, t(299)=0.95, $p = 4.12 * 10^{-10}, 9.67 * 10^{-9}, 0.343$ for Mturk, Prolific, and Onsite, respectively; SI vs NI t-test paired: t(889)=11.32, t(779)=11.20, t(299)=2.93, $p = 7.71 * 10^{-28}, 4.14 * 10^{-27}, 0.003$ for Mturk, Prolific, and Onsite, respectively; WI vs NI t-test paired: t(889)=11.21, t(779)=9.30, t(299)=1.91, $p = 2.32 * 10^{-27}, 1.34 * 10^{-19}, 0.058$ for Mturk, Prolific, and Onsite respectively; SI vs WI t-test paired: t(889)=-0.50, t(779)=1.83, t(299)=1.11, $p = 0.619, 0.068, 0.269$ for Mturk, Prolific, and Onsite, respectively; I vs NI t-test paired: t(889)=12.89, t(779)=11.91, t(299)=2.73, $p = 5.46 * 10^{-35}, 3.56 * 10^{-30}, 0.006$ for Mturk, Prolific, and Onsite, respectively).

To examine the behavioral characteristics captured by each model, we also performed posterior predictive checks. Specifically, we analyzed the role of the slow integration component in choice behavior by comparing simulated results from the fRL model with forgetting and slow integration against the fRL model with only forgetting. The models were simulated 10 times for each participant using individually fitted parameters, and the simulated results were analyzed identically to the actual data.

First, the model with the slow integration component was able to reproduce the actual performance differences among different block types (see the third and fourth columns of Fig. 3.10). For instance, the simulated performances were significantly higher in strongly or weakly informative blocks than in non-informative blocks, while the performance differences between strongly and weakly informative blocks were insignificant (SI vs NI t-test paired: t(889)=11.32, t(779)=11.20, t(299)=2.93, $p = 7.71*10^{-28}, 4.14*10^{-27}$, 0.003 for Mturk, Prolific, and Onsite, respectively; WI vs NI t-test paired: t(889)=11.21, t(779)=9.30, t(299)=1.91, $p = 2.32*10^{-27}$, $1.34*10^{-19}$, 0.058 for Mturk, Prolific, and Onsite respectively; SI vs WI t-test paired: t(889)=-0.50, t(779)=1.83, t(299)=1.11, $p = 0.619, 0.068, 0.269$ for Mturk, Prolific, and Onsite, respectively). Additionally, when choices in informative blocks were pooled together, performance in these blocks was significantly higher than in non-informative blocks (I vs NI t-test: t(889)=12.89, t(779)=11.91, t(299)=2.73, $p = 5.46 * 10^{-35}, 3.56 * 10^{-30}$, 0.006 for Mturk, Prolific, and Onsite, respectively).

In contrast, the model without the glia-like component failed to reproduce these patterns (see the third and fourth columns of Fig. 3.11). For example, in the Prolific data, the model showed a significant performance difference between strongly and weakly informative blocks. In the onsite data, the model could not reproduce the performance differences between informative and non-informative blocks (SI vs NI t-test paired: t(889)=11.56, t(779)=7.61, t(299)=2.06, $p = 6.97 * 10^{-29}, 7.80 * 10^{-14}$, 0.837 for Mturk, Prolific, and Onsite, respectively; WI vs NI t-test paired: t(889)=9.45, t(779)=4.70, t(299)=1.18, $p = 2.85 * 10^{-20}, 3.08 * 10^{-6}$, 0.118 for Mturk, Prolific, and Onsite, respectively; SI vs WI t-test paired: t(889)=1.67, t(779)=2.50, t(299)=0.10, p=0.094 0.012, 0.923 for Mturk, Prolific, and Onsite, respectively; I vs NI t-test paired: t(889)=12.07, t(779)=7.32, t(299)=0.15, $p = 1.63 * 10^{-6}, 6.27 * 10^{-13}$, 0.880 for Mturk, Prolific, and Onsite, respectively). It is also noteworthy that the model with the glia-like component successfully capture the initial bias toward choosing slot machines with high cues in online datasets but not in the onsite data (see the second column of Fig. 10 and 14, fRL+f+S model's initial exploration bias t-test: t(889)=6.32, t(779)=5.80, t(299)=0.95, $p = 4.12 * 10^{-10}, 9.67 * 10^{-9}$, 0.343 for Mturk, Prolific, and Onsite, respectively) On the other hand, the model without the glia-like component reproduced the actual initial exploration bias patterns (fRL+f model's initial exploration bias t-test: t-test: t(889)=4.83, t(779)=6.59, t(299)=7.09, $p = 1.63 * 10^{-6}, 7.88 * 10^{-11}, 3.71 * 10^{-12}$ for Mturk, Prolific, and Onsite, respectively).

Second, both tested models struggled to reproduce the actual patterns of exploration bias and performance increase in the later blocks. For example, exploration bias decreased in the simulation results of both models for the MTurk data and Onsite data while the actual data showed the increasing pattern in MTurk and nonsignificant effect in Onsite data (See Fig.3.5,3.12 and 3.13). The initial choices in informative blocks were rather decreasing (All informative blocks, fRL+f t-test paired : t(889)=-1.64, t(299)=-2.86, $p = 0.101, 4.30 * 10^{-3}$ for Mturk, and Onsite, respectively; fRL+f+s t-test paired: t(889)=-2.69, t(299)=0.12, $p = 0.007, 0.907$ for for Mturk and Onsite respectively; Strongly informative blocks only, fRL+f: t(889)=-0.23, t(299)=-1.19, p=0.815, 0.236 for Mturk, and Onsite, respectively; fRL+f+s t-test paired: t(889)=-1.92, t(299)=0.56, p=0.055, 0.578 for Mturk, and Onsite, respectively). However, both models showed an increase of exploration bias in the prolific data where the actual data showed significant increase of it in the strongly informative blocks only (All informative blocks, fRL+f t-test paired: t(779)=2.60, $p = 9.44 * 10^{-3}$; fRL+f+s t-test paired: t(779)=3.77, $p = 1.67 * 10^{-4}$ Strongly informative blocks only, fRL+f t-test paired: t(779)=1.46, p=0.145; fRL+f+s t-test paired: t(779)=1.80, p=0.073)

On the other hand, the results of the online datasets' simulations from both models showed higher performance in the last informative blocks, while the performance increase in the actual data was significant only in the Prolific data (see Fig.3.6,3.14 and 3.15). For instance, in the comparison of the first vs. last informative blocks, the fRL with forgetting showed significant improvements (informative first vs last block, t-test paired: t(889)=1.69, t(779)=4.40, $p = 0.092, 1.26 * 10^{-5}$ for Mturk and Prolific, respectively; strongly informative first vs last block, t-test paired: t(889)=5.19, t(779)=5.13, $p = 2.57 * 10^{-7}, 3.76 * 10^{-7}$ for Mturk and Prolific respectively). Similarly, ANOVA analyses on the informative vs. non-informative blocks showed that the fRL with forgetting model captured these differences in Prolific data but not in MTurk data (ANOVA on first vs last block & I vs NI: F(1)=0.38, 7.84, $p = 0.539, 0.005$ for for Mturk and Prolific, respectively; ANOVA first vs last block & on SI vs NI: F(1)=4.58, 14.19, $p = 0.032, 1.68 * 10^{-4}$ for Mturk and Prolific, respectively). The fRL model with forgetting and slow integration also showed similar patterns (informative first vs last block, t-test paired: t(889)=1.84, t(779)=5.20, $p = 0.066, 2.59 * 10^{-7}$ for Mturk and Prolific, respectively; ANOVA first vs last block & on I vs NI: F(1)=0.24, 8.65, $p = 0.627, 0.003$ for Mturk and Prolific respectively; strongly informative first vs last block, t-test paired: t(889)=3.84, t(779)=8.01, $p = 1.35 * 10^{-4}, 4.10 * 10^{-15}$ for Mturk and Prolific, respectively;

ANOVA on first vs last block & SI vs NI: F(1)=1.41, 30.13, $p = 0.235, 4.36 * 10^{-8}$ for Mturk and Prolific, respectively).

However, when analyzing the onsite data, both models struggled to reproduce the actual performance patterns. For example, instead of showing increased performance in the last informative blocks, as observed in the actual data, the fRL with forgetting showed decreased performance (see Fig.3.6 and 3.15, fRL+f Informative first vs last block, t-test paired: t(299)=-5.71, $p = 1.82 * 10^{-8}$; fRL+f ANOVA first vs last block & on I vs NI: F(1)=8.00, $p = 4.73 * 10^{-3}$. Note that the direction was opposite; fRL+f strongly informative first vs last block, t-test paired: t(299)=-1.99, $p = 0.047$; fRL+f ANOVA first vs last block & on SI vs NI: F(1)=0.94, $p = 0.333$. Note that the direction was opposite). Similar patterns were found in the fRL model with forgetting and slow integration, where performance decreased rather than increased (see Fig.12, fRL+f+s informative first vs last block, t-test paired: t(299)=-3.08, $p = 0.002$; fRL+f+s ANOVA first vs last block & on I vs NI: F(1)=4.68, $p = 0.031$. Note that the direction was opposite; fRL+f+s strongly informative first vs last block, t-test paired: t(299)=-1.61, $p = 0.110$; fRL+f+s ANOVA first vs last block & on SI vs NI: F(1)=2.34, $p = 0.126$. Note that the direction was opposite).

Interestingly, positive correlations between exploration bias and task performance in informative blocks were observed only in the onsite data simulation using the model without the glia-like slow integration component. In contrast, the actual onsite data showed an insignificant correlation (see Fig.3.7, 3.16, and 3.17; fRL+f Person correlation: r=-0.004, -0.050, 0.280, $p = 0.901, 0.166, 8.68 * 10^{-7}$ for Mturk, Prolific, and Onsite data, respectively, fRL+f+s Person correlation: r=0.067, -0.104, 0.059, $p = 0.045, 0.004, 0.328$ for Mturk, Prolific, and Onsite data, respectively). Notably, while the actual correlation was most significant in the Prolific data, the simulated correlations from both models were negative.

To further analyze the effect of the glia-like slow integration component on transfer learning, we examined model-based variables and the average likelihood per trial index. First, the proportion of the slow integration effect in decision-making was most prominent in initial trials (see the first row of Fig.3.18, see methods for the details). Additionally, the relative likelihood of the fRL model with forgetting and slow integration compared to the RL or fRL model with forgetting was higher in early trials of each block (see the second and third rows of Fig.3.18). These patterns collectively suggest that the knowledge transferred across different blocks was maintained and supported by the glia-like slow integration component.

3.6 Deep RL models

In addition to conventional cognitive modeling, we also performed deep-neural-network-based behavior modeling that included recurrency but lacked the slow integration component. We optimized a variant of a recurrent neural network called a Gated Recurrent Unit (GRU) to predict the actual action given the history of state, action, reward, and a binary variable indicating the start of a new block (see Fig.3.19 and methods for the details on model implementation) (Cho, 2014; Duan et al., 2016). The GRU model received a concatenated vector of the current and previous several trials' states, actions, rewards, and block start indicators as input, with further previous trial information passed through the previous hidden vector via recurrent connections. To determine the optimal context size, we tested different context lengths using leave-one-subject-out (LOSO) validation. The model was trained on all data except for one MTurk participant, who was used as the test subject. The LOSO log likelihood was found to be maximum when the context size was 3, suggesting that participants might maintain recent trial experiences in working memory, while further past experiences were kept implicitly (see Fig.3.20) (Keresztes et al., 2018).

Using the optimal context size of 3, the RNN model was trained and tested on the entire dataset using the LOSO method and compared with the cognitive models discussed in the previous section. The log likelihoods of each cognitive model for each dataset were summed, and the LOSO log likelihood of the RNN model was calculated for the entire dataset. This approach was more stringent than the other cognitive models, where models were fit to each individual and the maximized log likelihood was used as the metric. Even with this stringent metric, the RNN model demonstrated the best performance in explaining actual choices (see the second row of Fig.3.20).

To further validate the model's performance, we conducted simulation studies for each participant 10 times (see Fig.3.21,3.22 and 3.23 for the simulations on Mturk, Prolific and Onsite data, respectively). The simulated performance differences between conditions showed similar patterns to the actual MTurk and Prolific data. However, the simulation on the Onsite data failed to reproduce the actual pattern. For instance, the simulated performance in weakly informative blocks was lower than in non-informative blocks (t-test paired, t(299)=-2.79, p = 0.006). Additionally, the expected performance difference between informative and non-informative blocks was not observed (t-test paired, t(299)=-0.04, p = 0.966) While the model

successfully reproduced the actual initial action selection bias toward the slot machine with higher cues in the online datasets, it failed to do so in the Onsite dataset (t(889)=13.74, $p = 4.19 * 10^{-39}$; t(779)=4.98, $p = 7.69 * 10^{-7}$; t(299)=-1.31, $p = 0.189$ for MTurk, Prolific, and Onsite data, respectively).

However, the RNN model struggled to reproduce the increase in initial exploration bias in the later blocks. In the MTurk and onsite data simulations, the difference between the first and the last blocks' initial exploration biases in informative blocks was insignificant (t-test paired; t(889)=1.43, p=0.154; t(299)=1.43, p=0.154 for MTurk, and Onsite data each, also see Fig. 21 and 23). Moreover, in the MTurk data simulation, the initial exploration bias was even larger in the first strongly informative block than in the last strongly informative block (t-test paired; t(889)=-4.86, $p = 1.38 * 10^{-6}$). Nonetheless, the model was able to successfully reproduce the increasing trends of initial choice bias toward high cues in the Prolific data (see Fig. 22, all informative blocks t-test paired, t(779)=3.72, $p = 2.05 * 10^{-4}$; strongly informative blocks only, t(779)=3.01, $p = 2.68 * 10^{-3}$).

The correlation between initial exploration bias and performance was significantly positive in online datasets' simulations but not in the Onsite data simulation (r=0.218, 0.097, 0.037, $p = 4.44 * 10^{-11}$, 0.006, 0.528 for MTurk, Prolific and onsite data each). It is noteworthy that the actual correlation was significantly positive in the Prolific data and showed positive trends in the MTurk and Onsite data as well (see Fig.3.7). The simulated correlation value and the actual correlation value was similar in the MTurk dataset (r=0.18 and 0.218 for simulation and actual data each), but the simulation on other datasets showed discrepancies (r=0.35 and 0.097 for simulation and actual Prolific data each, r=0.15 and 0.037 for simulation and actual Onsite data each)

Despite these successes, the RNN model significantly failed to reproduce the increase in performance in the later blocks (see the last row of Fig.3.21,3.22 and 3.23). In the simulations of the online datasets, performance was significantly lower in the last blocks compared to the first (MTurk First vs Last all informative blocks, t-test paired t(899)=-50.08, $p \leq 1 * 10^{-5}$; ANOVA on first vs last block & I vs NI, F(1)=260.36, $p = 1.91 * 10^{-58}$; First vs Last strongly informative blocks, t-test paired t(889)=-26.62, $p = 2.78 * 10^{-154}$; ANOVA on first vs last block & SI vs NI, F(1)=47.74, $p = 4.92 * 10^{-12}$; Prolific First vs Last all informative blocks, t-test paired t(779)=-29.93, $p = 1.93 * 10^{-193}$; ANOVA on first vs last block & I vs NI, F(1)=61.21, $p = 5.24 * 10^{-15}$; First vs Last strongly informative blocks t-test

paired t(779)=-13.20, $p = 1.16 * 10^{-39}$; ANOVA on first vs last block & SI vs NI, F(1)=179.19, $p = 8.59 * 10^{-41}$). In the Onsite data simulation, the performance in the strongly informative blocks was higher in the last block compared to the first (See Fig.23, First vs Last strongly informative blocks, t-test paired, t(299)=1.69, $p = 0.091$; ANOVA on first vs last block & SI vs NI, F(1)=6.43, $p = 0.001$). However, when both strongly and weakly informative blocks were analyzed together, the RNN model showed the opposite pattern from the actual data (First vs Last all informative blocks t(299)=-1.18, p=0.236; ANOVA on first vs last block & I vs NI, F(1)=0.14, p=0.713).

3.7 Discussion

In this work, we aimed to explore the computational mechanisms of human transfer learning, particularly how individuals leverage previously acquired knowledge to solve novel environments with shared features. The behavioral findings from the experiments and the computational modeling provide insights into the mechanistic description about the human transfer learning.

First, behavioral results demonstrated that participants effectively learned the information about the feature value and transferred the knowledge across different blocks of the task. For example, the performance in blocks with informative cues was higher than non-informative blocks. Moreover, the initial exploration bias was significant towards slot machines with high-reward cues, and the performance and exploration bias tend to be positively correlated. These findings suggest that participants were not simply learning each block independently but were utilizing information from prior experiences to make more informed decisions in new situations.

The temporal dynamics observed in the behavioral study, particularly the increasing exploration bias in later blocks within strongly informative blocks of the online datasets, indicate that the efficacy of transfer learning improves with continued exposure to similar tasks. In addition, we could observe significant performance increase in the last informative blocks compared to the initial blocks. These suggest that as participants gain more experience, they become more adept at understanding meaning of the cues, leading to more efficient learning and decision-making.

The computational models used in this study, particularly the reinforcement learning (RL) models with glia-inspired slow integration components, were able to partially capture the transfer learning behavior observed in the participants. In addition, model-based variable analysis and trial-by-trial likelihood analysis suggest

that components that mimic the slow accumulation of information and retaining the learned information, akin to the role of glial cells in the brain, are crucial for effective transfer learning (Kofuji and Araque, 2021; Mu et al., 2019; Porto-Pazos et al., 2011). This slow integration mechanism allows for the retention of learned values over longer timescales, preventing the rapid forgetting that often plagues artificial systems.

On the other hand, the recurrent neural network (RNN) models showed strong performance in predicting participant choices, surpassing the other tested models. However, they struggled to replicate the temporal dynamics of transfer learning, particularly the increase in exploration bias and performance over time. These suggest that while deep learning models with recurrent connections are powerful tools for predicting behaviors, they may lack the nuanced temporal processing capabilities necessary to implement transfer learning like a human (Dezfouli et al., 2019a; Miller et al., 2024; Ji-An et al., 2023a). Moreover, it suggests that modeling the computational power of the brain requires not only recurrent connections but also more temporally heterogeneous components (Hihi and Bengio, 1995; Botvinick, 2012).

The findings of this study highlight the ongoing struggle to replicate how humans learn and transfer knowledge in artificial intelligence.This cognitive flexibility, which is effortless for humans, remains a challenge for artificial intelligence despite the exponential advancements in deep neural network-based models. The naive application of these models to learn new tasks often results in a degradation of performance on previously trained tasks, a phenomenon known as catastrophic forgetting (Kirkpatrick et al., 2017; Wang et al., 2024). The currently suggested techniques are rather heuristic than a fundamental solution, such as having different models for individual tasks and applying and tuning them based on the context. (Ma et al., 2018; Hazimeh et al., 2021) or having a subset of weights for each task (Kirkpatrick et al., 2017). However, our results suggest that incorporating biological insights, such as slow integration mechanisms inspired by glial cells, could offer a pathway to more robust transfer learning capabilities in AI. By mimicking the brain's ability to maintain learned information over time, artificial systems could potentially avoid the pitfalls of catastrophic forgetting and achieve more human-like cognitive flexibility (Botvinick, 2012; Botvinick and Weinstein, 2014; Ribas-Fernandes et al., 2019; Eckstein and Collins, 2020).

In addition, these results of computational modeling suggest future direction of the

research on the role of astrocyte glia cells in decision making and learning. So far, the research on glia cell in decision making has focused on rather basic components in decision making such as information accumulation and its correlation with the glial activity (Mu et al., 2019) or causal relation between the glia and the value-based learning performance (Mederos et al., 2021; Wang et al., 2017; Kofuji and Araque, 2021). However, our research is the first formal evidence that showed the potential crucial role of glia-like components in feature-based transfer learning. The simple form of context-based bandit tasks that can be tested in model animals will allow us to understand the precise casual relationship of the astrocyte glia and transfer learning.

Furthermore, even though there are no tools available to directly measure glial activities in the human brain, the computational model can be indirectly tested in humans using fMRI. Multiple studies have shown that the BOLD signal is closely related to astrocyte-glial activities (Takata et al., 2018; Kahali et al., 2021; Howarth et al., 2021). In addition, neural activities modeled with tripartite synapses, where synaptic efficacy is modulated by glial cells, may better predict actual neural activities (Perea et al., 2009a). Therefore, simulated glia activities or neural activities can be used for model-based analyses of fMRI analyses. The effectiveness of including the glia in the computational modeling of behavior can be tested by employing the model-based neural model comparisons as well.

Although this study provides valuable insights about the mechanisms of transfer learning, it also highlights several limitations. The discrepancy between the online and onsite data, and even between online datasets, suggests that environmental factors and task design can significantly influence the efficacy of transfer learning. This warrants further investigation into how different contexts, settings, and task order affect the transferability of learned knowledge.

Future research should explore the integration of other cognitive components, such as working or declarative memory and attention mechanisms, into computational models (Poldrack et al., 2001; Reber et al., 1996; Knowlton and Squire, 1993; Gershman and Daw, 2017; Collins and Frank, 2012; Niv et al., 2015; Leong et al., 2017). For example, the learning of multiple features simultaneously and generalization of the meaning of individual features to their compound is related to the literature on configural learning and multi-dimensional decision making (Farashahi et al., 2017; Kahnt et al., 2011a; Zysset et al., 2006; Soto et al., 2014; Pelletier and Fellows, 2019). The literature showed that the learning of the configural cue starts from the

elemental learning of the meaning of individual cues to the more thorough meaning of each configuration of cues (Gluck et al., 2002). Incorporating these elements into the model may better replicate the complex processes underlying human transfer learning.

In conclusion, this study contributes to the understanding of human transfer learning and offers potential directions for enhancing transfer learning in artificial systems. By bridging cognitive science and artificial intelligence, we move closer to replicating the cognitive flexibility that is so effortlessly demonstrated by humans but remains a challenge for machines.

3.8 Methods

Participants

Three behavioral datasets were collected: two online and one onsite. The first online dataset was collected from Amazon Mechanical Turk (MTurk) between September and October in 2020. The second dataset was collected onsite from October to December in 2022, and the third dataset was collected from Prolific in February 2024. I recruited 139 participants from MTurk, 30 participants from the onsite study, and 84 participants from Prolific. I applied exclusion criteria such that the proportion of choosing the most rewarding action, averaged across blocks, was not significantly higher than the chance level of 0.5, using a one-sample t-test. After exclusion, 89 participants from Mturk(38 female, age 40 ± 12.2), 30 participants from the onsite (16 female, age 32±9.2) and 78 participants from Prolific (33 female, age 34 ± 9.4) were included for the further analyses. Participants from Mturk and Prolific were paid a base rate of $10 or $12 respectively, plus a performance-based bonus up to $3. The onsite participants were paid $20 per hour, plus a performance bonus up to $3. All participants provided their informed consent, and the study was approved by the Institutional Review Board of California Institute of Technology.

Task

Participants completed multiple blocks of 2-armed bandit tasks, with each block presenting a novel bandit task. Before starting each block, participants were informed they were entering a new environment, which was indicated by displaying the name of the new environment. To maintain novelty, the slot machines featured unique paintings at their bases in each new block. To elicit transfer learning, two visual cues were displayed on each slot machine, shared across environments. Participants' goal was to collect as many coins as possible by identifying the better slot

machine in each block, and they were informed that the cues could help them more quickly determine the better slot machine. Rewards were given as a binary outcome: winning a coin or not.

There were 6 cues, categorized into 3 groups based on the contingency with the winning slot machine : High cue 1/2 (H1/H2), Neutral cue 1/2 (N1/N2) and Low cue 1/2 (L1/L2). For the MTurk participants, the cues consisted of fractal images. In contrast, for the Onsite and Prolific participants, AI-generated, abstract-art-stylized images of fruits were used (model: Stable Diffusion v1-4; prompt: abstract art, <fruit name>). This approach aimed to match the cues' features with the art paintings shown, andw make it easier for participants to remember the cues.

The online tasks consist of 36 blocks, while the onsite task had 72 blocks. There were three types of blocks, categorized based on the cues presented on the slot machines: Strongly informative (SI), weakly informative (WI), and noninformative (NI) blocks. In the online tasks, SI blocks featured one slot machine with H1 and H2 cues, and the other with L1 and L2 cues. WI blocks had each slot machine paired with H1, N1 cues, and L1, N2 cues, respectively. In NI blocks, one slot machine came with H2, L2 cues, and the other with N1, N2 cues. For the onsite task, there were two variations of SI blocks: one with H1, N1 (or H2, N2) cues on a slot machine, and the other with N2, L2 (or N1, L1) cues. Similarly, two types of WI blocks featured H2, N2 (or H1, N1) cues on one slot machine, and H1, L2 (or H2, L1) cues on the other. In NI blocks, slot machines were assigned with H1, L2 (or N1, L1) cues and H2, L1 (or N2, L2) cues, respectively.

In the online experiments, the reward probabilities for each slot machine were set at [0.3, 0.7], while in onsite experiments, they were [0.15, 0.55]. For the SI blocks, which totaled 12 in the online and 24 in the onsite experiments, the slot machines displaying H1, H2 (or H1, N1, or H2, N2 in the onsite) had the higher reward probability in 10 (or 20 in the onsite) of these blocks. In the WI blocks, totaling the same number, the slot machines with H1, N1 (or H2, N2, or H1, N1 in the onsite) were the better slot machine in 8 (or 16 in the onsite) blocks. For the NI blocks, in 6 (or 12 in the onsite) out of the total blocks, the slot machines with H2, L2 (or H2, L2, or N1, L1 in the onsite) were the winning machines. Therefore, in the online tasks, out of the 24 slot machines associated with H1, 18 were winning machines. The counts for the other cues were as follows: H2 led to 16 winning machines, N1 to 14, N2 to 10, L2 to 8, and L1 to 6. In the onsite task, among the 48 slot machines displaying H1, 28 were better slot machines. The winning machine counts for the

other cues were: H2 with 28, N1 with 26, N2 with 26, L2 with 18, and L1 also with 18. The length of each block was randomized. In the online tasks, lengths varied between 14-18 trials and in the onsite task, they they ranged from 7 to 9 trials. Consequently, the total number of trials across all tasks was 576.

In Prolific and onsite experiments, the order of the cues within each slot machine was randomized across blocks but remained constant within a block. In contrast, for Mturk, the order of the cues within slot machines was not randomized to simplify the task for participants. In all experiments, the left-right order of the slot machines was randomized across trials.

Given this task structure, participants were able to learn the contingencies of the cues related to the better slot machines through trial and error, and apply this learned knowledge to more quickly identify the winning slot machine in novel environments.

After completing the 2-armed bandit tasks, participants were asked to complete two additional post-task surveys. In one survey, participants were presented with a cue or a pair of cues and reported how likely it was that the cue (or pair of cues) shown was associated with an increased or decreased chance of winning a reward. Their responses were registered using a sliding bar, with the leftmost end labeled 'Decreased chance' and the rightmost end labeled 'Increased chance'. Responses were collected on a 0 to 100 scale. In the onsite experiment, each cue (6 kinds) and all possible permutations of cue pairs (15 kinds) were shown twice. In the online experiments, each cue (6 kinds) and the cue pairs that were presented during the main task (6 kinds) were shown once.

In the other survey, participants chose the better cue or cue pair based on their experience from among the two cues or two pairs of cues presented to them. In the onsite experiment, all possible permutations of cue pairs (15 kinds) and permutations of pairs of cue pairs (45 kinds) were shown twice each. In the online experiments, all possible permutations of cue pairs (15 kinds) and the permutations of pairs of cue pairs (15 kinds) that were shown during the main task were shown once each. The order of the two surveys was randomized across participants.

Psychopy 2021.2.0 was used for the onsite experiment and jsPsych 6.1.0 was used for the online experiment.

Behavioral analysis

Computational modeling of behavior

We tested conventional cognitive models that are based on the reinforcement learning (RL) algorithm. Inspired from the importance of the neuro-glia interaction in neuro-computation and accumulation and maintaining of information, glia-like computational components was added on top of the RL algorithm. The models were fit to each subject data using the Computatiaonal Behavioral Modeling (CBM) toolkit which calculates model evidence and likelihoood of the data using Laplace approximation and estimates parameters using maximum-a-posteriori (MAP) estimation (Piray et al., 2019b). Mean 0 and variance 66.25 Gaussian priors were used for estimating the parameters. Since CBM assumes that the parameters follow a normal distribution, we used transformation functions to adjust them: a sigmoid function for parameters that range between 0 and 1, and an exponential function for positive parameters. We opted not to use the hierarchical Bayesian inference function from the toolkit. Instead, we conducted Bayesian model selection by utilizing the log evidence from the CBM outputs for Bayesian model comparison (Stephan et al., 2009b).

Reinforcement learning model. The RL model is the simplest model we tested that does not consider the visual features consisting slot machines, but only models value of the slot machines themselves. Values Q for each slot machines are initialized at 0.

$$Q_0(s) = 0$$

and the probability of choosing a slot machine on the left of the screen s_l is a softmax function of the values of slot machines shown

$$P_t(s_l) = \frac{exp(\beta Q_t(s_l))}{exp(\beta Q_t(s_l))) + exp(\beta Q_t(s_r)))}$$

where $\beta > 0$ is an inverse temperature parameter. The value $Q_t(s)$ for the chosen slot machine s_t at trial t is updated based on the delta rule

$$RPE_t = r_t - Q_t(s_t)$$

$$Q_{t+1}(s_t) \leftarrow Q_t(s_t) + \alpha \times RPE_t$$

where the reward prediction error (RPE) is the difference between the reward r_t (0 or 1 depending on the realized outcome) at the trial t and the action value. $0 \leq \alpha \leq 1$ is a learning rate parameter. There were 2 free parameters (β, α) in this model.

Feature-based RL (fRL) model. The fRL model learns the values of visual features consisting slot machines. So the value of the slot machine (s) is defined as the weighted sum of the values of visual cues (c_1, c_2) on the top of slot machines and paintings (p) on the bottom of slot machines.

$$Q(s) = w(Q(c_1) + Q(c_2)) + (1 - w)Q(p)$$

where $w \in [0, 1]$. The learning rule was

$$RPE_t = r_t - Q'_t(s_t)$$

$$Q_{t+1}(c_1) \leftarrow Q_t(c_1) + \alpha * RPE_t$$

$$Q_{t+1}(c_2) \leftarrow Q_t(c_2) + \alpha * RPE_t$$

$$Q_{t+1}(p) \leftarrow Q_t(p) + \alpha * RPE_t$$

and decision making rule followed the rules in the RL model. There were 3 free parameters (β, α, w) in this model.

Forgetting model The forgetting was implemented as the exponential decay of learned value every trial. For all the learned visual cues and paintings, in every trial, the value was decayed with the following rule

$$Q \leftarrow \alpha' Q$$

where $\alpha' \in [0, 1]$. There fore, fRL with forgetting model has 4 free parameters ($\beta, \alpha, \alpha', w$).

Slow integration model. The computational component that resembles astrocyte-glia was implemented based on the properties of real biological glia cells. The astrocyte cells integrate neuronal activity with slower temporal scale than neurons, and respond to neurotransmitters in a nonlinear way. Also the astrocyte activities control the efficacy of the synapses (Perea et al., 2009b). Consequently, the slow integration component that resembles glia were implemented using the following rules. For each feature value coding neuron Q, corresponding slow integration component G (from glia) follows

$$G_{t+1} \leftarrow G_t + \alpha''(Q_t - G_{t-1})$$

where $\alpha'' \in [0, 1]$. This updating rule allows us to approximate the slower temporal dynamics of astrocyte-glia cells. Then the efficacy of the output of value coding neuron Q is updated to Q' by

$$Q' = Q + \beta_2 tanh(\beta_1 G + \beta_0)$$

which resemble the property of biological glia that controls the efficacy of synapses only when there was a strong neuronal activities. Then the neuron that represents the decision variable for a slot machine is

$$Q'(s) = w(Q'(c_1) + Q'(c_2)) + (1 - w)Q'(p)$$

Then the probability of choosing a left slot machine s_l is a softmax function of the Q'. The value coding neurons Q were updated using

$$RPE_t = r_t - Q'_t(s_t)$$

$$Q_{t+1}(c_1) \leftarrow Q_t(c_1) + \frac{\alpha}{3} * RPE_t$$

$$Q_{t+1}(c_2) \leftarrow Q_t(c_2) + \frac{\alpha}{3} * RPE_t$$

$$Q_{t+1}(p) \leftarrow Q_t(p) + \frac{\alpha}{3} * RPE_t$$

therefore, the fRL model with glia-like slow integration has 7 free parameters $(\alpha, \alpha'', \beta, \beta_0, \beta_1, \beta_2, w)$.

Forgetting and slow integration model. Forgetting and slow integration model incorporates all the components from the forgetting model and slow integration model. Therefore this model has 8 free parameters $(\alpha, \alpha', \alpha'', \beta, \beta_0, \beta_1, \beta_2, w)$.

Effect of slow integration in decision-making. To analyze the effect of slow integration in decision-making, the trial-by-trial model-based variables were generated using the fRL with forgetting and slow integration with fit parameters. Then the analyses of the first row in Fig.19 was done by calculating the relative size of the absolute difference of model-generated slow integration components $(|\sum_{s\in\{c_1,c_2,p\}}(\beta_2 tanh(\beta_1 G(s_{left})+\beta_0) - \beta_2 tanh(\beta_1 G(s_{right})+\beta_0))|)$ to the absolute difference of model-generated decision variables $(|\sum_{s\in\{c_1,c_2,p\}}(\beta_2 tanh(\beta_1 G(s_{left})+\beta_0) - \beta_2 tanh(\beta_1 G(s_{right})+\beta_0))| + |\sum_{s\in\{c_1,c_2,p\}}(Q(s_{left}) - Q(s_{right}))|)$.

Recurrent neural network model. Recurrent neural network (RNN) models were

trained to predict trial-by-trial responses based on the history of states, actions, and rewards. Each trial was represented by the cue identities of the left and right slot machines, the chosen action, the outcome received, and an indicator marking the start of a new block. The RNN model used the trial representations from the past T trials as input, with the actual human response to the current trial serving as the target for prediction.

Before processing by the RNN, the elements of the input were embedded into vectors with dimension d. For example, the two cue identities of the left slot machine were embedded into two vectors, which were then summed element-wise to represent the left slot machine as a whole. The cues in the right slot machine were processed similarly to represent the right slot machine. Additionally, the chosen action, the given outcome, and the block start indicator were embedded into d-dimensional vectors. The state of a single trial was thus represented by the concatenation of these five d-dimensional vectors. Therefore, the input to the RNN was a concatenated vector of the left and right slot machines, actions, outcomes from the past T trials, and the block start indicator for the current trial, resulting in a $(4T + 1) \times d$-dimensional input vector.

The neural network was trained using leave-one-subject-out cross-validation across all three datasets (n=197), with the model trained on data from all but one subject and tested on the held-out subject. Also mapping between cues and their embedding were randomized across training batch to model the cue randomization across participants and prevent the model from knowing the meaning of cues from the begining of making predictions on the one held-out participant. The neural network was optimized using cross-entropy loss with the Adam optimizer, set at a learning rate of 0.001. Four hyper parameters (embedding dimension, hidden layer size, number of layers and context size) were optimized using the grid search, resulting in the final selection of an embedding dimension of 16, hidden layer size of 256, three layers, and a context size of 3. Both GRU and LSTM were tested for the RNN unit, but the GRU showed better log likelihood than the LSTM. The implementation was done using Python 3.9.0 and PyTorch 2.1.1 with CUDA 11.8.

3.9 Acknowledgments

Funding: This work was supported by a Department of Defense Multidisciplinary

University Research Initiative 2021 to J.P.O. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562 (Towns et al., 2014). Computational resources, including PSC Bridges-2 Storage and PSC Bridges-2 Regular Memory, provided through XSEDE under allocation SOC210002, were utilized by S.Y. **Author contributions:** S.Y. and J.P.O. conceived and designed the study, S.Y. performed experiments and S.Y. and J.P.O. analyzed and discussed results. S.Y. and J.P.O. wrote the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The data, code and analysis results utilized in this manuscript will be available online at the time of publication.

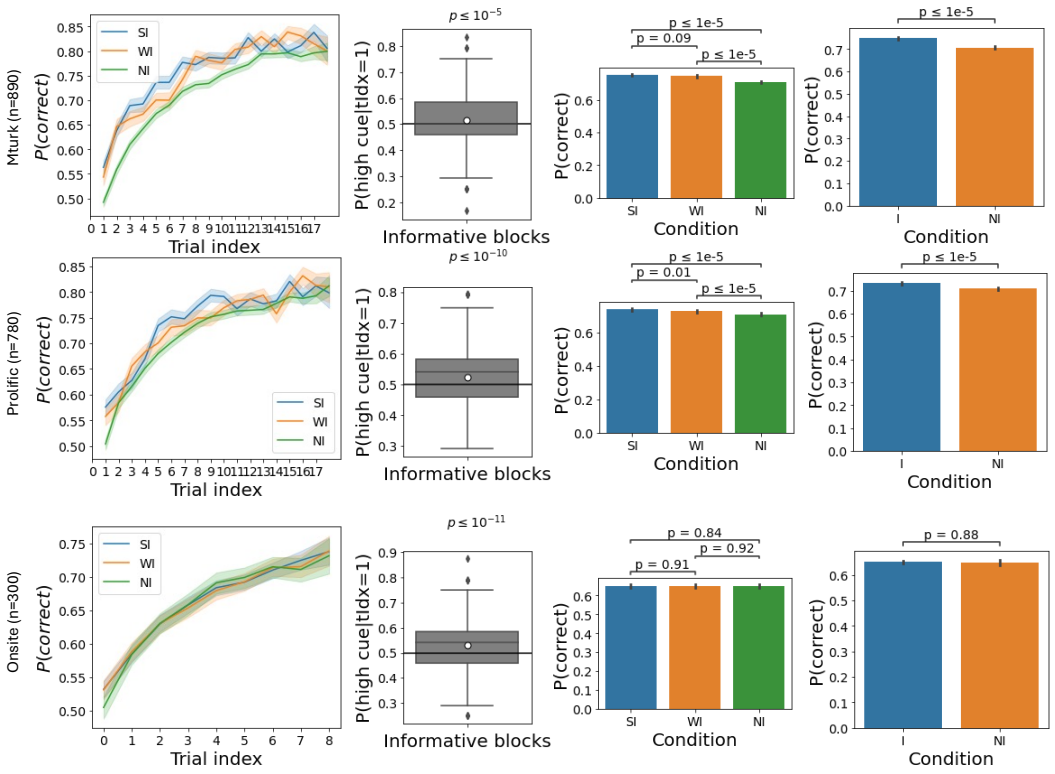


Figure 3.11: Posterior predictive check results of the transfer learning effect in exploration using the feature RL with forgetting only. The figures correspond to Fig 10. It cant be observed from the second column of the simulation results that the Initial exploration bias was significant in all three data (t-test: t(889)=4.83, t(779)=6.59, t(299)=7.09, $p = 1.63 * 10^{-6}, 7.88 * 10^{-11}, 3.71 * 10^{-12}$ for Mturk, Prolific, and Onsite, respectively). The simulations could capture the better performance level in the informative blocks only in the online datasets (I vs NI t-test paired: t(889)=12.07, t(779)=7.32, t(299)=0.15, $p = 1.63 * 10^{-6}, 6.27 * 10^{-13}, 0.880$ for Mturk, Prolific, and Onsite respectively). Moreover, unlike the simulation using the feature RL with forgetting and glia, the model without glia model failed in replicating the indifference in the performance between strongly and weakly informative blocks in the Prolific, data (SI vs NI t-test paired: t(889)=11.56, t(779)=7.61, t(299)=2.06, $p = 6.97 * 10^{-29}, 7.80 * 10^{-14}, 0.837$ for Mturk, Prolific, and Onsite respectively; WI vs NI t-test paired: t(889)=9.45, t(779)=4.70, t(299)=1.18, $p = 2.85 * 10^{-20}, 3.08 * 10^{-6}, 0.118$ for Mturk, Prolific, and Onsite respectively; SI vs WI t-test paired: t(889)=1.67, t(779)=2.50, t(299)=0.10, $p = 0.0940.012, 0.923$ for Mturk, Prolific, and Onsite, respectively).

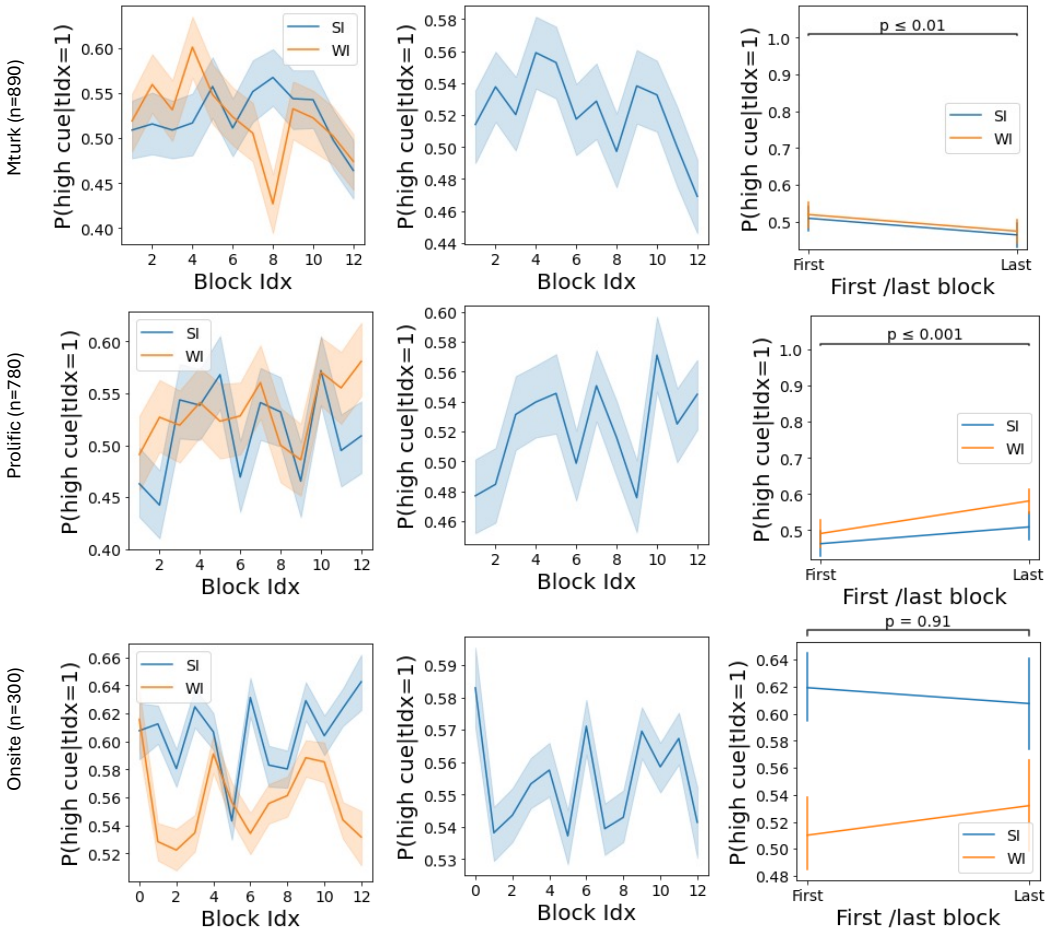


Figure 3.12: Posterior predictive check results of the transfer learning effect in terms of exploration bias using the feature RL with forgetting and slow integration. These plots corresponds to the plots in Fig 5. The frequency of choosing the slot machine with the higher cues was only significantly increasing in simulation using the Prolific data (All informative blocks, t-test paired: t(889)=-2.69, t(779)=3.77, t(299)=0.12, $p = 0.007, 1.67 * 10^{-4}, 0.907$ for Mturk, Prolific, and Onsite, respectively; Strongly informative blocks only, t-test: t(889)=-1.92, t(779)=1.80, t(299)=0.56, $p = 0.055, 0.073, 0.578$ for Mturk, Prolific, and Onsite, respectively).

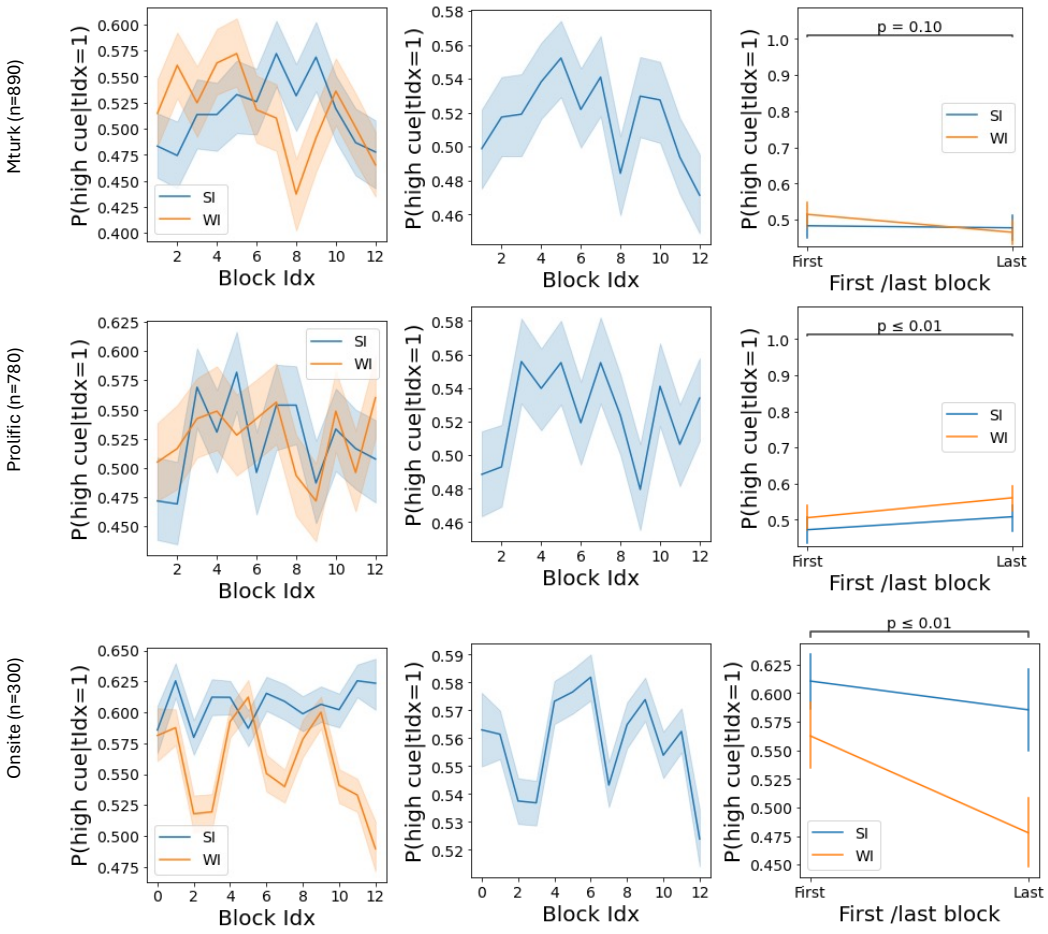


Figure 3.13: Posterior predictive check of the transfer learning effect in terms of exploration bias using the feature RL with forgetting. These graphs correspond to the Fig 11. Similar to the simulation results with the feature RL with forgetting and glia, the frequency of choosing the slot machine with the higher cues was only significantly increasing in simulation using the Prolific data. Moreover, the simulation results on the onsite data showed the significant opposite trend from the actual data (All informative blocks, t-test: t(889)=-1.64, t(779)=2.60, t(299)=-2.86, $p = 0.101, 9.44 * 10^{-3}, 4.30 * 10^{-3}$ for Mturk, Prolific, and Onsite, respectively; Strongly informative blocks only, t-test: t(889)=-0.23, t(779)=1.46, t(299)=-1.19, $p = 0.815, 0.145, 0.236$ for Mturk, Prolific, and Onsite, respectively).

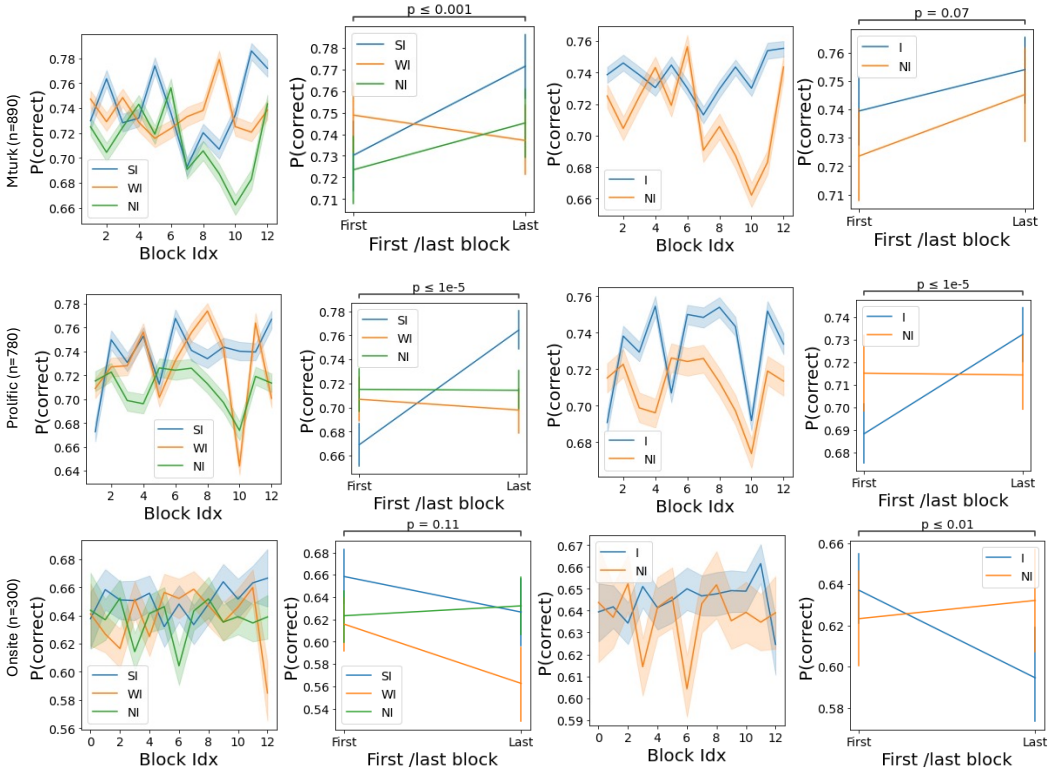


Figure 3.14: Simulated task performance by block indices using the feature RL with forgetting and slow integration. The plots correspond to the fig 6. The simulation results using online datasets showed the increasing performance in informative blocks, in particular strongly informative blocks. (Informative first vs last block, t-test: t(889)=1.84, t(779)=5.20, t(299)=-3.08, $p = 0.066, 2.59 * 10^{-7}, 0.002$ for Mturk, Prolific, and Onsite, respectively; ANOVA on I vs NI: F(1)=0.24, 8.65, 4.68, $p = 0.627, 0.003, 0.031$ for for Mturk, Prolific, and Onsite, respectively. Note that the direction was opposite in the onsite data; Strongly informative first vs last block, t-test: t(889)=3.84, t(779)=8.01, t(299)=-1.61, $p = 1.35 * 10^{-4}, 4.10 * 10^{-15}, 0.110$ for Mturk, Prolific, and Onsite, respectively; ANOVA on SI vs NI: F(1)=1.41, 30.13, 2.34, $p = 0.235, 4.36 * 10^{-8}, 0.126$ for for Mturk, Prolific, and Onsite, respectively. Note that the direction was opposite in the onsite data).

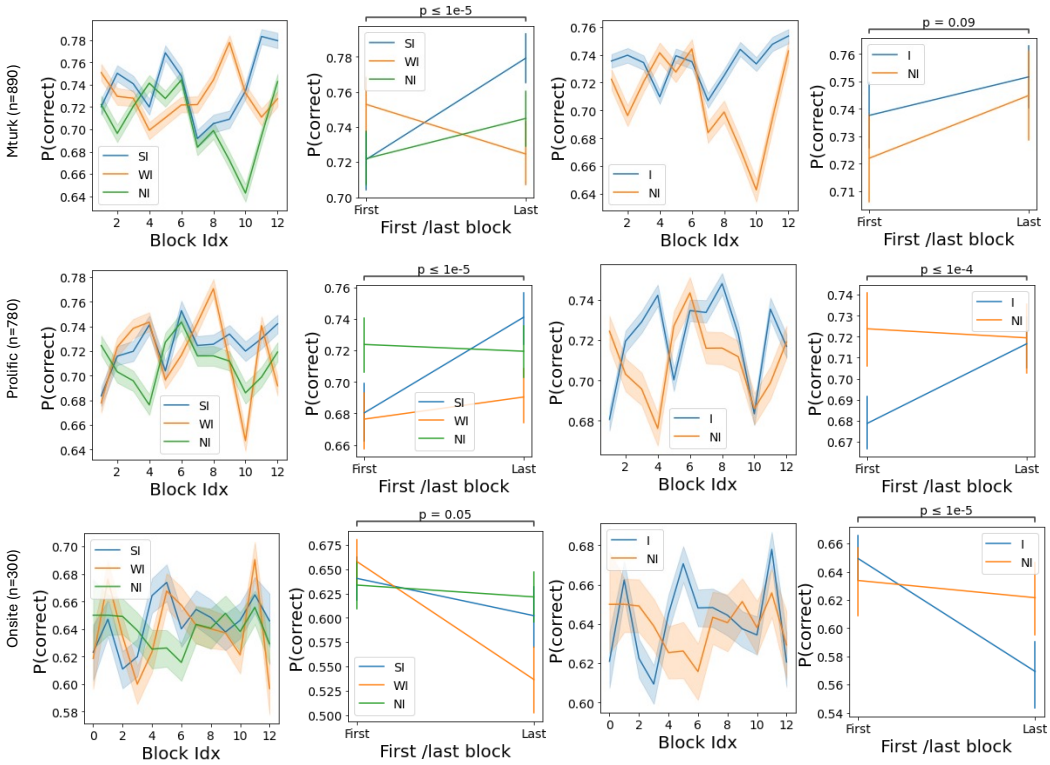


Figure 3.15: Simulated task performance by block indices using the feature RL with forgetting only. The plots correspond to the fig 12. The simulation results using online datasets showed the increasing performance in informative blocks, in particular strongly informative blocks. (Informative first vs last block, t-test: t(889)=1.69, t(779)=4.40, t(299)=-5.71, $p = 0.092, 1.26 * 10^{-5}, 1.82 * 10^{-8}$ for Mturk, Prolific, and Onsite, respectively; ANOVA on I vs NI: F(1)=0.38, 7.84, 8.00, $p = 0.539, 0.005, 4.73 * 10^{-3}$ for for Mturk, Prolific, and Onsite, respectively. Note that the direction was opposite in the onsite data; Strongly informative first vs last block, t-test: t(889)=5.19, t(779)=5.13, t(299)=-1.99, $p = 2.57 * 10^{-7}, 3.76 * 10^{-7}, 0.047$ for Mturk, Prolific, and Onsite, respectively; ANOVA on SI vs NI: F(1)=4.58, 14.19, 0.94, $p = 0.032, 1.68 * 10^{-4}, 0.333$ for for Mturk, Prolific, and Onsite, respectively. Note that the direction was opposite in the onsite data).

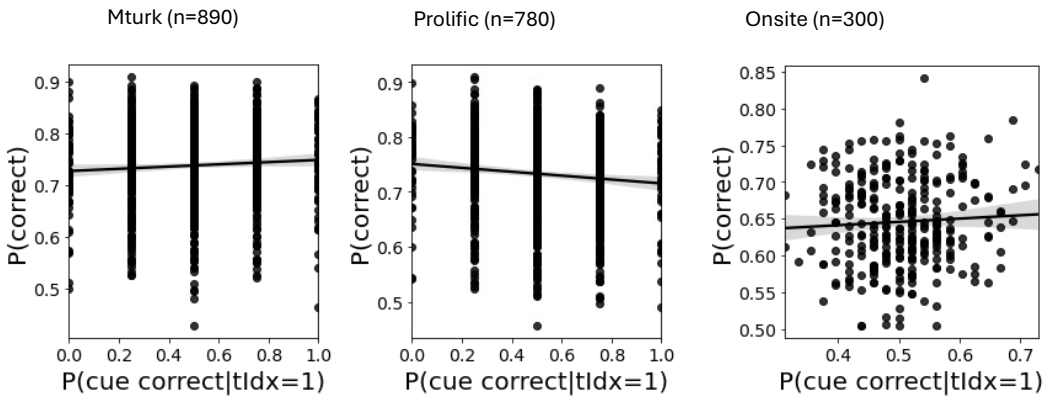


Figure 3.16: Simulated correlations between the exploration bias and task performance in informative blocks using the feature RL with forgetting and slow integration model. The figure corresponds to Fig.3.7. The posterior predictive check results did not show significant positive correlations between the two metrics (Person correlation: $r = 0.067, -0.104, 0.059$, $p = 0.045, 0.004, 0.328$ for Mturk, Prolific and Onsite data respectively).

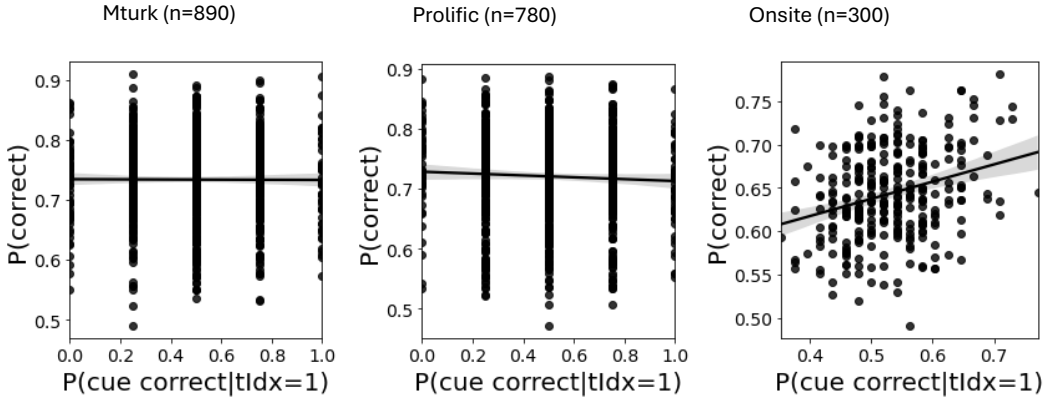


Figure 3.17: Simulated correlations between the exploration bias and task performance in informative blocks using the feature RL with forgetting only. The figure corresponds to Fig3.16. The posterior predictive check results did not show significant positive correlations between the two metrics except the onsite data simulation (Person correlation: $r = -0.004, -0.050, 0.280$, $p = 0.901, 0.166, 8.68 * 10^{-7}$ for Mturk, Prolific, and Onsite data, respectively)

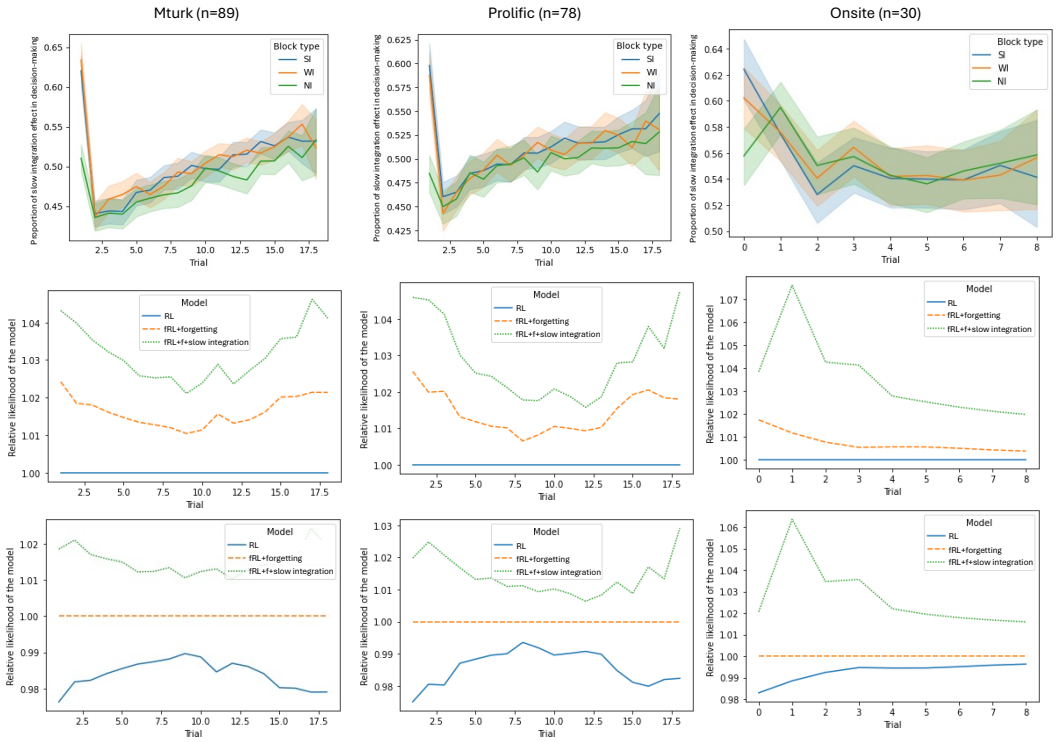


Figure 3.18: Effect of slow integration in initial exploration bias. The first row on the plots shows proportion of slow integration output in the difference of decision variables which are the sums of the outputs of neuron and slow integration component. The second and third row show the likelihood of each trial by models. The second row normalize the likelihood of each trial using the likelihood from the RL model and the third row normalize the likelihood using the feature RL with forgetting.

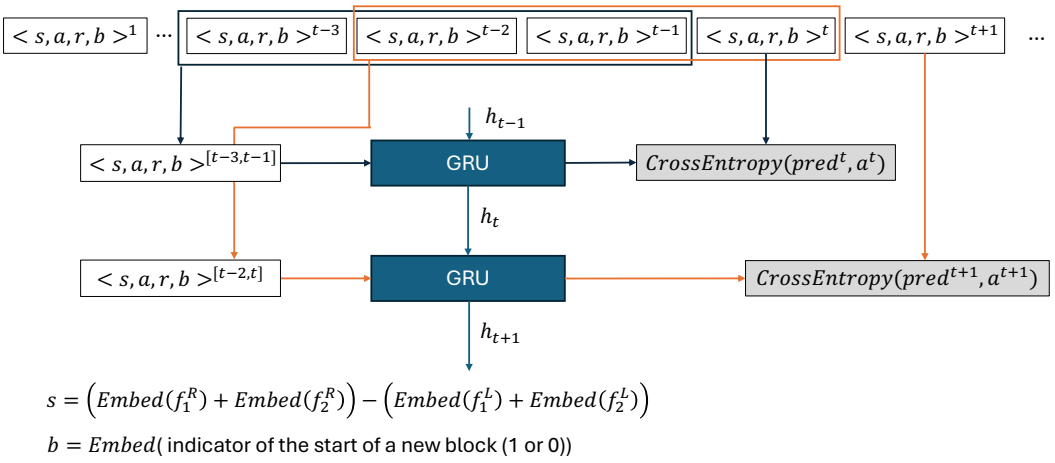


Figure 3.19: Deep neural network architecture. The neural network takes the embedding vectors of state, action, reward and a binary indicator that represent the start of a new block of previous certain number of trials and predict the current trial's action. The example plot shows when previous 3 trials were used to predict the action. The network was optimized to minimize the cross entropy between the action prediction and the actual action at each trial. The state representation of each trial was the difference between the sum of the embeddings of right slot machine features and left slot machine features.

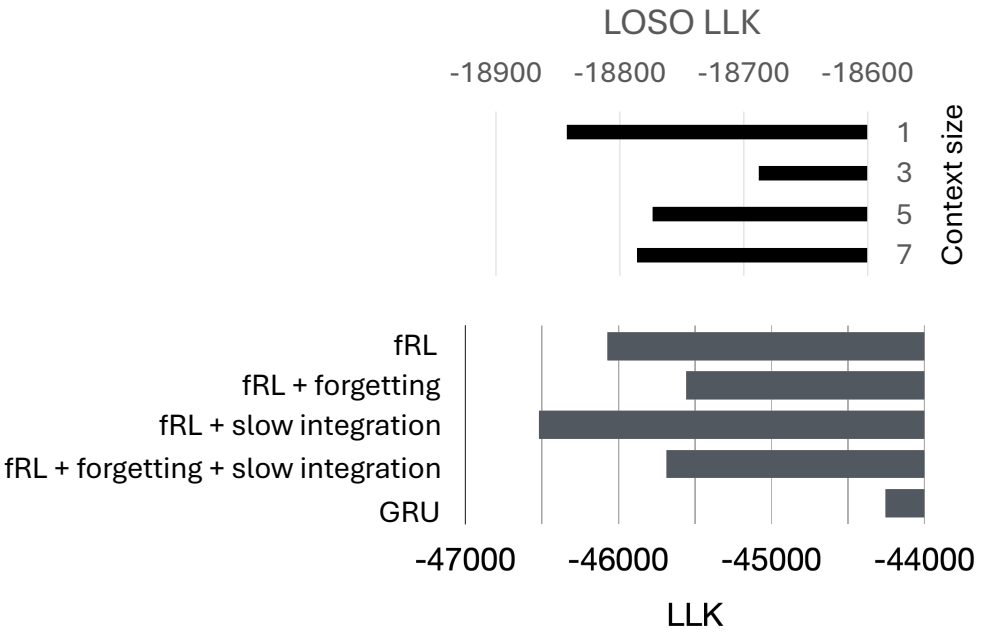


Figure 3.20: Model comparison of the deep neural network model. The upper panel shows the network performance tested on the MTurk data by the length of context of the input. Embedding size 16, hidden layer size 256 and 3 layered GRU was used. Using 3 previous trials showed the best performance in terms of leave-one-subject-out(LOSO) log likelihood. The lower panel shows the neural network model performance compared to the RL models sum over all three datasets. It is noteworthy that the neural network model used leave-one-subject-out log likelihood while other models' metrics are log likelihood on the data which were used to optimize the hyper parameters.

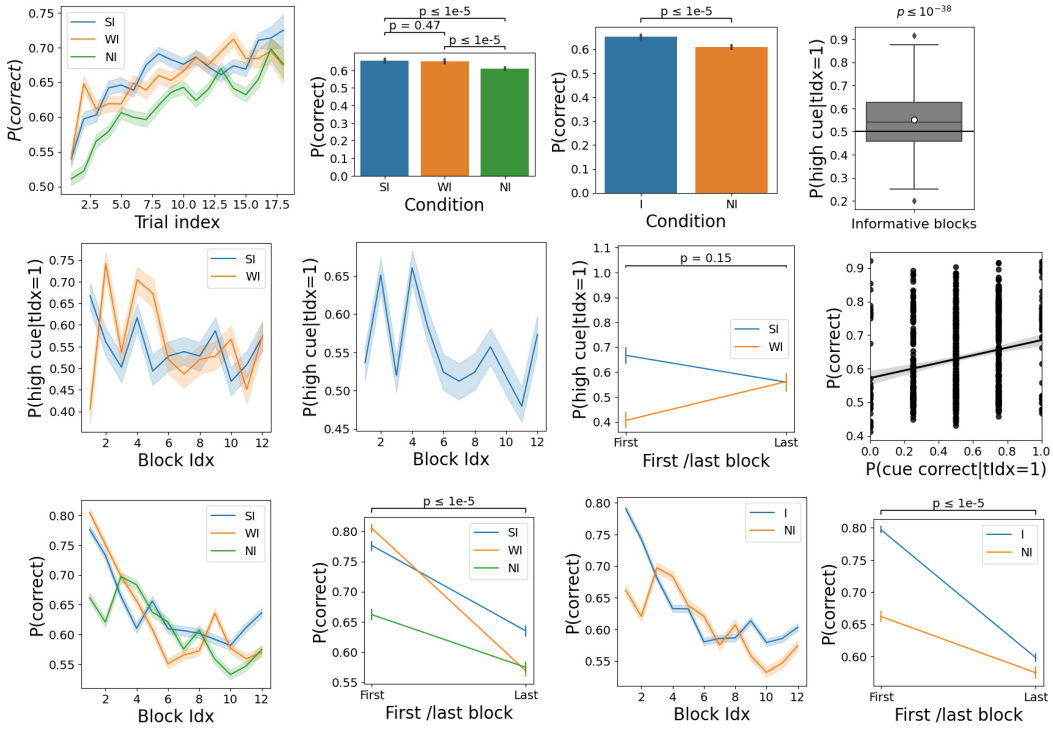


Figure 3.21: Posterior predictive check on Mturk data using the RNN model (n=89*10). The model was able to reproduce the actual initial exploration bias and performance differences between conditions (the first row, t-test paired, SI vs NI t(889)=1.18, $p = 4.73 * 10^{-30}$; WI vs NI t(889)=9.89, $p = 5.94 * 10^{-22}$; SI vs NI t(889)=0.72, $p = 0.472$; I vs NI t(889)=13.00, $p = 1.62 * 10^{-35}$; frequency of choosing the high cue slot machine in the initial trial t(889)=13.74, $p = 4.19 * 10^{-39}$) However, the simulated exploration bias in the later blocks were not increasing in the latter blocks (the second row, t-test paired, all informative blocks t(889)=1.43, p=0.154). Moreover, the exploration bias in the strongly informative blocks showed the opposite trend (t(889)=-4.86, $p = 1.38 * 10^{-6}$). The positive correlation between the exploration bias and the performance were reproduced from the simulations (r=0.218, $p = 4.44 * 10^{-11}$). However, the RNN model was failing in capturing the increasing performance level in the later blocks and rather generated the opposite trends(the last row, First vs Last all informative blocks t-test paired, t(899)=-50.08, $p \leq 1 * 10^{-5}$; ANOVA on I vs NI, F(1)=260.36, $p = 1.91 * 10^{-58}$; First vs Last strongly informative blocks t-test paired, t(889)=-26.62, $p = 2.78 * 10^{-154}$; ANOVA on SI vs NI, F(1)=47.74, $p = 4.92 * 10^{-12}$).

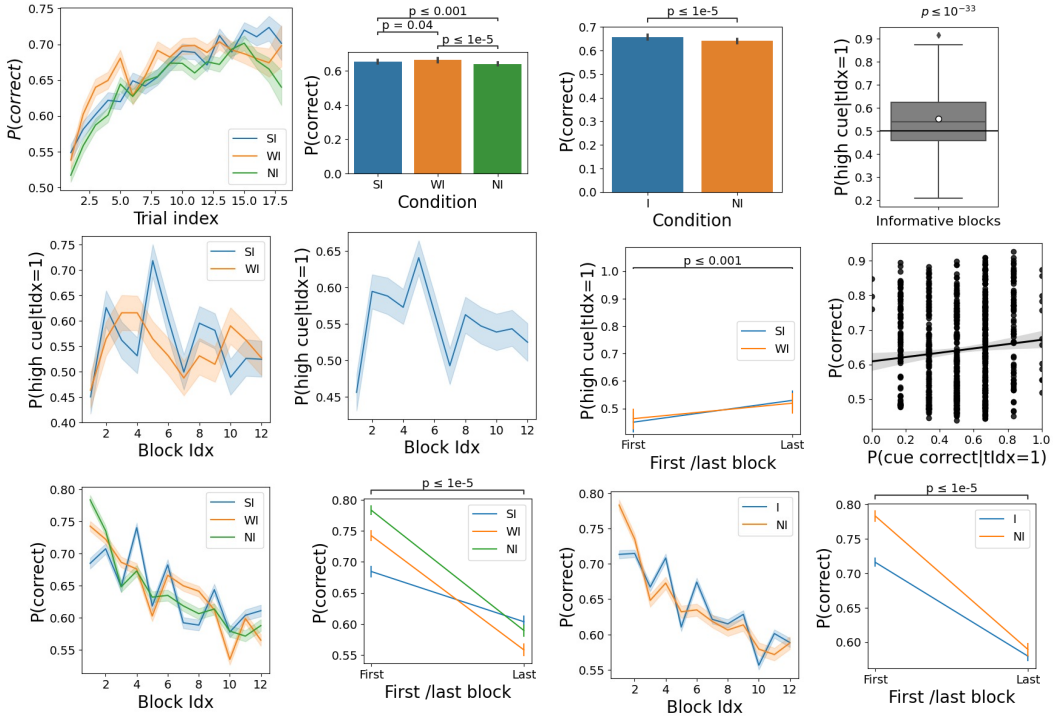


Figure 3.22: Posterior predictive check on Prolific data using the RNN model (n=78*10) .The model was able to reproduce the actual initial exploration bias and performance differences between conditions (the first row, t-test paired, SI vs NI t(779)=3.45, $p = 5.33 * 10^{-4}$; WI vs NI t(779)=5.27, $p = 1.77 * 10^{-7}$; SI vs NI t(779)=-2.02, $p = 0.044$; I vs NI t(779)=4.98, $p = 7.69 * 10^{-7}$; frequency of choosing the high cue slot machine in the initial trial t(779)=4.98, $p = 7.69 * 10^{-7}$). Moreover, the simulated exploration bias in the later blocks were increasing in the latter blocks (the second row, t-test paired, all informative blocks t(779)=3.72, $p = 2.05 * 10^{-4}$; strongly informative blocks only, t(779)=3.01, $p = 2.68 * 10^{-3}$). The positive correlation between the exploration bias and the performance were reproduced from the simulations as well (r=0.097, p=0.006). However, the RNN model was failing in capturing the increasing performance level in the later blocks and rather generated the opposite trends(the last row, First vs Last all informative blocks t-test paired, t(779)=-29.93, $p = 1.93 * 10^{-193}$; ANOVA on I vs NI, F(1)=61.21, $p = 5.24 * 10^{-15}$; First vs Last strongly informative blocks t-test paired, t(779)=-13.20, $p = 1.16 * 10^{-39}$; ANOVA on SI vs NI, F(1)=179.19, $p = 8.59 * 10^{-41}$).

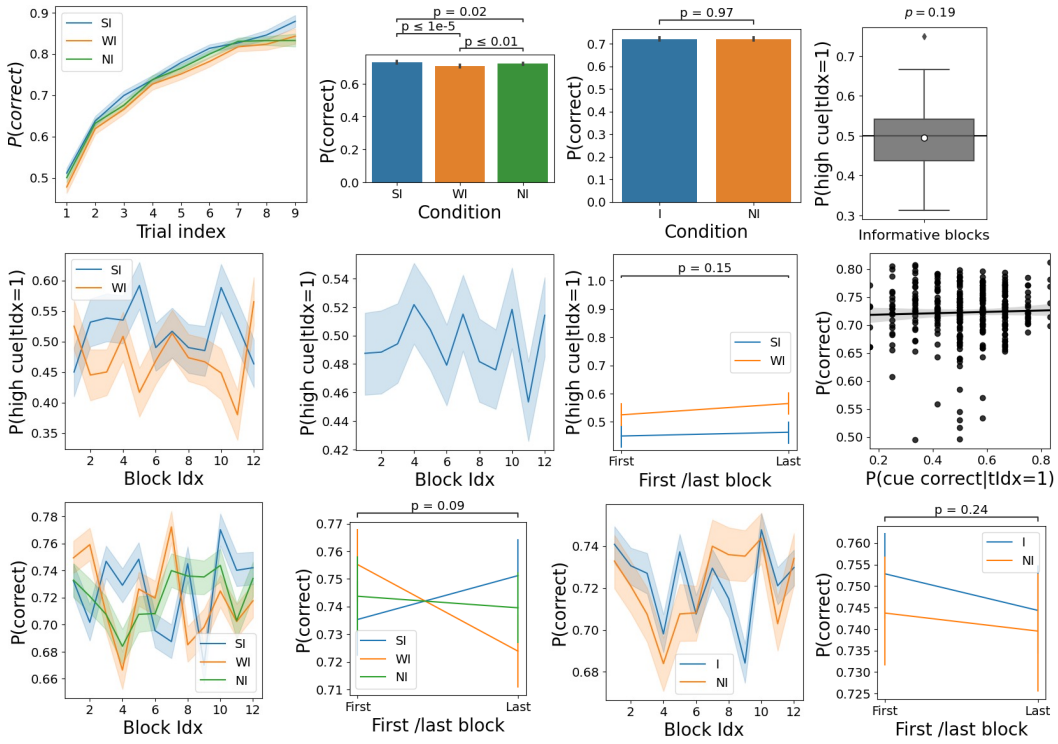


Figure 3.23: Posterior predictive check on Onsite data using the RNN model (n=30*10) .The model had a trouble in reproducing the actual initial exploration bias and performance differences between conditions (the first row, t-test paired, SI vs NI t(299)=2.26, $p = 0.024$; WI vs NI t(299)=-2.79, $p = 0.006$; SI vs NI t(299)=5.28, $p = 2.50 * 10^{-7}$; I vs NI t(299)=-0.04, $p = 0.966$; frequency of choosing the high cue slot machine in the initial trial t(299)=-1.31, $p = 0.189$). Moreover, the simulated exploration bias in the later blocks were not significantly increasing in the latter blocks (the second row, t-test paired, all informative blocks t(299)=1.43, $p = 0.154$; strongly informative blocks only, t(299)=0.49, $p = 0.632$). The correlation between the exploration bias and the performance were not significant in simulations as well (r=0.037, p=0.528). Furthermore, the RNN model was failing in capturing the increasing performance level in the later blocks (the last row, t-test paired, First vs Last all informative blocks t(299)=-1.18, $p = 0.236$; ANOVA on I vs NI, F(1)=0.14, $p = 0.713$). However, when we focused on the strongly informative blocks only, we could observe a greater increase in performance compared to noninformative blocks (First vs Last strongly informative blocks t-test paired, t(299)=1.69, $p = 0.091$; ANOVA on SI vs NI, F(1)=6.43, p=0.001).

Chapter 4

NEURO-COMPUTATIONAL MECHANISM OF INFLUENCE OF AFFORDANCE IN VALUE-LEARNING

The following chapter is adapted from Yi and O'Doherty, 2023 and modified according to Caltech Thesis format.

Sanghyun Yi and John P. O'Doherty. Computational and neural mechanisms underlying the influence of action affordances on value learning. *BioRxiv*, 2024. doi: https://doi.org/10.1101/2023.07.21.550102.

4.1 Abstract

When encountering a novel situation, an intelligent agent needs to find out which actions are most beneficial for interacting with that environment. One purported mechanism for narrowing down the scope of possible actions is the concept of action affordance. Here, we delve into the neuro-computational mechanisms accounting for how action affordance shapes value-based learning in a novel environment by utilizing a novel task alongside computational modeling of behavioral and fMRI data collected in humans. Our findings indicate that rather than simply exerting an initial or persistent bias on value-driven choices, action affordance is better conceived of as an independent system that concurrently guides action-selection alongside value-based decision-making. These two systems engage in a competitive process to determine final action selection, governed by a dynamic meta controller. We find that the pre-supplementary motor area and anterior cingulate cortex plays a central role in exerting meta-control over the two systems while the posterior parietal cortex integrates the predictions from these two controllers of what action to select, so that the action-selection process dynamically takes into account both the expected value and appropriateness of particular actions for a given scenario.

4.2 Introduction

In order to interact successfully with situations as they occur in the world, humans and other animals need to select particular actions from a very large set of possible actions based on which actions are most appropriate to the situation. One fundamental guiding principle for action selection, is that actions should be selected based on their expected value, that is by how much a particular action might in-

crease an individual's access to rewards, or decrease potential exposure to aversive outcomes (Glimcher and Fehr, 2013; Rangel et al., 2008; Sanfey et al., 2006). A large literature has shed light on the neural and computational underpinnings of value-based action selection, including reinforcement-based mechanisms for learning which actions to select based on the future expected rewards they engender (Montague et al., 1996; Schultz et al., 1997; O'Doherty et al., 2003). However, when encountering a stimulus for the first time in a particular context, the value of that stimulus is largely unknown and the brain needs to have a strategy to reduce the extremely large set of possible actions that could be selected to a tractable set of possible actions that could form the basis of subsequent trial and error learning. One putative mechanism for this is visual affordance (Gibson, 2014; Cisek, 2007; Pezzulo and Cisek, 2016). Action affordances are features of stimuli in the world that suggest the appropriateness of particular actions, which can make the action selection problem more tractable. For example, a computer keyboard might suggest pressing or poking actions, a watering can might suggest a grabbing or clenching action, a pair of chopsticks might suggest a pinching action.

It has previously been found that affordance automatically potentiates particular actions that are compatible with properties an object or scene presented (Ellis and Tucker, 2000; Grèzes et al., 2003; Zhang et al., 2021; Cisek, 2007). Particularly, it has been suggested that the selection of visually guided actions is supported directly by affordance and that this affordance mechanism is in turn biased by other decision variables such as the value of choice options (Pastor-Bernier and Cisek, 2011; Cisek and Pastor-Bernier, 2014). However, the role that affordances might play in actually guiding learning during value-based decision-making is essentially unknown. A natural hypothesis is that affordance acts as an initial prior on value learning by, for example, subtly inflating the value of the afforded action during initial choice behavior so as to guide exploration, or alternatively, affordance might act as a constant yet moderate tug on action selection, persistently increasing the likelihood that an affordance-based action is selected in spite of the effects of value learning. Yet another possibility is that affordance-based choice and value-based choice operate as independent systems, competing for access to behavior. As we will see, our findings rather surprisingly support this latter possibility and suggest a role for a dynamic arbitration between independent expert systems implementing affordance and value-based choice.

To answer these questions we designed a novel behavioral task to probe the ef-

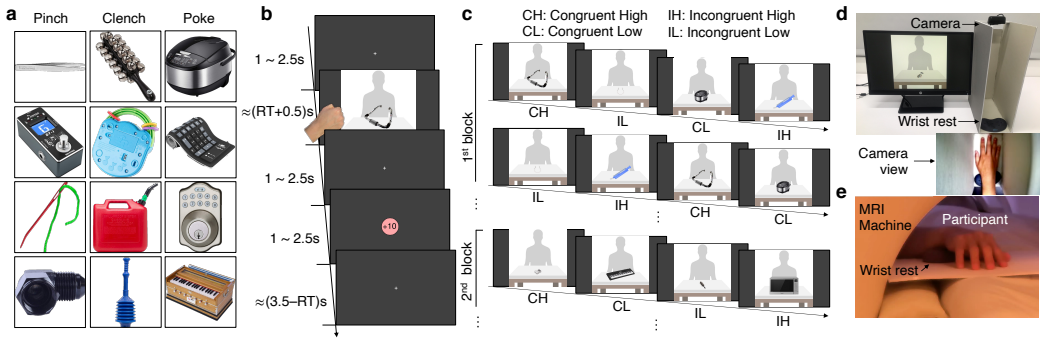


Figure 4.1: Decision-making task using naturalistic hand gestures and affordance-conferring stimuli. (a) Example stimuli associated with particular hand gesture affordances, determined from a separate online survey. (b) Example trial structure. A stimulus was displayed until an executed naturalistic hand gesture was registered. Each object was shown with a human silhouette background to give a naturalistic sense of the size of the object displayed. (c) Example block structure. Four objects were shown in each block and each object was associated with one of the 4 experimental conditions: congruent high (CH), incongruent high (IH), congruent low (CL), and incongruent low (IL) (see Methods for the details). (d) Experiment setting of the behavioral task. Participants made hand gestures inside the apparatus with a camera on top of it which sent a video stream to a server that classified hand movements into one of the three allowed hand gesture types in real time. (e) Implementation of the fMRI experiment. Participants lied down on the bed and mimed hand gestures on the plane which was placed on top of their bodies.

fects of affordance on decision-making. We adopted a computational model-driven approach in which we specified a series of computational models to capture the different possible mechanisms by which affordance affects value learning, which we then systematically tested against human behavior. Further, we then measured brain activity with fMRI in order to identify the neural correlates of these putative computational processes. The overarching goal of the present study is thus to investigate how action affordances might potentially interact with value-based action-learning at behavioral, computational and neural levels.

In the behavioral task, participants were presented with pictures of an array of different visual objects. These objects were pre-selected to have specific action affordances based on the actions a separate group of participants rated to be most appropriate for a given object, out of three possible actions: pinch, poke, or clench. In the task, when participants saw an object, they could make one of those three actions in response, using a naturalistic (right) hand gesture (Fig. 4.1; Methods). We implemented a computer vision-based approach for the behavioral experiment,

which classified a live video stream of the participant's hand movements into one of the three gesture classes in real time (Fig. 1d; Methods). For a given object, one of these actions was associated with a higher probability of reward (winning money), while the other actions were associated with a lower probability of winning money. So for example, upon seeing a picture of a watering can, if the participant makes a poking gesture they would be most likely to win money, whereas if they make a pinching gesture they are less likely to win. Four objects (out of a set of 24 or 48 depending on the study), were encountered in a particular block of trials of 80 duration on average. Participants are instructed that their goal is for each object to select the action associated with the greatest amount of reward in order to obtain as many rewards as possible. Thus for each object, participants had to learn which action to select in order to maximize rewards. Crucially, for some objects, the most rewarded action was also the action that had the greatest affordance (congruent), while for other objects, the most rewarded action was not the action that had the greatest affordance (incongruent) (Fig. 1c). Furthermore, we included an orthogonal manipulation that controlled the reward probability such that half of the stimuli were associated with a high reward probability condition and the other half with a low reward probability condition, with the latter offering half the reward probability of the former (though in both conditions one of the actions available still had a higher reward probability than the other two). Through these manipulations, we could therefore assess the role that action affordance plays in guiding action selection alongside expected value, as well as characterizing how these two processes might interact. We ran two studies using this paradigm. An initial behavioral study (n=19) was followed by an fMRI study (n=30). Together these studies allowed us to investigate the behavioral effects of affordance on value-based learning and to then utilize computational modeling to gain insight into the computational mechanisms underpinning these interactions. Finally, our fMRI study allowed us to characterize the neural mechanisms implementing these computations.

4.3 Results

Effects of affordance on reaction times and action selection

We first ascertained to what extent did the affordance properties of a stimulus influence response reaction times (RTs). We expected that choice of an action compatible with the dominant action affordance for an object would be associated with shorter RTs than choice of an affordance incompatible action regardless of the presence of ongoing value learning (Ellis and Tucker, 2000; Grèzes et al., 2003;

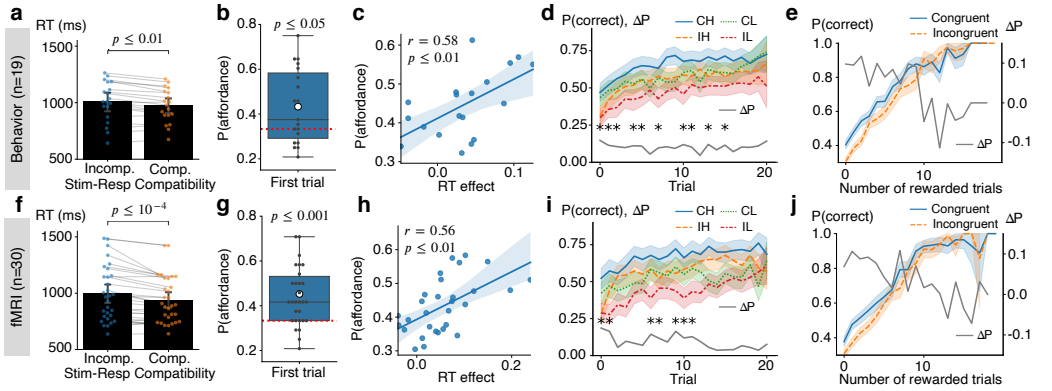


Figure 4.2: Behavioral effects of affordance during value learning and decision making. Upper-row plots show results from the behavioral study while the lower-row plots are from the fMRI study. (a and f). RTs were faster when the response was compatible with the affordance compared to when it was incompatible (paired-t tests; $t(18) = 3.62$ for the behavioral, $t(29) = 4.68$ for the fMRI experiment). RTs for each trial were averaged for each participant and trial types and the dots represent individual participants. (b an g) Initial responses to objects were significantly biased toward affordance-compatible actions. (t tests; $t(18) = 2.62$ for the behavioral, $t(29) = 4.48$ for the fMRI experiment). Dots represent the average frequencies for each participant and the red line is the chance level 1/3. (c and h) Correlations between the RT effect and the frequency of choosing affordance-compatible actions. (Pearson r) Dots represent each participant. (d and i) Learning curves averaged across blocks by experimental conditions. The star annotation presents the statistical significance of those difference between congruent and incongruent conditions in each trial (permutation tests; 5000 iterations each; *: $p < 0.05$ Bonferroni corrected) (e and j) Learning slopes analyzed as a function of the number of rewarded trials which were averaged across blocks. The gray lines in d,i,e, and j show the choice accuracy difference between congruent and incongruent conditions. All the errorbar shows 95% interval of estimated statistics.

Zhang et al., 2021). RTs were defined as the interval between the stimulus onset and initiation of movement, and were indeed found to be significantly shorter when making affordance-compatible actions compared to affordance-incompatible actions in both the behavioral and fMRI data even when participants were freely exploring the environment for finding the most rewarding hand gesture (Figs. 2a,f; for details on the reaction time measurement, see Methods). Moreover, the difference in RT remained consistent throughout. The average RT difference between the first 5 trials and the last 5 trials was not significantly different in either dataset (t(18)=-1.329, p=0.20 in the behavioral and t(29)=0.7592, p=0.45 in the fMRI data). We also found that the RTs were influenced by value-learning, in that participants were faster to

respond to stimuli with higher expected value than stimuli with lower expected value (Fig. 4.6). Thus, to disambiguate the effects of affordance on RTs from other possible confounding variables such as experimental condition, action type, choosing the most-rewarding action, and the progress of learning across trials, all of which are associated with value learning, we conducted mixed-effect linear regression which included each of these effects as potential confounding covariates. Even after accounting for these confounds, the effect of affordance on RT remained significant (Table 4.1; $\beta = -0.026, z = -2.66, p \leq 0.01$ for selecting affordance-compatible actions in the behavioral experiment, and $\beta = -0.029, z = -3.02, p \leq 0.01$ in the fMRI experiment, which indicates that RTs for affordance-compatible actions were about 2.7% shorter). These results not only confirm that our experimental paradigm and stimulus set reliably induce affordance-related response preparation effects, but also reveal that the reaction time effect due to stimulus-response compatibility persists regardless of the effects of value-learning.

Next, we aimed to examine the effect of affordance on which action was chosen on a given trial. Specifically, we hypothesized that the affordance associated with a particular object would bias choices in favor of the afforded action, independently of the expected value of that action. To test for this, we analyzed the initial responses participants made to each object, as those are the actions not affected by value learning. The probability of selecting affordance-compatible actions as the initial response was significantly higher than chance (1/3) in both datasets (Figs. 2b and 2g). Furthermore, those initial actions were biased toward the afforded action that each object was selected to confer based on the initial affordance ratings we previously obtained in a separate sample (Table 4.2; χ^2 test; null hypothesis was the probability distribution of choosing each action type independent of the affordance of objects calculated using the choice data; $\chi^2(2, N = 304) = 9.33, p \leq 0.01$ for pinch, $\chi^2(2, N = 304) = 16.02, p \leq 0.001$ for clench, $\chi^2(2, N = 304) = 21.12, p \leq 10^{-4}$ for poke in the behavioral experiment; $\chi^2(2, N = 240) = 17.91, p \leq 0.001$ for pinch, $\chi^2(2, N = 240) = 16.18, p \leq 0.001$ for clench, $\chi^2(2, N = 240) = 16.54, p \leq 0.001$ for poke in the fMRI experiment). Moreover, the initial selection bias remained constant throughout the task for all new objects introduced over the course of the experiment (Figs. 7a and 7c for details). This observation rules out an alternative explanation regarding the initial selection bias which posits that participants might have gradually inferred that the structure of the task is such that affordance-compatible actions are the most-rewarding action in half of the trials, and thus, the baseline probability for selecting affordance-compatible action would

be 1/2 rather than 1/3. According to this hypothesis, the initial selection bias toward affordance-actions would be expected to increase as the task progresses as participants become more aware of the task structure with increasing experience. However, the actual data contradict this notion by showing that the initial selection bias remained stable and did not increase as the task progressed in both the behavioral and fMRI studies.

On top of that, we observed significant positive correlations across participants between these two distinct affordance-related bias effects on choice. We computed the degree of RT bias as the additional reaction time for executing affordance-incompatible actions relative to affordance-compatible actions (RT effect = $\frac{RT_{incomp}-RT_{comp}}{RT_{comp}}$) for each participant and compared it to the frequency of choosing affordance-compatible actions throughout the task. Significant correlations were observed between the two metrics in both datasets (Figs. 2c,h), and moreover, the RT effect positively correlated with the frequency of selecting the affordance-compatible action as the initial response to each object (Figs. 7b,d; Pearson $r = 0.41, p = 0.08$ in the behavioral, Pearson $r = 0.58, p \leq 0.001$ in the fMRI experiment). These findings suggest that both forms of affordance-related biases on choice behavior are related and have a shared substrate.

Affordance influences value learning

The pronounced action selection bias toward the afforded action also influenced value-learning and contributed to the choice accuracy difference between congruent and incongruent conditions. To be specific, selection of the most rewarding action for each object was significantly greater in the congruent than in the incongruent conditions ($t(18) = 2.50, p \leq 0.05$ in the behavioral, $t(29) = 3.25, p \leq 0.01$ in the fMRI experiment). However, the bias toward selecting actions based on affordance was most evident in the early phase of the interaction with each specific object and diminished across subsequent trials involving that specific object. As demonstrated in Figs. 2d,i, statistical tests on the difference in choice accuracy between congruent and incongruent conditions in each trial revealed that the affordance effect was significant on early trials within a block but became less pronounced as learning progressed in both experiments.Because the exploration of choice options is biased by action affordance, participants are likely to need more trials to experience positive outcomes in incongruent than in congruent conditions. Therefore, the choice accuracy difference between congruent and incongruent conditions might be due to an affordance bias operating on the choice process rather than reflecting the effects

of impaired value learning.

To analyze the influence of affordance on learning further, we examined the choice accuracy difference between congruent and incongruent conditions as a function of the number of rewarded trials previously encountered. The analysis indicates that learning slopes were actually steeper in the incongruent conditions, as evidenced by a decreasing pattern in the difference between the two learning slopes from each condition (Figs. 2e,j). Mixed-effect GLM analyses confirmed that the choice accuracy difference between the conditions is decreasing and that the incongruent condition has a steeper slope than the congruent condition (Fixed-effect coefficients for the gradient of the choice accuracy difference ΔP : $\beta = -0.010, z = -3.16, p \leq 0.01$ in the behavioral, $\beta = -0.009, z = -2.97, p \leq 0.01$ in the fMRI experiment; Table 4.3).

As previously mentioned, two possible explanations might account for this effect: first, it is possible that participants go through a longer exploration phase between receiving rewards in the incongruent conditions, which might support counterfactual learning during exploration, leading to steeper learning (Camille et al., 2004; Lohrenz et al., 2007). In such scenarios, using a reinforcement learning (RL) model that incorporates negative learning signals to update the values of unchosen actions in the trials without rewards would be sufficient to account for the learning effect, as such a model could capture counterfactual learning. Alternatively, it is plausible that behavioral adaptation based on reward history is more sensitive in the incongruent conditions. This could result from having a higher learning rate in incongruent conditions, or it could be due to the exertion of a higher level of cognitive control on incongruent conditions resulting in decisions that are more heavily reliant on learned values (Verguts and Notebaert, 2008; Botvinick and Braver, 2015; Braem et al., 2019). As we will see, the computational modeling we implement supports the notion that cognitive control is allocated to balance the influence of affordance and value-learning in order to govern task performance (see below).

We also plotted the learning slope in high and low conditions as the function of the number of rewarded trials previously encountered. However mixed-effect GLM analyses revealed that the difference between high and low conditions does not decrease (Fixed-effect coefficients for the gradient of the choice accuracy difference ΔP : $\beta = -0.006, z = -1.30, p = 0.20$ in the behavioral, $\beta = -0.010, z = -0.66, p = 0.51$ in the fMRI experiment; Fig. 4.8).

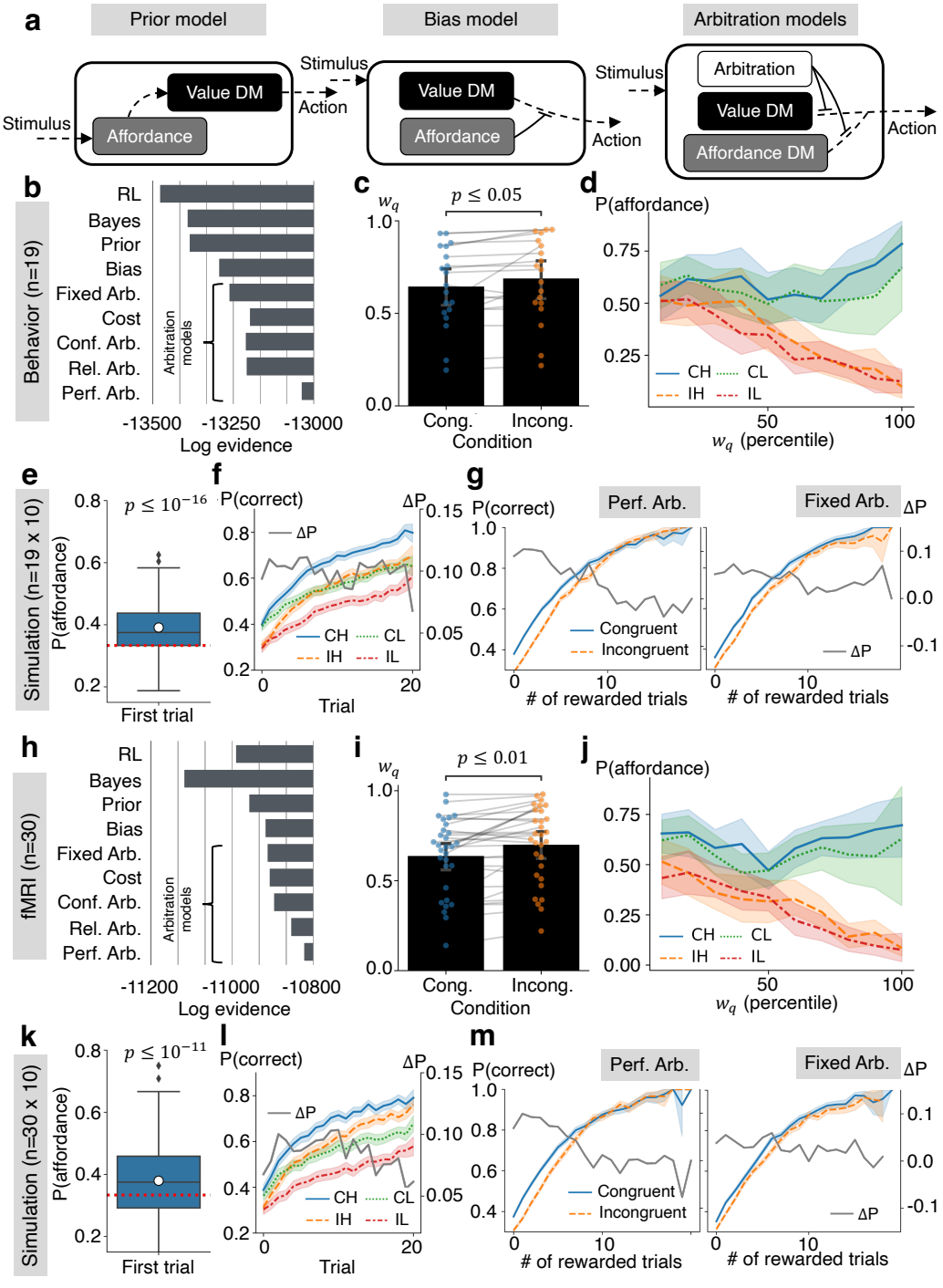

a
Prior model
Bias model
Arbitration models
Stimulus
Value DM
Affordance
Action
Arbitration
Affordance DM
b
Behavior (n=19)
RL
Bayes
Prior
Bias
Fixed Arb.
Cost
Conf. Arb.
Rel. Arb.
Perf. Arb.
Arbitration models
-13500
-13250
-13000
Log evidence
c
w_q
$p \leq 0.05$
Cong.
Incong.
Condition
d
P(affordance)
CH
CL
IH
IL
w_q (percentile)
e
Simulation (n=19 x 10)
$p \leq 10^{-16}$
P(affordance)
First trial
f
P(correct)
ΔP
Trial
g
Perf. Arb.
Fixed Arb.
Congruent
Incongruent
# of rewarded trials
h
fMRI (n=30)
-11200
-11000
-10800
i
$p \leq 0.01$
j
k
Simulation (n=30 x 10)
$p \leq 10^{-11}$
l
m

Figure 4.3: Computational model comparison, and the simulation results from the best model. Figs. 3b-g are from the behavioral study and their corresponding simulations while the Figs. 3h-m show the results from the fMRI study and their corresponding simulations. (a) Schematics of candidate computational mechanisms (Value DM: value-based decision-making; Affordance DM: affordance-based DM). See the main text and methods for details. (b and h) Log model evidence of the compared models. (c and i) Arbitration weights were affected by the affordance-value congruency so that the value-based decision-making was more favored in incongruent conditions. (paired-t tests; $t(18) = 2.28$ for the behavioral, $t(29) = 3.59$ for the fMRI experiment). Arbitration weights were calculated using the performance-based arbitration model for each trial and were averaged for each participant and condition. (d and j) Frequency of choosing affordance-compatible actions in the actual choice data is a decreasing function of the arbitration weight on the value-based decision-making. The arbitration weights were transformed into the percentiles within each participant. The tendency of choosing affordance-compatible actions more when the arbitration mechanism favors affordance-based decision-making was only evident in incongruent conditions as the responses based on affordance and value were indistinguishable in congruent conditions. (e and k) Simulated initial choice bias toward affordance-compatible action for each object. The performance-based arbitration models with individually estimated parameters were simulated 10 times each. (t tests; $t(189) = 9.29$ for the behavioral, $t(299) = 7.21$ for the fMRI experiment) (f and l) Simulated learning curves and their differences. (g and m) Simulated learning slopes and their differences as a function of the number of rewarded trials. All the error-bar shows 95% interval of estimated statistics.

Dynamic meta-level control merging affordance-based and value-based decision-making best explains behavior

To investigate the underlying computational mechanism responsible for these behavioral effects, we implemented various computational models, fit those to participant's behavioral data and then performed a formal model comparison (see Methods for details; Fig. 3a). We applied an RL model to capture value learning. We then modeled the degree of affordance assigned to each action for each object using the affordance-compatibility scores for the hand gestures for each object provided by the participants themselves after the completion of the value-based choice task.

One simple way in which affordance might influence choice is via a constant action-selection bias – essentially providing a constant push toward choosing the afforded action on each trial over and above other considerations such as value. Another possibility is that affordance acts as an initial prior operating on the initial value of the afforded action. If the afforded action is presumed to have a higher initial value than the other actions, this could produce a bias toward choosing that action more

often at the beginning of a block. We implement both of these possible biases in separate RL models, either as a bias in the decision variable (bias model), or as a bias in the initial values assigned to an action before the onset of reinforcement-learning (prior model). We also tested the possibility that value learning was supported by Bayesian inference, rather than RL (Bayes model) (Niv et al., 2015).

In addition, based on the observation that the steeper learning slope in incongruent conditions (Figs. 2e,j) could be captured by cognitive control resulting in a greater focus on value in the incongruent conditions, we explored models inspired by the concept of a mixture of experts (O'Doherty et al., 2021b). In this class of models, we assumed that two different decision-making systems are concurrently making predictions about the appropriate action for a given object. The first system is an affordance-based decision-making system, which simply makes choices in a manner proportional to the degree of affordance attached to particular actions, while the second system is a value-based decision-making system, by which actions get selected based on their learned expected values. Additionally, we assumed a meta-level controlling mechanism that arbitrates between the two systems and mediates the influence of the two systems in selecting actions. Four candidate arbitration mechanisms were tested.

The first model assumes that the outputs from the two systems are mixed with a fixed weight, which is conceptually similar to the bias model (Fixed arbitration model). We also tested a model that assumed affordance acts as a cost that makes the outcome from selecting affordance-incompatible actions less rewarding, thereby hindering selection of affordance-incompatible actions. This component was added on top of the fixed arbitration model (Cost model) (Shenhav et al., 2013). The second type of arbitration model gave a boost in control to the value-based decision-making system, when the two systems made conflicting predictions, which mainly happened in incongruent conditions (Conflict-based arbitration model) (Matsumoto and Tanaka, 2004; Botvinick et al., 2001; Yeung et al., 2004a). The third model type assigned a greater degree of control to the system that had the lower level of prediction errors, or a higher level of reliability in its predictions, so that participants could minimize the uncertainty in predicting the values of each action. The reliabilities of each system were estimated using the absolute value of reward prediction errors (RPE) or affordance prediction errors (APE), which is the difference between the outcome and the affordance compatibility of the chosen action (Reliability-based arbitration model) (Lee et al., 2014a; Charpentier et al., 2020; O'Doherty et al., 2021b).

The last type of model allowed for a larger influence from a system that had higher expected outcomes when the decision maker followed that particular system in making a decision (Botvinick and Braver, 2015; Shenhav et al., 2013). By doing so, the decision maker can maximize the outcome they can collect (see Methods for details). The performance of each system, or the expected outcome by following the specific decision-making system to a given object, was estimated using a method called inverse propensity scoring (Precup et al., 2000; Horvitz and Thompson, 1952), which was implemented in a form of delta rule supported by performance prediction errors (PPE) within each system. (Performance-based arbitration model; see Methods for details).

As illustrated in Figs. 3b,h, the performance-based arbitration model was found to best explain the actual choice data in terms of group-level log model evidence. Additionally, the Bayesian model selection results showed that the posterior model frequency and protected exceedance probabilities were in agreement with the above model comparison results (Fig. 4.9). It is noteworthy that the task design used in this study had sufficient power to differentiate between the various decision-making models tested. For example, the performance-based arbitration model could be well recovered when the actual data generative process was based on itself. When the performance-based model provided the best fit for a given individual's data, the probability of the performance-based model being the true generative model of the data was found to be 0.89 (Table 4.4; Methods).

In addition, the variables extracted from the best performing model suggest that the congruency between affordance and value determines how the meta-level controller weighs each system. Specifically, the arbitration weight on value-based decision-making was found to be higher in incongruent conditions (Figs. 3c,i; $t(18) = 2.28, p \leq 0.05$ in the behavioral; $t(29) = 3.59, p \leq 0.01$ in the fMRI data). We also observed that how often rewards were given on recent trials increased the arbitration weight toward value-based decision-making (Shenhav et al., 2013). In the high conditions where rewards were given more frequently, the weight on the value-based decision-making system was higher (Figs. 10a,f; $t(18) = 2.53, p \leq 0.05$ in the behavioral study; $t(29) = 3.77, p \leq 0.01$ in the fMRI study). Moreover, those trials with less weight on value-based decision-making were also trials in which the actual choice was compatible with the afforded action more often (Figs. 3d,j). We also observed that those participants who exhibited higher accuracy in selecting the most rewarding hand gestures were the individuals who also had a higher average

arbitration weight assigned to the value-based decision-making system (Figs. 10b,g; Pearson $r = 0.62, p \leq 0.01$ in the behavioral, Pearson $r = 0.65, p \leq 0.001$ in the fMRI experiment).

Furthermore, through model simulations utilizing the estimated parameters, we observed that the performance-based arbitration model could very closely replicate behavioral patterns found in the real choice data. In the simulated data, initial action selection was biased toward affordance-compatible actions and this bias was consistent across the blocks (Figs. 3e,k; Fig. 4.12) as found in the real data described earlier. The simulated learning curves also exhibited a close correspondence with the actual learning curves in each condition, and the choice accuracy gap between congruent and incongruent conditions showed a tendency to be larger in the earlier trials of each object, which was consistent with the actual data (Figs. 3f,l; Fixed-effect coefficients for the gradient of the choice accuracy difference ΔP : $\beta = -0.001, z = -2.26, p \leq 0.05$ in the behavioral, $\beta = -0.001, z = -1.50, p = 0.13$ in the fMRI data simulation; Table 4.5).

It is notable that, apart from the performance-based arbitration model, simulations using other types of arbitration model did not show an initial action selection bias or a gradual decrease in the choice accuracy gap between congruent and incongruent conditions (Figs. 3e,f,k,l; Fig. 4.13). Although the cost and the conflict-based arbitration models could exhibit several properties of the actual choice data with specific ranges of free parameters, simulations using the fitted parameters could not reproduce the patterns from actual choices (Fig. 4.13). For example, the choice accuracy gap was increasing in the cost model, which is the opposite pattern from the real data. The conflict-based arbitration model did not show a bias toward affordance compatible actions in the initial trial. Moreover, the simulated choice accuracy in incongruent condition was even better or comparable in the conflict-based arbitration model which contradicts the actual data ($P(correct|incongruent) - P(correct|congrunet)$ of the conflict-based arbitration model; $t(189) = -3.12, p \leq 0.01$ in the behavioral data simulation, $t(299) = -0.69, p = 0.49$ in the fMRI data simulation). The fixed and the reliability-based arbitration models could reproduce the initial action selection bias, but the gradient of the choice accuracy gap between congruent and incongruent conditions was marginal compared to that from the performance-based arbitration model (Table 4.5).

Additionally, when model-simulated learning curves were plotted as a function of

the number of rewarded trials similar to Figs. 2e,j, the performance-based and reliability-based arbitration models could reproduce the real behavioral patterns, but not the fixed arbitration model (See Figs. 3g,m; Fig. 4.14 and Table 4.6). For example, the difference between the congruent and incongruent conditions in terms of frequency of choosing correct actions as a function of the number of rewarded trials was only decreasing in the simulated data using the performance-based and the reliability-based arbitration models, but not in the fixed arbitration model's simulated choices (Fixed-effect coefficients for the gradient of the choice accuracy difference ΔP: performance-based arbitration model's $\beta = -0.007, z = -6.64, p \leq 0.001$ in the behavioral, $\beta = -0.008, z = -8.81, p \leq 0.001$ in the fMRI data simulation; reliability-based arbitration model's $\beta = -0.003, z = -3.08, p \leq 0.01$ in the behavioral, $\beta = -0.004, z = -4.28, p \leq 0.001$ in the fMRI data simulation; fixed arbitration model's $\beta = -0.001, z = -0.91, p = 0.36$ in the behavioral, $\beta = 0.001, z = 0.62, p = 0.54$ in the fMRI data simulation).

Nonetheless, the steeper learning slope we observed might be due to a potentially higher learning rate in the incongruent conditions. To explore this possibility, we fit the data with a fixed-arbitration model that incorporates two distinct learning rate parameters, one for incongruent conditions and another for congruent conditions. However, our analysis revealed no significant difference in learning rates between congruent and incongruent conditions ($t(18) = -1.16, p = 0.26$ in the behavioral, $t(29) = 0.73, p = 0.47$ in the fMRI data). Moreover, we could eliminate the possibility that the presence of counterfactual learning alone is sufficient to demonstrate the steeper learning slope. With the implementation of counterfactual learning through the updating of values for unchosen actions in unrewarded trials, it is clear that counterfactual learning on its own falls short of replicating the actual behavioral data (see Methods and Fig. 4.15). These results collectively indicate that the observed steeper learning slope in incongruent conditions is due to a dynamic arbitration mechanism that boosts the better working system for a given situation, rather than being a result of counterfactual learning in unrewarded trials or the effects of a larger learning rate in the incongruent conditions.

In addition, the non-decreasing difference between high and low conditions in terms of the frequency of choosing correct actions as a function of the number of rewarded trials could be captured by all tested arbitration models (Fig. 4.16, Table4.7 and Supplementary Note 2)

Consequently, through rigorous model comparisons and simulations, we identi-

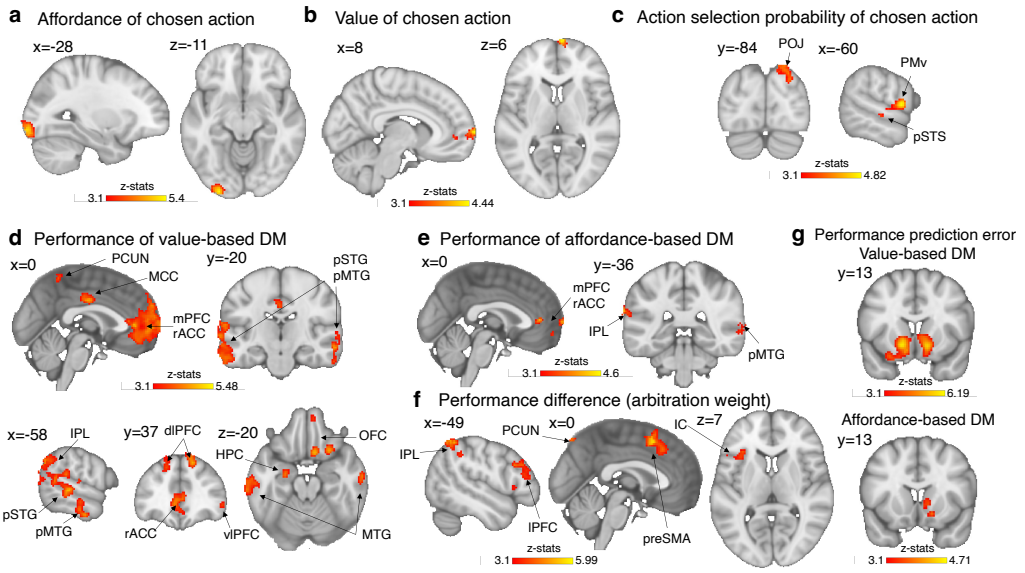


Figure 4.4: Neural implementation of performance-based arbitration. (a) Affordance-compatibility scores of the chosen action correlated with the high-level visual area including V3 and V4 in the left occipital lobe (b) The chosen action value was significantly identified in the mPFC. (c) Action selection probabilities of the chosen action were found in various regions of the cortical grasping circuit. (d) Performance of the value-based decision-making signals (e) BOLD signals of the performance of the affordance-based decision-making. (f) The fMRI correlates of the difference between the performances of two systems ($Perf_{aff} - Perf_q$), which is directly related to the arbitration weight. (g) PPE signals for tracking performances of the two systems were identified in the striatum. All the results were cluster-corrected $p < 0.05$ with the cluster defining threshold $z = 3.1$.

fied performance-based arbitration as the best candidate model for explaining the behavioral data, both quantitatively and qualitatively.

Next, we explored the neural implementation of affordance-based action-selection and its influence on value-based action-selection by utilizing computational variables from this model in the analysis of the fMRI data.

Neural correlates of affordance, value-learning, and action selection

We conducted a GLM analysis that included the chosen action's affordance- compatibility score, action value and action selection probability from the performance-based model as parametric regressors to identify the regions associated with affordance and value-based decision-making, respectively, as well as to uncover an action selection region responsible for integrating the predictions from these two systems to guide action-selection (GLM1; See Methods for details).

First of all, we found that the affordance-compatibility of the chosen action on each trial was encoded in the higher-level ventral visual stream such as V3 and V4 in the left occipital lobe which is the region that has been suggested to be responsible for affordance-perception or object recognition and processing physcial properties such as shape and size that are necessary for hand gesture control (Fig. 4a)(Bonner and Epstein, 2017; Sakagami and Pan, 2007; Kravitz et al., 2013; Polanen and Davare, 2015) The chosen action value from the performance-based aribtration model was found in medial prefrontal cortex (mPFC) which is the region that has been implicated in the encoding of learned value (Fig. 4b) (Dolan and Dayan, 2013; Wunderlich et al., 2009; Hare et al., 2011).

Regions correlating with the action selection probability of the chosen action, which is the integration of the predictions from value-based and affordance-based decision-making systems were located in the posterior parietal cortex (PPC) near the parieto-occipital junction (POJ), ventral premotor cortex (PMv), and the posterior superior temporal sulcus (pSTS) all of which are parts of the cortical grasping circuit (Fig. 4c) (Polanen and Davare, 2015; Davare et al., 2011; Rizzolatti and Luppino, 2001; Borra et al., 2017). Notably, these identified regions are responsible for not only performing the hand gesture itself but also integrating different sensory modalities related to manual actions (Andersen, 1997).

To ensure that regions correlating with the probability of the chosen action were not merely reflecting the effects of action execution vigor, we conducted an identical GLM analysis but incorporated RT as an additional parametric regressor (Shadmehr et al., 2019). These regions remained significant (at an uncorrected threshold of $p < 0.001$) suggesting that the regions are associated with the weighted sum of affordance-based and value-based decision-making systems rather than merely reflecting movement vigor (Fig. 4.18).

Moreover, we extracted coordinates of the right-hand joints from the recorded hand videos and included them in the GLM as a parametric regressor in order to control for any potential confounding effects arising from the effects of hand movements per se (See Methods for details). While we found action selection related signals in various regions of the cortical grasping circuit, the actual motor implementation of the hand gesture was correlated with activity in the left primary motor cortex, left premotor cortex, left primary somatosensory cortex, and bilateral inferior parietal lobule (IPL) (Figs. 19c-f).

Another GLM analysis using the second best fitting reliability-based arbitration

model was conducted as well, which identified mostly overlapping regions for each of the cognitive variables (GLM2; See Figs. 17a-c, and 4.18).

Neural implementation of the meta-level controller

Next, we probed for brain regions responsible for the meta-level computations involved in arbitrating between value and affordance-based decision-making systems. To achieve this, we utilized arbitration-related variables extracted from the performance model as parametric regressors in a GLM (GLM3; See Methods for details).

As illustrated in Figs. 4d,e, we found that the mPFC, rostral anterior cingulate cortex (rACC), posterior division of middle temporal gyrus (pMTG), and IPL to be associated with the performances of both systems. The posterior division of superior temporal gyrus (pSTG), dorso and ventro lateral prefrontal cortex (dlPFC, vlPFC), mid-cingulate cortex (MCC), orbitofrontal cortex (OFC), precuneous cortex (PCUN), and hippocampus (HPC) were correlated with the performance of value-based decision-making but not with the performance of affordance-based decision-making.

In addition, we observed that the signal corresponding to the difference in performance between the two decision-making systems ($Perf_{aff} - Perf_q$), which determines the arbitration weight, was present in the pre-supplementary motor area (preSMA), lateral prefrontal cortex (lPFC), and insular cortex (IC) (Fig. 4f). We also found that the IPL and PCUN were regions correlated with the difference in performance between the affordance and value-based decision-making systems. These regions have been shown to be associated with the implementation of cognitive control (Botvinick and Braver, 2015; Nachev et al., 2008; Botvinick et al., 2004; Miller, 2000; Behrmann et al., 2004). Interestingly, our findings indicate that the regions identified were less activated when the value-based decision-making system was allocated more weight.

Similar to the RPE signals (Fig. 19a), the variables responsible for updating arbitration weights, which are also outcome-dependent prediction error signals, were found in the striatum (Fig. 4g). Specifically, the PPE of the value-based decision-making system was prominently encoded in the ventral striatum, while the PPE of the affordance-based decision-making system was found in the dorsal striatum.

In addition to investigating neural correlates for the performance-based arbitration model, we also tested for regions correlating with reliability-based arbitration which was the second best fitting model in our model comparison (GLM4). We also

found clear neural correlates of reliability-based arbitration signals (Figs. 17d-g and Fig. 19b). In particular, we found the regions that correlate with the difference between the reliabilities of the two systems largely coincide with those identified using the performance-based arbitration model, particularly the preSMA. It is note worthy that the average correlation between the arbitration variables of the two models were small($r = 0.196$ in the fMRI data and $r = 0.159$ in the behavioral data). Therefore, the most plausible explanation is that each model captures distinct variance in activity in the preSMA. These findings indicate that the brain keeps track not only of the performance of the different systems but also of their reliability, suggesting that both variables might ultimately be taken into account during the arbitration process.

Better task performance is linked to a more robust representation of arbitration variables in meta-level controller regions

Next we explored the relationship between individual differences in task performance and representation of the arbitration variables. We found that individuals with higher task performance demonstrated more robust representations of the arbitration variables such as the performances of the two systems and the difference in performance between the two systems in brain regions involved in encoding arbitration-related variables. To be specific, we found positive correlations between each participant's propensity to choose the most rewarding actions and the extent to which BOLD activity in the identified arbitration regions can be explained by our model-derived arbitration variables.

For instance, in those participants who tend to be more accurate in choosing the correct actions, the computational variable corresponding to the difference in performance between the two decision-making systems provided a better account of BOLD activity in the arbitration regions shown in Fig. 4f ($r = 0.47, p \leq 0.01$; Fig. 5a, See Methods for details). Moreover, in those participants who achieved higher accuracy in choosing the most rewarding actions, brain regions involved in representing the computational variables corresponding to the performance of each individual system were more correlated with the model-estimated performance variables ($r = 0.38, p \leq 0.05$ for the performance of the affordance-based system; $r = 0.43, p \leq 0.05$ for the performance of the value-based system; Fig. 5a). However, the strength of the neural representation of the *within* system computational variables produced by each system relevant for decision-making did not correlate with the participants' accuracy ($r = 0.23, p = 0.22$ for the affordance;

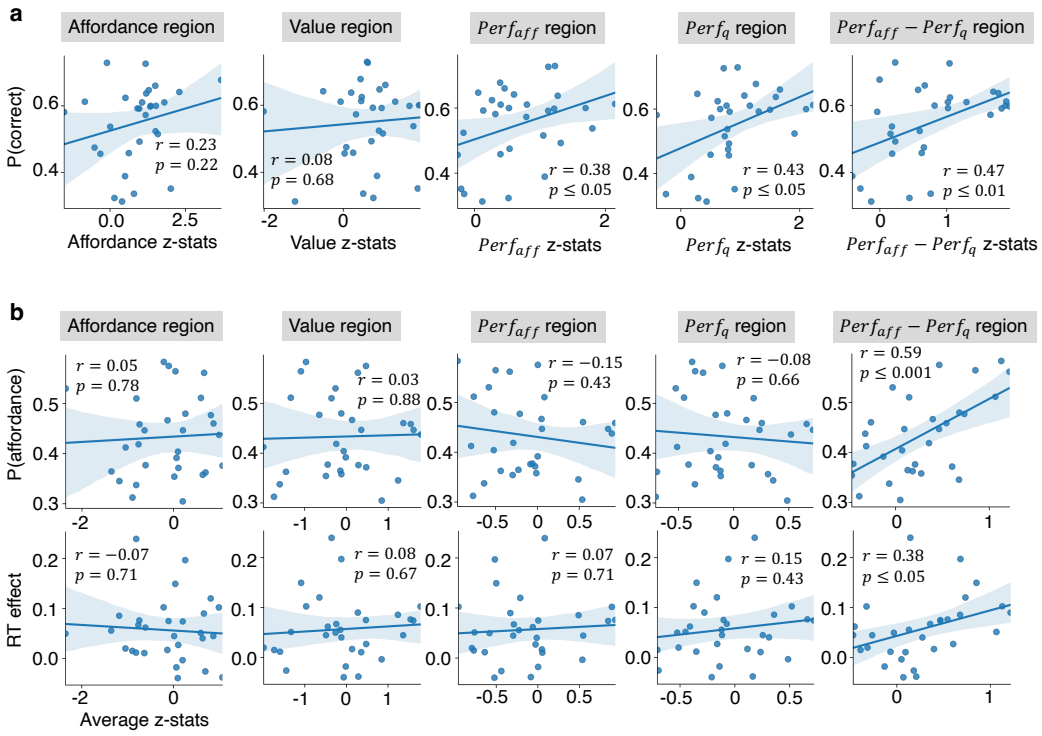


Figure 4.5: Correlations between behavioral effects and neural representations. (a) Correlations between the tendency to select the most rewarding actions and the strength of the neural representation of cognitive variables by functionally defined regions of interest. (b) Correlation between the affordance effect on the behavior and the increased BOLD activity when executing affordance-incompatible action by functionally defined regions of interest. The x-axis is the average z-statistics of the contrast between the affordance-incompatible and affordance-compaitble action execution from the GLM5 across the voxels within the ROIs. RT effect $= \frac{RT_{incomp} - RT_{comp}}{RT_{comp}}$. Each dot represents an individual participant. See Methods for the details.

$r = 0.08$, $p = 0.68$ for the value; Fig. 5a). These findings, suggest that more robust neural representations of performance-based arbitration are associated with better behavioral performance on the task, providing additional evidence in support of the role for an arbitration process in mediating effective interactions between value-based and affordance-based action selection (Analogous results were also found for the reliability-based arbitration model; see Fig. 20).

We also found that those individuals with a stronger behavioral affordance effect exhibited increased activation of performance-based arbitration regions when executing affordance-incompatible actions. Such individuals likely have to exert

stronger behavioral control to suppress the prepotent affordance compatible action compared to individuals with a lower overall affordance effect in their behavior. To test for this, we looked at the relationship between the overall proportion of affordance compatible actions chosen by each individual and activation on trials where the affordance incompatible action was chosen compared to when it was not chosen (GLM5; See Methods for details). A significant correlation was observed in performance-based arbitration regions ($Perf_{aff} - Perf_q$ regions in Fig. 4f; $r = 0.59, p \leq 0.001$; Fig. 5b). A similar effect was found when examining the relationship between the reaction time increase when executing affordance-incompatible actions compared to affordance compatible reaction times, and activity in those areas ($r = 0.38, p \leq 0.05$; Fig. 5b; See Methods for details). These findings suggest that regions involved in arbitration including the preSMA, have an important role in suppressing the prepotent affordance response, especially in those individuals who have a stronger overall bias toward affordance effects (Nachev et al., 2008).

4.4 Discussion

We provide evidence that action affordance plays a key role in shaping value-based action-selection and learning in humans. Action affordance and value-based decision-making were found to interact to guide behavior. Affordances were found to both influence reaction times, such that actions compatible with the affordance are selected more rapidly regardless of the effects of ongoing value learning, as well as biasing choice toward the afforded action , an effect that was most prominent in the early trials.

In order to understand the mechanism by which action affordance influences instrumental value learning, we implemented and tested a series of computational models. We found that the effects of affordance on choice behavior is not well described by either a simple bias in action-selection or an initial prior boosting the value of afforded actions. Instead, a form of dynamic arbitration was found to best explain behavior on the task. According to this framework, two separate systems operate in parallel, a value-based system and an affordance-based system, and the determination of which system contributes most to behavior at any one time is based on estimates of the relative performance of the two systems. This neuro-computational mechanism provides a potential explanation for the problem of how human learns appropriate actions in the absence of clear information about their value. Given that a very large family of actions could be implemented in any given situation, it is highly beneficial for the set of actions available for selection to be constrained by

properties of the visual environment such as object affordance. This can help make the action-selection problem more tractable, and provide a scaffold for selecting actions so as to explore and learn about value during initial exploration.

In the brain, consistent with our computational model-based findings, we found evidence for the existence of distinct affordance-based and value-based decision-making systems. While the value of the chosen action was found in medial prefrontal cortex, the affordance of the action chosen on a given trial showed a positive correlation with activity in the visual pathways of the occipital cortex, including regions V3 and V4 (Bonner and Epstein, 2017; Dolan and Dayan, 2013; Wunderlich et al., 2009; Hare et al., 2011). This finding suggests that the occipital visual areas, crucial for recognizing an object's physical properties and identity, are more engaged when an action compatible with the object's affordance is chosen (Polanen and Davare, 2015; Sakagami and Pan, 2007; Kravitz et al., 2013). Conversely, when an action irrelevant to the object is selected, these characteristics are less involved in the final action selection and motor execution process.

Moreover, our study revealed that the overall integrated choice probability, which is the integration of value and the compatibility of certain actions to the shown object, is associated with activity in the PPC. The PPC is known for representing object-associated hand movements and has been identified as a key area in reward-based decision making, especially concerning action values (Chen et al., 2016; Castiello, 2005; Johnson-Frey, 2004; Fagg and Arbib, 1998; Platt and Glimcher, 1999; Dorris and Glimcher, 2004; Wunderlich et al., 2009). Thus, our finding aligns with previous findings and supports the perspective that the PPC plays a critical role in integrating multimodal information necessary for movement execution (Andersen and Cui, 2009). It also unifies two different literatures on visually guided motor control and economic decision making, emphasizing the PPC's comprehensive function in these processes.

In addition to identifying regions correlating with each strategy and their integration, we also looked for evidence of brain regions involved in meta-control over the two underlying action-selection systems. Specifically, we identified a network of brain regions tracking and comparing the performance of both strategies, including the anterior cingulate, pre-SMA, lateral prefrontal cortex, and inferior parietal lobule, regions that have been found to be involved in cognitive control in a number of previous studies (Botvinick and Braver, 2015; Nachev et al., 2008; Botvinick et al., 2004; Miller, 2000; Behrmann et al., 2004). Moreover, activity within the

arbitration regions served as an indicator of the likelihood that participants would choose affordance-compatible actions, as well as the extent of the stimulus-response compatibility effect on reaction times. Additionally, we found that the more activity in regions representing arbitration-related variables correlated with the relevant arbitration variables across participants, the more optimal a participant's behavior was in terms of selection of the most rewarding actions. This finding therefore suggest a crucial role for the arbitration process between affordance and value-based decision-making in guiding optimal behavioral outcomes.

The finding of a role for an arbitration scheme governing the contribution of value-based and affordance-based choices to behavior is consistent with a mixture of experts framework (O'Doherty et al., 2021b). In the present study, we found evidence for a role of performance-based arbitration over and above reliability-based arbitration, which we have previously investigated in relation to its role in allocating control between other strategies such as model-based and model-free RL and different forms of observational learning (Lee et al., 2014a; **?**). While these two arbitration concepts are closely related as they are both concerned with how well a particular expert system is doing in making predictions, the precise computational variables underpinning the arbitration process is an important research question. In the present study, reliability-based arbitration was the second best performing model, outperformed only by the performance-based arbitration model. Moreover, we found evidence for reliability-based arbitration signals in the brain alongside performance-based arbitration signals and those two arbitration signals capture distinct variance in the activity of the identified arbitration regions which includes the preSMA. It is possible that both performance and reliability are playing a role in guiding arbitration between these different expert systems.

It is important to note that the arbitration process described here can be seen as a form of cognitive control. Central to our model is the concept of directing control towards the system that promises a higher expected return. This idea aligns with the Expected Value of Control (EVC) model which allocates cognitive control based on the control's effectiveness such as expected value and cost (Shenhav et al., 2013). According to the model, the demand for control and deciding the control intensity can be evaluated by tracking the conflict between task relevant and irrelevant information as well (Botvinick et al., 2001). However, our findings suggest that a conflict-based arbitration model, rooted in conflict monitoring theory, does not most effectively explain choice behavior in value learning contexts, despite the proposed significance

of conflict signals in control allocation. In contrast, our model emphasizes allocating control towards the system with a higher expected return, a concept resonating with the EVC model's focus on control effectiveness.

Furthermore, the well-established link between perception and action extends beyond the concept of affordance. The automatic potentiation of likely actions has also been explored from an information processing viewpoint, indicating that motor and visual representations share a common representational space (Hommel et al., 2001; Hommel, 2004), and from attention theory, which suggests that attention to a spatial location primes corresponding action plans toward that location (Rizzolatti et al., 1987; Craighero and Rizzolatti, 2005). While efforts have been made to differentiate the affordance effect from the Simon effect (Symes et al., 2005; Iani et al., 2011; Pappas, 2014) the necessity of affordance in explaining stimulus-response compatibility is still debated (Proctor and Miles, 2014). Notably, the affordance effect is significant primarily when motor representation, aligned with an object's affordance, is triggered by movement intention rather than solely by the visual stimulus (Bub et al., 2021; Ferguson et al., 2021). Our experiment which required movement intention for completing each trial, is particularly relevant in this context. Recent studies also suggest that affordance, or plausible actions triggered by a stimulus, impact attention allocation and other cognitive processes such as working memory and visual perception. This indicates that explaining affordance through spatial attention alone is insufficient, and that a bidirectional process must be considered (Heuer et al., 2020; Olivers and Roelfsema, 2020; Ede, 2020). Our computational approach was not focused on the mechanisms behind affordance-induced stimulus-response compatibility but instead on how automatically activated actions influence learning.

Another pertinent question concerns how affordance-based action-selection relates to other forms of action-selection that are thought to exist alongside value-based decision-making. The most obvious comparison is with habits. Habits are suggested to be formed via repeated reinforcement of stimulus-response associations (Dickinson, 1985). Here, the affordances associated with particular objects were not acquired in the experiment through trial and error reinforcement, because they were manifested on the very first trial that each object was encountered and were by design kept orthogonal to the reward contingencies. Thus, they are not "habits" in the traditional sense as typically studied in the lab. However, it is possible that affordances do correspond to a type of stimulus-response association that has been

historically acquired over the course of development as individuals interact with objects in their environment and learn to implement specific physical actions in response to them. In that sense, these affordances could be thought of as very well-learned habits. However, our fMRI study revealed evidence for the neural implementation of affordances and their influence on action-selection in occipital and parietal cortices, and not in areas traditionally associated with habitual action-selection such as the posterior putamen (Tricomi et al., 2009a; McNamee et al., 2015). It is possible that such extremely well learned habits come to eventually depend on the cortex and not the basal ganglia. However, within the dorsal striatum, we did find learning update signals related to meta-control that steers individuals to favor affordance-based decisions — suggesting that the basal ganglia may play a role in updating control signals related to the arbitration process between strategies, over and above its contributions to implementing individual strategies.

Alternatively, affordance-based influences on action-selection could be implemented via a much more dynamic visuomotor computation, in which specific visual features of an object guide on-line action-selection computations in which the most relevant actions for interacting with a particular object are decided upon and implemented (Wolpert et al., 2011). Further investigation of the specific neural computations unfolding during affordance-based action selection could help discriminate the underlying mechanisms.

We acknowledge a potential limitation in our study arising from the use of object images instead of real objects. Images might primarily convey the object's semantics or 'stable affordance,' constructed from people's familiarity with the object, and not fully represent the physical properties like the object's orientation and proximity to participants, which are also important factors in affordance perception (Sakreida et al., 2016). Additionally, there is a possibility that the different neural circuitry has been involved in the current study compared to using the real object. For example, researches have shown that interacting with real objects and images involve distinct visuo-motor circuitry and object pictures are rather processed conceptually or as the object word (Marini et al., 2019; Freud et al., 2018; Martin, 2007). Furthermore, fMRI studies have identified parieto-occipital cortex regions sensitive to the physical properties of real objects (Gallivan et al., 2011; Rice et al., 2007; Valyear et al., 2006; Gallivan et al., 2009; Symes et al., 2007).

Nevertheless, our data suggest that the effects we found about hand gestures associated with objects was influenced not just by participants' familiarity and the

semantics of objects, but also by their physical properties inferred from visual features of the displayed objects. The familiarity scores participant reported about stimuli used in the behavioral and fMRI tasks varied significantly (66.45 ± 24.62 on a 0 to 100 scale, Fig. 4.21) and the difference in probabilities of choosing affordance-compatible actions in the initial trial upon seeing unfamiliar objects (familiarity score < 66.45) and familiar objects (familiarity score ≥ 66.45) was not significant ($t(18)=1.529$, $p=0.144$ in the behavioral, $t(29)=0.001$, $p=0.996$ in the fMRI), despite a general bias towards the affordance-compatible actions in the first trials. This indicates that the choices were influenced by their physical properties discernible in the images, beyond just familiarity or object identity. Furthermore, the stimuli were not arbitrarily selected but annotated by online participants viewing the same images as the on-site participants. From approximately 1000 stimuli, 48 objects significantly associated with specific hand gestures were chosen for our experiments. Thus, the affordance labels used in our study can be considered stable, taking into account the object's orientation and virtual proximity as displayed in the stimuli.

Additionally, the automatic potentiation of action and its behavioral and neural effects have been observed even with photographic representations of objects (Tucker and Ellis, 1998; Grèzes et al., 2003). While planar presentation may affect visuomotor processing, stable affordance aspects such as mechanical and functional knowledge about an object are likely unaffected (Osiurak et al., 2017). Consequently, our findings remain valid within the context of studying the effect of stable affordance, predominantly derived from object semantics, in value learning. However, future research utilizing real objects will be helpful for studying affordance in more ecologically valid and realistic settings.

In conclusion, the present study provides evidence that human value learning is guided not only by the value of particular actions, but also by the visual affordance of objects. Rather than affordance acting as a simple bias in decision-making or prior in value learning, instead we find that affordance and value-based decision-making are best viewed as distinct expert systems that interact by means of a determination about which system is performing best in obtaining rewards.

4.5 Methods

Participants. We recruited 21 and 32 healthy participants for the behavioral-only experiment and the fMRI experiment, respectively. 2 participants were excluded

from the behavioral and another 2 from the fMRI experiment because they did not complete the study. Therefore, data from the remaining 19 (8 females, 18 right-handed; 1 ambidextrous, 18 ~ 24 years: 6; 25 ~ 34 years: 10; 35 ~ 44 years: 1; 45 years or above: 2) and 30 (13 females, 28 right-handed; 2 ambidextrous, 18 ~ 24 years: 12; 25 ~ 34 years: 12; 35 ~ 44 years: 4; 45 years or above: 2) were used for the analyses. Before taking part in the experiment, all participants were assessed to ensure that those with the history of neurological or psychiatric illness were excluded. All participants provided their informed consent, and the study was approved by the Institutional Review Board of California Institute of Technology.

Stimuli. To implement the paradigm, we began by creating a set of stimuli. The stimulus set consisted of various object images, each of which was designed to engender a specific hand configuration among four hand movement classes known to be employed during object interaction: pinch, clench, poke, and palm (e.g., a button-shaped object that affords poking)(Klatzky et al., 1987). We first obtained 1000 images of around 900 unique objects and built a set of visual stimuli by superimposing those object images onto a human silhouette image. We designed the stimuli in this way to ensure that the visual stimuli retain information about the object's size.

Then we used those edited photographs in an online task on Amazon Mechanical Turk (M-Turk). Each image was displayed 4 times during the task to collect annotations on the suitability, or affordance-compatibility score, of 4 hand movements to the object displayed. Within a trial, the image with the human silhouette was displayed for 2 seconds and the object's zoomed-in image was shown for one second along with a text of one of the 4 hand gestures. Then while the zoomed-in image was on the screen for additional 8 seconds, M-Turk participants responded how suitable it is to pinch, clench, poke or palm the object shown by sliding a bar ranging from "Very unsuitable" to "Very suitable." Additionally, we asked about their familiarity with the object 4 times. The affordance-compatibility and familiarity scores were both converted to 0 to 100 scale later. Throughout the task, 4 catch questions were used to check the participants' attention.

We recruited 227 participants online, but only 160 of them were included in the data analysis as we excluded those that did not pass at least 2 catch trials. Each of the participants was asked to annotate 50 objects. On average, each object was annotated by 7.2 individuals.

Then we chose objects based on their 0-100 affordance-compatibility scores. For example, if an objects' average affordance-compatibility score of pinch was greater than 50, and the mean score of pinch was significantly larger than the mean scores of other actions (independent t-tests, $p < 0.05$ uncorrected), the object was chosen as a pinch-affording object. By doing so we obtained 21 pinch-, 102 clench- and 16 poke-compatible objects. Then we randomly selected 16 items from each set, yielding 48 stimuli for the main task (Fig. 1a).

Task. We designed a variant of a 3-armed bandit task in which the available actions are natural pinch, clench, or poke gestures made by the right hand, which yields binary outcomes. We used 48 visual object stimuli for the behavioral experiment and 24 object stimuli (a subset of the 48 stimuli) for the fMRI experiment. On each trial, one of the stimuli showing an object with a human silhouette was displayed, and a participant mimed one of the three hand gestures while maintaining their wrist position on the provided wrist pad (Figs. 1a,e). The images used here (including the human silhouettes) were identical in form to the selected sub-set of images initially presented in the online survey. Participants were provided with instructional videos showcasing example hand movements corresponding to each gesture type to ensure that they had a clear understanding of the hand gestures under consideration. Participants were instructed that the reward probability of each hand motion is different for each object, and that they need to figure out the most rewarding hand gesture for each object by trying different hand motions. Each category of hand postures corresponded to actions with different reward probabilities, and an object was shown in a trial to evoke one of the three hand gesture affordances. There were 4 experimental conditions in this paradigm: congruent high, congruent low, incongruent high, and incongruent low conditions. In a congruent trial, an action that was consistent with the presented stimulus' affordance had a high reward probability, whereas in an incongruent trial, the opposite applied. Each hand position had a reward probability of 0.8, 0.2, or 0.2 in the high condition and 0.4, 0.1, or 0.1 in the low condition.

The task consisted of a 2-day experiment for the behavioral-only study and a 1-day experiment for the fMRI study, with 76~84 trials (on average 80 trials) per block, and 6 blocks per day, totaling 960 trials for the behavioral study and 480 trials for the fMRI study. In each block, 4 distinct objects were displayed using the event-related design. Each object corresponded to one of the 4 conditions and was shown for an average of 20 trials (ranging from 19 to 21 trials). All objects were

displayed once in every 4 trials in random order. Everyone in the behavioral-only and fMRI experiment was shown the same set of 48 or 24 objects, respectively, but the associations between experimental conditions and stimuli was randomized across participants. For the counterbalancing, at most two objects within a block had the same affordance and at most two objects within a block had the same most-rewarding response. Also, the two congruent condition objects within a block had different affordances and the two incongruent condition objects had different affordances and different correct responses from each other.

Each trial lasted between 7 to 11.5 seconds (9.25s on average). The trial timing is detailed in Fig. 1b. Each trial started with a jittered fixation cross (1-2.5s), followed by the stimulus display. The stimulus was shown at most for 4 seconds. After a hand gesture was recognized, there was another jittered fixation cross (1-2.5s) and a jittered reward feedback (1-2.5s). Following the reward feedback, another fixation cross was displayed for the duration of the difference between 4s and the reaction time plus the processing time for response registration. Eye-tracking data were also collected during both the behavioral and fMRI experiments but is not analyzed herein.

Following the main task, the behavioral and fMRI participants did the same online survey that the M-Turk participants completed to annotate affordance-compatibility scores and familiarity for each object. However, in this phase, the participants annotated only the 48 (or 24 in the fMRI case) stimuli that were used for the main experiment.

Response decoding. In the behavioral experiment, participants' hand gestures were tracked by a consumer web camera (30 FPS) and decoded in real-time using a computer vision algorithm, Openpose (Cao et al., 2019; Simon et al., 2017; Cao et al., 2017; Wei et al., 2016), and a fully connected neural network (FCNN) that classifies such movement, frame by frame. For the real-time processing of the video stream, frames were first resized to 192×108 pixels and then inputted into BODY_25 model of Openpose with a net resolution setting -1×112 and an output resolution -1×80. The FCNN classified a participant's hand position in a frame using an 81-dimensional vector that was created based on Openpose estimations of the 2-dimensional coordinates of 21 key points on the right hand (4 from each finger and 1 from the wrist), as well as the confidence scores of the estimations for each key point.

Specifically, the x and y coordinates of each key point were centered using the wrist location, and then scaled based on the x-directional distance between the left-most and right-most key points, as well as the y-directional distance between the bottom-most and the top-most key points in the current frame. The 81-dimensional input to the FCNN was then built using those centered and scaled x, y coordinates of each key point except for the wrist (20×2 dimensions), their sum of squares representing the distances between each key point and the wrist (20 dimensions), and confidence scores of the 21 key point estimations (21 dimensions).

The neural network classifier comprised of 3 hidden layers (128, 64, and 32-dimensions, respectively, from low to high layers) and a 4-dimensional softmax output layer in which each unit was associated with one of four hand gestures: pinch, clench, poke, or palm. The deep network classifier was trained using videos of hand motions that were labeled frame-by-frame. The training videos featured hand gestures of 2 males and 2 females, and the total duration of the videos were 40240 frames. We used `MLPClassifier` function from Sklearn 0.23.1 with `lbfgs` optimizer and learning rate 10^{-5}.

A response in a trial was registered as a valid action only when all frames for 500ms were classified with a probability greater than 95% into one specific hand gesture. To ensure consistency in the starting hand position across trials, it was required that each trial begins with a hand position classified as the palm position by the deep network. The classifier classified responses with 93.2% accuracy and misclassified responses were manually corrected after the experiment (and we then used the correctly decoded action in the subsequent analyses – those ~7% errors trials would have produced outcomes with differing probabilities to that intended according to the experimental design, but this variation would not have been noticeable to the participants and was fully accounted for in subsequent data analysis involving the computational models). Reaction times were measured by calculating the time difference between the stimulus onset and the frame where the hand gesture was initiated from the resting palm position using the recorded videos of hand gestures.

In the fMRI experiment, decoding of participant's actions was found to be much less reliable with the machine-learning algorithm we used successfully in the behavioral study, because participant's hand position was occluded by the MRI bore. Consequently, we resorted instead to manually decoding each gesture in real-time. Participants' hand gestures were monitored by the experimenter using a low light USB camera (See3CAM_CU30, 40 FPS), and the footage was displayed in the

control room in real-time. The experimenter manually but rapidly categorized the movements into one of 3 hand gesture categories (pinch, clench, or poke) thereby enabling participants to interact with the task in real-time. The registered responses were double-checked after each experiment by replaying the recorded video stream, revealing that less than 0.1% of the responses were mistakenly classified during the task performance phase. These error trials were corrected prior to the subsequent data analysis to reflect participant's intended responses. RTs were calculated using the identical method used for analyzing RTs in the behavioral data.

Computational modeling of behavior. We tested 9 different models to identify the computational mechanism that best explains the behavioral data. The models were fit to each subject data separately using the Computational Behavioral Modeling (CBM) toolkit `cbm_lap` function which calculates model evidence and likelihood of the data using Laplace approximation and estimates parameters using maximum-a-posteriori (MAP) estimation (Piray et al., 2019a). We used mean 0 and variance 6.25 Gaussian priors for calculating the parameter estimates. Because CBM assumes the parameters to be normally distributed, we applied transformation functions to the parameters: a sigmoid function to model parameters which ranges are between 0 and 1, or an exponential function to model positive parameters. We did not use the hierarchical Bayesian inference function of the toolkit because model simulations using the hierarchically estimated parameters could not reproduce the characteristics of the real data. Instead, we performed Bayesian model selection (Stephan et al., 2009a) using the log evidence from the CBM outputs to do the Bayesian model comparison. The models were developed solely based on the behavioral data and later tested on the fMRI data.

In order to model affordance perception, we employed the affordance-compatibility score derived from the post-task survey of each participant, which was scaled from 0 to 1. We used individually derived affordance-compatibility scores to fit the model to the behavioral experiment data. However, in the case of fitting the model to each fMRI participant, we used the average scores of annotations from all participants as such a strategy showed a better fit to the data. Notably, the choice of using either individual scores or average scores did not affect the results of the model comparison for either dataset.

Reinforcement learning model. The reinforcement learning (RL) model is the simplest model we tested that does not consider the effect of affordance, but only

models value learning. Action values Q for each object are initialized at 0.

$$Q_0(s, a) = 0$$

And the probability of choosing an action a given a stimulus s is a softmax function of action value

$$P_t(a|s) = softmax(\beta Q_t(s, a) + b_a) = \frac{exp(\beta Q_t(s, a) + b_a)}{\sum_{a'} exp(\beta Q_t(s, a') + b_{a'})}$$

where $\beta > 0$ is an inverse temperature parameter and $b_a \in \mathbb{R}$ models action selection bias toward the action independent to the stimulus shown. The action value $Q_t(s, a)$ for the chosen action a_t to the stimulus s_t at trial t is updated based on the delta rule

$$RPE_t = r_t - Q_t(s_t, a_t)$$

$$Q_{t+1}(s_t, a_t) \leftarrow Q_t(s_t, a_t) + \alpha \times RPE_t$$

where the reward prediction error (RPE) is the difference between the reward r_t (0 or 1 depending on the realized outcome) at the trial t and the action value. $0 \leq \alpha \leq 1$ is a learning rate parameter. There were 5 free parameters ($\beta, \alpha, b_{pinch}, b_{clench}$ and b_{poke}) in this model.

Bayesian learning model. In this model, instead of using the reinforcement learning rule, value learning is modeled as Bayesian inference Niv et al. (2015). The model tracks the probability of reward of each action given an object using a beta prior and a binomial likelihood. Specifically, the prior on the reward probability of an action a to an object s is

$$P_0(R = 1|s, a) \sim Beta(\alpha_a, \beta_a),\ \beta_a = \frac{\alpha_a(1 - \gamma)}{\gamma}$$

where $\alpha_a > 0$ and $0 < \gamma < 1$ are free parameters and the prior mean is γ. Given a scenario where an object s shown n times, with an action a being chosen n_a times, resulting in r_a positive outcomes, the posterior reward probability of the action a to the object s is

$$P_n(R = 1|s, a) \sim Beta(\alpha_a + r_a, \beta_a + n_a - r_a),\ n = \sum_a n_a$$

which posterior mean is $\frac{\alpha_a + r_a}{\alpha_a + \beta_a + n_a}$.

We then assumed affordance elicits its effect as a bias in the action selection. We used 0-to-1 scaled affordance-compatibility score of action a to object s from the

post-task survey as affordance-compatibility score $Aff(s,a)$ and used the softmax function,

$$P_t(a|s) = softmax(\beta(Q_t(s,a) + \ k \times Aff(s,a)) + b_a)$$

as the action selection policy where $Q_t(s,a)$ is the posterior mean of $P(R=1|s,a)$ at trial t and $\beta > 0, k > 0$ and $b_a \in \mathbb{R}$ are free parameters that model inverse temperatures, weight on affordance and action selection bias, respectively. In total, this model has 9 free parameters (α_a and b_a for each action type, γ, β and k)

We also tested models incorporating the affordance into the prior mean instead of including the affordance-based selection bias term in the softmax policy. However, these models exhibited an even poorer fit to the data compared to the model described above.

Affordance as a prior model. The prior model uses the affordance-compatibility score $Aff(s,a)$ as the initial action value of an action a to an object s.

$$Q_0(s,a) = k \times Aff(s,a), k > 0$$

The decision probability and learning rules are identical to the RL model, which results in having one additional free parameter than the RL model ($\beta, \alpha, b_{pinch}, b_{clench}, b_{poke}$ and k). This model can also be interpreted as implementing a single decision-making system based on affordance and in which the affordance is being relearned based on task experience.

Affordance as a bias model. The bias model is identical to the RL model except for the calculation of the action selection probability. In the affordance as a bias model, action selection is a softmax function of the linear summation of the action values and the affordance-compatibility scores.

$$P_t(a|s) = softmax(\beta(Q_t(s,a) + k \times Aff(s,a)) + b_a), \ k > 0$$

This model has 6 free parameters ($\beta, \alpha, b_{pinch}, b_{clench}, b_{poke}$ and k).

Fixed arbitration model. In this model, we assumed there are two different decision-making strategies, value-based decision-making and affordance-based decision-making. Value-based decision-making and value learning is modeled using the RL model previously described.

$$P_{q,t}(a|s) = softmax(\beta Q_t(s,a) + b_a)$$

Action selection by the affordance-based decision-making system is modeled as a softmax function of the affordance-compatibility score.

$$P_{aff}(a|s) = softmax(\beta(k \times Aff(s,a)) + b_a)$$

The action selection policy is calculated as a weighted sum of these two policies

$$P_t(a|s) = w_q P_{q,t}(a|s) + (1 - w_q) P_{aff}(a|s)$$

where $0 \leq w_q \leq 1$ is the fixed arbitration weight between the two decision-making strategies. This model has 7 free parameters in total ($\beta, \alpha, b_{pinch}, b_{clench}, b_{poke}, k$ and w_q).

Affordance as a cost model. The cost model is an extension of the fixed arbitration model where we assumed that the response-affordance-incompatibility acts as a cost that is contrasted against the reward signal (Botvinick et al., 2009). Specifically, we modified the delta learning rule to include a cost term which results in a larger reward when an affordance-compatible action is selected so that,

$$RPE_t = (r_t + c \times k \times Aff(s_t, a_t)) - Q_t(s_t, a_t)$$

$$Q_{t+1}(s_t, a_t) \leftarrow Q_t(s_t, a_t) + \alpha \times RPE_t$$

where $c > 0$ is an additional free parameter that models the effect of affordance in learning as a cost. For example, the additional reward will be smaller when the chosen action has a low affordance-compatibility score compared to when it has a high affordance-compatibility which can be interpreted as there was a cost in selecting affordance-incompatible action.

The other components of the model remain identical to the fixed arbitration model, resulting in a cost model with 8 free parameters ($\beta, \alpha, b_{pinch}, b_{clench}, b_{poke}, k, w_q$ and c).

Conflict-based arbitration model. Like the fixed arbitration model, we assumed that the value-based policy and the affordance-based policy are mixed for the action selection but in this model the arbitration weight dynamically changes trial by trial for each stimulus. The arbitration weight $w_{q,t}(s)$ at trial t is a logistic function of a conflict variable $c_t(s)$ which tracks the conflict between the value-based and affordance-based policies given the stimulus s. Thus,

$$w_{q,t}(s) = \frac{1}{1 + exp(-\eta_1 c_t(s) + \eta_0)}$$

where $\eta_1 > 0$ and $\eta_0 \in \mathbb{R}$ are free parameters. We assume that the conflict signal between the two policies causes an increase in cognitive control which pushes behavior toward value-based decision-making in this paradigm (Matsumoto and Tanaka, 2004; Botvinick et al., 2001). The conflict variable $c_t(s)$ is calculated as the square root of the Jensen-Shannon Divergence (JSD) between the value-based and affordance-based policies.

$$c_t(s) = \sqrt{JSD(P_{q,t}(s,\cdot), P_{aff}(s,\cdot))}$$

The calculation of value-based and affordance-based polices of the model are identical to the fixed arbitration model which makes this model have 8 free parameters ($\beta, \alpha, b_{pinch}, b_{clench}, b_{poke}, k, \eta_0$ and η_1).

We also tested other design choices such as the energy from a Hopfield network(Yeung et al., 2004a) as the proxy of conflict or using a conflict tracking variable that is updated gradually using trial-by-trial JSD. However, the model described above explained the data better than the other tested options.

Reliability-based arbitration model. This model uses reliabilities (Lee et al., 2014a; **?**; O'Doherty et al., 2021b) of the two strategies as the driving force of arbitration between the value-based and affordance-based decision-making. The reliabilities are calculated using an RL-like updating rule based on unsigned prediction errors (Lee et al., 2014a). For example, the reliability of value-based decision-making given a stimulus s_t is updated using the following delta rule.

$$\chi_{q,t+1}(s_t) \leftarrow \chi_{q,t}(s_t) + \alpha_\chi((1 - |RPE_t|) - \chi_{q,t}(s_t)),\ 0 \le \alpha_\chi \le 1$$

$$RPE_t = r_t - Q_t(s_t, a_t)$$

Therefore, the reliability of value-based decision-making will increase when the unsigned RPE is small. Similarly, we defined the reliability of affordance-based decision-making given a stimulus s_t is updated based on

$$\chi_{aff,t+1}(s_t) \leftarrow \chi_{aff,t}(s_t) + \alpha_\chi((1 - |APE_t|) - \chi_{aff,t}(s_t))$$

$$APE_t = r_t - k \times Aff(s_t, a_t)$$

where APE is the affordance prediction error and k is between 0 to 1 to ensure the APE is between 0 and 1. The reliabilities were initialized to 0. Then the arbitration weight given a stimulus s is the logistic function of these two reliabilities. Therefore,

$$w_{q,t}(s) = \frac{1}{1 + exp(-\eta_q \chi_{q,t}(s) + \eta_{aff}\chi_{aff,t}(s) + \eta_0)}$$

where $\eta_q, \eta_{aff} > 0$ and $\eta_0 \in \mathbb{R}$. Identical to the fixed and conflict-based arbitration model, the value-based and affordance-based policies are softmax functions of action values, or affordance-compatibility scores and the final action selection probability is the reliability-based arbitration-weighted sum of the two policies. As a consequence, this model has 10 free parameters ($\beta, \alpha, b_{pinch}, b_{clench}, b_{poke}, k, \alpha_\chi, \eta_0, \eta_q$ and η_{aff}).

Performance-based arbitration model. Here, we are proposing a novel arbitration framework that aims to optimize the arbitration weight in terms of maximizing the return. The mathematical derivation of the proposed model is detailed in the supplementary note1. Specifically, we defined value, or performance, of a strategy given a stimulus as the expected returns that can be collected by following that particular strategy in response to the stimulus (Botvinick and Braver, 2015; Shenhav et al., 2013). Then, the arbitration weight on value-based decision-making given a stimulus s is a logistic function of the performances of affordance-based and value-based policies. Thus,

$$w_{q,t}(s) = \frac{1}{1 + exp(-\eta_1(Perf_{q,t}(s) - Perf_{aff,t}(s)) + \eta_0)}, \eta_1 > 0, \eta_0 \in \mathbb{R}$$

The value-based policy's performance $Perf_{q,t}(s)$ and affordance-based policy's performance $Perf_{aff,t}(s)$ given the stimulus s are updated trial by trial by the following delta rules which estimate the performances using inverse propensity scoring from the off-policy policy evaluation literature (Precup et al., 2000; Horvitz and Thompson, 1952).

$$Perf_{q,t+1}(s_t) \leftarrow Perf_{q,t}(s_t) + \alpha_p(\frac{P_{q,t}(a_t|s_t)}{P_t(a_t|s_t)}r_t - Perf_{q,t}(s_t))$$

$$Perf_{aff,t+1}(s_t) \leftarrow Perf_{aff,t}(s_t) + \alpha_p(\frac{P_{aff}(a_t|s_t)}{P_t(a_t|s_t)}r_t - Perf_{aff,t}(s_t))$$

$$Perf_{q,0}(s) = Perf_{aff,0}(s) = 0$$

$$P_t(a|s) = w_{q,t}(s)P_{q,t}(a|s) + (1 - w_{q,t}(s))P_{aff}(a|s)$$

where r_t denotes the realized outcome at the trial t, $P_{q,t}(a_t|s_t)$ is the probability of selecting the chosen action under the value-based policy at the trial t. Similarly, $P_{aff}(a_t|s_t)$ represents the probability of selecting the chosen action under the affordance-based policy at the trial t, and $P_t(a_t|s_t)$ denotes the probability of selecting the chosen action. $0 \leq \alpha_p \leq 1$ is a free parameter that denotes the performance

updating rate. We called $\frac{P_{q,t}(a_t|s_t)}{P_t(a_t|s_t)}r_t - Perf_{q,t}(s_t)$, and $\frac{P_{aff}(a_t|s_t)}{P_t(a_t|s_t)}r_t - Perf_{aff}(s_t)$ as the performance prediction error of value-based and affordance-based decision-making, respectively. Analogous to the other arbitration models, value-based and affordance-based policies are calculated in the same manner as the fixed arbitration model. The final action selection probability is determined by a performance-based arbitration-weighted sum of the two policies. Notably, we set the performance updating rate α_p to be identical to the value learning rate α. As a result, the performance-based arbitration model has a total of 8 free parameters ($\beta, \alpha, b_{pinch}, b_{clench}, b_{poke}, k, \eta_0$ and η_1).

Models with counterfactual value update. We also tested models that update values of unchosen actions when there was no reward. These were implemented using the following learning rule, while keeping the other components the same.

$$RPE_t = r_t - Q_t(s_t, a_t)$$

$$Q_{t+1}(s_t, a_t) \leftarrow Q_t(s_t, a_t) + \alpha \times RPE_t$$
$$Q_{t+1}(s_t, a'_t) \leftarrow Q_t(s_t, a'_t) - \frac{\alpha}{2} \times RPE_t \quad \text{(only when } r_t = 0\text{)}$$

where the reward prediction error (RPE) is the difference between the reward r_t (0 or 1 depending on the realized outcome) at the trial t and the action value of chosen action a_t at t. a'_t represents the unchosen action at t and $0 \leq \alpha \leq 1$ is a learning rate parameter. The model comparison and simulation results are shown in Fig.4.15

Model recovery and parameter recovery analyses. We conducted a model recovery analysis across the nine models to ensure that the underlying cognitive mechanism that generated data can be reliably identified using the task design. Using the fMRI version of the task, which has half the number of trials compared to the behavior-only version, each of the nine models was simulated 60 times with the estimated parameters from the real fMRI participants' data. We then fit the models to each simulated behavior using CBM toolkit and did model comparisons based on log evidence which take into account the number of parameters. Each simulated data was labeled with the best-fitting model, and we calculated the probability of the identified model being the true generative model (Table 4.4). It is noteworthy that the probability of the best-fit model being the true generative model for the performance-based and the reliability-based arbitration models were 0.89 and 0.79 each which implies that these models were reliably distinguishable through the model comparison process.

Additionally, we conducted a parameter recovery analysis on the performance-based and the reliability-based arbitration models which were employed for fMRI analyses, to ensure that the model-based variables reliably reflect the underlying cognitive mechanisms. Using the fMRI version of the task which has 480 trials in total, we generated 100 simulated data from each model, with parameters sampled from Gaussian distributions based on sample means and variances of the estimated parameters from the actual fMRI participants data. We estimated the model parameters using CBM toolkit and calculated correlations between the true generative parameters and the estimated parameters. The performance-based arbitration model had a mean Pearson correlation of $r = 0.76$ between true and estimated parameters, while the reliability-based arbitration model had a mean Pearson correlation of $r = 0.53$.

Behavior general linear model analyses. We conducted mixed-effect GLM analyses on the RT data from the behavior and fMRI studies to examine the effect of choosing affordance-compatible actions on RTs (Figs. 2a,f). As the RT is influenced by both the affordance-compatibility of the response and the value learning process, we incorporated a set of independent variables in the analyses: indicator variables for the affordance-compatibility of the response, whether the response was correct, whether the trial was in a congruent or incongruent condition, whether the trial was in a high or low condition, a categorical variable representing the movement type of the response (i.e., pinch, clench, or poke) and a continuous variable representing the trial index. The dependent variable was RT, which was logarithmically transformed to make it normally distributed. The slopes and intercepts were estimated as random effects for each participant and the fixed-effect coefficients are reported in (Table 4.1)

In addition, we performed mixed-effect GLM analyses to evaluate the impact of affordance-value congruency on the learning curves (Figs. 2d,e,i,j and Figs. 3f,g,l,m). The dependent variable was the difference between congruent and incongruent conditions in terms of frequencies of choosing correct actions, which were extracted for each trial index or the number of rewarded trials so far, for each subject. We included a continuous variable representing the trial index or the number of rewarded trials so far, as regressors in the analyses. The slopes and intercepts were estimated as random effects for each subject and the fixed-effect coefficients are reported in (Table 4.3)

We used the `mixedlm` function from the statsmodels 0.12.2 package in Python 3.7.7.

The package calculated z statistics of each coefficient by dividing the estimates of coefficients with standard errors. Using the z statistics, the p-values were calculated with respect to a standard normal distribution.

fMRI data acquisition. fMRI data were acquired at the Caltech Brain Imaging Center (Pasadena, CA), using a Siemens Prisma 3T scanner with a 32-channel radio-frequency coil. The functional scans were conducted using a multi-band echo-planar imaging (EPI) sequence with 72 slices, –30 degrees slice tilt from AC-PC line, 192 mm × 192 mm field of view, 2 mm isotropic resolution, repetition time (TR) of 1.12 s, echo time (TE) of 30ms, multi-band acceleration of 4, 54-degree flip angle, in-plane acceleration factor 2, echo spacing of 0.56 ms, and EPI factor of 96. Following each run, both positive and negative polarity EPI-based field maps were collected using similar parameters to the functional sequence, but with a single band, TR of 5.13 s, TE of 41.40 ms, and 90-degree flip angle. T1-weighted and T2-weighted structural images were also acquired for each participant with 0.9 mm isotropic resolution and 230 mm × 230 mm field of view. For the T1-weighted scan, TR of 2.55 s, TE of 1.63 ms, inversion time (TI) of 1.15 s, flip angle of 8 degrees, and in-plane acceleration factor 2 were used. The T2-weighted scan was acquired with TR of 3.2 s, TE of 564 ms, and in-plane acceleration factor of 2.

fMRI data preprocessing. Results included in this manuscript come from preprocessing performed using *fMRIPrep* 20.2.6 (Esteban et al., 2018b,a, RRID:SCR_016216), which is based on *Nipype* 1.7.0 (Gorgolewski et al., 2011, 2018, RRID:SCR_002502).

Anatomical data preprocessing. A total of 1 T1-weighted (T1w) images were found within the input BIDS dataset.The T1-weighted (T1w) image was corrected for intensity non-uniformity (INU) with `N4BiasFieldCorrection` Tustison et al. (2010), distributed with ANTs 2.3.3 (Avants et al., 2008, RRID:SCR_004757). The T1w-reference was then skull-stripped with a *Nipype* implementation of the `antsBrainExtraction.sh` workflow (from ANTs), using OASIS30ANTs as target template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using `fast` (Zhang et al., 2001, FSL 5.0.9, RRID:SCR_002823,). Brain surfaces were reconstructed using `recon-all` (Dale et al., 1999, FreeSurfer 6.0.1, RRID:SCR_001847,), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived

segmentations of the cortical gray-matter of Mindboggle (Klein et al., 2017, RRID:SCR_002438,). Volume-based spatial normalization to one standard space (MNI152NLin2009cAsym) was performed through nonlinear registration with `antsRegistration` (ANTs 2.3.3), using brain-extracted versions of both T1w reference and the T1w template. The following template was selected for spatial normalization: *ICBM 152 Nonlinear Asymmetrical template version 2009c* (Fonov et al., 2009, RRID:SCR_008796; TemplateFlow ID:MNI152NLin2009cAsym).

Functional data preprocessing. For each of the 6 BOLD runs per subject (across all tasks and sessions), the following preprocessing was performed. First, a reference volume and its skull-stripped version were generated by aligning and averaging 1 single-band references (SBRefs). A B0-nonuniformity map (or *fieldmap*) was estimated based on two (or more) echo-planar imaging (EPI) references with opposing phase-encoding directions, with `3dQwarp` Cox and Hyde (1997) (AFNI 20160207). Based on the estimated susceptibility distortion, a corrected EPI (echo-planar imaging) reference was calculated for a more accurate co-registration with the anatomical reference. The BOLD reference was then co-registered to the T1w reference using `bbregister` (FreeSurfer) which implements boundary-based registration (Greve and Fischl, 2009). Co-registration was configured with six degrees of freedom. Head-motion parameters with respect to the BOLD reference (transformation matrices, and six corresponding rotation and translation parameters) are estimated before any spatiotemporal filtering using `mcflirt` (Jenkinson et al., 2002, FSL 5.0.9,). BOLD runs were slice-time corrected to 0.52s (0.5 of slice acquisition range 0s-1.04s) using `3dTshift` from AFNI 20160207 (Cox and Hyde, 1997, RRID:SCR_005927). First, a reference volume and its skull-stripped version were generated using a custom methodology of *fMRIPrep*. The BOLD time-series (including slice-timing correction when applied) were resampled onto their original, native space by applying a single, composite transform to correct for head-motion and susceptibility distortions. These resampled BOLD time-series will be referred to as *preprocessed BOLD in original space*, or just *preprocessed BOLD*. The BOLD time-series were resampled into standard space, generating a *preprocessed BOLD run in MNI152NLin2009cAsym space*. First, a reference volume and its skull-stripped version were generated using the custom methodology of *fMRIPrep*. Several confounding time-series were calculated based on the *preprocessed BOLD*: framewise displacement (FD), DVARS and three region-wise global signals. FD was computed using two formulations following Power (absolute sum

of relative motions, Power et al. (2014)) and Jenkinson (relative root mean square displacement between affines, Jenkinson et al. (2002)). FD and DVARS are calculated for each functional run, both using their implementations in *Nipype* (following the definitions by Power et al. (2014)). The three global signals are extracted within the CSF, the WM, and the whole-brain masks. Additionally, a set of physiological regressors were extracted to allow for component-based noise correction (Behzadi et al., 2007, *CompCor*). Principal components are estimated after high-pass filtering the *preprocessed BOLD* time-series (using a discrete cosine filter with 128s cut-off) for the two *CompCor* variants: temporal (tCompCor) and anatomical (aCompCor). tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For aCompCor, three probabilistic masks (CSF, WM and combined CSF+WM) are generated in anatomical space. The implementation differs from that of Behzadi et al. in that instead of eroding the masks by 2 pixels in BOLD space, the aCompCor masks subtract a mask of pixels that likely contain a volume fraction of GM. This mask is obtained by dilating a GM mask extracted from FreeSurfer's *aseg* segmentation, and it ensures components are not extracted from voxels containing a minimal fraction of GM. Finally, these masks are resampled into BOLD space and binarized by thresholding at 0.99 (as in the original implementation). Components are also calculated separately within the WM and CSF masks. For each CompCor decomposition, the *k* components with the largest singular values are retained, such that the retained components' time series are sufficient to explain 50 percent of variance across the nuisance mask (CSF, WM, combined, or temporal). The remaining components are dropped from consideration. Head-motion estimates calculated in the correction step were also placed within the corresponding confounds file. The confound time series derived from head motion estimates and global signals were expanded with the inclusion of temporal derivatives and quadratic terms for each Satterthwaite et al. (2013). Frames that exceeded a threshold of 0.5 mm FD or 1.5 standardised DVARS were annotated as motion outliers. All resamplings were performed with *a single interpolation step* by combining all the pertinent transformations (i.e., head-motion transform matrices, susceptibility distortion correction when available, and co-registrations to anatomical and output spaces). Gridded (volumetric) resamplings were performed using `antsApplyTransforms` (ANTs), configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964). Non-gridded (surface) resamplings were performed using `mri_vol2surf` (FreeSurfer).

Many internal operations of *fMRIPrep* use *Nilearn* 0.6.2 (Abraham et al., 2014,

RRID:SCR_001362), mostly within the functional processing workflow. For more details of the pipeline, see the section corresponding to workflows in *fMRIPrep*'s documentation.

Whole brain general linear model analysis. FSL (FMRIB Software Library) 6.0 software package was used to analyze the fMRI data. Following the preprocessing using fMRIPrep, the fMRI data was convolved with a 3D Gaussian smoothing kernel (8.0mm FWHM) using FSL SUSAN. Subsequently, BOLD signals for each voxel were scaled to ensure that the mean BOLD signal of each voxel is 100. Before running GLM analyses, first 4 volumes were discarded and a high-pass filter with a cutoff 128s was applied. Framewise displacement, global signal, global signal derivative, 6 head-motion parameters (x, y, z translations and rotations), and temporal CompCor from fMRIPrep were included as nuisance variables in every analysis. Stick regressors that represent onsets of stimulus, fixation cross and outcome were also included in the GLMs. The response, or the hand movement, was modeled using a parametric regressor detailed in the following section. We used the fixed-effect higher-level modeling for individual-level inference that averages across blocks and FSL FLAME1 for group-level inference.

Hand movement parametric regressor. In order to minimize potential confounding effects that may arise from using natural hand gestures as the action in the experiment, we designed a hand movement regressor based on 2-dimensional 21 key points for the right hand extracted from the hand video using Openpose. This regressor was included in all of the analyses.

First, video frames were resized into 162×108 pixels to fit the Openpose net resolution -1×122. By utilizing the BODY_25 model from the Openpose, key points were extracted frame by frame with the output resolution setting -1×80. Next, the key points were weighted-moving-averaged with a window of 20 frames (about 500ms) to reduce the noise stemming from Openpose's estimation error. The confidence scores of Openpose estimations were utilized as the weights in order to mitigate the impact of uncertain estimations. Then, the first principal component of the weighted-moving-averaged 42-dimensional data (21×2) was calculated and used as the parametric regressor that models the hand movement. In all of our GLM analyses, the hand movement regressor consistently identified the left primary motor cortex, left premotor cortex, left primary somatosensory cortex, and bilateral inferior parietal lobule as the regions associated with the right-hand movement (Fig.

4.19).

Identifying decision-making regions. A GLM analysis was conducted to investigate the regions associated with the affordance-based and value-based decision-making, as well as the action selection which is the integration of those two systems (GLM1). The chosen action's affordance-compatibility score ($Aff(s_t, a_t)$), chosen action's value ($Q_t(s_t, a_t)$), the action selection probability of chosen action ($P_t(a_t|s_t)$) from the performance-based arbitration model models were included as parametric regressors at the onsets of stimuli. A parametric regressor of the reward prediction error was also included at the onsets of outcomes. Another GLM analysis using model-based parameters from the reliability-based arbitration model was conducted as well (GLM2). The results were cluster corrected with a cluster-defining threshold of $z = 3.1$.

Identifying arbitration regions. Two GLM analyses were conducted to identify regions related to arbitration between value-based and affordance-based decision-making. One GLM analysis utilized arbitration-related parameters obtained from the performance-based arbitration model (GLM3) and the other one used those from the reliability-based arbitration model (GLM4).

GLM3 included performances of the value-based and affordance-based decision-making systems ($Perf_{q,t}(s_t), Perf_{aff,t}(s_t)$), and the difference between those two ($Perf_{aff,t}(s_t) - Perf_{q,t}(s_t)$) as parametric regressors at the onsets of stimuli. Parametric regressors of the performance prediction errors of value-based and affordance-based policies each were also included at the onsets of outcomes in the GLM3.

In GLM4, reliabilities of the value-based and affordance-based decision-making systems ($\chi_{q,t}(s_t), \chi_{aff,t}(s_t)$), and the difference between those two ($\chi_{aff,t}(s_t) - \chi_{q,t}(s_t)$) were included as parametric regressors at the onsets of stimuli. A parametric regressor of the affordance prediction error was included as well at the onsets of outcomes in the GLM4. The results were cluster corrected with a cluster-defining threshold $z = 3.1$.

Neural correlates of affordance-incompatible action execution. To identify neural correlates that are activated more when executing affordance-incompatible actions compared to affordance-compatible actions (Grèzes et al., 2003; Zhang et al.,

2021), a model-free GLM analysis was conducted (GLM5). GLM5 included stick regressors representing the stimuli identities at the stimulus onsets, 2 regressors representing whether the affordance and response were compatible or not at the response onsets, and a parametric regressor of reward at the outcome onsets. While the contrast between the affordance response compatibility regressors have not survived the cluster correction with a cluster-defining threshold $z = 3.1$, Fig. 19g shows the uncorrected results with a threshold set at $p < 0.001$, corresponding to $z \geq 3.1$.

Neural representation strength analysis To calculate the extent to which the identified regions from GLM analyses can be explained by the corresponding cognitive variables, we computed the mean z-statistics for those specific variables across the voxels within the corresponding group-level regions. The regions were defined using the survived clusters from GLM1,2,3 and 4 (For the details, see Fig. 4.4 and Fig. 17). We then compared the mean z-statistics for each individual with the frequency of choosing the most rewarding action.

Similarly, the increased BOLD activity for executing affordance-incompatible actions compared to executing affordance-compatible actions was calculated using the contrast between the affordance-response compatibility regressors from GLM5. The z-statistics of the contrast were averaged across the voxels within functionally-defined regions of interest (ROI). The ROIs were defined in the same way as in the previous analyses. We then compared the average z-statistics for each individual with the frequency of choosing the affordance compatilble action or the reaction time effect which is the relative increase in reaction time when choosing affordance-incompatible actions compared to affordance-compatible actions (RT effect = $\frac{RT_{incomp} - RT_{comp}}{RT_{comp}}$).

4.6 Acknowledgments

We thank Shinsuke Shimojo, Jeff Cockburn, Vincent Man and Seungyong Moon for discussion and suggestions, Weilun Ding, Thomas Henning, Seokyoung Min, Jihong Min, Areum Kim, HyeongChan Jo, and Serim Ryou for their help in implementing the task. **Funding:** This work was supported by a graduate innovator grant award from Caltech's Tianqiao and Chrissy Chen Institute for Neuroscience to S.Y. **Author contributions:** S.Y. and J.P.O. conceived and designed the study, S.Y. performed experiments and S.Y. and J.P.O. analyzed and discussed results. S.Y. and J.P.O. wrote the manuscript. **Competing interests:** The authors declare that they have no

competing interests. **Data and materials availability:** The data, code and analysis results utilized in this manuscript will be available online at the time of publication.

4.7 Supplementary information

Note1: Derivation of performance-based arbitration model

The objective of the performance-based arbitration model in the context of the given research paradigm, or for general contextual bandit problems, is to maximize the return with respect to the arbitration weight. More precisely, we consider a scenario in which a behavior policy, denoted as π, is a linear mixture of multiple component policies, represented as π_i.

$$\pi(a|s) = \sum_i w_i(s)\pi_i(a|s),\ \sum_i w_i = 1,\ 0 \leq w_i \leq 1 \tag{4.1}$$

In our experiment, we assumed the final action selection probability corresponds to the behavior policy and the value-based and affordance-based decision-makings are the component policies.

If we define a value, or performance, of a component policy π_i with respect to a given context, or stimulus, s as the expected return obtained by following the policy π_i when the stimulus s is shown:

$$Perf_{\pi_i}(s) := \sum_a \pi_i(a|s)R(s,a) \tag{4.2}$$

Then the performance of the policy π at the state s can be decomposed into the linear summation of the performances of the component policies.

$$Perf_\pi(s) := \sum_a \pi(a|s)R(s,a) = \sum_a \sum_i w_i(s)\pi_i(a|s)R(s,a) \tag{4.3}$$

$$= \sum_i w_i(s) \sum_a \pi_i(a|s)R(s,a) = \sum_i w_i(s)Perf_{\pi_i}(s) \tag{4.4}$$

Therefore, determining arbitration weights in terms of maximize the performance of the behavior policy π can be formulated as identifying one of the component policies that yields the highest performance and assigning weight of 1 to it.

$$max_{w_i(s)}(Perf_\pi(s)) = max_i(Perf_{\pi_i}(s)) \tag{4.5}$$

Therefore, the arbitration problem can be solved if the performances of the component policies can be estimated. Given a trajectory of actions, and outcomes generated by the behavior policy π at a state s, the approximation of the performance of a component policy π_i can be achieved by employing various methods

from the off-policy evaluation literature. One such method is the inverse propensity scoring estimator (Precup et al., 2000; Horvitz and Thompson, 1952), which can be written in the following form:

$$Perf_{\pi_i}(s) \approx \frac{1}{T}\sum_{t=1}^{T}\frac{\pi_i(a_t|s)}{\pi(a_t|s)}r_t \tag{4.6}$$

where T is the number of trials at the state s in the trajectory and r_t is the realized outcome at the t-th trial at the state s. In our scenario of having a value-based policy as one of the component policies, the behavior policy and the value-based policy are nonstationary and get updated every trial. Therefore, it is more practical to assign greater weight to recent events while estimating the performance of component policies. In this context, the above equation can be approximated trial by trial by the following delta rule.

$$Perf_{\pi_i,t+1}(s_t) \leftarrow Perf_{\pi_i,t}(s_t) + \alpha_p(\frac{\pi_i(a_t|s_t)}{\pi(a_t|s_t)}r_t - Perf_{\pi_i,t}(s_t)),\ 0 \leq \alpha_p \leq 1 \tag{4.7}$$

where we called $(\frac{\pi_i(a_t|s_t)}{\pi(a_t|s_t)}r_t - Perf_{\pi_i,t}(s_t))$ the performance prediction error of π_i and s_t is the stimulus shown at the trial t. However, in practical scenarios, accurately determining the performance of each component policy is often unfeasible. Instead, the optimal weighting can be approximated by weighting each component policy with the probability that the policy's performance is larger than any other component policies.

$$max_{w_i(s)}(Perf_\pi(s)) \approx \sum_i P(Perf_{\pi_i}(s) > Perf_{\pi_j}(s), \forall j \neq i) Perf_{\pi_i}(s) \tag{4.8}$$

The ideal approach for calculating the probability $P(Perf_{\pi_i}(s) > Perf_{\pi_j}(s), \forall j \neq i)$ would involve approximating the full distribution of the performance of each component policy which can be implemented using, for example, distributional RL (Dabney et al., 2020). However, a simpler approach to approximate such probability would be to employ a softmax function of the performances estimated by the learning rule (Eq. 4.7) which approximates means of performances.

It is noteworthy that the above framework for the multi-armed bandit problem can be extended to a general Markov decision process. For example, the performance of a component policy π_i at a state s can be defined as the expected return by executing π_i at the state s and following π thereafter.

$$Perf_{\pi_i}(s) := \sum_a \pi_i(a|s)(R(s,a) + \gamma \sum_{s'} P(s'|s,a) Perf_\pi(s')) \tag{4.9}$$

$$Perf_\pi(s) := \sum_a \pi(a|s)(R(s,a) + \gamma \sum_{s'} P(s'|s,a) Perf_\pi(s')) = V_\pi(s) \quad (4.10)$$

where $P(s'|s,a)$ is a state transition probability from the current state s to the subsequent state s' after selecting an action a, and γ is a discounting rate. Then, the performance of the behavior policy can be expressed as the linear summation of component policies' performances.

$$Perf_\pi(s) = \sum_i w_i(s) Perf_{\pi_i}(s) \quad (4.11)$$

As a result, the problem of maximizing the behavior policy's performance with respect to arbitration weights can be reduced to the task of identifying the component policy with the highest performance.

Because $\sum_j w_j(s') Perf_{\pi_j}(s') \approx max_j Perf_{\pi_j}(s')$, the performance of π_i can be written in Bellman equation form:

$$Perf_{\pi_i}(s) = \sum_a \pi_i(a|s)(R(s,a) + \gamma \sum_{s'} P(s'|s,a) \sum_j w_j(s') Perf_{\pi_j}(s')) \quad (4.12)$$

$$\approx \sum_a \pi_i(a|s)(R(s,a) + \gamma \sum_{s'} P(s'|s,a) max_j Perf_{\pi_j}(s')) \quad (4.13)$$

Analogous to Eq. 4.7, the performances of component policies can be estimated by using the following delta rule:

$$Perf_{\pi_i,t+1}(s_t) \leftarrow Perf_{\pi_i,t}(s_t) + \alpha_p (\frac{\pi_i(a_t|s_t)}{\pi(a_t|s_t)} (r_t + \gamma max_j Perf_{\pi_j,t}(s_{t+1})) - Perf_{\pi_i,t}(s_t)) \quad (4.14)$$

Note2: Additional model simulation results

In the main text, we demonstrated that the learning slope in the low conditions is not steeper than it in the high conditions when the learning slopes were plotted as the function of the number of rewarded trials previously encountered. To explore this observation further, a fixed-arbitration model incorporating two distinct learning rate parameters was fit to the data. This additional analysis also revealed no significant difference in learning rates between high and low conditions ($t(18) = -1.57, p = 0.13$ in the behavioral, $t(29) = -0.47, p = 0.65$ in the fMRI data).

There is a potential that participants adopted a simpler strategy, such as win-stay lose-shift, instead of value learning. However, this possibility can be discounted based on the estimated learning rates derived from model fitting. If participants were using a win-stay lose-shift strategy, the estimated learning rates would be expected to be

near 1. In contrast, the learning rates obtained using the performance-based model were 0.227±0.206 and 0.269±0.173 for the behavioral and fMRI data, respectively.

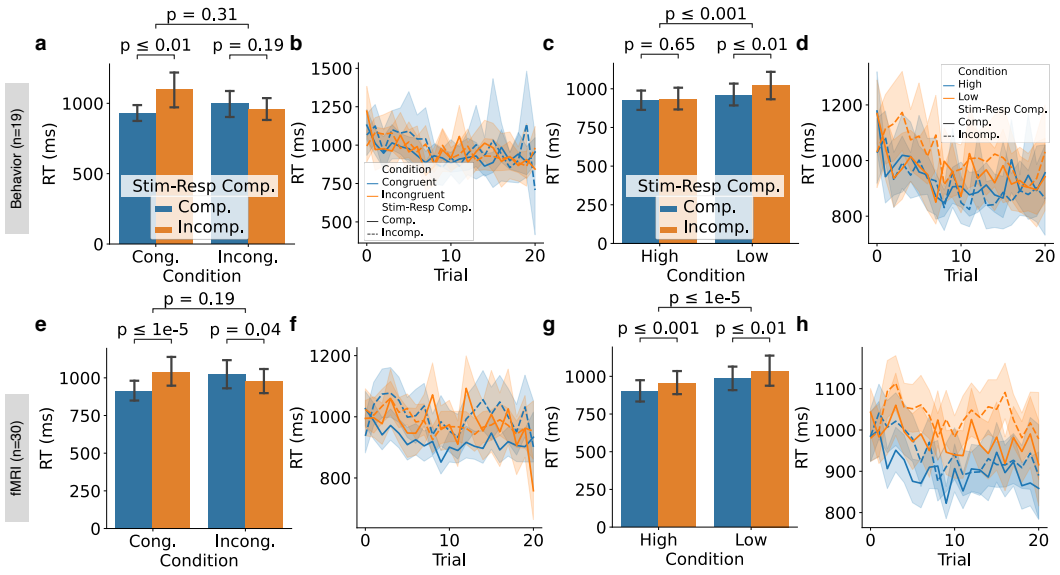


Figure 4.6: Effects of stimulus-response compatibility and value learning on reaction times. (a and e) The RTs for making affordance-compatible actions were significantly faster in the congruent condition, whereas the trend was in the opposite direction in the incongruent condition. This is because affordance-compatible actions have higher action values in the congruent condition, but not in the incongruent condition (paired t tests, $t(18) = -3.30$ and $t(18) = 1.36$ for comparisons within congruent and incongruent conditions, respectively, in the behavioral, $t(29) = -5.87$ and $t(29) = 2.17$ for the same comparisons in the fMRI experiment). But there was no significant overall RT difference on average between congruent and incongruent conditions (paired t tests, $t(18) = 1.02$ and $t(29) = -1.32$ for the behavioral and fMRI experiments, respectively). (b and f) Similar to plots A and E, but the RTs were averaged across blocks. Across learning, as action value learning progresses, RTs become faster. (c and g) RT contrasts plotted similarly to plots A and E but now examined separately within high and low value conditions. The RTs were slower in the low value condition compared to the high value condition (paired t tests, $t(18) = -3.79$ and $t(29) = -7.60$ for the behavioral and fMRI experiments, respectively). Also, RTs for executing affordance compatible actions were faster in general regardless of the value manipulation (paired t tests, $t(18) = -0.46$ and $t(18) = -3.58$ for comparisons within high and low conditions, respectively, in the behavioral, $t(29) = -3.99$ and $t(29) = -3.07$ for the same comparisons in the fMRI experiment). (d and h) Analogous to B and F, the RTs depicted in plots C and G were averaged across blocks.

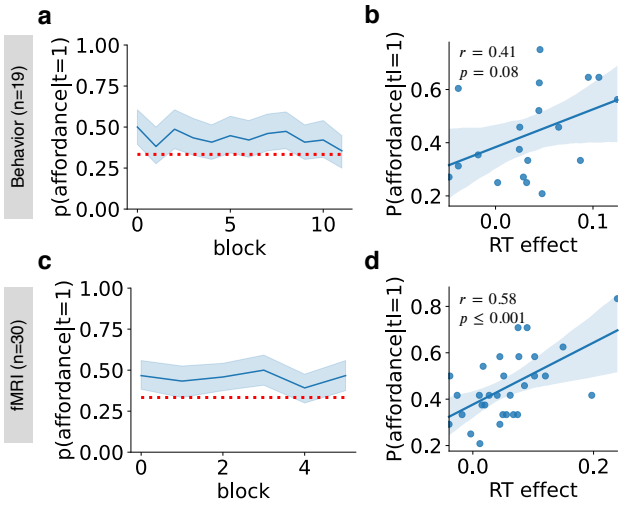
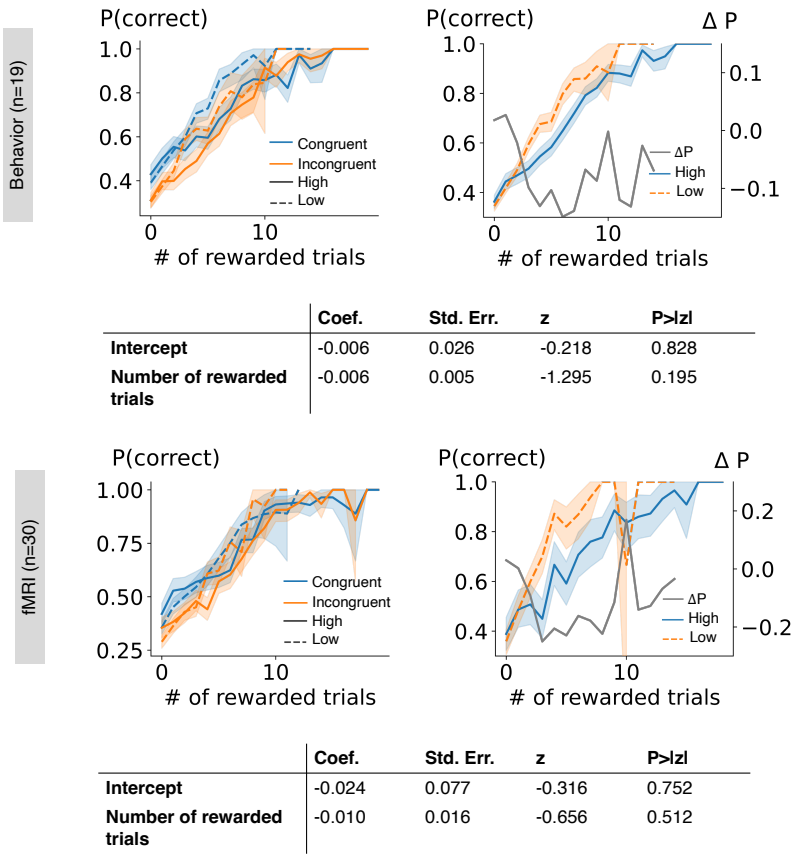


Figure 4.7: Behavioral effects of affordance during value learning and decision making. Upper-row plots depict results from the behavioral study while the lower-row plots are from the fMRI study. (a and c) Initial action selection bias toward the affordance-compatible action over blocks. The red dotted lines show the chance level of 1/3. The mixed linear regression of initial selection bias on block index showed non-significant slopes in both datasets (Fixed-effect coefficients for the slope: β = -0.005, z=-1.115 and p=0.265 in the behavioral data, β=-0.002, z=-0.232 and p=0.817 in the fMRI data).(b and d) The additional RT for executing an affordance-incompatible action relative to an affordance-compatible action, or RT effect, is positively correlated with the participant's tendency to select the affordance-compatible action as the initial response to each object.



	Coef.	Std. Err.	z	P>|z|
Intercept	-0.006	0.026	-0.218	0.828
Number of rewarded trials	-0.006	0.005	-1.295	0.195

	Coef.	Std. Err.	z	P>|z|
Intercept	-0.024	0.077	-0.316	0.752
Number of rewarded trials	-0.010	0.016	-0.656	0.512

Figure 4.8: Learning slopes by conditions and their differences between high and low conditions. The table shows the fixed-effect coefficients obtained from the mixed-effect GLM analysis on the difference in the choice accuracy between high and low conditions, with the number of rewarded trials previously encountered being the regressor. $\Delta P = P(correct|high) - P(correct|low)$

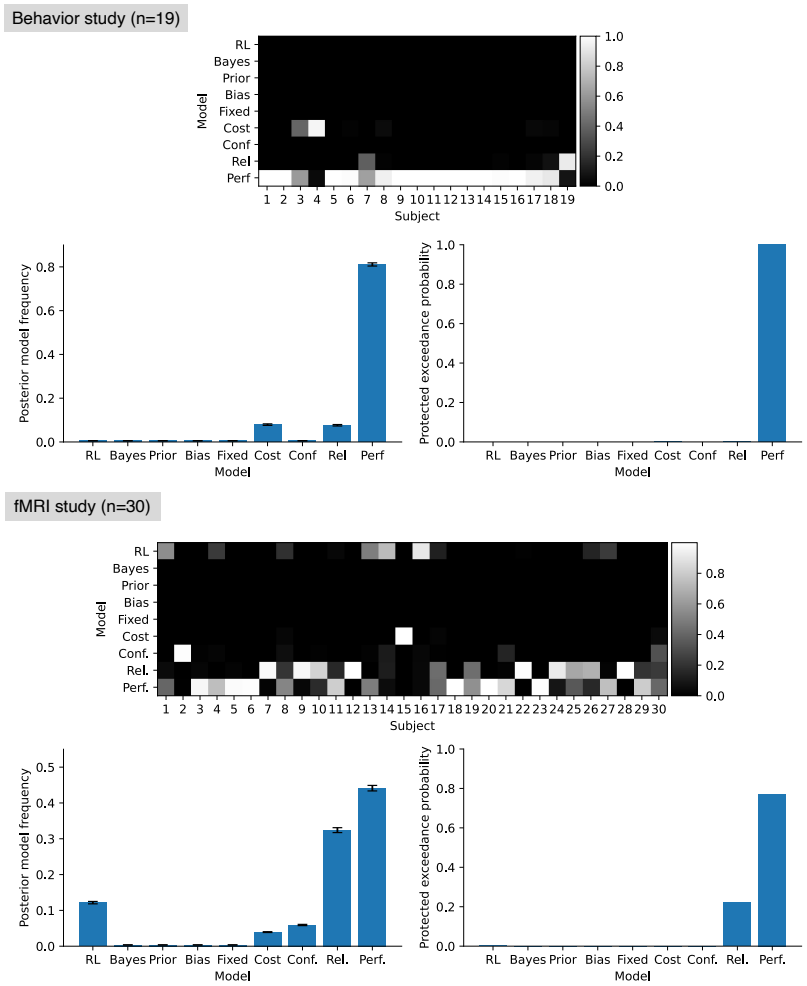


Figure 4.9: Bayesian model selection results. The upper panel displays the posterior probability of each model to best explain each participant. The lower-left panel shows model frequencies over models, while the lower-right panel shows protected exceedance probabilities of models. Overall, the performance-based arbitration model best fit the datasets.

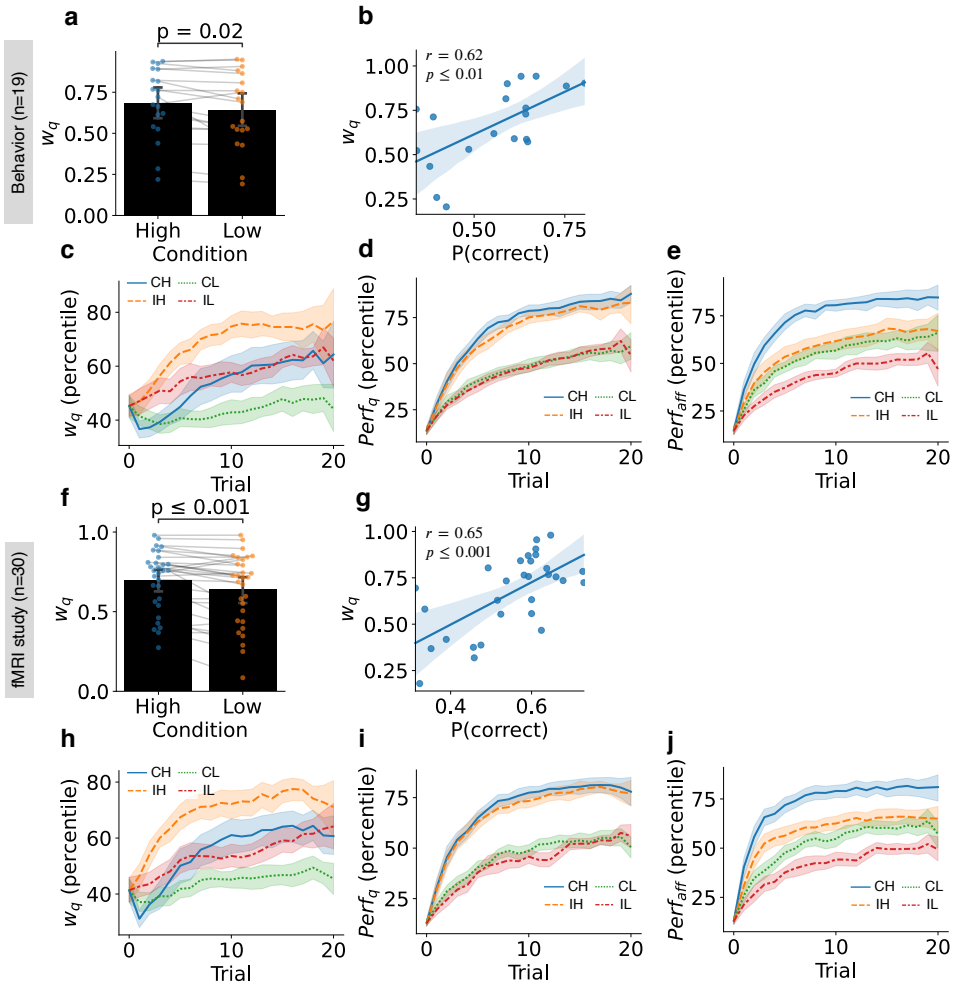


Figure 4.10: Arbitration variables extracted from the performance-based arbitration model. (a and f) Modulation of arbitration by the reward probability manipulation. Arbitration weights are calculated using the performance-based arbitration model for each trial and averaged for each participant and condition. The results suggest that value-based decision-making is more favored in high-value conditions (paired t tests, $t(18) = 2.53, t(29) = 3.77$ each). (b and g) Participants who have a larger average weight toward the value-based decision-making across trials exhibited better accuracy in choosing the most rewarding action. (c and h) Modulation of the performance-based arbitration weight over trials by conditions. CH: congruent high; CL: congruent low; IH: incongruent high; IL: incongruent low. The arbitration weights were transformed into the percentiles within each subject. (d and i) Modulation of the performance of the value-based decision-making system over trials by conditions. (e and j) Modulation of the performance of the affordance-based decision-making system over trials by conditions. The performances were transformed into the percentiles within each subject. All the curves were averaged across blocks and the error-bar shows 95% interval of estimated statistics.

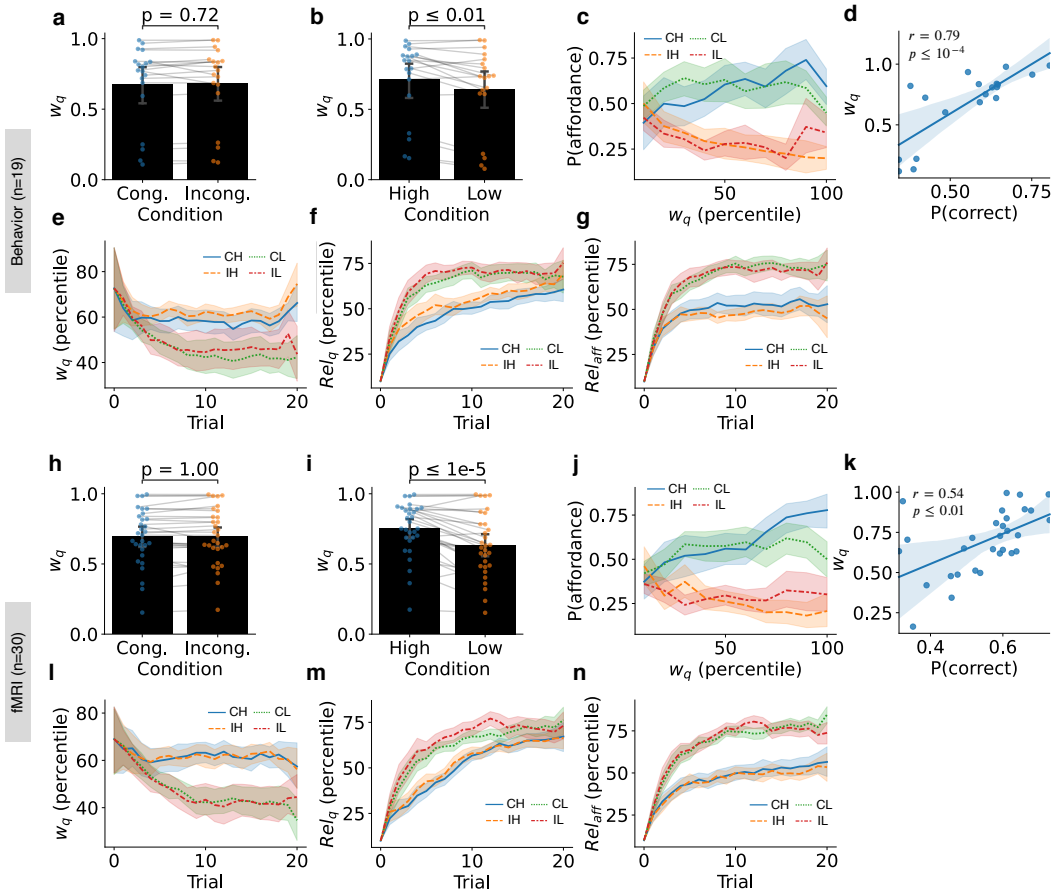

Behavior (n=19)
fMRI (n=30)
a
p = 0.72
b
p ≤ 0.01
c
d
r = 0.79
p ≤ 10^−4
e
f
g
h
p = 1.00
i
p ≤ 1e-5
j
k
r = 0.54
p ≤ 0.01
l
m
n
Cong.
Incong.
Condition
High
Low
CH
CL
IH
IL
Trial
P(affordance)
P(correct)

Figure 4.11: Arbitration variables extracted from the reliability-based arbitration model. (a and h) Modulation of arbitration by affordance-value congruency. Arbitration weights were calculated using the reliability-based arbitration model for each trial and averaged for each participant and condition. The congruency between affordance and value did not affect the reliability-base arbitration weight. (paired t tests, $t(18) = 0.36, t(29) = 0.00$ each) (b and i) Modulation of arbitration by the reward probability manipulation. These results suggest that value-based decision-making is more favored in the high-value condition. (paired t tests, $t(18) = 3.86, t(29) = 5.44$ each) (c and j) Frequency of choosing affordance-compatible actions in the data as a function of the arbitration weight on value-based decision-making extracted from the reliability-based arbitration model. The arbitration weights were transformed into the percentiles within each participant. The results show that in the trials when the model preferred affordance-based decision-making, the actual choices were more biased toward affordance-compatible actions. Notably, this trend was only evident in incongruent conditions as the responses based on affordance and value were indistinguishable in congruent conditions. CH: congruent high; CL: congruent low; IH: incongruent high; IL: incongruent low. (d and k) Participants who have a larger average weight toward value-based decision-making across trials exhibited better accuracy in choosing the most rewarding action. (e and l) Modulation of the reliability-based arbitration weight over trials by conditions. The arbitration weights were transformed into percentiles within each subject. (f and m) Modulation of reliability of the value-based decision-making system over trials by conditions. (g and n) Modulation of reliability of the affordance-based decision-making system over trials by conditions. The reliabilities were transformed into the percentiles within each subject. All the curves were averaged across blocks and the error-bar shows 95% interval of estimated statistics.

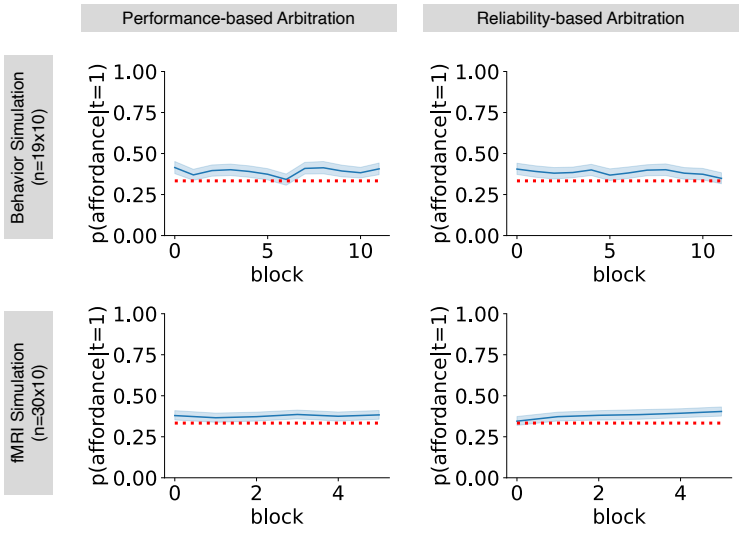


Figure 4.12: Simulated initial action selection bias toward the affordance-compatible action over blocks. The red dotted lines show the chance level 1/3.

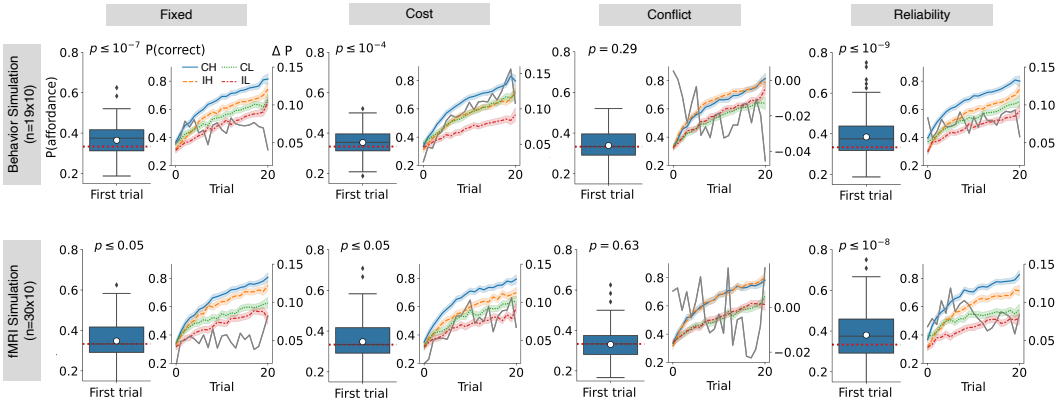


Figure 4.13: Simulated behaviors generated by various computational models. Each plot corresponds to Fig. 3e,k or 3f,l in which behaviors were generated by the computational model labeled on the top of each column. The t statistics of the initial response bias toward affordance compatible actions are $t(189) = 5.56, 4.07, 1.06,$ and 6.59 each for the simulations based on the behavioral experiment, $t(299) = 2.55, 2.45, -0.48,$ and 6.00 each for the simulations based on the fMRI experiment.

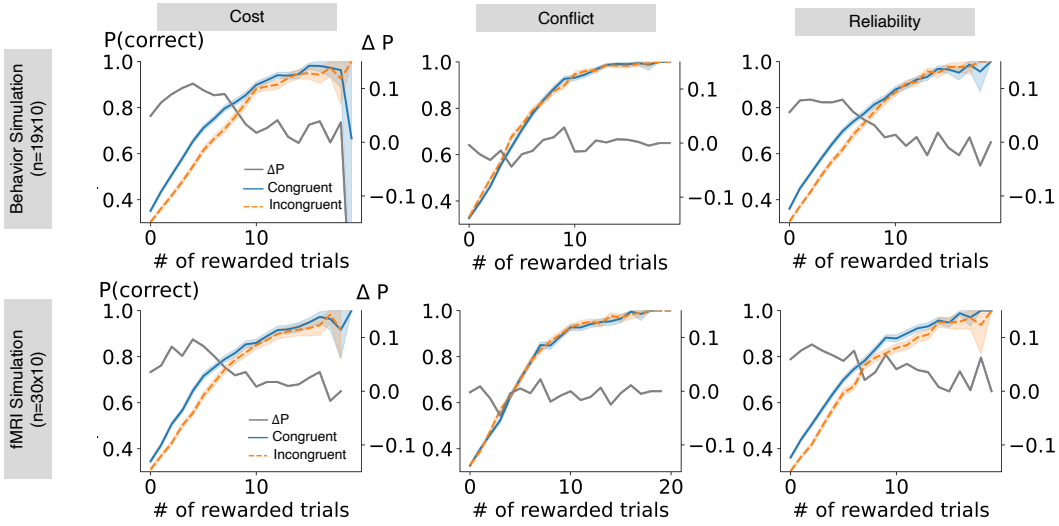


Figure 4.14: Simulated learning slopes and their differences produced by various computational models. Each plot corresponds to Fig. 3g or 3m in which behaviors were generated by the computational model labeled on top of each column.

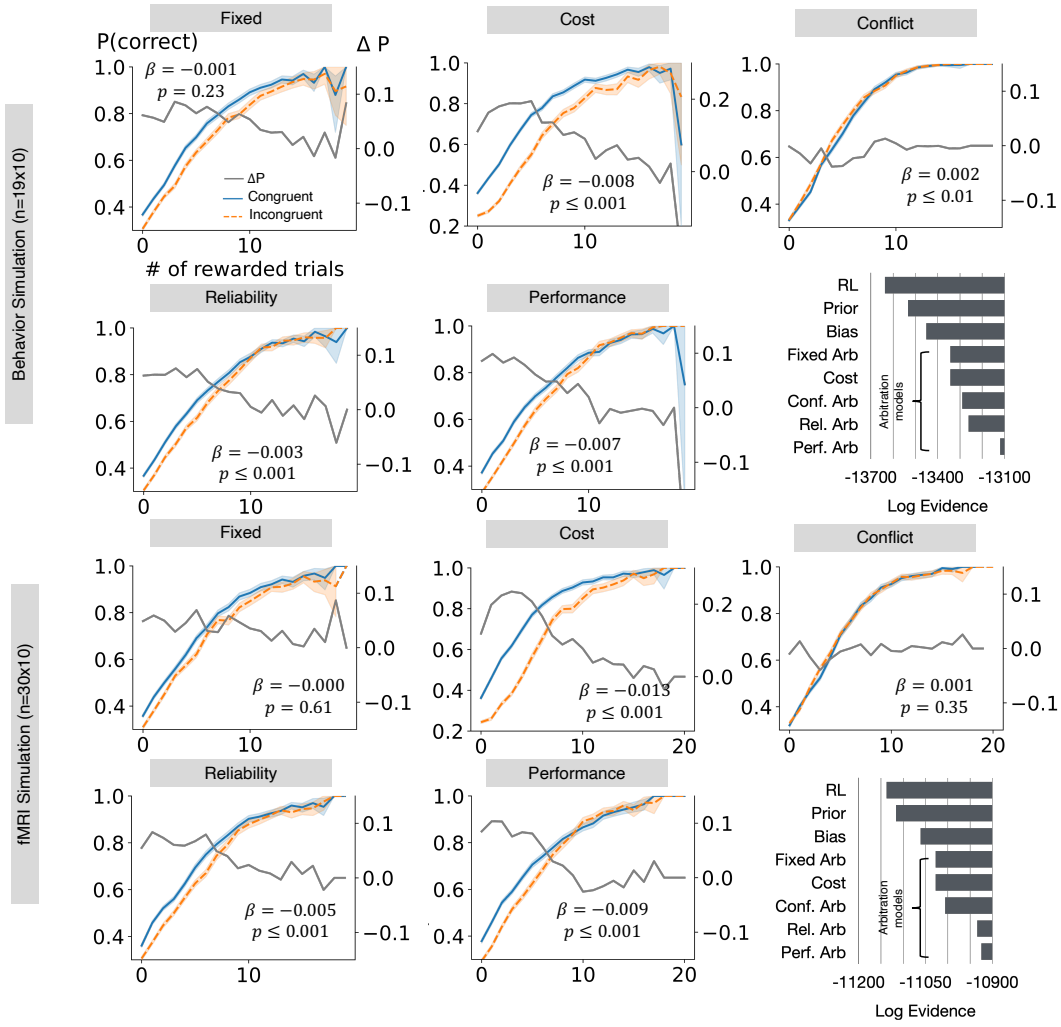


Figure 4.15: Simulated learning slopes and their differences produced by computational models that includes value updates for unchosen actions in unrewarded trials. Each plot corresponds to Figs. 3g or m in which behaviors were generated by the computational model labeled on top of each column. The β and $p-$value represent the fixed-effect coefficient of the slope of the difference between learning slopes and its corresponding $p-$value. The bar graphs represent the model fitting results for models incorporating value updates in unchosen actions. The simulation results from models with unchosen action value updates were generally comparable with the model simulation results without the unchosen action value update, except for the cost model which exhibited much steeper learning slope differences and different ranges for both the learning slopes and their differences compared to the actual data shown in Figs. 2e or j.

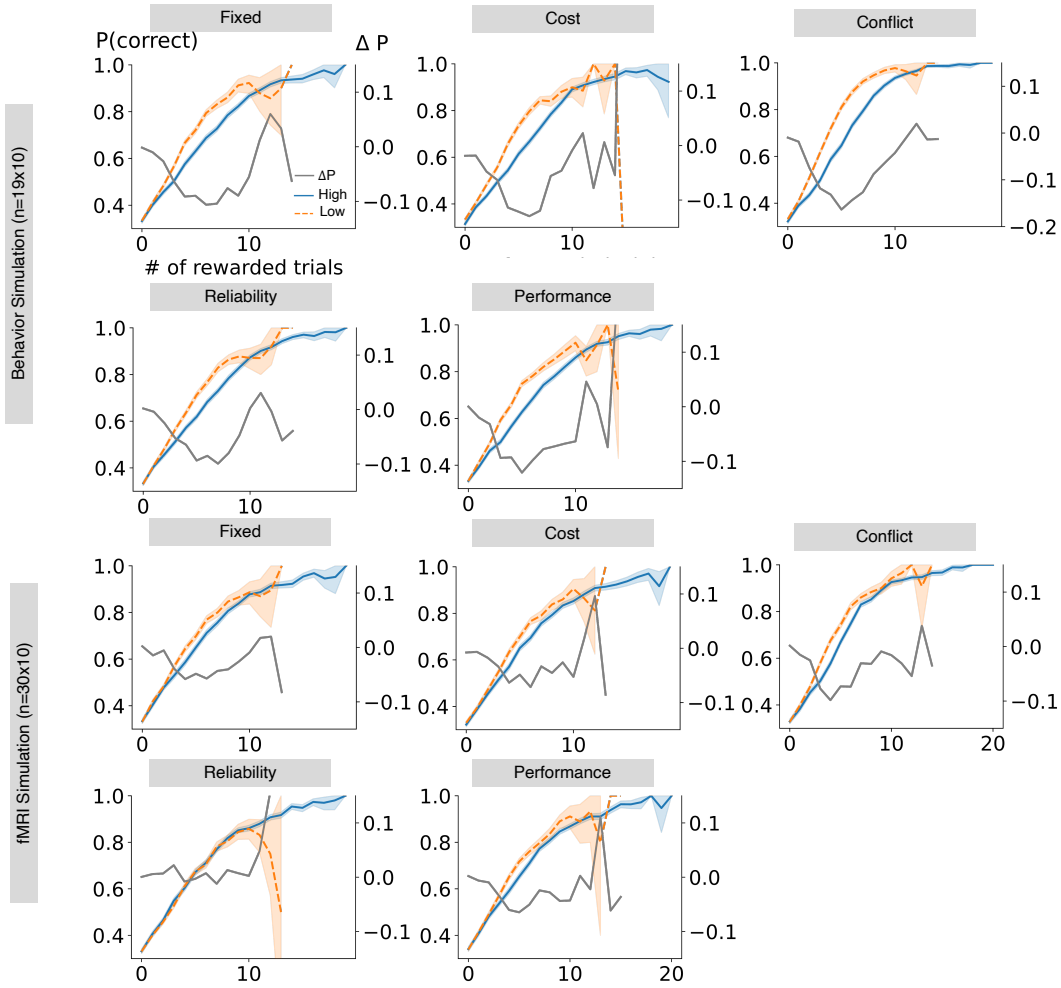


Figure 4.16: Simulated learning slopes and their differences produced by various computational models. Each plot corresponds to Fig.8 in which behaviors were generated by the computational model labeled on top of each subplot.

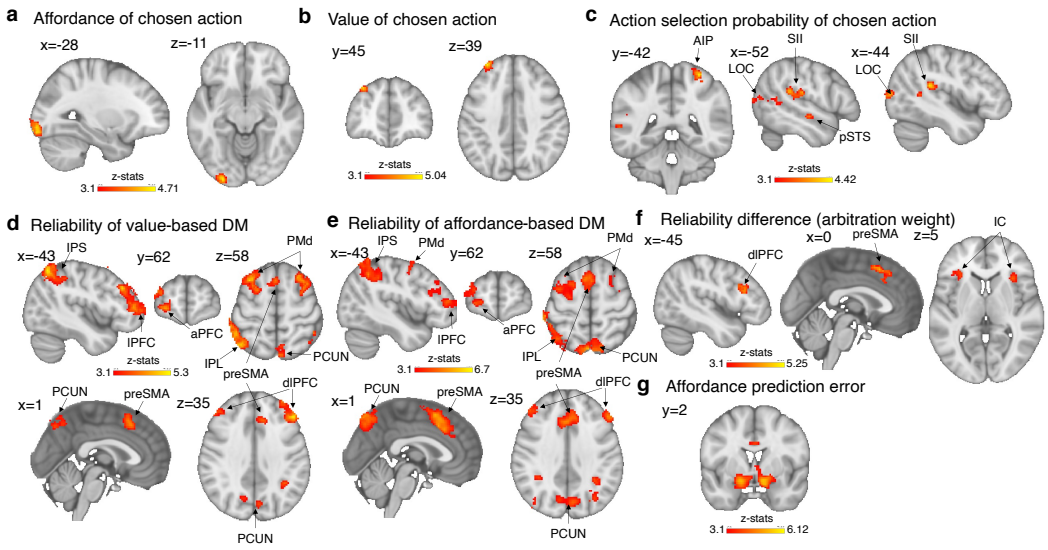


Figure 4.17: Neural implementation of the reliability-based arbitration model. (a) Affordance-compatibility scores of the chosen action correlated with the high-level ventral visual stream including V3 and V4 in the left occipital lobe (b) Action values from the reliability-based arbitration model correlated with activity in the dlPFC (c) Action selection probabilities of the chosen action from reliability models correlated with activity in AIP, SII, pSTS and LOC. (d and e) Reliabilities of both decision-making systems significantly correlated with the preSMA, IPS, lPFC, aPFC, PMd, PCUN, and IPL. (f) The difference between the reliabilities of the two systems ($Rel_{aff} - Rel_q$), which is directly related to the arbitration weight, was identified in the preSMA, dlPFC, and IC. (g) Affordance prediction error (APE) signals which are putatively employed for tracking the reliability of the affordance-based decision-making system were identified in the ventral striatum. All the results were cluster-corrected $p < 0.05$ with the cluster defining threshold $z = 3.1$.

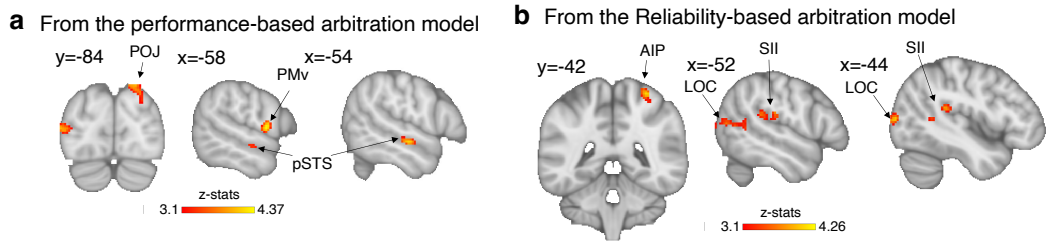


Figure 4.18: Neural correlates of action selection probability of chosen action after controlling reaction times. (a) Results from GLM1 with RT (uncorrected with the threshold $p < 0.001$). (b) Reuslts from GLM2 with RT (cluster corrected with the threshold $p < 0.05$ with the cluster defining threshold $z = 3.1$.)

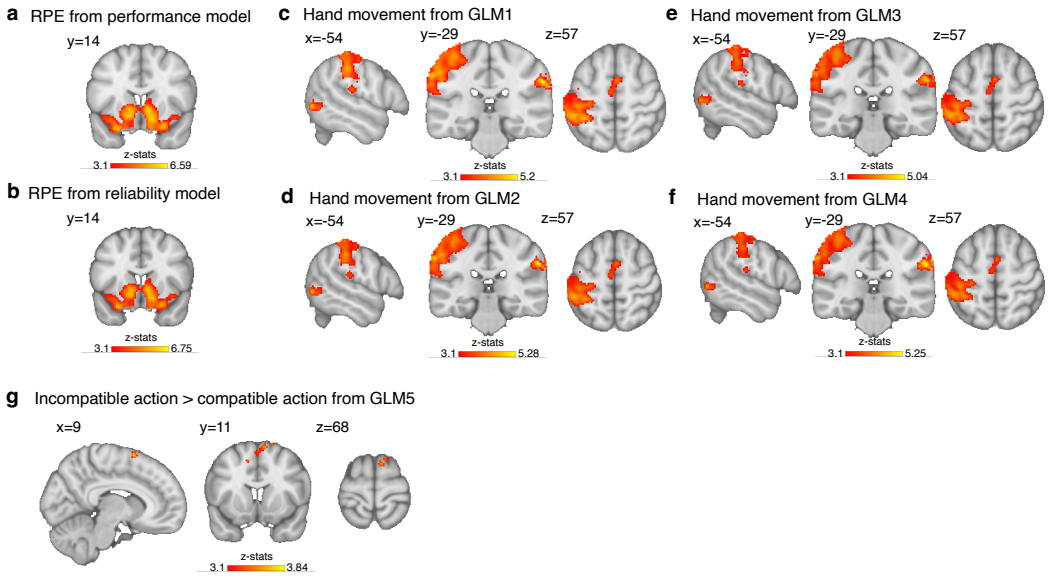


Figure 4.19: Neural correlates of reward prediction error (RPE), motor control and stimulus response incompatibility. (a and b) RPE signals from both models correlated with the ventral striatum. (c to f) The motor control information extracted from the recorded hand videos identified left primary motor cortex and adjacent regions. GLM1 and 3 are the GLMs utilizing variables extracted from the performance-based arbitration model, GLM2 and 4 are the GLMs using variables from the reliability-based arbitration model. All the results were cluster-corrected $p < 0.05$ with the cluster defining threshold $z = 3.1$. (G) The analysis utilizing GLM5 revealed that the dorsal premotor area (PMd) and preSMA exhibit greater activation when the affordance of an object and the selected response are incompatible, compared to when an affordance-compatible action is chosen (uncorrected, $p < 0.001$). See Methods for the details on GLMs.

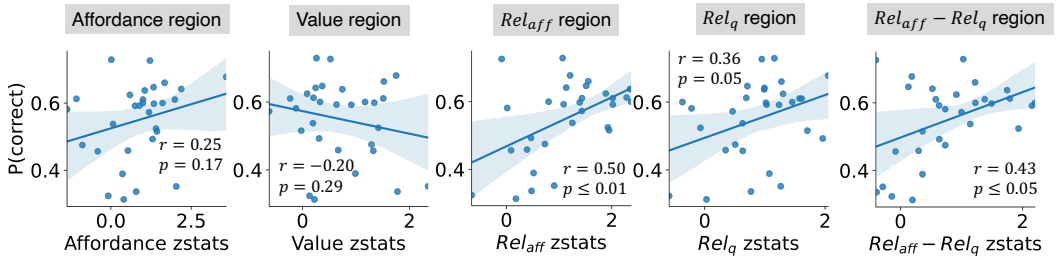


Figure 4.20: Correlations between the strength of the neural representation of cognitive variables and the tendency to select the most rewarding actions. The plots relate to those in Fig. 4.5 with regions identified using the GLMs based on the reliability-based arbitration model, and the variables of interest are from the reliability-based arbitration model. Each dot represents an individual participant.

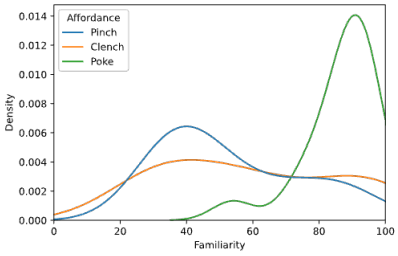


Figure 4.21: Distributions of familiarity scores by the type of affordances of stimuli.

Behavior study (n=19)	Coef.	Std. Err.	z	P>|z|
Intercept	6.960	0.056	124.404	0.000
Choosing pinch	-0.006	0.018	-0.327	0.744
Choosing clench	-0.003	0.010	-0.260	0.795
Choosing affordance	-0.026	0.010	-2.664	0.008
Choosing correct	-0.054	0.017	-3.143	0.002
Trial index	-0.006	0.001	-5.970	0.000
In congruent condition	0.006	0.008	0.689	0.491
In high condition	-0.052	0.010	-5.289	0.000

fMRI study (n=30)	Coef.	Std. Err.	z	P>|z|
Intercept	6.915	0.051	135.278	0.000
Choosing pinch	-0.006	0.010	-0.593	0.553
Choosing clench	-0.018	0.015	-1.174	0.241
Choosing affordance	-0.029	0.010	-3.016	0.003
Choosing correct	-0.058	0.014	-4.214	0.000
Trial index	-0.003	0.001	-2.215	0.027
In congruent condition	-0.004	0.009	-0.471	0.638
In high condition	-0.075	0.010	-7.362	0.000

Table 4.1: Fixed-effect coefficients of the mixed-effect general linear models on the reaction time. The trial-by-trial reaction time, which is the dependent variable, was logarithmically transformed.

Behavior study (n=19)			
Initial response	**Pinch object**	**Clench object**	**Poke object**
Pinch	102 (**23**)	79 (0)	56 (-23)
Clench	93 (-15)	138 (**30**)	93 (-15)
Poke	109 (-8)	87 (-30)	155 (**38**)

fMRI study (n=30)			
Initial response	**Pinch object**	**Clench object**	**Poke object**
Pinch	98 (**28.33**)	62 (-7.67)	49 (-20.67)
Clench	61 (-19.67)	109 (**28.33**)	72 (-8.33)
Poke	81 (-8.67)	69 (-20.67)	119 (**29.33**)

Table 4.2: Initial responses sorted by the affordance of the object. The numbers indicate the sum across all initial trials where a particular action was selected as the initial response to the specific object type. The numbers in the parentheses represent the deviation of the number in the cell from the expected number of choosing a particular action type, regardless of the object type which was calculated as the average across columns.

Behavior study (n=19)

	Coef.	Std. Err.	z	P>|z|
Intercept	0.116	0.033	3.474	0.001
Number of rewarded trials	-0.010	0.003	-3.160	0.002

fMRI study (n=30)

	Coef.	Std. Err.	z	P>|z|
Intercept	0.113	0.035	3.216	0.001
Number of rewarded trials	-0.009	0.003	-2.968	0.003

Table 4.3: Fixed-effect coefficients from mixed-effect general linear models on the difference in the frequency of choosing the most-rewarding action between congruent and incongruent conditions. The tables are from the GLMs where the number of rewarded trials previously encountered is the regressor.

Behavior-generative model \ **Identified model**	RL	Bayesian	Prior	Bias	Fixed	Cost	Conflict	Rel	Perf
RL	**0.24**	0	0.11	0.06	0.03	0	0.03	0	0
Bayesian	0.01	**0.92**	0	0	0	0	0	0	0
Prior	0.13	0.03	**0.58**	0.26	0.09	0	0	0.12	0
Bias	0.17	0.03	0	**0.36**	0.03	0.13	0	0	0.02
Fixed	0.11	0.02	0.05	0.06	**0.43**	0.13	0.31	0.03	0.04
Cost	0.11	0	0.11	0.11	0.29	**0.75**	0.26	0.03	0.04
Conflict	0.25	0	0.05	0.02	0.03	0	**0.03**	0.03	0
Rel	0.08	0	0.05	0.06	0.06	0	0.29	**0.79**	0
Perf	0.04	0	0.05	0.06	0.06	0	0.09	0	**0.89**

Table 4.4: Probability of the identified model being the actual model that generated the behavior.

Behavior Simulation (n=19x10)

Fixed Arb.	Coef.	Std. Err.	z	P>|z|	Cost	Coef.	Std. Err.	z	P>|z|	Conf. Arb.	Coef.	Std. Err.	z	P>|z|
Intercept	0.068	0.008	8.401	0.000	Intercept	0.047	0.007	6.483	0.000	Intercept	-0.012	0.007	-1.750	0.080
Trial index	-0.000	0.001	-0.301	0.763	Trial index	0.005	0.001	8.027	0.000	Trial index	-0.001	0.001	-1.209	0.226
Rel. Arb.	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**Perf. Arb.**	**Coef.**	**Std. Err.**	**z**	**P>|z|**					
Intercept	0.071	0.009	8.151	0.000	Intercept	0.109	0.010	11.015	0.000					
Trial index	0.001	0.001	1.016	0.310	Trial index	-0.001	0.001	-2.258	0.024					

fMRI Simulation (n=30x10)

Fixed Arb.	Coef.	Std. Err.	z	P>|z|	Cost	Coef.	Std. Err.	z	P>|z|	Conf. Arb.	Coef.	Std. Err.	z	P>|z|
Intercept	0.042	0.009	4.748	0.000	Intercept	0.044	0.009	5.134	0.000	Intercept	0.004	0.008	0.497	0.619
Trial index	0.001	0.001	1.556	0.120	Trial index	0.003	0.001	4.603	0.000	Trial index	-0.001	0.001	-1.214	0.225
Rel. Arb.	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**Perf. Arb.**	**Coef.**	**Std. Err.**	**z**	**P>|z|**					
Intercept	0.089	0.009	9.413	0.000	Intercept	0.094	0.009	10.184	0.000					
Trial index	-0.001	0.001	-0.943	0.346	Trial index	-0.001	0.001	-1.500	0.134					

Table 4.5: Results of mixed-effect general linear models on the simulated behaviors. Fixed-effect coefficients obtained from the mixed-effect GLM on the difference in the choice accuracy between congruent and incongruent conditions are reported, with the trial index being the regressor.

Behavior Simulation (n=19x10)

Fixed Arb.	Coef.	Std. Err.	z	P>|z|	Cost	Coef.	Std. Err.	z	P>|z|	Conf. Arb.	Coef.	Std. Err.	z	P>|z|
Intercept	0.056	0.008	6.691	0.000	Intercept	0.086	0.009	10.144	0.000	Intercept	-0.025	0.007	-3.513	0.000
Number of rewarded trials	-0.001	0.001	-0.908	0.364	Number of rewarded trials	-0.002	0.001	-2.351	0.019	Number of rewarded trials	0.003	0.001	3.775	0.000
Rel. Arb.	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**Perf. Arb.**	**Coef.**	**Std. Err.**	**z**	**P>|z|**					
Intercept	0.075	0.008	9.996	0.000	Intercept	0.099	0.009	11.612	0.000					
Number of rewarded trials	-0.003	0.001	-3.077	0.002	Number of rewarded trials	-0.007	0.001	-6.644	0.000					

fMRI Simulation (n=30x10)

Fixed Arb.	Coef.	Std. Err.	z	P>|z|	Cost	Coef.	Std. Err.	z	P>|z|	Conf. Arb.	Coef.	Std. Err.	z	P>|z|
Intercept	0.037	0.008	4.410	0.000	Intercept	0.072	0.008	8.871	0.000	Intercept	-0.005	0.007	-0.786	0.432
Number of rewarded trials	0.001	0.001	0.618	0.537	Number of rewarded trials	-0.004	0.001	-4.453	0.000	Number of rewarded trials	-0.000	0.001	-0.054	0.957
Rel. Arb.	**Coef.**	**Std. Err.**	**z**	**P>|z|**	**Perf. Arb.**	**Coef.**	**Std. Err.**	**z**	**P>|z|**					
Intercept	0.078	0.008	9.510	0.000	Intercept	0.097	0.008	11.536	0.000					
Number of rewarded trials	-0.004	0.001	-4.280	0.000	Number of rewarded trials	-0.008	0.001	-8.814	0.000					

Table 4.6: Results of mixed-effect general linear models on the simulated behaviors. Fixed-effect coefficients obtained from the mixed-effect GLM on the difference in the choice accuracy between congruent and incongruent conditions are reported, with the number of rewarded trials previously encountered being the regressor.

Behavior Simulation (n=19x10)

Fixed Arb.	Coef.	Std. Err.	z	P>|z|
Intercept	-0.024	0.007	-3.263	0.001
Number of rewarded trials	-0.000	0.002	-0.209	0.834

Cost	Coef.	Std. Err.	z	P>|z|
Intercept	-0.051	0.007	-7.192	<0.001
Number of rewarded trials	0.003	0.002	2.243	0.025

Conf. Arb.	Coef.	Std. Err.	z	P>|z|
Intercept	-0.072	0.009	-8.335	<0.001
Number of rewarded trials	0.006	0.001	5.557	<0.001

Rel. Arb.	Coef.	Std. Err.	z	P>|z|
Intercept	-0.022	0.007	-2.967	0.003
Number of rewarded trials	0.003	0.001	1.820	0.069

Perf. Arb.	Coef.	Std. Err.	z	P>|z|
Intercept	-0.038	0.008	-4.762	<0.001
Number of rewarded trials	0.002	0.002	1.106	0.269

fMRI Simulation (n=30x10)

Fixed Arb.	Coef.	Std. Err.	z	P>|z|
Intercept	-0.019	0.008	-2.402	0.016
Number of rewarded trials	0.003	0.001	2.248	0.025

Cost	Coef.	Std. Err.	z	P>|z|
Intercept	-0.011	0.008	-1.257	0.209
Number of rewarded trials	-0.000	0.001	-0.209	0.834

Conf. Arb.	Coef.	Std. Err.	z	P>|z|
Intercept	-0.036	0.010	-3.647	<0.001
Number of rewarded trials	0.005	0.002	3.282	0.001

Rel. Arb.	Coef.	Std. Err.	z	P>|z|
Intercept	0.001	0.008	0.109	0.913
Number of rewarded trials	0.007	0.002	4.694	<0.001

Perf. Arb.	Coef.	Std. Err.	z	P>|z|
Intercept	-0.013	0.008	-1.599	0.110
Number of rewarded trials	0.003	0.002	1.672	0.094

Table 4.7: Results of mixed-effect general linear models on the simulated behaviors shown in the Fig.4.16. Fixed-effect coefficients obtained from the mixed-effect GLM on the difference in the choice accuracy between high and low conditions are reported, with the number of rewarded trials previously encountered being the regressor.

Chapter 5

CONCLUSION

5.1 Summary of results

In this thesis, I explore the neuro-computational mechanisms of human cognitive capacities that support decision-making in novel environments through three interconnected studies.

The first study demonstrated that value computation for one of the most complex visual stimuli, art paintings, can be explained through feature-based computation. Interestingly, the features that explain human aesthetic prefrence were found to emerge automatically from a convolutional neural network trained on object recognition. By analyzing the hidden activations of the CNN, we developed a framework to categorize these features into high-level and low-level visual categories, a classification which was further validated by functional MRI data. This study provides a mechanistic description of feature-based value computation, enabling the implementation of value computation processes even for novel stimuli, which is essential for making informed decisions in unfamiliar environments. The results show that value computation is supported by a hierarchical representation of features in the brain, along the rostro-caudal axis, which are integrated into a value calculation process occurring in the prefrontal cortex.

The second study explored the computational mechanisms underlying human transfer learning. A novel task was designed, consisting of multiple environments that shared features, allowing participants to use these features as cues for making better choices even before interacting with a new environment, provided they had learned the meaning of the features from previous contexts. By analyzing large-scale online behavioral data and on-site data, we identified the most plausible computational model of transfer learning. While recurrent neural network model had the highest predictive power in predicting the human actual choice, it was failing in modeling the characteristics of transfer learning of humans such as increasing bias toward the choice option with features that have higher contingency to the better choices in previous experience. However, the model that best reproduced human behavior patterns included a component that slowly tracks the learned value of each feature. This slow temporal dynamics allowed the model to retain previously learned in-

formation, facilitating decision-making in novel environments, and reproducing the pattern of gradually increasing choice bias toward the good features. Notably, this component mirrors the properties of astrocyte glial cells, which are the predominant cell type in the brain. This study not only suggests future research directions for understanding the role of astrocyte glial cells in decision-making but also offers insights into developing next-generation neural networks capable of flexible transfer learning without catastrophic forgetting.

The third study shifted focus to more naturalistic settings, examining the role of affordance in everyday decision-making. Utilizing cutting-edge computer vision techniques, we implemented a novel task involving naturalistic hand gestures as inputs in a three-armed bandit task, where participants had to determine the most rewarding hand gestures through trial and error. This naturalistic approach allowed us to study the influence of affordance on decision-making from a more integrated perspective, as both affordance and value were concurrently affecting the decisions during the task. Participants exhibited a bias toward affordance-compatible actions, but surprisingly, they adapted more quickly to the most rewarding action when the affordance was not the optimal choice for a given object. A thorough model comparison revealed that the underlying computational mechanism involves two systems: affordance-based and value-based decision-making. The faster adaptation to the most rewarding action was explained by a hierarchical reinforcement learning model with a higher-level arbitrator that assigns weights to each decision-making system based on their performance.

Model-based fMRI analyses further identified that high-level visual areas correlated with the compatibility of the chosen action to the object's affordance. The value inferred from the model fitting was observed in the ventromedial prefrontal cortex. Interestingly, signals related to arbitration, such as the performance of the value-based and affordance-based decision-making systems, were represented across the medial prefrontal cortex. The variable directly determining the arbitration weight, or the difference in performance between the two systems, was localized in the pre-supplementary motor area. The integrated final action selection signal could be identified in the posterior parietal cortex.

This third study provides a comprehensive perspective on real-world decision-making, which is influenced not only by value, but also by the physical characteristics of the environment and the physical constraints of the decision-maker. The hierarchical RL framework proposed for modeling the data is grounded in the

normative idea that the overall goal of decision-making is to greedily maximize collected rewards rather than aiming for global maximization. This framework, which aligns with recent advances in deep RL that have surpassed human performance in multiple tasks, suggests a different perspective: that human behavior might be better explained by near-sighted maximization of performance rather than global optimization of value.

These studies collectively contribute to a deeper understanding of the neuro-computational processes that govern human decision-making in novel environments. The findings highlight the importance of feature-based hierarchical processing in value computation, the role of slow integration mechanisms in knowledge transfer, and the dynamic interplay between affordance-based and value-based systems in action selection. Together, these results offer a unified framework for understanding how the brain integrates sensory information, retains and applies learned knowledge, and adapts to new challenges.

5.2 Broad implications and future directions

While the first study provides a mechanistic description of aesthetic preference, it is important to recognize its limitation in that the model primarily explains the influence of visual features in the valuation process. This approach, while valuable, does not fully capture the multifaceted nature of aesthetic experience, which involves a complex interplay of sensory, cognitive, and emotional factors (Chatterjee and Cardilo, 2021; Zeki, 2002).

In real life, preferences for art are shaped by multiple dimensions that extend beyond visual features. The context in which the art is displayed, such as the environment and the lighting, plays a significant role in influencing perception. Additionally, factors related to the creator, including their identity, intent, and the historical or cultural context of the artwork, contribute to how it is valued. Personal experiences of the viewer, such as life events that resonate with the themes of the artwork, previous exposure to art education, and even the viewer's current mood, further modulate their aesthetic preferences. These individual differences highlight the complexity of aesthetic judgment and suggest that future models should incorporate a broader range of factors to better reflect the multidimensional nature of art appreciation (Leder et al., 2004; Pelowski et al., 2016).

The multidimensionality of real aesthetic experience presents a significant challenge in scientifically studying the holistic nature of the process. Capturing the full spec-

trum of factors that contribute to art appreciation requires more than just traditional methods. However, advances in experimental equipment, particularly in virtual reality (VR), offer promising avenues for addressing these complexities. VR technology enables the creation of immersive environments where the entire art appreciation experience can be simulated, allowing researchers to manipulate various contextual factors in a controlled manner (Bohil et al., 2011). Simultaneously, physiological measurements collected via wearable devices, such as heart rate monitors and skin conductance sensors, can be used to infer emotional states, providing real-time data on the viewer's affective responses (Fairclough, 2009).

Furthermore, the integration of large-scale data from social media platforms offers a novel approach to studying naturalistic valuation processes. User behavior data, including liking patterns, dwell time on posts, and the semantics and sentiments of consumed content, can provide insights into how various contexts influence aesthetic preferences (Kosinski et al., 2013). By combining these rich data sources with VR and physiological measurements, researchers can develop more comprehensive models that better capture the complexity of real-world art appreciation.

The neural network model used to simulate the valuation process, while effective in some respects, fails to capture the complexity of the brain's real visual stream. The human visual system is primarily divided into two pathways: the ventral stream, often referred to as the "what" pathway, responsible for object recognition, and the dorsal stream, known as the "where" pathway, which processes spatial information and guides movement (Mishkin et al., 1983; Goodale and Milner, 1992; Milner and Goodale, 2008). These streams are not isolated; they are highly interconnected, allowing for the integration of object identity and spatial context, which is crucial for robust visual perception (Kravitz et al., 2011). Moreover, there are extensive top-down connections from the frontal cortex to earlier visual areas that support visual recognition by incorporating prior knowledge and expectations, thereby enhancing the accuracy and efficiency of visual processing (Gilbert and Li, 2013; Miller and Cohen, 2001). Incorporating these neurobiological properties into artificial neural networks has led to significant improvements in predicting neural activities in the visual pathways (Kar et al., 2019).

Although correlated neural representations can be identified in the brain using the hidden activations of convolutional neural networks (CNNs), to achieve a closer approximation of real brain function, it is essential to design neural networks that are more biologically plausible. Initiatives such as BrainScore, which aim to eval-

uate and identify neural network models that best explain the ventral visual stream, are crucial steps toward this goal Schrimpf et al. (2018). By focusing on models that accurately capture the intricacies of human vision, we can advance our understanding of both artificial intelligence and the neural mechanisms underlying visual perception.

The second study on transfer learning focused on one specific subset within the broad field of transfer learning: feature-based transfer learning. Transfer learning strategies in machine learning can generally be categorized into four approaches: instance-based, parameter-based, relational-based, and feature-based (Zhuang et al., 2020). The instance-based approach involves memorizing decisions made in previous events along with their outcomes and reusing these successful decisions when encountering similar events. This approach is analogous to episodic memory in humans, where specific experiences are recalled to inform future decisions (Tulving, 2002).

The parameter-based approach, on the other hand, involves reusing learned parameters from one task to apply them to a different but related task. For example, a neural network trained on image classification might be repurposed for image captioning or generation tasks, utilizing shared features learned during the initial training (Yosinski et al., 2014; Radford et al., 2021). Relational-based transfer learning involves understanding the structural relationships within an environment and applying that knowledge to achieve new goals. This approach is similar to the concept of cognitive maps in the brain, where individuals learn the layout of an environment and reuse this knowledge for navigation and goal-directed behavior in novel contexts (Whittington et al., 2020; Behrens et al., 2018).

To fully understand human transfer learning, it is crucial to investigate how these different strategies are implemented in the brain. Research on memory systems, such as the hippocampus and prefrontal cortex, which are involved in decision-making, model-based reasoning, and neural reuse, will be instrumental in advancing our understanding of this process (Shohamy and Daw, 2015; Botvinick et al., 2019).

For implementing feature-based transfer learning, we proposed that a temporally hierarchical neural network could be an effective solution, as the slower temporal components can retain learned information over extended periods, thereby enhancing the network's ability to apply this knowledge in new contexts. This approach aligns with suggestions in the machine learning field, where temporally layered models have been used to enhance the retention of long-term dependencies (Hihi and Bengio,

1995; Chung et al., 2016; Neil et al., 2016; Vezhnevets et al., 2017). Incorporating the properties of astrocyte glial cells into artificial neural networks could further improve the performance of hierarchical networks. Glial cells, which play a crucial role in modulating synaptic efficacy could inspire novel architectures that better mimic the brain's natural learning processes (Fields et al., 2014; Alvarez-Gonzalez et al., 2023).

The question of how feature-based learning is implemented in the brain remains open. Future research could address this by designing tasks that produce more robust and larger effect sizes than those used in previous studies, allowing for more precise investigation of the underlying neural mechanisms. These tasks should be coupled with advanced neuroimaging techniques, such as functional MRI (fMRI) or magnetoencephalography (MEG), to observe the brain regions involved in real-time. One of the possible neural substrates for feature-based transfer learning is the prefrontal cortex, a region critical for value integration and decision-making. The prefrontal cortex is known to support a range of cognitive functions that are likely involved in transfer learning, such as working memory and cognitive flexibility, which allow the brain to adapt previously learned information to new contexts. By targeting these regions in future studies, we can gain deeper insights into how the brain implements feature-based learning and transfer across different tasks Miller and Cohen (2001).

The final study explored the effect of affordance on value learning using images of everyday objects and naturalistic hand gestures, offering a more ecologically valid setting compared to traditional button-press tasks with abstract stimuli. However, the task still has limitations in achieving true ecological validity. Specifically, the event-related design, which presents stimuli on a trial-by-trial basis, creates an artificially discrete experience for participants, disconnecting them from the continuous nature of real-world interactions (Kingstone et al., 2008). Additionally, each trial within the MRI scanner took nearly 10 seconds, which raises concerns that the brain might engage different processes when solving the task compared to more continuous, real-time situations.

To address these limitations, several solutions could be considered. One approach is to use VR technology, which allows for the naturalistic display of visual stimuli and hand gestures as inputs. Affordance is not only determined by the shape of an object but also by factors such as the distance from the actor and the orientation of the object relative to the person (Gibson, 2014). For instance, if an object is placed out-

side of reachable distance, the action-stimulus compatibility effect from affordance decreases significantly (Costantini et al., 2010) This effect is also evident in cases where an object has a dominant orientation, such as a cup with a single handle—if the handle is on the left side, a left-handed individual will experience a stronger affordance effect (Tucker and Ellis, 1998). Such variables can be systematically controlled using a VR system, providing a more ecologically valid environment for future studies. Moreover, instead of using a remote controller to interact with the VR environment, advanced hand gesture tracking technologies—such as computer vision, depth sensors, or electromyography (EMG)—could be employed to capture naturalistic interactions (Cipresso et al., 2018). Additionally, incorporating a 360-degree VR treadmill could enable more comprehensive studies of affordance by allowing participants to navigate and interact with the environment in real-time, further enhancing ecological validity.

Furthermore, intracranial electroencephalography (iEEG) or single-neuron recordings could be utilized to study affordance and decision-making processes with greater temporal and spatial precision. A key research question to explore is whether affordance-based decision-making overrides value-based decision-making or vice versa. If automatic processes are temporally faster than deliberate ones, it is likely that affordance-based decisions occur first, represented in regions such as the posterior parietal cortex, and are gradually overridden by value-based decision-making signals (Cisek and Kalaska, 2010). Analyzing the temporal dynamics of neural recordings from action-selection regions, including the ventromedial prefrontal cortex, high-level visual areas, and posterior parietal cortex, as well as arbitration regions like the pre-supplementary motor area, would provide further insights into how the multiple decision-making systems are integrated dynamically (Hare et al., 2011).

Additionally, the collected data could be used to study the computational processes underlying affordance perception in everyday objects. Affordance compatibility annotations for four different hand gestures, covering the majority of gestures used in object interaction, were collected for 1,000 object images from a large online participant pool. Along with fMRI data from participants viewing 24 objects multiple times, a similar approach to that used in the art valuation study—utilizing deep learning and feature-based methods—could be applied to this dataset to elucidate the computational mechanisms of affordance perception.

The novel computational framework I propose for modeling value-based decision-

making and learning under the influence of affordance offers an adaptive method capable of selecting the best-performing policy in real-time. Unlike traditional RL frameworks, which aim to maximize cumulative rewards over an infinite horizon, the performance-based arbitration framework I propose focuses on maximizing performance. In this context, performance is defined as the expected sum of rewards obtained after executing a particular component policy at the current time step, followed by a mixture of policies that maximize subsequent performance. This approach is akin to the Bellman equation used for recursive value calculation but differs in that it utilizes two distinct policies: a component policy of the mixture of policies for the current time step and the mixture of policies that maximizes performance thereafter. The framework operates under the assumption that selecting the best policy at each time step is more practical for addressing continual learning problems, where estimating the global maximum is often impractical (Thrun and Mitchell, 1995; Khetarpal et al., 2022).

Additionally, this framework offers an automatic way to balance exploration and exploitation. If we assume that the policies include a random policy and a value-based decision-making policy, the arbitration mechanism would weigh between these two based on their relative performance in the current situation. The performances of the two policies are continuously tracked and updated using the delta rule, enabling the framework to automatically manage the exploration-exploitation trade-off, which is typically tuned using hyperparameters in traditional RL approaches.

Moreover, the performance-based arbitration framework may address a critical gap in current studies on arbitration between different RL systems, which often rely on the uncertainty or reliability of value estimates from each policy (Daw et al., 2005; Lee et al., 2014b; Charpentier et al., 2020). If we conceptualize arbitration as a higher-level decision-making process between policies, using uncertainty or reliability of value estimates may minimize variance in value predictions. However, if the objective is to maximize the total collected reward, it may be more effective to use performance as the metric for making decisions. Furthermore, inspired by the upper confidence bound (UCB) algorithm, one could envision arbitration mechanisms that incorporate both performance and uncertainty or reliability for exploitation-exploration trade-offs among policies (Auer, 2002) Such an approach could significantly advance our understanding of how cognitive functions are selected and executed.

Overall, the three studies collectively contribute to a comprehensive understanding

of how the brain integrates sensory information, retains and applies learned knowledge, and adapts to new challenges. The findings presented in this thesis highlight the dynamic interplay between different cognitive systems, such as value-based and affordance-based decision-making, in guiding behavior, offering a deeper understanding of human decision-making. Furthermore, this thesis provides foundational insights for improving AI models by aligning them more closely with human cognitive processes, particularly in novel or complex environments.

Bibliography

Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gael Varoquaux. Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 8, 2014. ISSN 1662-5196. doi: 10.3389/fninf.2014.00014. URL https://www.frontiersin.org/articles/10.3389/fninf.2014.00014/full.

Sara Alvarez-Gonzalez, Francisco Cedron, Alejandro Pazos, and Ana B. Porto-Pazos. Artificial glial cells in artificial neuronal networks: A systematic review. *Artificial Intelligence Review*, 56(Suppl 2):2651–2666, 2023.

Richard A. Andersen. Multimodal integration for the representation of space in the posterior parietal cortex. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 352(1360):1421–1428, 1997. ISSN 0962-8436. doi: 10.1098/rstb.1997.0128.

Richard A. Andersen and He Cui. Intention, Action planning, and decision making in parietal-rrontal circuits. *Neuron*, 63(5):568 – 583, 09 2009. doi: 10.1016/j.neuron.2009.08.028.

Richard A. Andersen, Lawrence H. Snyder, David C. Bradley, and Jing Xing. Multimodal representation of space in the posterior parietal cortex and its use in planning movements. *Annual Review of Neuroscience*, 20(1):303–330, 1997.

Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Brian B. Avants, Charles L. Epstein, Murray Grossman, and James C. Gee. Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain. *Medical image analysis*, 12(1):26–41, 2008.

Dominik R. Bach, Mkael Symmonds, Gareth Barnes, and Raymond J. Dolan. Whole-brain neural dynamics of probabilistic reward prediction. *Journal of Neuroscience*, 37(14):3789–3798, 2017.

Adrià Puigdomènech Badia, Bilal Piot, Steven Kapturowski, Pablo Sprechmann, Alex Vitvitskyi, Zhaohan Daniel Guo, and Charles Blundell. Agent57: Outperforming the atari human benchmark. In *International Conference on Machine Learning*, pages 507–517. The Proceedings of Machine Learning Research, 2020.

Joan S. Baizer, Leslie G. Ungerleider, and Robert Desimone. Organization of visual inputs to the inferior temporal and posterior parietal cortex in macaques. *Journal of Neuroscience*, 11(1):168–190, 1991.

Moshe Bar and Maital Neta. Humans prefer curved visual objects. *Psychological Science*, 17(8):645–648, 2006.

Deanna M. Barch, Gregory C. Burgess, Michael P. Harms, Steven E. Petersen, Bradley L. Schlaggar, Maurizio Corbetta, Matthew F. Glasser, Sandra Curtiss, Sachin Dixit, Cindy Feldt, et al. Function in the human connectome: Task-fmri and individual differences in behavior. *Neuroimage*, 80:169–189, 2013.

Helen C. Barron, Raymond J. Dolan, and Timothy E. J. Behrens. Online evaluation of novel choices by simultaneous representation of multiple memories. *Nature Neuroscience*, 16(10):1492, 2013.

Timothy E. J. Behrens, Timothy H. Muller, James C. R. Whittington, Shirley Mark, Alon B. Baram, Kimberly L. Stachenfeld, and Zeb Kurth-Nelson. What is a cognitive map? organizing knowledge for flexible behavior. *Neuron*, 100(2): 490–509, 2018.

Marlene Behrmann, Joy J. Geng, and Sarah Shomstein. Parietal cortex and attention. *Current Opinion in Neurobiology*, 14(2):212–217, 2004. ISSN 0959-4388. doi: 10.1016/j.conb.2004.03.012.

Yashar Behzadi, Khaled Restom, Joy Liau, and Thomas T. Liu. A component based noise correction method (CompCor) for BOLD and perfusion based fmri. *NeuroImage*, 37(1):90–101, 2007. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2007.04.042. URL `http://www.sciencedirect.com/science/article/pii/S1053811907003837`.

Yoshua Bengio. Deep learning of representations for unsupervised and transfer learning. In *Proceedings of ICML Workshop on Unsupervised and Transfer Learning*, pages 17–36, 2012.

Irving Biederman and Edward A. Vessel. Perceptual pleasure and the brain: A novel theory explains why the brain craves information and seeks it through the senses. *American Scientist*, 94(3):247–253, 2006.

Christopher M. Bishop. *Pattern recognition and machine learning*. Springer, 2006.

Corey J. Bohil, Bradly Alicea, and Frank A. Biocca. Virtual reality in neuroscience research and therapy. *Nature Reviews Neuroscience*, 12(12):752–762, 2011.

Michael F. Bonner and Russell A. Epstein. Coding of navigational affordances in the human visual system. *Proceedings of the National Academy of Sciences of the United States of America*, 114(18):4793 – 4798, 05 2017. doi: 10.1073/pnas.1618228114.

Elena Borra, Marzio Gerbella, Stefano Rozzi, and Giuseppe Luppino. The macaque lateral grasping network: A neural substrate for generating purposeful hand actions. *Neuroscience & Biobehavioral Reviews*, 75:65–90, 2017. ISSN 0149-7634. doi: 10.1016/j.neubiorev.2017.01.017.

Matthew Botvinick and Todd Braver. Motivation and cognitive control: From behavior to neural mechanism. *Annual Review of Psychology*, 66(1):83–113, 2015. ISSN 0066-4308. doi: 10.1146/annurev-psych-010814-015044.

Matthew Botvinick and Ari Weinstein. Model-based hierarchical reinforcement learning and human action control. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655):20130480, 2014. ISSN 0962-8436. doi: 10.1098/rstb.2013.0480.

Matthew Botvinick, Sam Ritter, Jane X. Wang, Zeb Kurth-Nelson, Charles Blundell, and Demis Hassabis. Reinforcement learning, fast and slow. *Trends in Cognitive Sciences*, 23(5):408–422, 2019.

Matthew M. Botvinick, Todd S. Braver, Deanna M. Barch, Cameron S. Carter, and Jonathan D. Cohen. Conflict monitoring and cognitive control. *Psychological Review*, 108(3):624, 2001.

Matthew M. Botvinick, Jonathan D. Cohen, and Cameron S. Carter. Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences*, 8 (12):539–546, 2004. ISSN 1364-6613. doi: 10.1016/j.tics.2004.10.003.

Matthew M. Botvinick, Stacy Huffstetler, and Joseph T. McGuire. Effort discounting in human nucleus accumbens. *Cognitive, Affective, & Behavioral Neuroscience*, 9(1):16–27, 2009. ISSN 1530-7026. doi: 10.3758/cabn.9.1.16.

Matthew Michael Botvinick. Hierarchical reinforcement learning and decision making. *Current Opinion in Neurobiology*, 22(6):956–962, 2012.

Senne Braem, Julie M. Bugg, James R. Schmidt, Matthew J.C. Crump, Daniel H. Weissman, Wim Notebaert, and Tobias Egner. Measuring adaptive control in conflict tasks. *Trends in Cognitive Sciences*, 23(9):769–783, 2019. ISSN 1364-6613. doi: 10.1016/j.tics.2019.07.002.

David Brieber, Marcos Nadal, and Helmut Leder. In the white cube: Museum context enhances the valuation and memory of art. *Acta Psychologica*, 154: 36–42, 2015.

Daniel N. Bub, Michael E. J. Masson, and Maria van Noordenne. Motor representations evoked by objects under varying action intentions. *Journal of Experimental Psychology: Human Perception and Performance*, 47(1):53–80, 2021. ISSN 0096-1523. doi: 10.1037/xhp0000876.

Charles F. Cadieu, Ha Hong, Daniel L. K. Yamins, Nicolas Pinto, Diego Ardila, Ethan A. Solomon, Najib J. Majaj, and James J. DiCarlo. Deep neural networks rival the representation of primate it cortex for core visual object recognition. *PLoS Computational Niology*, 10(12):e1003963, 2014.

Nathalie Camille, Giorgio Coricelli, Jerome Sallet, Pascale Pradat-Diehl, Jean-René Duhamel, and Angela Sirigu. The involvement of the orbitofrontal cortex in the experience of regret. *Science*, 304(5674):1167–1170, 2004.

Zhe Cao, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Realtime multi-person 2d pose estimation using part affinity fields. In *CVPR*, 2017.

Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields, 2019. URL `https://arxiv.org/abs/1812.08008`.

Umberto Castiello. The neuroscience of grasping. *Nature Reviews Neuroscience*, 6 (9):726–736, 2005. ISSN 1471-003X. doi: 10.1038/nrn1744.

Camilo J. Cela-Conde, Gisele Marty, Fernando Maestú, Tomás Ortiz, Enric Munar, Alberto Fernández, Miquel Roca, Jaume Rosselló, and Felipe Quesney. Activation of the prefrontal cortex in the human visual aesthetic perception. *Proceedings of the National Academy of Sciences*, 101(16):6321–6325, 2004.

Caroline J. Charpentier, Kiyohito Iigaya, and John P. O'Doherty. A neuro-computational account of arbitration between choice imitation and goal emulation during human observational learning. *Neuron*, 106(4):687–699, 2020.

Anjan Chatterjee. Prospects for a cognitive neuroscience of visual aesthetics. *Bulletin of Psychology and the Arts*, 4, 2003.

Anjan Chatterjee. Neuroaesthetics: A coming of age story. *Journal of Cognitive Neuroscience*, 23(1):53–62, 2011.

Anjan Chatterjee and Eileen Cardilo. *Brain, beauty, and art: Essays bringing neuroaesthetics into focus*. Oxford University Press, 2021.

Anjan Chatterjee and Oshin Vartanian. Neuroaesthetics. *Trends in Cognitive Sciences*, 18(7):370–375, 2014.

Anjan Chatterjee, Amy Thomas, Sabrina E. Smith, and Geoffrey K. Aguirre. The neural response to facial attractiveness. *Neuropsychology*, 23(2):135, 2009.

Anjan Chatterjee, Page Widick, Rebecca Sternschein, William B. Smith, and Bianca Bromberger. The assessment of art attributes. *Empirical Studies of the Arts*, 28 (2):207–222, 2010.

Quanjing Chen, Frank E. Garcea, and Bradford Z. Mahon. The representation of object-directed action and function knowledge in the human brain. *Cerebral Cortex*, 26(4):1609 – 1618, 03 2016. doi: 10.1093/cercor/bhu328.

Vikram S. Chib, Antonio Rangel, Shinsuke Shimojo, and John P. O'Doherty. Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience*, 29(39):12315–12320, 2009.

Kyunghyun Cho. Learning phrase representations using rnn encoder–decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014.

Junyoung Chung, Sungjin Ahn, and Yoshua Bengio. Hierarchical multiscale recurrent neural networks. *arXiv preprint arXiv:1609.01704*, 2016.

Pietro Cipresso, Irene Alice Chicchi Giglioli, Mariano Alcañiz Raya, and Giuseppe Riva. The past, present, and future of virtual and augmented reality research: A network and cluster analysis of the literature. *Frontiers in Psychology*, 9:2086, 2018.

Paul Cisek. Cortical mechanisms of action selection: The affordance competition hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1485):1585 – 1599, 09 2007. doi: 10.1098/rstb.2007.2054.

Paul Cisek and John F. Kalaska. Neural correlates of reaching decisions in dorsal premotor cortex: Specification of multiple direction choices and final selection of action. *Neuron*, 45(5):801–814, 2005.

Paul Cisek and John F. Kalaska. Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience*, 33(1):269–298, 2010.

Paul Cisek and Alexandre Pastor-Bernier. On the challenges and mechanisms of embodied decisions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 369(1655):20130479, 2014. ISSN 0962-8436. doi: 10.1098/rstb.2013.0479.

Anne G. E. Collins and Michael J. Frank. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, 35(7):1024 – 1035, 04 2012. doi: 10.1111/j.1460-9568.2011.07980.x. URL https://onlinelibrary.wiley.com/doi/epdf/10.1111/j.1460-9568.2011.07980.x.

Marcello Costantini, Ettore Ambrosini, Gaetano Tieri, Corrado Sinigaglia, and Giorgia Committeri. Where does an object trigger an action? an investigation about affordances in space. *Experimental Brain Research*, 207:95–103, 2010.

Robert W. Cox and James S. Hyde. Software tools for analysis and visualization of fmri data. *NMR in Biomedicine: An International Journal Devoted to the Development and Application of Magnetic Resonance In Vivo*, 10(4-5):171–178, 1997.

Laila Craighero and Giacomo Rizzolatti. Chapter 31 — the premotor theory of attention. In Laurent Itti, Geraint Rees, and John K. Tsotsos, editors, *Neurobiology of Attention*, pages 181–186. Academic Press, Burlington, 2005. ISBN 978-0-12-375731-9. doi: https://doi.org/10.1016/B978-012375731-9/50035-5. URL https://www.sciencedirect.com/science/article/pii/B9780123757319500355.

Logan Cross, Jeff Cockburn, Yisong Yue, and John P. O'Doherty. Using deep reinforcement learning to reveal how the brain encodes abstract state-space representations in high-dimensional environments. *Neuron*, 109(4):724–738, 2021.

Will Dabney, Zeb Kurth-Nelson, Naoshige Uchida, Clara Kwon Starkweather, Demis Hassabis, Rémi Munos, and Matthew Botvinick. A distributional code for value in dopamine-based reinforcement learning. *Nature*, 577(7792):671–675, 2020.

Anders M. Dale, Bruce Fischl, and Martin I. Sereno. Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*, 9(2):179–194, 1999. ISSN 1053-8119. doi: 10.1006/nimg.1998.0395. URL http://www.sciencedirect.com/science/article/pii/S1053811998903950.

Marco Davare, Alexander Kraskov, John C. Rothwell, and Roger N. Lemon. Interactions between areas of the cortical grasping network. *Current Opinion in Neurobiology*, 21(4):565–570, 2011. ISSN 0959-4388. doi: 10.1016/j.conb.2011.05.021.

Nathaniel D. Daw, Yael Niv, and Peter Dayan. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience*, 8(12):1704–1711, 2005.

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on Computer Vision and Pattern Recognition*, pages 248–255. Ieee, 2009.

Amir Dezfouli, Kristi Griffiths, Fabio Ramos, Peter Dayan, and Bernard W Balleine. Models that learn how humans learn: The case of decision-making and its disorders. *PLoS Computational Biology*, 15(6):e1006903 – 33, 06 2019a. doi: 10.1371/journal.pcbi.1006903.

Amir Dezfouli, Kristi Griffiths, Fabio Ramos, Peter Dayan, and Bernard W. Balleine. Models that learn how humans learn: The case of decision-making and its disorders. *PLoS Computational Biology*, 15(6):e1006903, 2019b.

Anthony Dickinson. Actions and habits: The development of behavioural autonomy. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 308(1135):67–78, 1985.

Ray J. Dolan and Peter Dayan. Goals and Habits in the Brain. *Neuron*, 80(2):312 – 325, 10 2013. doi: 10.1016/j.neuron.2013.09.007.

Michael C. Dorris and Paul W. Glimcher. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron*, 44(2): 365–378, 2004.

Yan Duan, John Schulman, Xi Chen, Peter L. Bartlett, Ilya Sutskever, and Pieter Abbeel. Rl2: Fast reinforcement learning via slow reinforcement learning. *arXiv preprint arXiv:1611.02779*, 2016.

Celia Durkin, Eileen Hartnett, Daphna Shohamy, and Eric R. Kandel. An objective evaluation of the beholder's response to abstract and figurative art based on construal level theory. *Proceedings of the National Academy of Sciences*, 117 (33):19809–19815, 2020.

Maria K. Eckstein and Anne G. E. Collins. Computational evidence for hierarchically structured reinforcement learning in humans. *Proceedings of the National Academy of Sciences*, 117(47):201912330, 2020. ISSN 0027-8424. doi: 10.1073/pnas.1912330117.

Freek van Ede. Visual working memory and action: Functional links and bidirectional influences. *Visual Cognition*, 28(5-8):401–413, 2020. ISSN 1350-6285. doi: 10.1080/13506285.2020.1759744.

Rob Ellis and Mike Tucker. Micro-affordance: The potentiation of components of action by seen objects. *British Journal of Psychology*, 91(4):451–471, 2000.

Oscar Esteban, Ross Blair, Christopher J. Markiewicz, Shoshana L. Berleant, Craig Moodie, Feilong Ma, Ayse Ilkay Isik, Asier Erramuzpe, Mathias Kent, James D. andGoncalves, Elizabeth DuPre, Kevin R. Sitek, Daniel E. P. Gomez, Daniel J. Lurie, Zhifang Ye, Russell A. Poldrack, and Krzysztof J. Gorgolewski. fmriprep. *Software*, 2018a. doi: 10.5281/zenodo.852659.

Oscar Esteban, Christopher Markiewicz, Ross W. Blair, Craig Moodie, Ayse Ilkay Isik, Asier Erramuzpe Aliaga, James Kent, Mathias Goncalves, Elizabeth DuPre, Madeleine Snyder, Hiroyuki Oya, Satrajit Ghosh, Jessey Wright, Joke Durnez, Russell Poldrack, and Krzysztof Jacek Gorgolewski. fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, 2018b. doi: 10.1038/s41592-018-0235-4.

Andre Esteva, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542(7639):115, 2017.

Andrew H. Fagg and Michael A. Arbib. Modeling parietal-premotor interactions in primate control of grasping. *Neural Networks : The official journal of the International Neural Network Society*, 11(7-8):1277 – 1303, 1998. doi: 10.1016/s0893-6080(98)00047-1.

Stephen H. Fairclough. Fundamentals of physiological computing. *Interacting with Computers*, 21(1-2):133–145, 2009.

Jiajun Fan, Yuzheng Zhuang, Yuecheng Liu, Jianye Hao, Bin Wang, Jiangcheng Zhu, Hao Wang, and Shu-Tao Xia. Learnable behavior control: Breaking atari human world records via sample-efficient behavior selection. *arXiv preprint arXiv:2305.05239*, 2023.

Shiva Farashahi, Katherine Rowe, Zohra Aslami, Daeyeol Lee, and Alireza Soltani. Feature-based learning improves adaptability without compromising precision. *Nature Communications*, 8(1):1 – 16, 11 2017. doi: 10.1038/s41467-017-01874-w.

Gustav Theodor Fechner. *Vorschule der aesthetik*, volume 1. Breitkopf & Härtel, 1876.

Daniel J. Felleman and D. C. Essen Van. Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex*, 1(1):1–47, 1991.

Thomas D. Ferguson, Daniel N. Bub, Michael E. J. Masson, and Olave E. Krigolson. The role of cognitive control and top-down processes in object affordances. *Attention, Perception, & Psychophysics*, 83:2017–2032, 2021.

Fernando Fernández and Manuela Veloso. Policy reuse for transfer learning across tasks with different state and action spaces. In *ICML Workshop on Structural Knowledge Transfer for Machine Learning*. Citeseer, 2006.

Fernando Fernández, Javier García, and Manuela Veloso. Probabilistic policy reuse for inter-task transfer learning. *Robotics and Autonomous Systems*, 58(7):866–871, 2010.

R. Douglas Fields, Alfonso Araque, Heidi Johansen-Berg, Soo-Siang Lim, Gary Lynch, Klaus-Armin Nave, Maiken Nedergaard, Ray Perez, Terrence Sejnowski, and Hiroaki Wake. Glial biology in learning and cognition. *The Neuroscientist*, 20(5):426–431, 2014.

Thomas H. B. FitzGerald, Karl J. Friston, and Raymond J. Dolan. Characterising reward outcome signals in sensory cortex. *Neuroimage*, 83:329–334, 2013.

Timo Flesch, Andrew Saxe, and Christopher Summerfield. Continual task learning in natural and artificial agents. *Trends in Neurosciences*, 46(3):199–210, 2023.

Vladimir S. Fonov, Alan C. Evans, Robert C. McKinstry, C. Robert Almli, and D. L. Collins. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47:S102, 2009.

Erez Freud, Scott N. Macdonald, Juan Chen, Derek J. Quinlan, Melvyn A. Goodale, and Jody C. Culham. Getting a grip on reality: Grasping movements directed to real objects and images rely on dissociable neural representations. *Cortex*, 98: 34–48, 2018. ISSN 0010-9452. doi: 10.1016/j.cortex.2017.02.020.

Jason P. Gallivan, Cristiana Cavina-Pratesi, and Jody C. Culham. Is that within reach? fmri reveals that the human superior parieto-occipital cortex encodes objects reachable by the hand. *Journal of Neuroscience*, 29(14):4381–4391, 2009.

Jason P. Gallivan, Adam McLean, and Jody C. Culham. Neuroimaging reveals enhanced activation in a reach-selective brain area for objects located within participants' typical hand workspaces. *Neuropsychologia*, 49(13):3710–3721, 2011. ISSN 0028-3932. doi: 10.1016/j.neuropsychologia.2011.09.027.

Leon A. Gatys. A neural algorithm of artistic style. *arXiv preprint ArXiv:1508.06576*, 2015.

Leon A. Gatys, Alexander S. Ecker, and Matthias Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 2414–2423, 2016.

Samuel J. Gershman and Nathaniel D. Daw. Reinforcement learning and episodic memory in humans and animals: an integrative framework. *Annual Review of Psychology*, 68(1):101–128, 2017.

James J. Gibson. *The ecological approach to visual perception: Classic edition*. Psychology Press, 2014.

Charles D. Gilbert and Wu Li. Top-down influences on visual processing. *Nature Reviews Neuroscience*, 14(5):350–363, 2013.

Jan Gläscher, Alan N. Hampton, and John P. O'Doherty. Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making. *Cerebral Cortex*, 19(2):483–495, 2008.

Paul W. Glimcher and Ernst Fehr. *Neuroeconomics: Decision making and the brain*. Academic Press, 2013.

Paul W. Glimcher and Aldo Rustichini. Neuroeconomics: The consilience of brain and decision. *Science*, 306(5695):447–452, 2004.

Mark A. Gluck, Daphna Shohamy, and Catherine Myers. How do people solve the "weather prediction" task?: Individual variability in strategies for probabilistic category learning. *Learning & Memory*, 9(6):408 – 418, 2002. doi: 10.1101/lm.45202.

Joshua I. Gold and Michael N. Shadlen. The neural basis of decision making. *Annual Review of Neuroscience*, 30(1):535–574, 2007.

Melvyn A. Goodale and A. David Milner. Separate visual pathways for perception and action. *Trends in Neurosciences*, 15(1):20–25, 1992.

Krzysztof Gorgolewski, Christopher D. Burns, Cindee Madison, Dav Clark, Yaroslav O. Halchenko, Michael L. Waskom, and Satrajit S. Ghosh. Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in python. *Frontiers in Neuroinformatics*, 5:12318, 2011.

Krzysztof J. Gorgolewski, Oscar Esteban, Christopher J. Markiewicz, Erik Ziegler, David Gage Ellis, Michael Philipp Notter, Dorota Jarecka, Hans Johnson, Christopher Burns, Alexandre Manhães-Savio, Carlo Hamalainen, Benjamin Yvernault, Taylor Salo, Kesshi Jordan, Mathias Goncalves, Michael Waskom, Daniel Clark, Jason Wong, Fred Loney, Marc Modat, Blake E. Dewey, Cindee Madison, Matteo Visconti di Oleggio Castello, Michael G. Clark, Michael Dayan, Dav Clark, Anisha Keshavan, Basile Pinsard, Alexandre Gramfort, Shoshana Berleant, Dylan M. Nielson, Salma Bougacha, Gael Varoquaux, Ben Cipollini, Ross Markello, Ariel Rokem, Brendan Moloney, Yaroslav O. Halchenko, Demian Wassermann, Michael Hanke, Christian Horea, Jakub Kaczmarzyk, Gilles de Hollander, Elizabeth DuPre, Ashley Gillman, David Mordom, Colin Buchanan, Rosalia Tungaraza, Wolfgang M. Pauli, Shariq Iqbal, Sharad Sikka, Matteo Mancini, Yannick Schwartz, Ian B. Malone, Mathieu Dubois, Caroline Frohlich, David Welch, Jessica Forbes, James Kent, Aimi Watanabe, Chad Cumba, Julia M. Huntenburg, Erik Kastman, B. Nolan Nichols, Arman Eshaghi, Daniel Ginsburg, Alexander Schaefer, Benjamin Acland, Steven Giavasis, Jens Kleesiek, Drew Erickson, René Küttner, Christian Haselgrove, Carlos Correa, Ali Ghayoor, Franz Liem, Jarrod Millman, Daniel Haehn, Jeff Lai, Dale Zhou, Ross Blair, Tristan Glatard, Mandy Renfro, Siqi Liu, Ari E. Kahn, Fernando Pérez-García, William Triplett, Leonie Lampe, Jörg Stadler, Xiang-Zhen Kong, Michael Hallquist, Andrey Chetverikov, John Salvatore, Anne Park, Russell Poldrack, R. Cameron Craddock, Souheil Inati, Oliver Hinds, Gavin Cooper, L. Nathan Perkins, Ana Marina, Aaron Mattfeld, Maxime Noel, Lukas Snoek, K Matsubara, Brian Cheung, Simon Rothmei, Sebastian Urchs, Joke Durnez, Fred Mertz, Daniel Geisler, Andrew Floren, Stephan Gerhard, Paul Sharp, Miguel Molina-Romero, Alejandro Weinstein, William Broderick, Victor Saase, Sami Kristian Andberg, Robbert Harms, Kai Schlamp, Jaime Arias, Dimitri Papadopoulos Orfanos, Claire Tarbert, Arielle Tambini, Alejandro De La Vega, Thomas Nickson, Matthew Brett, Marcel Falkiewicz, Kornelius Podranski, Janosch Linkersdörfer, Guillaume Flandin, Eduard Ort, Dmitry Shachnev, Daniel McNamee, Andrew Davison, Jan Varada,

Isaac Schwabacher, John Pellman, Martin Perez-Guevara, Ranjeet Khanuja, Nicolas Pannetier, Conor McDermottroe, and Satrajit Ghosh. Nipype. *Software*, 2018. doi: 10.5281/zenodo.596855.

Fabian Grabenhorst and Edmund T. Rolls. Value, pleasure and choice in the ventral prefrontal cortex. *Trends in Cognitive Sciences*, 15(2):56–67, 2011.

Daniel Graham and David Field. Statistical regularities of art images and natural scenes: Spectra, sparseness and nonlinearities. *Spatial Vision*, 21(1-2):149–164, 2008.

Douglas N. Greve and Bruce Fischl. Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, 48(1):63–72, 2009. ISSN 1095-9572. doi: 10.1016/j.neuroimage.2009.06.060.

Julie Grèzes, Mike Tucker, Jorge Armony, Rob Ellis, and Richard E. Passingham. Objects automatically potentiate action: An fmri study of implicit processing. *European Journal of Neuroscience*, 17(12):2735–2740, 2003.

Umut Güçlü and Marcel A. J. van Gerven. Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of Neuroscience*, 35(27):10005–10014, 2015.

Alan N. Hampton and John P. O'doherty. Decoding the neural substrates of reward-related decision making with functional mri. *Proceedings of the National Academy of sciences*, 104(4):1377–1382, 2007.

Todd A. Hare, Colin F. Camerer, and Antonio Rangel. Self-control in decision-making involves modulation of the vmpfc valuation system. *Science*, 324(5927): 646–648, 2009.

Todd A. Hare, Wolfram Schultz, Colin F. Camerer, John P. O'Doherty, and Antonio Rangel. Transformation of stimulus value signals into motor commands during simple choice. *Proceedings of the National Academy of Sciences*, 108(44):18120–18125, 2011. ISSN 0027-8424. doi: 10.1073/pnas.1109322108.

Hussein Hazimeh, Zhe Zhao, Aakanksha Chowdhery, Maheswaran Sathiamoorthy, Yihua Chen, Rahul Mazumder, Lichan Hong, and Ed Chi. Dselect-k: Differentiable selection in the mixture of experts with applications to multi-task learning. *Advances in Neural Information Processing Systems*, 34:29335–29347, 2021.

Matthias Hein and Thomas Bühler. An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse pca. In *Advances in Neural Information Processing Systems*, pages 847–855, 2010.

Paul Hekkert and Piet C. W. van Wieringen. The impact of level of expertise on the evaluation of original and altered versions of post-impressionistic paintings. *Acta Psychologica*, 94(2):117–131, 1996.

Anna Heuer, Sven Ohl, and Martin Rolfs. Memory for action: A functional view of selection in visual working memory. *Visual Cognition*, 28(5-8):388–400, 2020. ISSN 1350-6285. doi: 10.1080/13506285.2020.1764156.

Salah Hihi and Yoshua Bengio. Hierarchical recurrent neural networks for long-term dependencies. *Advances in neural information processing systems*, 8, 1995.

Shaul Hochstein and Merav Ahissar. View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5):791–804, 2002.

Thomas Hofmann, Bernhard Schölkopf, and Alexander J. Smola. Kernel methods in machine learning. *The Annals of Statistics*, pages 1171–1220, 2008.

Bernhard Hommel. Event files: Feature binding in and across perception and action. *Trends in Cognitive Sciences*, 8(11):494–500, 2004. ISSN 1364-6613. doi: 10.1016/j.tics.2004.08.007.

Bernhard Hommel, Jochen Müsseler, Gisa Aschersleben, and Wolfgang Prinz. The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences*, 24(5):849–878, 2001. ISSN 1469-1825. doi: 10.1017/s0140525x01000103.

Ha Hong, Daniel L. K. Yamins, Najib J. Majaj, and James J. DiCarlo. Explicit information for category-orthogonal object properties increases along the ventral stream. *Nature Neuroscience*, 19(4):613, 2016.

Daniel G. Horvitz and Donovan J. Thompson. A generalization of sampling without replacement from a finite universe. *Journal of the American statistical Association*, 47(260):663–685, 1952.

James D. Howard and Jay A. Gottfried. Configural and elemental coding of natural odor mixture components in the human brain. *Neuron*, 84(4):857–869, 2014.

Clare Howarth, Anusha Mishra, and Catherine Hall. More than just summed neuronal activity: how multiple cell types shape the BOLD response. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 376(1815):20190630–20190630, 2021. ISSN 0962-8436. doi: 10.1098/rstb.2019.0630.

Peter J. Huber. Robust estimation of a location parameter. *Ann. Math. Statist.*, 35 (1):73–101, 03 1964. doi: 10.1214/aoms/1177703732.

Cristina Iani, Giulia Baroni, Antonello Pellicano, and Roberto Nicoletti. On the relationship between affordance and Simon effects: Are the effects really independent? *Journal of Cognitive Psychology*, 23(1):121–131, 2011. ISSN 2044-5911. doi: 10.1080/20445911.2011.467251.

Kiyohito Iigaya, Tobias U Hauser, Zeb Kurth-Nelson, John P. O'Doherty, Peter Dayan, and Raymond J. Dolan. The value of what's to come: Neural mechanisms coupling prediction error and the utility of anticipation. *Science Advances*, 6(25): eaba3828, 2020a.

Kiyohito Iigaya, Sanghyun Yi, Iman A. Wahle, Koranis Tanwisuth, and John P. O'Doherty. Aesthetic preference for art emerges from a weighted integration over hierarchically structured visual features in the brain. *bioRxiv*, 2020b.

Kiyohito Iigaya, Sanghyun Yi, Iman A. Wahle, Koranis Tanwisuth, and John P. O'Doherty. Aesthetic preference for art can be predicted from a mixture of low- and high-level visual features. *Nature Human Behaviour*, 5(6):743–755, 2021.

Kiyohito Iigaya, Sanghyun Yi, Iman A. Wahle, Sandy Tanwisuth, Logan Cross, and John P. O'Doherty. Neural mechanisms underlying the hierarchical construction of perceived aesthetic value. *Nature Communications*, 14(1):127, 2023.

Tomohiro Ishizu and Semir Zeki. The brain's specialized systems for aesthetic and perceptual judgment. *European Journal of Neuroscience*, 37(9):1413–1420, 2013.

Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, 17(2):825–841, 2002. ISSN 1053-8119. doi: 10.1006/nimg.2002.1132. URL http://www.sciencedirect.com/science/article/pii/S1053811902911328.

Li Ji-An, Marcus K. Benna, and Marcelo G. Mattar. Automatic Discovery of Cognitive Strategies with Tiny Recurrent Neural Networks. *bioRxiv*, page 2023.04.12.536629, 2023a. doi: 10.1101/2023.04.12.536629.

Li Ji-An, Marcus K Benna, and Marcelo G. Mattar. Automatic discovery of cognitive strategies with tiny recurrent neural networks. *bioRxiv*, pages 2023–04, 2023b.

Scott H. Johnson-Frey. The neural bases of complex tool use in humans. *Trends in Cognitive Sciences*, 8(2):71 – 78, 2004. doi: 10.1016/j.tics.2003.12.002.

Joseph W. Kable and Paul W. Glimcher. The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, 10(12):1625, 2007.

Sayan Kahali, Marcus E Raichle, and Dmitriy A Yablonskiy. The role of the human brain neuron–glia–synapse composition in forming resting-state functional connectivity networks. *Brain Sciences*, 11(12):1565, 2021.

Daniel Kahneman. Thinking, fast and slow. *Farrar, Straus and Giroux*, 2011.

Daniel Kahneman and Amos Tversky. Choices, values, and frames. *American Psychologist*, 39(4):341, 1984.

Daniel Kahneman and Amos Tversky. Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I*, pages 99–127. World Scientific, 2013.

Thorsten Kahnt, Jakob Heinzle, Soyoung Q Park, and John-Dylan Haynes. Decoding different roles for vmPFC and dlPFC in multi-attribute decision making. *NeuroImage*, 56(2):709 – 715, 05 2011a. doi: 10.1016/j.neuroimage.2010.05.058.

Thorsten Kahnt, Jakob Heinzle, Soyoung Q. Park, and John-Dylan Haynes. Decoding different roles for vmpfc and dlpfc in multi-attribute decision making. *Neuroimage*, 56(2):709–715, 2011b.

Immanuel Kant. *Critique of judgment*. Hackett Publishing, 1987.

Kohitij Kar and James J. DiCarlo. Fast recurrent processing via ventrolateral prefrontal cortex is needed by the primate ventral stream for robust core visual object recognition. *Neuron*, 2020.

Kohitij Kar, Jonas Kubilius, Kailyn Schmidt, Elias B. Issa, and James J. DiCarlo. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience*, 22(6):974–983, 2019.

Hideaki Kawabata and Semir Zeki. Neural correlates of beauty. *Journal of Neurophysiology*, 91(4):1699–1705, 2004.

Yan Ke, Xiaoou Tang, and Feng Jing. The design of high-level features for photo quality assessment. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 1, pages 419–426. IEEE, 2006.

Alexander J. E. Kell, Daniel L. K. Yamins, Erica N. Shook, Sam V. Norman-Haignere, and Josh H. McDermott. A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy. *Neuron*, 98(3):630–644, 2018.

Attila Keresztes, Chi T. Ngo, Ulman Lindenberger, Markus Werkle-Bergner, and Nora S. Newcombe. Hippocampal Maturation Drives Memory from Generalization to Specificity. *Trends in Cognitive Sciences*, 22(8):676–686, 2018. ISSN 1364-6613. doi: 10.1016/j.tics.2018.05.004.

Seyed-Mahdi Khaligh-Razavi and Nikolaus Kriegeskorte. Deep supervised, but not unsupervised, models may explain it cortical representation. *PLoS Computational Biology*, 10(11):e1003915, 2014.

Khimya Khetarpal, Matthew Riemer, Irina Rish, and Doina Precup. Towards continual reinforcement learning: A review and perspectives. *Journal of Artificial Intelligence Research*, 75:1401–1476, 2022.

Alan Kingstone, Daniel Smilek, and John D. Eastwood. Cognitive ethology: A new approach for studying human cognition. *British Journal of Psychology*, 99(3): 317–340, 2008.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

Roberta L. Klatzky, Brian McCloskey, Sally Doherty, James Pellegrino, and Terence Smith. Knowledge About Hand Shaping and Knowledge About Objects. *Journal of Motor Behavior*, 19(2):187–213, 1987. ISSN 0022-2895. doi: 10.1080/00222895.1987.10735407.

Arno Klein, Satrajit S. Ghosh, Forrest S. Bao, Joachim Giard, Yrjö Häme, Eliezer Stavsky, Noah Lee, Brian Rossa, Martin Reuter, Elias Chaibub Neto, and Anisha Keshavan. Mindboggling morphometry of human brains. *PLOS Computational Biology*, 13(2):e1005350, 2017. ISSN 1553-7358. doi: 10.1371/journal.pcbi.1005350. URL `http://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1005350`.

B J Knowlton and L R Squire. The learning of categories: parallel brain systems for item memory and category knowledge. *Science*, 262(5140):1747 – 1749, 12 1993. doi: 10.1126/science.8259522.

Paulo Kofuji and Alfonso Araque. Astrocytes and Behavior. *Annual Review of Neuroscience*, 44(1):1–19, 2021. ISSN 0147-006X. doi: 10.1146/annurev-neuro-101920-112225. Thorough review of astrocytes' role in behavior.

Christina S. Konen and Sabine Kastner. Two hierarchically organized neural systems for object information in human visual cortex. *Nature Neuroscience*, 11(2):224, 2008.

Michal Kosinski, David Stillwell, and Thore Graepel. Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802–5805, 2013.

Leo Kozachkov, Ksenia V. Kastanenka, and Dmitry Krotov. Building transformers from neurons and astrocytes. *Proceedings of the National Academy of Sciences*, 120(34):e2219150120, 2023.

Dwight J. Kravitz, Kadharbatcha S. Saleem, Chris I. Baker, and Mortimer Mishkin. A new neural framework for visuospatial processing. *Nature Reviews Neuroscience*, 12(4):217–230, 2011.

Dwight J. Kravitz, Kadharbatcha S. Saleem, Chris I. Baker, Leslie G. Ungerleider, and Mortimer Mishkin. The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences*, 17(1): 26–49, 2013. ISSN 1364-6613. doi: 10.1016/j.tics.2012.10.011.

Nikolaus Kriegeskorte. Deep neural networks: A new framework for modeling biological vision and brain information processing. *Annual Review of Vision Science*, 1(1):417–446, 2015.

C. Lanczos. Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics Series B Numerical Analysis*, 1(1):76–85, 1964. ISSN 0887-459X. doi: 10.1137/0701007. URL http://epubs.siam.org/doi/10.1137/0701007.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521 (7553):436, 2015.

Helmut Leder and Marcos Nadal. Ten years of a model of aesthetic appreciation and aesthetic judgments: The aesthetic episode–developments and challenges in empirical aesthetics. *British Journal of Psychology*, 105(4):443–464, 2014.

Helmut Leder, Benno Belke, Andries Oeberst, and Dorothee Augustin. A model of aesthetic appreciation and aesthetic judgments. *British Journal of Psychology*, 95(4):489–508, 2004.

Sang Wan Lee, Shinsuke Shimojo, and John P. O'doherty. Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, 81 (3):687–699, 2014a.

Sang Wan Lee, Shinsuke Shimojo, and John P. O'doherty. Neural computations underlying arbitration between model-based and model-free learning. *Neuron*, 81 (3):687–699, 2014b.

Yuan Chang Leong, Angela Radulescu, Reka Daniel, Vivian DeWoskin, and Yael Niv. Dynamic Interaction between Reinforcement Learning and Attention in Multidimensional Environments. *Neuron*, 93(2):451 – 463, 01 2017. doi: 10.1016/j.neuron.2016.12.040.

Moshe Leshno, Vladimir Ya Lin, Allan Pinkus, and Shimon Schocken. Multilayer feedforward networks with a nonpolynomial activation function can approximate any function. *Neural Networks*, 6(6):861–867, 1993.

Dino J. Levy and Paul W. Glimcher. The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, 22(6):1027–1038, 2012.

Congcong Li and Tsuhan Chen. Aesthetic visual quality assessment of paintings. *IEEE Journal of selected topics in Signal Processing*, 3(2):236–252, 2009.

Seung-Lark Lim, John P. O'Doherty, and Antonio Rangel. Stimulus value signals in ventromedial pfc reflect the integration of attribute value signals computed in fusiform gyrus and posterior superior temporal gyrus. *Journal of Neuroscience*, 33(20):8729–8741, 2013.

Alberto Llera, Thomas Wolfers, Peter Mulders, and Christian F. Beckmann. Inter-individual differences in human brain structure and morphology link to variation in demographics and behavior. *Elife*, 8:e44443, 2019.

Terry Lohrenz, Kevin McCabe, Colin F. Camerer, and P. Read Montague. Neural signature of fictive learning signals in a sequential investment task. *Proceedings of the National Academy of Sciences*, 104(22):9493–9498, 2007. ISSN 0027-8424. doi: 10.1073/pnas.0608842104.

Jiaqi Ma, Zhe Zhao, Xinyang Yi, Jilin Chen, Lichan Hong, and Ed H Chi. Modeling task relationships in multi-task learning with multi-gate mixture-of-experts. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1930–1939, 2018.

Birgit Mallon, Christoph Redies, and Gregor Uwe Hayn-Leichsenring. Beauty in abstract paintings: Perceptual contrast and statistical properties. *Frontiers in Human Neuroscience*, 8:161, 2014.

Valerio Mante, David Sussillo, Krishna V. Shenoy, and William T. Newsome. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature*, 503(7474):78, 2013.

Francesco Marini, Katherine A. Breeding, and Jacqueline C. Snow. Distinct visuo-motor brain dynamics for real-world objects versus planar images. *NeuroImage*, 195:232–242, 2019. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2019.02.026.

Shirley Mark, Rani Moran, Thomas Parr, Steve W. Kennerley, and Timothy E. J. Behrens. Transferring structural knowledge across cognitive maps in humans and models. *Nature Communications*, 11(1):4783, 2020.

Alex Martin. The representation of object concepts in the brain. *Annual Review of Psychology*, 58(1):25 – 45, 2007. doi: 10.1146/annurev.psych.57.102904.190143.

Kenji Matsumoto and Keiji Tanaka. Conflict and cognitive control. *Science*, 303 (5660):969–970, 2004. ISSN 0036-8075. doi: 10.1126/science.1094733.

Daniel McNamee, Mimi Liljeholm, Ondrej Zika, and John P. O'Doherty. Characterizing the associative content of brain structures involved in habitual and goal-directed actions in humans: a multivariate fmri study. *Journal of Neuroscience*, 35(9):3764–3771, 2015.

Sara Mederos, Cristina Sánchez-Puelles, Julio Esparza, Manuel Valero, Alexey Ponomarenko, and Gertrudis Perea. GABAergic signaling to astrocytes in the prefrontal cortex sustains goal-directed behaviors. *Nature Neuroscience*, 24(1): 82–92, 2021. ISSN 1097-6256. doi: 10.1038/s41593-020-00752-x.

Earl K. Miller. The prefontral cortex and cognitive control. *Nature Reviews Neuroscience*, 1(1):59–65, 2000. ISSN 1471-003X. doi: 10.1038/35036228.

Earl K. Miller and Jonathan D. Cohen. An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24(1):167–202, 2001.

Kevin Miller, Maria Eckstein, Matt Botvinick, and Zeb Kurth-Nelson. Cognitive model discovery via disentangled rnns. *Advances in Neural Information Processing Systems*, 36, 2024.

A. David Milner and Melvyn A. Goodale. Two visual systems re-viewed. *Neuropsychologia*, 46(3):774–785, 2008.

Mortimer Mishkin, Leslie G. Ungerleider, and Kathleen A. Macko. Object vision and spatial vision: two cortical pathways. *Trends in Neurosciences*, 6:414–417, 1983.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.

P. R. Montague, P. Dayan, and T. J. Sejnowski. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *The Journal of Neuroscience*, 16(5):1936–1947, 1996. ISSN 0270-6474. doi: 10.1523/jneurosci.16-05-01936.1996.

Yu Mu, Davis V. Bennett, Mikail Rubinov, Sujatha Narayan, Chao-Tsung Yang, Masashi Tanimoto, Brett D. Mensh, Loren L. Looger, and Misha B Ahrens. Glia accumulate evidence that actions are futile and suppress unsuccessful behavior. *Cell*, 178(1):27–43, 2019.

Naila Murray and Albert Gordo. A deep architecture for unified aesthetic prediction. *arXiv preprint arXiv:1708.04890*, 2017.

Naila Murray, Luca Marchesotti, and Florent Perronnin. Ava: A large-scale database for aesthetic visual analysis. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 2408–2415. IEEE, 2012.

Parashkev Nachev, Christopher Kennard, and Masud Husain. Functional role of the supplementary and pre-supplementary motor areas. *Nature Reviews Neuroscience*, 9(11):856–869, 2008. ISSN 1471-003X. doi: 10.1038/nrn2478.

Thomas Naselaris, Kendrick N. Kay, Shinji Nishimoto, and Jack L. Gallant. Encoding and decoding in fmri. *Neuroimage*, 56(2):400–410, 2011.

Daniel Neil, Michael Pfeiffer, and Shih-Chii Liu. Phased lstm: Accelerating recurrent network training for long or event-based sequences. *Advances in neural information processing systems*, 29, 2016.

Yael Niv, Reka Daniel, Andra Geana, Samuel J. Gershman, Yuan Chang Leong, Angela Radulescu, and Robert C. Wilson. Reinforcement learning in multidimensional environments relies on attention mechanisms. *The Journal of Neuroscience*, 35(21):8145 – 8157, 05 2015. doi: 10.1523/jneurosci.2978-14.2015.

Richard Nock and Frank Nielsen. Statistical region merging. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(11):1452–1458, 2004.

M. P. Noonan, M. E. Walton, T. E. J. Behrens, J. Sallet, MJ Buckley, and MFS Rushworth. Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex. *Proceedings of the National Academy of Sciences*, 107(47):20547–20552, 2010.

John O'Doherty, Peter Dayan, Johannes Schultz, Ralf Deichmann, Karl Friston, and Raymond J. Dolan. Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science*, 304(5669):452–454, 2004.

John P. O'Doherty, Peter Dayan, Karl Friston, Hugo Critchley, and Raymond J. Dolan. Temporal difference models and reward-related learning in the human brain. *Neuron*, 38(2):329–337, 2003.

Christian N.L. Olivers and Pieter R. Roelfsema. Attention for action in visual working memory. *Cortex*, 131:179–194, 2020. ISSN 0010-9452. doi: 10.1016/j.cortex.2020.07.011.

François Osiurak, Yves Rossetti, and Arnaud Badets. What is an affordance? 40 years later. *Neuroscience & Biobehavioral Reviews*, 77:403–417, 2017. ISSN 0149-7634. doi: 10.1016/j.neubiorev.2017.04.014.

John P. O'Doherty. The problem with value. *Neuroscience & Biobehavioral Reviews*, 43:259–268, 2014.

John P. O'Doherty, Sang Wan Lee, Reza Tadayonnejad, Jeff Cockburn, Kyo Iigaya, and Caroline J. Charpentier. Why and how the brain weights contributions from a mixture of experts. *Neuroscience & Biobehavioral Reviews*, 123:14–23, 2021a.

John P. O'Doherty, Sangwan Lee, Reza Tadayonnejad, Jeff Cockburn, Kyo Iigaya, and Caroline J. Charpentier. Why and how the brain weights contributions from a mixture of experts. *Neuroscience & Biobehavioral Reviews*, 123:14–23, 2021b. ISSN 0149-7634. doi: 10.1016/j.neubiorev.2020.10.022.

John P. O'Doherty, Ueli Rutishauser, and Kiyohito Iigaya. The hierarchical construction of value. *Current Opinion in Behavioral Sciences*, 41:71–77, 2021c.

Camillo Padoa-Schioppa and John A. Assad. Neurons in the orbitofrontal cortex encode economic value. *Nature*, 441(7090):223–226, 2006.

Stephen E. Palmer, Karen B. Schloss, and Jonathan Sammartino. Visual aesthetics and human preference. *Annual Review of Psychology*, 64:77–107, 2013.

Zissis Pappas. Dissociating Simon and affordance compatibility effects: Silhouettes and photographs. *Cognition*, 133(3):716–728, 2014. ISSN 0010-0277. doi: 10.1016/j.cognition.2014.08.018.

German I. Parisi, Ronald Kemker, Jose L. Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural Networks*, 113:54–71, 2019.

Alexandre Pastor-Bernier and Paul Cisek. Neural correlates of biased competition in premotor cortex. *The Journal of Neuroscience*, 31(19):7083 – 7088, 05 2011. doi: 10.1523/jneurosci.5681-10.2011.

Gabriel Pelletier and Lesley K. Fellows. A critical role for human ventromedial frontal lobe in value comparison of complex objects based on attribute configuration. *Journal of Neuroscience*, 39(21):4124–4132, 2019.

Matthew Pelowski, Patrick S. Markey, Jon O. Lauring, and Helmut Leder. Visualizing the impact of art: An update and comparison of current psychological models of art experience. *Frontiers in Human Neuroscience*, 10:160, 2016.

Gertrudis Perea, Marta Navarrete, and Alfonso Araque. Tripartite synapses: astrocytes process and control synaptic information. *Trends in Neurosciences*, 32(8): 421–431, 2009a. ISSN 0166-2236. doi: 10.1016/j.tins.2009.05.001.

Gertrudis Perea, Marta Navarrete, and Alfonso Araque. Tripartite synapses: Astrocytes process and control synaptic information. *Trends in Neurosciences*, 32(8): 421–431, 2009b.

Giovanni Pezzulo and Paul Cisek. Navigating the affordance landscape: Feedback control as a process model of behavior and cognition. *Trends in Cognitive Sciences*, 20(6):414–424, 2016.

Payam Piray, Amir Dezfouli, Tom Heskes, Michael J. Frank, and Nathaniel D. Daw. Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. *PLoS Computational Biology*, 15(6):e1007043 – 34, 2019a. doi: 10.1371/journal.pcbi.1007043.

Payam Piray, Amir Dezfouli, Tom Heskes, Michael J. Frank, and Nathaniel D. Daw. Hierarchical bayesian inference for concurrent model fitting and comparison for group studies. *PLoS Computational Biology*, 15(6):e1007043, 2019b.

Hilke Plassmann, John O'doherty, and Antonio Rangel. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. *Journal of Neuroscience*, 27(37):9984–9988, 2007.

Michael L. Platt and Paul W. Glimcher. Neural correlates of decision variables in parietal cortex. *Nature*, 400(6741):233–238, 1999.

Vonne van Polanen and Marco Davare. Interactions between dorsal and ventral streams for controlling skilled grasp. *Neuropsychologia*, 79(Pt B):186–191, 2015. ISSN 0028-3932. doi: 10.1016/j.neuropsychologia.2015.07.010.

R. A. Poldrack, J. Clark, E. J. Paré-Blagoev, D. Shohamy, J. Creso Moyano, C. Myers, and M. A. Gluck. Interactive memory systems in the human brain. *Nature*, 414 (6863):546 – 550, 11 2001. doi: 10.1038/35107080.

Ana B. Porto-Pazos, Noha Veiguela, Pablo Mesejo, Marta Navarrete, Alberto Alvarellos, Oscar Ibáñez, Alejandro Pazos, and Alfonso Araque. Artificial astrocytes improve neural network performance. *PloS One*, 6(4):e19109, 2011.

Jonathan D. Power, Anish Mitra, Timothy O. Laumann, Abraham Z. Snyder, Bradley L. Schlaggar, and Steven E. Petersen. Methods to detect, characterize, and remove motion artifact in resting state fmri. *NeuroImage*, 84 (Supplement C):320–341, 2014. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2013.08.048. URL http://www.sciencedirect.com/science/article/pii/S1053811913009117.

Doina Precup, Richard S. Sutton, and Satinder Singh. Eligibility traces for off-policy policy evaluation. In Pat Langley, editor, *Proceedings of the Seventeenth International Conference on Machine Learning (ICML 2000), Stanford University, Stanford, CA, USA, June 29 - July 2, 2000*, pages 759–766. Morgan Kaufmann, 2000.

Robert W. Proctor and James D. Miles. Does the concept of affordance add anything to explanations of stimulus–response compatibility effects? 60:227–266, 2014.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*, pages 8748–8763. PMLR, 2021.

Vilayanur S. Ramachandran and William Hirstein. The science of art: A neurological theory of aesthetic experience. *Journal of Consciousness Studies*, 6(6-7): 15–51, 1999.

Antonio Rangel, Colin Camerer, and P. Read Montague. A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9(7):545–556, 2008.

S. Chenchal Rao, Gregor Rainer, and Earl K. Miller. Integration of what and where in the primate prefrontal cortex. *Science*, 276(5313):821–824, 1997.

Paul J. Reber, Barbara J. Knowlton, and Larry R. Squire. Dissociable properties of memory systems: Differences in the flexibility of declarative and nondeclarative knowledge. *Behavioral Neuroscience*, 110(5):861 – 871, 1996. doi: 10.1037//0735-7044.110.5.861. URL https://www.researchgate.net/.

José J. F. Ribas-Fernandes, Danesh Shahnazian, Clay B. Holroyd, and Matthew M. Botvinick. Subgoal-and goal-related reward prediction errors in medial prefrontal cortex. *Journal of Cognitive Neuroscience*, 31(1):8–23, 2019.

Nichola J. Rice, Kenneth F. Valyear, Melvyn A. Goodale, A. David Milner, and Jody C. Culham. Orientation sensitivity to graspable objects: An fMRI adaptation study. *NeuroImage*, 36:T87–T93, 2007. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2007.03.032.

Mattia Rigotti, Omri Barak, Melissa R Warden, Xiao-Jing Wang, Nathaniel D. Daw, Earl K. Miller, and Stefano Fusi. The importance of mixed selectivity in complex cognitive tasks. *Nature*, 497(7451):585, 2013.

Giacomo Rizzolatti and Giuseppe Luppino. The Cortical Motor System. *Neuron*, 31(6):889–901, 2001. ISSN 0896-6273. doi: 10.1016/s0896-6273(01)00423-8.

Giacomo Rizzolatti, Lucia Riggio, Isabella Dascola, and Carlo Umiltá. Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention. *Neuropsychologia*, 25(1):31–40, 1987. ISSN 0028-3932. doi: 10.1016/0028-3932(87)90041-8.

Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut: Interactive foreground extraction using iterated graph cuts. In *ACM transactions on graphics (TOG)*, volume 23, pages 309–314. ACM, 2004.

MFS Rushworth, Mark E Walton, Steve W Kennerley, and DM Bannerman. Action sets and decisions in the medial frontal cortex. *Trends in Cognitive Sciences*, 8 (9):410–417, 2004.

Masamichi Sakagami and Xiaochuan Pan. Functional role of the ventrolateral prefrontal cortex in decision making. *Current Opinion in Neurobiology*, 17(2): 228–233, 2007. ISSN 0959-4388. doi: 10.1016/j.conb.2007.02.008.

Katrin Sakreida, Isabel Effnert, Serge Thill, Mareike M. Menz, Doreen Jirak, Claudia R. Eickhoff, Tom Ziemke, Simon B. Eickhoff, Anna M. Borghi, and Ferdinand Binkofski. Affordance processing in segregated parieto-frontal dorsal stream sub-pathways. *Neuroscience & Biobehavioral Reviews*, 69:89–112, 2016. ISSN 0149-7634. doi: 10.1016/j.neubiorev.2016.07.032.

Mohamed Ben Salah, Amar Mitiche, and Ismail Ben Ayed. Multiregion image segmentation by parametric kernel graph cuts. *IEEE Transactions on Image Processing*, 20(2):545–557, 2010.

Alan G. Sanfey, George Loewenstein, Samuel M. McClure, and Jonathan D. Cohen. Neuroeconomics: Cross-currents in research on decision-making. *Trends in Cognitive Sciences*, 10(3):108–116, 2006. ISSN 1364-6613. doi: 10.1016/j.tics.2006.01.009.

Theodore D. Satterthwaite, Mark A. Elliott, Raphael T. Gerraty, Kosha Ruparel, James Loughead, Monica E. Calkins, Simon B. Eickhoff, Hakon Hakonarson, Ruben C. Gur, Raquel E. Gur, and Daniel H. Wolf. An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *NeuroImage*, 64(1):240–256, 2013. ISSN 10538119. doi: 10.1016/j.neuroimage.2012.08.052.

Jeffrey D. Schall. Neural basis of deciding, choosing and acting. *Nature Reviews Neuroscience*, 2(1):33–42, 2001.

Martin Schrimpf, Jonas Kubilius, Ha Hong, Najib J. Majaj, Rishi Rajalingham, Elias B. Issa, Kohitij Kar, Pouya Bashivan, Jonathan Prescott-Roy, Franziska Geiger, et al. Brain-score: Which artificial neural network for object recognition is most brain-like? *BioRxiv*, page 407007, 2018.

Wolfram Schultz, Peter Dayan, and P. Read Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997.

John T. Serences. Value-based modulations in human visual cortex. *Neuron*, 60(6): 1169–1181, 2008.

Michael N. Shadlen and William T. Newsome. Neural basis of a perceptual decision in the parietal cortex (area lip) of the rhesus monkey. *Journal of Neurophysiology*, 86(4):1916–1936, 2001.

Reza Shadmehr, Thomas R. Reppert, Erik M. Summerside, Tehrim Yoon, and Alaa A. Ahmed. Movement vigor as a reflection of subjective economic utility. *Trends in Neurosciences*, 42(5):323–336, 2019.

Amitai Shenhav, Matthew M. Botvinick, and Jonathan D. Cohen. The Expected Value of Control: An Integrative Theory of Anterior Cingulate Cortex Function. *Neuron*, 79(2):217–240, 2013. ISSN 0896-6273. doi: 10.1016/j.neuron.2013.07.007.

Amitai Shenhav, Sebastian Musslick, Falk Lieder, Wouter Kool, Thomas L. Griffiths, Jonathan D. Cohen, and Matthew M. Botvinick. Toward a rational and mechanistic account of mental effort. *Annual Review of Neuroscience*, 40(1):99–124, 2017.

Arthur P. Shimamura and Stephen E. Palmer. *Aesthetic science: Connecting minds, brains, and experience*. OUP USA, 2012.

Daphna Shohamy and Nathaniel D. Daw. Integrating memories to guide decisions. *Current Opinion in Behavioral Sciences*, 5:85–90, 2015.

David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.

Tomas Simon, Hanbyul Joo, Iain Matthews, and Yaser Sheikh. Hand keypoint detection in single images using multiview bootstrapping. In *CVPR*, 2017.

Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

Philip L. Smith and Daniel R. Little. Small is beautiful: In defense of the small-n design. *Psychonomic Bulletin & Review*, 25(6):2083–2101, 2018.

Lawrence H. Snyder, Aaron P. Batista, and Richard A. Andersen. Coding of intention in the posterior parietal cortex. *Nature*, 386(6621):167–170, 1997.

Fabian A Soto, Samuel J Gershman, and Yael Niv. Explaining compound generalization in associative and causal learning through rational principles of dimensional generalization. *Psychological Review*, 121(3):526 – 558, 2014. doi: 10.1037/a0037018.

Larry R.. Squire and Stuart M. Zola. Structure and function of declarative and nondeclarative memory systems. *Proceedings of the National Academy of Sciences*, 93(24):13515–13522, 1996. ISSN 0027-8424. doi: 10.1073/pnas.93.24.13515.

Liviu Stănişor, Chris van der Togt, Cyriel M. A. Pennartz, and Pieter R. Roelfsema. A unified selection signal for attention and reward in primary visual cortex. *Proceedings of the National Academy of Sciences*, 110(22):9136–9141, 2013.

Klaas Enno Stephan, Will D. Penny, Jean Daunizeau, Rosalyn J. Moran, and Karl J. Friston. Bayesian model selection for group studies. *NeuroImage*, 46(4):1004–1017, 2009a. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2009.03.025.

Klaas Enno Stephan, Will D. Penny, Jean Daunizeau, Rosalyn J. Moran, and Karl J. Friston. Bayesian model selection for group studies. *Neuroimage*, 46(4):1004–1017, 2009b.

Richard S. Sutton. Reinforcement learning: An introduction. *A Bradford Book*, 2018.

Shinsuke Suzuki, Logan Cross, and John P. O'Doherty. Elucidating the underlying components of food valuation in the human orbitofrontal cortex. *Nature Neuroscience*, 20(12):1780, 2017a.

Shinsuke Suzuki, Logan Cross, and John P. O'Doherty. Elucidating the underlying components of food valuation in the human orbitofrontal cortex. *Nature Neuroscience*, 20(12):1780–1786, 2017b.

Ed Symes, Rob Ellis, and Mike Tucker. Dissociating object-based and space-based affordances. *Visual Cognition*, 12(7):1337–1361, 2005. ISSN 1350-6285. doi: 10.1080/13506280444000445.

Ed Symes, Rob Ellis, and Mike Tucker. Visual object affordances: Object orientation. *Acta Psychologica*, 124(2):238–255, 2007. ISSN 0001-6918. doi: 10.1016/j.actpsy.2006.03.005.

Norio Takata, Yuki Sugiura, Keitaro Yoshida, Miwako Koizumi, Nishida Hiroshi, Kurara Honda, Ryutaro Yano, Yuji Komaki, Ko Matsui, Makoto Suematsu, et al. Optogenetic astrocyte activation evokes bold fmri response with oxygen consumption without neuronal activity modulation. *Glia*, 66(9):2013–2023, 2018.

Chen Tessler, Shahar Givony, Tom Zahavy, Daniel Mankowitz, and Shie Mannor. A deep hierarchical approach to lifelong learning in minecraft. In *Proceedings of the AAAI conference on Artificial Intelligence*, volume 31, 2017.

Richard Thaler. Mental accounting and consumer choice. *Marketing Science*, 4(3): 199–214, 1985.

Sebastian Thrun and Tom M. Mitchell. Lifelong robot learning. *Robotics and autonomous systems*, 15(1-2):25–46, 1995.

Edward C. Tolman. Cognitive maps in rats and men. *Psychological Review*, 55(4): 189, 1948.

Momchil S. Tomov, Eric Schulz, and Samuel J. Gershman. Multi-task reinforcement learning in humans. *Nature Human Behaviour*, 104:1 – 12, 01 2021. doi: 10.1038/s41562-020-01035-y. UVFA: Q function approximation includes task information (here the w term) SF&GPI: learn successor feature representation, learn policies, Then using the SF and policies, get the most rewarding policy given a new task.

John Towns, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, et al. Xsede: accelerating scientific discovery. *Computing in Science & Engineering*, 16(5):62–74, 2014.

Elizabeth Tricomi, Bernard W. Balleine, and John P. O'Doherty. A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience*, 29(11):2225–2232, 2009a. ISSN 0953-816X. doi: 10.1111/j.1460-9568.2009.06796.x.

Elizabeth Tricomi, Bernard W Balleine, and John P. O'Doherty. A specific role for posterior dorsolateral striatum in human habit learning. *European Journal of Neuroscience*, 29(11):2225–2232, 2009b.

Mike Tucker and Rob Ellis. On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human perception and performance*, 24(3):830, 1998.

Endel Tulving. Episodic memory: From mind to brain. *Annual Review of Psychology*, 53(1):1–25, 2002.

N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee. N4itk: Improved n3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6):1310–1320, 2010. ISSN 0278-0062. doi: 10.1109/TMI.2010.2046908.

Amos Tversky. Intransitivity of preferences. *Psychological Review*, 76(1):31, 1969.

Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5:297–323, 1992.

Avinash R Vaidya, Marcus Sefranek, and Lesley K Fellows. Ventromedial frontal lobe damage alters how specific attributes are weighed in subjective valuation. *Cerebral Cortex*, pages 1–11, 2017.

Kenneth F. Valyear, Jody C. Culham, Nadder Sharif, David Westwood, and Melvyn A. Goodale. A double dissociation between sensitivity to changes in object identity and object orientation in the ventral and dorsal visual streams: A human fMRI study. *Neuropsychologia*, 44(2):218–228, 2006. ISSN 0028-3932. doi: 10.1016/j.neuropsychologia.2005.05.004.

David C. Van Essen and John H. R. Maunsell. Hierarchical organization and functional streams in the visual cortex. *Trends in Neurosciences*, 6:370–375, 1983.

Jorien Van Paasschen, Elisa Zamboni, Francesca Bacci, and David Melcher. Consistent emotions elicited by low-level visual features in abstract art. *Art & Perception*, 2(1-2):99–118, 2014.

Ashish Vaswani. Attention is all you need. *arXiv preprint arXiv:1706.03762*, 2017.

Tom Verguts and Wim Notebaert. Hebbian learning of cognitive control: dealing with specific and nonspecific adaptation. *Psychological review*, 115(2):518, 2008.

Edward A. Vessel and Nava Rubin. Beauty and the beholder: Highly individual taste for abstract, but not real-world images. *Journal of Vision*, 10(2):18–18, 2010.

Edward A. Vessel, Natalia Maurer, Alexander H. Denker, and G Gabrielle Starr. Stronger shared taste for natural aesthetic domains than for artifacts of human culture. *Cognition*, 179:121–131, 2018.

Alexander Sasha Vezhnevets, Simon Osindero, Tom Schaul, Nicolas Heess, Max Jaderberg, David Silver, and Koray Kavukcuoglu. Feudal networks for hierarchical reinforcement learning. In *International Conference on Machine Learning*, pages 3540–3549. Proceedings of Machine Learning Research, 2017.

Dirk Walther and Christof Koch. Modeling attention to salient proto-objects. *Neural Networks*, 19(9):1395–1407, 2006.

Jun Wang, Jie Tu, Bing Cao, Li Mu, Xiangwei Yang, Mi Cong, Aruna S. Ramkrishnan, Rosa H.M. Chan, Liping Wang, and Ying Li. Astrocytic l-Lactate Signaling Facilitates Amygdala-Anterior Cingulate Cortex Synchrony and Decision Making in Rats. *Cell Reports*, 21(9):2407–2418, 2017. ISSN 2211-1247. doi: 10.1016/j.celrep.2017.11.012.

Liang Wang, Ryan E. B. Mruczek, Michael J. Arcaro, and Sabine Kastner. Probabilistic maps of visual topography in human cortex. *Cerebral Cortex*, 25(10): 3911–3931, 2014.

Liyuan Wang, Xingxing Zhang, Hang Su, and Jun Zhu. A comprehensive survey of continual learning: theory, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

Elke U. Weber and Eric J. Johnson. Constructing preferences from memory. *The Construction of Preference, Lichtenstein, S. & Slovic, P.,(eds.)*, pages 397–410, 2006.

Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. In *Conference on Computer Vision and Pattern Recognition*, 2016.

James C. R. Whittington, Timothy H. Muller, Shirley Mark, Guifen Chen, Caswell Barry, Neil Burgess, and Timothy E. J. Behrens. The tolman-eichenbaum machine: Unifying space and relational memory through generalization in the hippocampal formation. *Cell*, 183(5):1249–1263, 2020.

G. Elliott Wimmer and Daphna Shohamy. Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science*, 338(6104):270–273, 2012.

Daniel M. Wolpert, Jörn Diedrichsen, and J. Randall Flanagan. Principles of sensorimotor learning. *Nature Reviews Neuroscience*, 12(12):739–751, 2011. ISSN 1471-003X. doi: 10.1038/nrn3112.

Klaus Wunderlich, Antonio Rangel, and John P. O'Doherty. Neural computations underlying action-based decision making in the human brain. *Proceedings of the National Academy of Sciences*, 106(40):17199 – 17204, 2009. doi: 10.1073/pnas.0901077106.

Daniel L. K. Yamins and James J. DiCarlo. Using goal-driven deep learning models to understand sensory cortex. *Nature Neuroscience*, 19(3):356–365, 2016.

Daniel L. K. Yamins, Ha Hong, Charles F. Cadieu, Ethan A. Solomon, Darren Seibert, and James J. DiCarlo. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *Proceedings of the National Academy of Sciences*, 111(23):8619–8624, 2014.

Nick Yeung, Matthew M. Botvinick, and Jonathan D. Cohen. The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review*, 111(4):931, 2004a.

Nick Yeung, Matthew M. Botvinick, and Jonathan D. Cohen. The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychological Review*, 111(4):931, 2004b.

Henry H. Yin and Barbara J. Knowlton. The role of the basal ganglia in habit formation. *Nature Reviews Neuroscience*, 7(6):464–476, 2006.

Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks? *Advances in Neural Information Processing Systems*, 27, 2014.

Semir Zeki. Inner vision: An exploration of art and the brain. *Journal of Aesthetics and Art Criticism*, 60(4):365–366, 2002.

Carey Y. Zhang, Tyson Aflalo, Boris Revechkis, Emily R. Rosario, Debra Ouellette, Nader Pouratian, and Richard A. Andersen. Partially mixed selectivity in human posterior parietal association cortex. *Neuron*, 95(3):697–708, 2017.

Y. Zhang, M. Brady, and S. Smith. Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57, 2001. ISSN 0278-0062. doi: 10.1109/42.906424.

Zuo Zhang, Peter Zeidman, Natalie Nelissen, Nicola Filippini, Jörn Diedrichsen, Stefania Bracci, Karl Friston, and Elisabeth Rounis. Neural correlates of hand–object congruency effects during action planning. *Journal of Cognitive Neuroscience*, 33(8):1487–1503, 2021.

Qiang Zhu, Mei-Chen Yeh, Kwang-Ting Cheng, and Shai Avidan. Fast human detection using a cascade of histograms of oriented gradients. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1491–1498. IEEE, 2006.

Chengxu Zhuang, Siming Yan, Aran Nayebi, Martin Schrimpf, Michael C. Frank, James J. DiCarlo, and Daniel L. K. Yamins. Unsupervised neural network models of the ventral visual stream. *Proceedings of the National Academy of Sciences*, 118(3):e2014196118, 2021a.

Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning, 2020. URL `https://arxiv.org/abs/1911.02685`.

Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning.

Proceedings of the IEEE, 109(1):43–76, 2021b. ISSN 0018-9219. doi: 10.1109/jproc.2020.3004555.

Stefan Zysset, Cornelia S. Wendt, Kirsten G. Volz, Jane Neumann, Oswald Huber, and D. Yves von Cramon. The neural implementation of multi-attribute decision making: A parametric fMRI study with human subjects. *NeuroImage*, 31(3): 1380–1388, 2006. ISSN 1053-8119. doi: 10.1016/j.neuroimage.2006.01.017.