The Neural Mechanisms of Value Construction

Thesis by
Logan M. Cross

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

Caltech

CALIFORNIA INSTITUTE OF TECHNOLOGY
Pasadena, California

2022
Defended March 30, 2022

© 2022

Logan M. Cross
ORCID: 0000-0002-5248-9499

All rights reserved

ACKNOWLEDGEMENTS

I am very grateful for the wonderful experience I have had in graduate school at Caltech, and I would not have made it through this process without the help of numerous tremendous people.

I would first like to thank my advisor John O'Doherty, who has been a fantastic mentor. I am very appreciative of the opportunities he gave me to pursue very ambitious projects and all of the advice and support he has given throughout these years. The O'Doherty lab is also a tremendous environment that will honestly be hard to leave.

I also want to thank my wonderful collaborators, Shinsuke Suzuki, Yisong Yue, Jeff Cockburn, and Ryan Webb for all of their crucial contributions to my projects. I have also had many instructive interactions with my committee members Ralph Adolphs and Dean Mobbs, and I am very grateful they have agreed to be on my committee.

I also am incredibly thankful for all the friends I've made in the O'Doherty lab and the guidance I have received from them at every step of the process. The work presented in this thesis was far from easy, and I was able to accomplish it due to an overwhelming amount of help and support from numerous O'Doherty lab members over the years including: Shinsuke Suzuki, Jeff Cockburn, Tomas Aquino, Vince Man, Caroline Charpentier, Kyo Iigaya, Wolgang Pauli, Eva Pool, Aniek Fransen, Sandy Tanwisuth, Jaron Colas, Sarah Oh, Weilun Ding, Omar Perez, Sanghyun Yi, Ryo Adachi, Tessa Rusch, Lisa Kluen.

Through the years, I have met many other friends, colleagues, and mentors at Caltech who have positively influenced my research and life in many ways, including Whitney Griggs, William Poole, Joe Marino, Giovanni Gentile, Bowen Fung, Shin Shimojo, Colin Camerer, Thanos Siapas, Pietro Perona, Mike Tyszka, and many more.

Maybe most importantly, I have been extremely lucky to be surrounded by a bunch of amazing people in my personal life. My parents have inspired me my entire life and pushed me to become the best man I could become, and they have been more than supportive about me pursuing my passion in neuroscience. For the past few years, Rina Gutierrez has been my rock, picking me up when I am most down and frustrated, while also caring for me in endless small and big ways with a level of compassion that inspires me greatly. Many other people have been incredibly influential to my growth as a person and as a scientist, including but not limited to: Lindsey Cross, Austin Gregersen, Kathryn Hack, Colin McGuire, Luke Petrarca, JT Dodd, Xavier Hernandez, Jake Orthwein, and many, many more.

I also want to thank Dalton Combs and Glenn Fox, who were incredible mentors to me at USC, and motivated me to go to graduate school in the first place. They helped lay the groundwork for my research career, and I will always be grateful for their guidance when I was an undergrad.

ABSTRACT

Research in decision neuroscience has characterized how the brain makes decisions by assessing the expected utility of each option in an abstract value space that affords the ability to compare dissimilar options. Experiments at multiple levels of analysis in multiple species have localized the ventromedial prefrontal cortex (vmPFC) and nearby orbitofrontal cortex (OFC) as the main nexus where this abstract value space is represented. However, much less is known about how this value code is constructed by the brain in the first place. By using a combination of behavioral modeling and cutting edge tools to analyze functional magnetic resonance imaging (fMRI) data, the work of this thesis proposes that the brain decomposes stimuli into their constituent attributes and integrates across them to construct value. These stimulus features embody appetitive or aversive properties that are either learned from experience or evaluated online by comparing them to previously experienced stimuli with similar features. Stimulus features are processed by cortical areas specialized for the perception of a particular stimulus type and then integrated into a value signal in vmPFC/OFC.

The project presented in Chapter 2 examines how food items are evaluated by their constituent attributes, namely their nutrient makeup. A linear attribute integration model succinctly captures how subjective values can be computed from a weighted combination of the constituent nutritive attributes of the food. Multivariate analysis methods revealed that these nutrient attributes are represented in the lateral OFC, while food value is encoded both in medial and lateral OFC. Connectivity between lateral and medial OFC allows this nutrient attribute information to be integrated into a value representation in medial OFC.

In Chapter 3, I show that this value construction process can operate over higher-level abstractions when the context requires bundles of items to be valued, rather than isolated items. When valuing bundles of items, the constituent items themselves become the features, and their values are integrated with a subadditive function to construct the value of the bundle. Multiple subregions of PFC including but not limited to vmPFC compute the value of a bundle with the same value code used to evaluate individual items, suggesting that these general value regions contextually adapt within this hierarchy. When valuing bundles and single items in interleaved trials, the value code rapidly switches between levels in this hierarchy by normalizing

to the distribution of values in the current context rather than representing all options on an absolute scale.

Although the attribute integration model of value construction characterizes human behavior on simple decision-making tasks, it is unclear how it can scale up to environments of real-world complexity. Taking inspiration from modern advances in artificial intelligence, and deep reinforcement learning in particular, in Chapter 4 I outline how connectionist models generalize the attribute integration model to naturalistic tasks by decomposing sensory input into a high dimensional set of nonlinear features that are encoded with hierarchical and distributed processing. Participants freely played Atari video games during fMRI scanning, and a deep reinforcement learning algorithm trained on the games was used as an end-to-end model for how humans evaluate actions in these high-dimensional tasks. The features represented in the intermediate layers of the artificial neural network were found to also be encoded in a distributed fashion throughout the cortex, specifically in the dorsal visual stream and posterior parietal cortex. These features emerge from nonlinear transformations of the sensory input that connect perception to action and reward. In contrast to the stimulus attributes used to evaluate the stimuli presented in the preceding chapters, these features become highly complex and inscrutable as they are driven by the statistical properties of high-dimensional data. However, they do not solely reflect a set of features that can be identified by applying common dimensionality reduction techniques to the input, as task-irrelevant sensory features are stripped away and task-relevant high-level features are magnified.

PUBLISHED CONTENT AND CONTRIBUTIONS

Cross, Logan, Jeff Cockburn, Yisong Yue, and John P. O'Doherty (2021). "Using deep reinforcement learning to reveal how the brain encodes abstract state-space representations in high-dimensional environments". In: *Neuron* 109.4, pp. 724–738. DOI: 10.1016/j.neuron.2020.11.021.
L.C. lead the project, developed experimental protocol, and collected data with J.C. L.C. analyzed the data and wrote the manuscript.

Suzuki, Shinsuke, Logan Cross, and John P. O'Doherty (2017). "Elucidating the underlying components of food valuation in the human orbitofrontal cortex". In: *Nature Neuroscience* 20.12, pp. 1780–1786. DOI: 10.1038/s41593-017-0008-x.
L.C. collected and analyzed data with S.S. and helped edit the manuscript. L.C. was particularly involved in the use of the cutting-edge multivariate pattern analysis methods of the fMRI data.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

Number *Page*

Chapter 1

INTRODUCTION

Decision Neuroscience

The human experience is fundamentally a series of moments, characterized by perception action loops, where we perpetually perceive what's happening in our immediate environment and take an action to causally affect that environment in the service of our goals. Taking the right actions in the right situations to achieve a diverse set of goals is the hallmark of human intelligence, yet exactly how this happens still remains a mystery. This process is mediated by the computations in the brain and broader nervous system, as the world is perceived through sensory input that projects to the brain and acted on by motor output that is sent by the brain. Due to the overwhelming complexity of this organ that houses 86 billion neurons, neuroscience has been studied with subdivisions that try to identify the neural mechanisms at different levels of analysis or different aspects of cognition. For example, visual neuroscience tries to understand how visual perception works in the brain. On the other side of the perception action loop lies decision neuroscience, a more recently developed subfield that focuses on how choices and actions are produced by the brain. For the past few decades, researchers have produced pioneering work that lays the foundation for the projects I outline in this dissertation. This work is heavily interdisciplinary, as it often involves mathematical modeling of the decision-making process. Below I will outline the progress made by two interdisciplinary subdivisions of decision neuroscience: neuroeconomics, which utilizes tools from economics to examine how value and choice emerges in the brain, and reinforcement learning, a framework originating from computer science and animal psychology that outlines how agents (biological or artificial) can select actions to maximize reward.

Neuroeconomics

As evidenced by its name, neuroeconomics draws heavily from the field of economics. In particular, economists have developed methods to characterize how people make choices. Neoclassical economists outlined normative frameworks for how rational actors should act to maximize utility (Glimcher, et al., 2009). These frameworks are built from simple axioms such as that individuals will make choices that are consistent and transitive. If a person chooses a chocolate bar over a banana in one situation, they should not choose the banana over a chocolate bar in another situation. Additionally, expected utility theory was developed (Morgenstern and Von Neumann, 1944) to model choices with uncertain outcomes with continuous utility functions. These axiomatic approaches form the basis of microeconomics and thus explain a wide variety of consumer behavior in aggregate and can be used to predict the consequences of policy decisions.

In the second half of the 20th century, a group of economists and psychologists moved past analyzing choice behavior from first principles and axioms, and took on a more descriptive approach to model how human behavior appeared in practice (Simon, 1957; Glimcher, et al., 2009). Naturally, this connected with elements of psychology and led to emergence of a new subfield, behavioral economics. As you might expect, humans are not always optimally rational agents, and our behavior is often biased in systematic ways that vary from neoclassical theories such as expected utility theory. Due to resource constraints and a neural architecture with quirks and limitations that are shaped by our evolutionary history (Simon, 1957; Gigerenzer and Selten, 2002), our behavior is not just noisy, but is predictably irrational (Ariely and Jones, 2008). One prominent contribution came from psychologists Daniel Kahneman and Amos Tversky (Kahneman and Tversky, 2013) called prospect theory. Prospect theory describes how individuals make choices involving risk and uncertainty, and how in practice their decisions diverge from the predictions of expected utility theory. For example, people weigh losses more than gains of equivalent monetary amounts, a phenomenon known as loss aversion. Additionally, people tend to overweight outcomes with low probabilities and underweight highly certain outcomes, a cognitive bias that has real-world implications for human behavior in areas like insurance, gambling, and risk assessment. The descriptive insights of prospect theory and other findings from behavioral economics have sparked considerable interest in uncovering the psychological and cognitive mechanisms of our decision-making systems. The field of neuroeconomics emerged from perspective, as reverse-engineering the mechanisms that produce choices is fundamentally a neuroscientific pursuit.

Lesion studies provided the first clue as to which neural regions are causally involved in choice. Patients with damage to their prefrontal cortex (PFC) consistently exhibit various decision-making deficits, stemming from an inability to evaluate the future consequences of their actions or to learn from feedback (Bechara et al., 1994). In the past few decades, further electrophysiological and neuroimaging research has helped illuminate the functions that PFC regions have in decision-making.

Researchers sought out to identify whether the economic concept of value is represented in the brain. Both the neoclassical and behavioral schools of economic thought are built on the concept of subjective value, where a decision between multiple options is made by comparing the values of the available options. Value is an abstract representation of the utility or reward one expects when choosing an option, and therefore value allows dissimilar stimuli to be compared in a common currency (ie. would you rather have a hamburger or a lottery ticket?). This abstraction is also divorced from its specific sensory or motor contingencies in the raw stimulus or experimental setup. The value of a banana, for example, should not vary depending on whether a subject chooses the banana with a button press or with an eye movement, whereas the motor signals involved in initiating the choice will diverge. In a series of electrophysiological experiments with monkeys, Padoa-Scioppa and Assad identified neurons in the orbitofrontal cortex (OFC; a subregion of PFC) that coded for economic value when choosing between juice rewards of various quantities (Padoa-Schioppa and Assad, 2006). OFC neurons coded for three different variables: the value of a particular juice (called offer value), the value of the chosen juice (chosen value), or a binary response when a particular juice was chosen (taste).

A multitude of functional imaging studies in humans have additionally identified neural correlates of subjective value in OFC and in the adjacent ventromedial prefrontal cortex (Chib et al., 2009; Hare, Camerer, Knoepfle, et al., 2010; Kable and Glimcher, 2007; De Martino et al., 2009; Plassmann, John P. O'Doherty, and Rangel, 2007; Tom et al., 2007). Several of these studies elicited subjective values of food items and consumer goods from subjects by using Becker-DeGroot-Marschack (BDM) auctions, which ask "how much would you be willing to pay for this item?" (Becker, DeGroot, and Marschak, 1964; McNamee, Rangel, and John P O'Doherty, 2013; Plassmann, John P. O'Doherty, and Rangel, 2007). At the end of an experiment, a price is randomly generated, and if the subjects reported a bid higher or equivalent to the randomly generated price, then they can buy the item at this random price. This procedure therefore incentivizes subjects to report their true willingness to pay price, and is thus a reliable way to obtain subjective values for items in an experiment and we use the BDM auction for the experiments presented in Chapters 2 and 3. Additionally, several fMRI studies indicate that vmPFC value signals reflect a common currency value code, with overlapping areas coding for value across a diverse set of stimuli, such as food, money, consumer goods (clothes, books, etc.), and social reward (Chib et al., 2009; Levy and Glimcher, 2011; Lin, Adolphs, and Rangel, 2012; McNamee, Rangel, and John P O'Doherty, 2013). However, distributed non-overlapping value codes for food items and consumer items were also found more ventrally in vmPFC with multivariate analyses (McNamee, Rangel, and John P O'Doherty, 2013).

These studies have been foundational for neuroeconomics by establishing a clear neural basis for value-based decision-making, but there are still many open questions. How are the value codes in OFC and vmPFC constructed? To compute a general value code that is abstracted from the specific features that constitute a stimulus, the neural system needs to first perceive the stimulus and infer its desirability from its features. This is one of the central questions of this thesis, and two projects designed to investigate this are presented in Chapters 2 and 3. Moreover, before making a decision, the brain has to construct a representation of the decision problem (Rangel, Camerer, and Montague, 2008). Internal and external states need to be perceived (for example: "I am hungry and I see food"), and the actions available to choose between need to be identified. This component of the decision-making process is relatively trivial in simple experimental settings where subjects choose between two options, but task representation becomes a much larger issue in real-world domains that are high-dimensional in the sensory space and the action space. Chapter 4 of this thesis addresses this question directly by using naturalistic decision-making tasks that are more representative of the high-dimensional decision-making scenarios the brain encounters in daily life.

Reinforcement Learning

Reinforcement learning (RL) is a theoretical framework that can be used to characterize any decision-making problem where an agent takes actions and receives reward or punishment based on these actions (Sutton and Barto, 2018). Due to this generality, RL can model how any type of agent can learn from feedback, regardless if the agent is an animal, a human, or an AI. An RL environment is typically modeled as a Markov Decision Process (MDP), where an agent traverses an environment by moving from state to state. In each state, an agent chooses an action which deterministically or stochastically puts the agent in another state, and reward is accumulated depending on the state an agent reaches. The objective is then for an agent to learn by trial and error which actions to take in each state to maximize reward.

The original inspiration to mathematically formalize trial and error learning comes from research on pavlovian and operant conditioning in animal psychology. Pavlovian conditioning occurs when an initially neutral stimulus (called a conditioned stimulus CS) is paired with a reward or punishment (unconditioned stimulus US), which was first observed when Ivan Pavlov rang a bell before delivering food to dogs. After training, the dogs would salivate to the bell even without the presence of food. This conditioned response is due to the predictive relationship the bell had to reward and thus illustrates a basic form of learning. A simple mathematical model explains this learning process (Rescorla, 1972). The value of a stimulus V(s) is equal to the amount of reward that stimulus predicts. Before learning, the stimulus has no value V(s), but when it is paired with reward, V(s) is compared with the reward received, with the difference between the two quantities called a prediction error (PE).

$$PE = r - V(s)$$

The prediction error is positive when more reward is received than expected from the previous V(s), and PE is then used to update V(s) with the following learning rule:

$$V(s) = V(s) + \alpha(PE).$$

Alpha (α) here refers to a learning rate, which is a parameter between 0 and 1 that controls the speed of learning by weighing the previous value estimate versus the current reward received. A large learning rate will put more weight on the current reward, and a small learning rate will put more weight on the previous estimate of V(s). After repeated trials where a neutral CS (ie. bell) predicts a rewarding US (ie. food), the value estimate of the CS gravitates towards the value of the reward, which makes the CS desirable itself and leads to an appetitive response (ie. salivation) when it is presented. This Rescorla Wagner model explains many important experimental effects in pavlovian conditioning, including the blocking effect where conditioning can be impaired when a CS is presented with a second CS that has already been trained to predict the US.

While Pavlovian conditioning only refers to reflexive behaviors that occur when an organism passively observes a stimulus associated with reward or punishment, extensions of this model can also account for the form of conditioning that is action-dependent: operant conditioning (also called instrumental conditioning). Operant conditioning occurs when a behavior itself is reinforced by reward (or weakened by punishment). For example, if a rat pressing a lever produces food pellets to be delivered, that behavior will be repeated in the future. Here, the stimulus-response (see lever -> press lever) becomes predictive of reward, and can be similarly learned as an action value estimate of a state-action pair Q(s,a). These action values can be similarly learned through error driven mechanisms mediated by prediction errors.

$$PE = r - Q(s', a')$$

$$Q(s, a) = Q(s, a) + \alpha(PE)$$

An agent can then select the action with the highest action value in order to maximize reward. Q-function models such as this explain animal and human choices well in operant conditioning paradigms (Daw and Tobler, 2014).

The error-driven learning framework has been expanded to more complex environments with multiple states and sequential state to state transitions. This generalization now encapsulates the majority of decision-making environments reinforcement learning researchers care about, ranging from a rat navigating a maze to an AI playing chess or backgammon. As outlined above, a state value V(s) (or action values Q(s,a)) can be computed at every state in the environment. The temporal difference algorithm (TD) allows you to incorporate future rewards into the state value function (Sutton and Barto, 2018), so the value of a state is not just the reward expected at that state but also the cumulative expected reward following that state.

$$V(s_t) = r(s_t) + \mathbb{E}[r(s_{t+1}) + \mathbb{E}[r(s_{t+2}) + ...|s_{t+1}]|s_t]$$

Future rewards can also be downweighted with a discount factor gamma in contexts where reward is less valuable when it is delayed.

$$V(s_t) = r(s_t) + \mathbb{E}[\gamma r(s_{t+1}) + \mathbb{E}[\gamma^2 r(s_{t+2}) + ...|s_{t+1}]|s_t]$$

The trick of temporal difference learning is to leverage the recursive structure of this equation to compactly represent all expected future rewards with the value of the next state V$_{t+1}$. Therefore the value at any state is equal to the reward received in that state plus the value of the successor state s$_{t+1}$.

$$V(s_t) = r(s_t) + \mathbb{E}[r(s_{t+1})|s_t]$$

Due to the recursion, the value of the s$_{t+1}$ encapsulates the sum of all future rewards at s$_{t+1}$, s$_{t+2}$. . . s$_{t+n}$. By bootstrapping from the value estimate of the next state, learning can efficiently propagate backwards from rewarding states to the states preceding them. Learning the value estimates are similarly mediated by prediction errors, but now prediction errors are equal to the difference between the V(s$_t$) and r$_t$ + V$_{t+1}$.

$$\delta_t = r(s_t) + V(s_{t+1}) - V(s_t)$$

By adding V$_{t+1}$ to the equation, TD learning has the ability to not only learn about the immediate reward but the reward accumulated in subsequent states as well. Thus, if an agent lands in a state that is better than expected, it can update its value of the preceding state similarly to if reward was unexpectedly received at that state. In chess for example, raw reward is not received until the game ends and the winner is declared, but if an agent estimates the value of every game state it encounters (which becomes a proxy for win probability), it can learn about the quality of decisions that lead to sharp changes in state value/win probability. To construct an action-selection policy from TD learning in this way, it is combined with Q-functions that estimate the value of state-action pairs as explained above.

$$Q(s_t, a_t) = r(s_t) + \mathbb{E}[r(s_{t+1})|s_t, a_t]$$

This algorithm, known as Q-learning, uses a similar update rule at TD learning, but replaces the value estimate of the successor state with the maximum q-value of the successor state.

$$\delta_t = r(s_t) + \max_{a_{t+1}} Q(s_{t+1}, a_{t+1}) - Q(s_t, a_t)$$

These models have been highly influential in neuroscience as well. In a series of landmark electrophysiological experiments in which monkeys undergo pavlovian conditioning, signaling of the dopamine neurons in the ventral tegmental area (VTA) and substantia nigra showed striking resemblance to reward prediction errors of a TD learning algorithm (W. Schultz, Dayan, and Montague, 1997). As in a typical conditioning protocol, an appetitive fruit juice was used as reward and paired with a preceding conditioned stimulus. Before learning, dopamine neurons fired directly after fruit juice was delivered. After learning however, the dopamine neurons began to fire at the onset of the conditioned stimulus and did not fire above baseline when reward was received. Additionally, if after learning the CS was presented but the reward failed to occur, then dopamine neurons fired at the presence of the CS and the firing rate went below baseline when the reward was absent. These dopaminergic firing patterns are neatly explained by TD models. When reward is initially unexpected, dopamine neurons only fire when a reward appears signaling a positive prediction error, and they do not fire at the presence of a stimulus that is not yet predictive of reward. Once the CS becomes associated with reward through learning, a positive prediction error occurs (via dopaminergic neurons increasing their firing rate) when the CS is presented and there is no prediction error at the time reward is received as expected. Moreover, if reward fails to be delivered, a negative prediction error is conveyed with dopaminergic firing below baseline. The dopaminergic neurons in the midbrain project to the striatum, the amygdala, and diffusely throughout the frontal cortex, which affords the opportunity for the prediction error signal to reinforce distributed motor pathways (Daw and Tobler, 2014). Voltammetry techniques have been used to confirm that dopamine is released in targets such as the ventral striatum in a manner consistent with the reward prediction error hypothesis (Day et al., 2007). Human fMRI experiments have additionally identified correlations between blood oxygen level dependent (BOLD) activation in the striatum and reward prediction errors in a wide variety of tasks (Abler et al., 2006; McClure, Berns, and Montague, 2003; John P. O'Doherty, Dayan, Friston, et al., 2003; Spicer et al., 2007; Tobler et al., 2007). Altogether, these results point to a tight correspondence between neural experimental results and computational theories of learning, thereby highlighting the roles of the dopamine system and the striatum in learning.

The reinforcement learning approaches discussed thus far do not explain the entirety of how humans learn however. Algorithms that implement trying and error learning like TD learning and Q-learning do remarkably well in modeling implicit learning tasks with a small set of stimuli, a type of learning that has been primarily investigated experimentally through bandit tasks, but they fall short of accounting for many cognitive capacities that make human intelligence powerful, such as reasoning and planning. This is because they make no attempt to model the environmental dynamics but only try to learn cached values about state-action pairs. They are therefore referred to as model-free algorithms. In contrast, model-based algorithms construct an internal model of the environment by learning about the structure of state to state transitions. This affords the ability to plan with this world model or to quickly adapt a policy when the reward structure of the environment changes (Daw and O'Doherty, 2014). A current hypothesis is that the brain contains both model-free and model-based systems and arbitrates between them (Lee, Shi- mojo, and O'Doherty, 2014). Much less is known about model-based learning in the brain however, other than the likely possibility that it is PFC dependent (Daw and O'Doherty, 2014).

The model-free algorithms discussed thus far encounter other challenges too when applied to tasks with a high-dimensional state space. The complexity of the learning process scales exponentially with the number of states to learn about, and therefore these algorithms fail to scale to environments with a large number of variables, such as ones involving vision. In the real-world, animals and humans use perception to identify what state they are in. In addition, we can encounter completely novel states (in fact the exact information that hits our retina is almost always unique) and instinctively know which actions to take by generalizing from our past experience of similar states. In classic RL however, a new state is defined any time a part of the visual space is novel, even if just one pixel is changed, and the action values at this state would have to be inefficiently learned by scratch without generalizing from previously learned states. This issue motivates the project presented in Chapter 4. Therefore, recent approaches in machine learning that solved the curse of dimensionality will be outlined in detail below.

Deep Reinforcement Learning

To combat the issues with classic RL outlined above, in 2015 Google DeepMind developed the deep Q-network (DQN) that could learn to play dozens of Atari video games at human to superhuman levels from scratch (Mnih et al., 2015). The algorithm takes in the pixels of the game as input and processes this input with a artificial neural network/deep neural network in order to approximate a Q-value function. The last layer of the network computes Q-values for all of the available actions in the game and takes the action with the highest Q-value as in classic Q-learning. DQN specifically uses a convolutional neural network (CNN) architecture to process the frames of the games. CNN architectures are very useful for processing images, and had previously been used with much success in object recognition (LeCun, Bottou, et al., 1998; Krizhevsky, Sutskever, and Hinton, 2012). CNNs are loosely inspired by how early visual cortex neurons only respond to visual input in a certain receptive field, as the artificial neurons are only locally connected to the neurons in tangential layers with similar receptive fields, in contrast to the full connectivity of multilayer perceptrons. The parameters in a CNN layer consist of a set of learnable filters that detect visual features in the input, which are then convolved across the input to produce a 2-dimensional activation map that gives the responses of that filter at every spatial position. As the receptive fields increase from early layers to later layers, the filters tend to encode visual features in a hierarchical fashion. Filters in early layers may encode an edge of a particular orientation, while filters in later layers represent complex textures or parts of objects such as eyes. DQN also introduced other improvements that improved the stability of approximating a Q-function with a nonlinear function. As previously described, the value of a state-action pair Q(s,a) represents the reward received by taking that action plus the max action value at the next state ($r + \gamma \max_{a'} Q(s', a')$), and the discrepancy between Q(s,a) and this TD target becomes a prediction error used to update Q(s,a) towards its optimal value.

$$Q(s,a) := Q(s,a) + \alpha \left(r + \gamma \max_{a'} Q(s',a') - Q(s,a) \right)$$

DQN turns this into a regression problem parametrized by the weights of the deep neural network, so the discrepancy between Q(s,a) and the TD target can be minimized with backpropagation and modern optimization tools.

$$\mathcal{L}(\theta) = \mathbb{E}_{s,a,r,s'} \left(r + \gamma \max_{a'} Q(s',a',\theta') - Q(s,a,\theta) \right)^2$$

Additionally, DQN uses a fixed target network with parameters θ' to improve the stability of the optimization process. The parameters of the target network are periodically updated to match the parameters of the online network. In order to optimize this function with stochastic gradient descent, DQN introduces one other method that takes biological inspiration from hippocampal replay mechanisms. For stochastic gradient descent to work well, the data needs to be independent and identically distributed (i.i.d.), however sequences of frames in an Atari video game or similar RL environment are highly correlated. An experience replay buffer was developed to tackle this issue, which involves storing the states of the game and uniformly sampling the stored samples during training to break this correlation.

This marriage of RL and deep learning was a breakthrough in AI that led to the emergence of the deep reinforcement learning field. Since DQN's development, numerous improvements and subfields have emerged, as deep RL has been used to beat human experts in Go and StarCraft (Silver et al., 2016; Vinyals et al., 2019), develop chatbots (Cuayáhuitl, Keizer, and Lemon, 2015), advance robotics (Tai et al., 2016), and much more (Li, 2017). A parallel line of deep RL algorithms do not use value functions, but seek to directly optimize a policy that maximizes the RL objective (Sutton and Barto, 2018). These methods use a general class of optimization that can be applied to many domains including environments with continuous action spaces. However, policy-based methods are sample inefficienct and suffer from high variance gradients. Many of the current state of the art algorithms combine these policy-based methods with the value-based methods to get the best of both worlds (Lillicrap et al., 2015; Schulman et al., 2017; Haarnoja et al., 2018; Espeholt et al., 2018), in a class of models known as actor-critics (policy: actor, critic: value function).

Neuroimaging

The projects presented in this thesis utilize functional magnetic resonance imaging (fMRI) methods to record human brain activity during the decision-making process. fMRI non-invasively records brain activity via blood oxygen-level dependent (BOLD) signals, which are caused by the compensatory blood flow that follows neural activity (Ogawa et al., 1992). BOLD signals increase the brightness of the volumetric pixels (voxels) in an MRI image, which allows cognitive neuroscientists to record from the entire brain at once in the scanner, with spatial resolution much better than other neuroimaging methods such as electroencephalography (EEG) and magnetoencephalography (MEG). In contrast to these methods, fMRI has a more limited temporal resolution due to the slow dynamics of the BOLD response (BOLD peaks around 5s after neural activity).

Traditionally, fMRI data has been analyzed with the general linear model (GLM), which implements mass univariate regressions to model every voxel independently as a function of the experimental variables (Penny et al., 2011). More recently, cutting-edge techniques have been developed to better link brain activity to the computational mechanisms that produce behavior. In model-based fMRI analysis, computational models are developed to capture the cognitive process humans use to behave on a task, and then components of the model are regressed against the BOLD signal (John P. O'Doherty, Hampton, and Kim, 2007). For example, this method has identified neural correlates of the prediction errors from a temporal difference learning model in the ventral striatum and OFC (John P. O'Doherty, Dayan, Friston, et al., 2003). Additionally, another fMRI analysis approach involves using methods from machine learning and data science to examine how information is distributed across voxels, in contrast to modeling every voxel independently. These multivariate pattern analysis (MVPA) tools are heavily utilized in each of the projects presented in this thesis. Therefore, I will outline these tools in more detail.

Multivariate Pattern Analysis (MVPA)

MVPA has revolutionized fMRI research by using decoding methods to examine what information is encoded in a region. A common approach is to use classification techniques from machine learning to distinguish patterns of voxel responses associated with different classes of stimuli. In 2001, Haxby and colleagues used a classifier to dissociate the brain's representation of faces, animals, and other objects by using voxel patterns in ventral temporal cortex as the features fed to the classifier (Haxby, Gobbini, et al., 2001). The idea is that viewing stimuli in these different categories does not necessarily lead to different regions being activated for each category, but that the ventral temporal cortex encodes information about all the categories with a population code distributed across hundreds or thousands of overlapping voxels. Each object category was found to have a distinct signature response profile across the voxel population, which could be picked up by a nearest neighbor classifier that identify pattern similarity within and between classes in the high-dimensional voxel space. Similar decoding analyses have been used to distinguish information held in working memory (Harrison and Tong, 2009), decode abstract cognitive states (Mitchell et al., 2004), and probe the encoding of variables related to value and choice (Kahnt et al., 2010; McNamee, Rangel, and John P

O'Doherty, 2013; McNamee, Liljeholm, et al., 2015).

The MVPA procedure first requires partitioning the data into independent training and test data sets. Voxels from a particular region or from even the entire brain are used as features (independent variables or predictors, in regression). The stimulus categories are used as labels to classify (dependent variables, in regression). A classifier, such as a Support Vector Machine or a logistic regression classifier, are trained to predict the labels from the features in the training set and tested on the labels in the test set. Performance can then be quantified with any accuracy metric such as classification accuracy, precision, recall, etc.. To make use of all the data available, cross-validation is most often used, where the data is partitioned into a number of folds, with each fold taking turns being included in the training and testing sets. Since training and testing sets should be independent, the fMRI run is most commonly used as the partition with leave-one-run-out cross-validation.

Representational Similarity Analysis (RSA)

Decoding approaches come with some disadvantages. Sophisticated classifiers can pick up on any type of information encoded in neural patterns that is related to the categories of experimental interest. Therefore, classifier performance may be high in regions that represent confounding variables that are correlated to the categories of interest, and thus MVPA can provide an incomplete or even misleading picture about what variables are represented in a region. Representational similarity analysis methods offer a more data-driven approach by empirically examining the geometry of neural patterns (Kriegeskorte, Mur, and Bandettini, 2008). This is done with representational dissimilarity matrices (also called RDMs or DSMs), which convert neural responses in each trial or condition to a response vector and compare response vectors between trials or conditions in this multidimensional vector space with a distance metric (like Euclidean or correlation distance). These pairwise comparisons then afford the researcher the ability to probe how similar or dissimilar neural responses are in a region between conditions. The neural DSMs can then be compared to model DSMs that represent how similar stimulus features or other variables are across trials, such as stimulus category, stimulus value, or variables of no interest like reaction time or fMRI run. This allows the empirical neural DSMs to be compared to multiple model DSMs without the bias of classifying one feature at a time.

In a seminal paper, monkeys and humans viewed images of real-world objects and responses in the inferior temporal cortex (IT) were measured (Kriegeskorte, Mur, Ruff, et al., 2008). With RSA, it was found that IT clusters categories of images similarly across species, with the largest dissimilarity between animate and inanimate objects. Within the animate category, faces are bodies were represented in another cluster with higher similarity within category than between category. RSA also allows comparison of the representational geometry across different data types and models with varying numbers of dimensions. As previously stated, comparisons can be made across species, imaging modalities (such as EEG, MEG, fMRI), and between biological neural networks and artificial neural networks as we show in Chapter 4.

Chapter 2

ELUCIDATING THE UNDERLYING COMPONENTS OF FOOD VALUATION IN THE HUMAN ORBITOFRONTAL CORTEX

Abstract

The valuation of food is a fundamental component of our decision-making. Yet little is known about how value signals for food and other rewards are constructed by the brain. Using a food-based decision task in human participants, we found that subjective values can be predicted from beliefs about constituent nutritive attributes of food: protein, fat, carbohydrates, and vitamin content. Multivariate analyses of functional MRI data demonstrated that, while food value is represented in patterns of neural activity in both medial and lateral parts of the orbitofrontal cortex (OFC), only the lateral OFC represents the elemental nutritive attributes. Effective connectivity analyses further indicate that information about the nutritive attributes represented in the lateral OFC is integrated within the medial OFC to compute an overall value. These findings provide a mechanistic account for the construction of food value from its constituent nutrients.

Introduction

There is accumulating evidence, from an array of studies using diverse methods in multiple species, of a key role for the OFC and adjacent medial prefrontal cortex (PFC) in representing the expected value or utility of options at the time of decision-making (Clithero and Rangel, 2014; Padoa-Schioppa and Assad, 2006; Rich and Wallis, 2016; Rudebeck and Murray, 2014; Grabenhorst and Rolls, 2011). It has been suggested that such value signals can serve as inputs into the decision process, thereby enabling individuals to choose actions yielding outcomes that maximize expected gains (Clithero and Rangel, 2014; Padoa-Schioppa and Assad, 2006). Value signals have been found in this region in response to cues or actions associated with many different types of potential outcomes, including food rewards, monetary rewards, consumer goods, and even more abstract goals such as pursuing imaginary leisure activities (Clithero and Rangel, 2014; McNamee, Rangel, and John P O'Doherty, 2013; Chikazoe et al., 2014; Howard, Gottfried, et al., 2015; Lebreton et al., 2009; Small et al., 2003; Kable and Glimcher, 2007; Stalnaker et al., 2014; Gross et al., 2014; Chib et al., 2009; Levy and Glimcher, 2011; Suzuki, Harasawa, et al., 2012; Suzuki, Adachi, et al., 2015). However, while value signals in OFC have been well characterized, much less is known about how it is that value signals are constructed in the first place.

In the present study, we focus on valuation for food rewards. The valuation of food is a fundamental component of the decision-making process that all humans complete on a daily basis. A dysfunctional food valuation process may result in the development of obesity and eating disorders (Foerde et al., 2015; Carnell et al., 2012). Recent human neuroimaging studies have begun to elaborate functional contributions of OFC in food value computations. Medial OFC encodes value signals independent of the identity of food rewards (Howard, Gottfried, et al., 2015), irrespective of whether the value information is acquired through direct experience or through imagining the consequences of a new experience (Barron, Raymond J. Dolan, and Behrens, 2013). On the other hand, lateral OFC encodes value in an identity-specific manner (Howard, Gottfried, et al., 2015; Klein-Flügge et al., 2013). However, the constituent attributes that underlie the construction of food value and how these constituent attributes are represented and integrated in the OFC remain elusive.

We hypothesized that the value of a food reward is at least in part computed by taking into account beliefs about the properties of the constituent nutritive attributes of a food item. We focused on beliefs about the amount of protein, carbohydrates, and fat, and we also included beliefs about the specifically sweet carbohydrates (sugar), sodium, and vitamin content contained in a food item. We further hypothesized that the OFC would play a role in representing these elemental attributes, which could thereby constitute precursor representations used to generate an integrated value signal.

In the human brain, value signals for food rewards have been reported throughout the orbital surface, most prominently in the medial OFC (Clithero and Rangel, 2014; Grabenhorst and Rolls, 2011; McNamee, Rangel, and John P O'Doherty, 2013; Chib et al., 2009; Levy and Glimcher, 2011). However, sensory inputs from the visual, auditory, gustatory, olfactory and somatosensory systems arrive into the OFC primarily in the lateral portions of the orbital surface (Öngür and Price, 2000). Thus, we hypothesized that more lateral parts of the OFC would be especially involved in encoding elemental attributes about a food outcome, in contrast to the medial OFC, which we hypothesized would be especially involved in encoding an overall subjective goal-value signal for the foods, as found in many previous reports (Clithero and Rangel, 2014; McNamee, Rangel, and John P O'Doherty, 2013; Gross et al., 2014).

Results

Experimental task and behavior

To test these hypotheses, we scanned 23 human participants using functional MRI (fMRI) while they reported their 'willingness to pay' (WTP; i.e., subjective value) for 56 food items (WTP task; Fig. 2.1a) (McNamee, Rangel, and John P O'Doherty, 2013). After the MRI scan, the participants provided subjective ratings about the constituent nutrient attributes for the same set of items. Specifically, we asked participants to rate the quantities of fat, sodium, carbohydrates, sugar, protein, and vitamins contained in the foods, as well as to provide an estimate of the overall caloric content (Tang, Fellows, and Dagher, 2014) (attribute-rating task; Fig. 2.1b). In this task, subjective ratings about the nutrient factors were found to be significantly correlated with the objective factors ($P<0.01$ for all factors; Fig. 2.1c). Moreover, while performing the WTP task in the scanner, the participants were not aware that they would be subsequently required to rate the nutrient attributes of the items, and thus they were not biased by experimenter-demand effects to artificially reflect on information about nutrient attributes during the food valuation phase.

We first conducted behavioral analyses to test our hypothesis that participants' ratings of the elemental nutritive attributes of a food would predict the subjective valuation of the food items. As some nutritive attribute ratings were tightly coupled with others (Supplementary Fig. 1), including all the attributes in the predictive model did not necessarily provide the best prediction of value. To specify which combinations of subjective nutrient factors provided the best prediction about subjective value, we performed a series of linear regression analyses (Methods). In the regression analyses, performance of the prediction was assessed by leave-one-item-out cross-validation. Comparing every possible combination of the six nutrient factors (i.e., $2^6 = 64$ models), we found that subjective value was best predicted by a model including the following four subjective nutrient factors: fat, carbohydrates, protein, and vitamin (Supplementary Table 1). Consistent with this result, among the best 10 models, protein, and vitamin appeared in all 10 models; fat and carbohydrate appeared in 8 and 6 models, respectively; and sodium and sugar were present only in 5 and 4 models, respectively (Supplementary Table 1).

Here we note that sugar content did not make a significant contribution to the food valuation, despite previous findings showing a role for sugar content in food intake behaviors (Zuker, 2015; De Araujo et al., 2008; Tellez et al., 2016). Given that sugar is a subcomponent of carbohydrates and that subjective ratings about the two factors were indeed highly correlated (Supplementary Fig. 2.1b), a reasonable interpretation of this result is that the effects of sugar content are subsumed under the more general carbohydrate category. This interpretation was further supported by an additional analysis demonstrating that including sugar instead of carbohydrate to the regression model significantly reduced the accuracy of the model for predicting subjective value ($P<0.05$).

The prediction performance of the best fitting model was better than chance-level (at $P<0.01$; Fig. 2.1d). Even when implementing Bonferroni corrections for every possible combination of variables we ran (n=64), the prediction performance of the best fitting model was nevertheless still significant at $P<0.01$. In addition to testing for the role of subjective beliefs about the nutritive content of the foods, we also extracted objective information about the nutritive content of the foods and used that information in a regression analysis similar to that performed using the subjective ratings. We found that the best fitting model with subjective nutrient factors outperformed the best fitting model with objective factors (Fig. 2.1d). Furthermore, the regression model that included subjective beliefs about all four nutritive factors also performed better than subjective or objective estimates of overall caloric content (Fig. 2.1d).

We further validated these results by using logistic regression analyses with categorical binary predicted variables (constructed by splitting subjective value into low and high categories based on a median split). That is, the model providing the best prediction in the original linear regression analyses outperformed the other models also in the logistic regression analyses (Fig. 2.1e and Supplementary Table 1). Collectively, these behavioral analyses support the notion that food value is computed through integrating information about the subjective beliefs about the nutrient factors of fat, carbohydrates, protein, and vitamin content.

Representation of subjective value in the OFC

Having established that the subjective value of food items can be predicted in part from subjective beliefs about nutritive content, we set out to replicate previous findings of a role for OFC in encoding the subjective value of the food items, using multivoxel pattern analyses (MVPA) (Haynes, 2015) with leave-one-run-out cross-validation. In this analysis, a linear classifier was trained on patterns of fMRI response to categorize food items as being either high or low in subjective value based on each participant's ratings (Methods).

Consistent with our hypothesis, value representations could be decoded from medial parts of the OFC at the time of valuation. In addition, subjective value codes were also found in parts of lateral OFC, consistent with other previous reports (Chikazoe et al., 2014; Howard, Gottfried, et al., 2015). Specifically, subjective value could be decoded above chance from patterns of fMRI response within anatomically defined medial as well as lateral OFC regions of interest (ROIs; $P<0.01$ for both ROIs, t test and permutation test; Fig. 2.2a; and see Supplementary Fig. 2a for information about the ROIs). Value information could also be decoded both at the time of bidding and at the time of feedback (Supplementary Fig. 2b). A searchlight analysis (Haynes, 2015) also identified significant codes of subjective value in both medial and lateral OFC (P<0.05, family-wise error rate small-volume corrected (FWE SVC); Fig. 2.2b). Furthermore, we found that for each classifier, the classification weights of the voxels were broadly distributed across the range of negative to positive values (Supplementary Fig. 2c,d), suggesting that the subjective value codes in the OFC are multivariate in nature.

Representation of nutrient factors in the OFC

We then tested whether, while evaluating a food item for decision-making (i.e., during the WTP task), the OFC represented information about the four subjective nutrient factors identified as predictors of the value. To this end, we applied the same MVPA procedure used for value coding (see above) to each of the four subjective nutrient factor ratings. Consistent with our initial hypothesis, information about the subjective nutrient factors could be significantly decoded at the time of valuation in the lateral OFC ROI (P<0.05, conjunction test against the conjunction null (Nichols et al., 2005) based on t and permutation tests; Fig. 2.3a; see Methods for detailed information about the conjunction test; classification scores are plotted as functions of subjective nutrient factors in Supplementary Fig. 3) but not in the medial OFC ROI (P>0.05, conjunction test; Fig. 2.3b). On the other hand, at the time of bidding or feedback, we found no significant decoding of the subjective nutrient factors either in lateral or medial OFC (P>0.05, conjunction test; Supplementary Fig. 4a,b), suggesting that the lateral OFC represents information about the nutrient factors only at the timing of valuation. Searchlight analyses confirmed encoding of information for each of the subjective nutrient factors at the time of valuation in various loci within the lateral OFC (Fig. 2.3c), with clusters encoding fat, protein, and carbohydrate content all significant at P<0.05 under voxel-level multiple-comparison correction within the anatomically defined lateral OFC ROI (i.e., FWE SVC), while the cluster encoding vitamin content bordered on significance (P=0.080 FWE SVC; Fig. 2.3c). Moreover, the distributions of the classification weights across the voxels were not highly biased toward negative or positive values (Supplementary Fig. 4c,d), consistent with the notion that the representations of subjective nutrient factors are multivariate. In sum, these results suggest that, during food valuation, information about subjective nutrient factors is encoded in the lateral OFC but not in the medial OFC.

We also examined whether linear classifiers can decode the subjective nutrient factors of novel food items. Given that, in our experiment, half of the items were presented in run 1 and 3 while the other half were presented in runs 2 and 4, we trained the classifiers on the data from runs 1 and 3 and tested them on runs 2 and 4 (and vice versa; accuracy scores were averaged; c.f. the leave-one-run-out cross-validation used in the above main analyses). The analysis revealed that, in the lateral OFC, decoding accuracies were significantly greater than chance for fat, carbohydrates and vitamins (P<0.05; Supplementary Fig. 4e), while the accuracy for protein was at the trend level (P<0.10; Supplementary Fig. 4e).

Here we note that the differential decoding performance between the lateral and medial OFC cannot be attributed to any differences in the number of voxels contained in the ROIs (although the lateral and the medial OFC ROIs contained 2,325 and 533 voxels, respectively). To demonstrate this, we randomly resampled the same number of adjacent voxels from the lateral OFC as found in the medial OFC ROI (i.e., forming a continuous cluster consisting of 553 voxels), and we then tested whether information about the subjective nutrient factors could still be decoded within this reduced ROI (Supplementary Fig. 4f). The analysis demonstrated that, even with the smaller number of voxels, we could still decode information about the subjective nutrient factors from patterns of fMRI activity in the lateral OFC (P<0.05, conjunction test; Supplementary Fig. 4f).

Furthermore, we examined whether distinct patterns of voxel activity in the lateral OFC represent each of the four subjective nutrient factors. Since the classification weights of each voxel were correlated across some combinations of the nutrient factors (Supplementary Fig. 4g), the alternative possibility would be that the same patterns of voxel activity represent two or more nutrient factors. To exclude the alternative possibility, we performed a cross-decoding MVPA procedure across the nutrient factors. In this analysis, for each pair of the four nutrient factors, we trained a classifier on one factor and tested it on the other factor (and the reverse; decoding accuracy was assessed by the average across both directions). Here we reasoned that if the subjective nutrient factors were coded in different patterns, the cross-decoding analysis would not provide any significant results. The analysis indeed revealed that the cross-category decoding accuracy was not significantly different from chance (P>0.05; Supplementary Fig. 4h), except for only one pair: fat and vitamins. In the fat–vitamins pair, decoding accuracy was significantly below chance (P<0.01; Supplementary Fig. 4h). There are at least two possible interpretations for the negative accuracy. One is that, in the lateral OFC, a highly similar pattern of fMRI response codes for fat and vitamins in the opposite directions of a multivariate decision boundary. The other possibility is that distinct patterns code for the two factors, but these are not dissociable in our dataset, given the highly negative correlation between subjective fat and vitamins in the behavioral ratings (r=–0.44±0.15, mean±s.d. across participants; Supplementary Fig. 1b) and the neural classifiers' weights (r=–0.41±0.18, mean±s.d.; Supplementary Fig. 4g). To tease apart these two possibilities, we conducted the following additional analysis: (i) 42 food items were randomly resampled from the original set of 56 items to ensure that the fat and vitamins were less correlated (mean r > –0.3); (ii) MVPA was then performed on the resampled data; and (iii) the above procedure was repeated ten times (accuracies were averaged). On the resampled data, we found that, consistent with results from the original dataset (Fig. 2.3a), a classifier trained on fat (or vitamins) could decode information about fat (or vitamins; P<0.05; Supplementary Fig. 4i). On the other hand, in the cross-decoding (i.e., a classifier was trained on fat and tested on vitamins, and vice versa), the accuracy was not significantly different from chance (P>0.05; Supplementary Fig. 4i). These findings together suggest that, in the lateral OFC, different patterns of voxel activity represent information about different subjective nutrient factors.

While we have so far focused on the four subjective nutrient factors identified as value predictors in our behavioral analyses, we also nevertheless tested for evidence of representations of the other factors we included in our experiment (but which were found to not be significantly associated with value): subjective sodium and sugar content. Information about subjective sugar content could be significantly decoded in the lateral OFC (P<0.01; Supplementary Fig. 4j) but not in the medial OFC (P=0.100; Supplementary Fig. 4k). On the other hand, neither the lateral nor the medial OFC showed significant decoding of sodium content (P=0.474 and P=0.557, respectively; Supplementary Fig. 4j,k).

We also investigated the extent to which objective (as opposed to subjective) nutrient content could be decoded from the OFC by training the MVPA classifiers on labels extracted from the objective nutrient content as opposed to the subjective content. This analysis identified a weaker overall effect of objective nutrient factors in the lateral and medial OFC (i.e., no significant conjunction effect at P>0.05; Supplementary Fig. 4l), although a subset of the individual objective factors could be significantly decoded in the lateral OFC. These results suggest that subjective nutrient factors are more robustly represented in the OFC than objective factors.

Representation of the relative content of the subjective nutrient factors in the OFC

To further explore how nutritive information is represented in the OFC, we implemented a representational similarity analysis (RSA) (Kriegeskorte and Kievit, 2013) to examine the extent to which the pattern of subjective ratings of the nutritive factors was related to encoding of these factors in the orbitofrontal cortex. In the RSA, we compared the voxel-wise similarity structure obtained from the fMRI data with the similarity structure of the subjective nutritive components for each item (Supplementary Fig. 5 and Methods). In this analysis, the voxel-wise similarity is defined as the correlation across voxel activity for each pair of items (Supplementary Fig. 5a), while the nutritive similarity is defined as the correlation in bundles of the four subjective nutrient factors (fat, carbohydrates, protein, and vitamins) for each item pair (Supplementary Fig. 5b). In other words, because correlation distance is employed to measure similarity, the nutritive similarity between two items is defined in terms of the relative content of the nutrient factors. The RSA revealed that the similarity of fMRI responses significantly reflected the similarity of the relative content of the nutrient factors in the lateral OFC ROI ($P<0.01$; Supplementary Fig. 2.5c) but not in the medial OFC. We also conducted a searchlight RSA and found a significant association between the voxel-wise fMRI and the subjective nutritive similarity across diffuse regions of the lateral OFC ($P<0.05$ FWE SVC; Supplementary Fig. 5d). These results suggest that there is a representation in lateral OFC of the relative content of each nutritive attribute.

Representation of low-level visual features in the OFC

Here we aimed to rule out the possibility that the lateral OFC contains information about low-level visual features such as luminance and contrast, which could potentially be detected by the classifier if there were inadvertent correlations between such low-level sensory features and value and/or subjective nutrient factors. For this purpose, we extracted eight low-level visual features (luminance, contrast, red intensity, green intensity, blue intensity, hue, saturation and brightness) from the food images presented to participants, and then examined whether the visual features could be decoded in the OFC. Moreover, as a positive control, we also tested for the primary visual cortex (V1, Brodmann area 17). These analyses showed that neither the lateral nor the medial OFC contains significant information about low-level visual features (P>0.05 for all features; Fig. 2.4a), while, as would be expected, primary visual cortex did indeed contain significant information about low-level visual features (P<0.05, conjunction test; Fig. 2.4b). These results suggest that the lateral OFC encodes information about value and subjective nutrient factors but not low-level visual information about the food images.

Effective connectivity between OFC subregions at the time of valuation

By leveraging MVPA on fMRI data, we were able to demonstrate that during food valuation, lateral OFC contains information about the elemental nutritive attributes of food. However, to compute an overall subjective value, the individual nutritive representations need to be integrated. We hypothesized that this integration of the individual nutritive representations would occur in regions found to encode subjective value in either the medial or lateral subregions of OFC. To test which of the value-encoding OFC subregions is primarily involved in the integration process, we performed an effective connectivity analysis: a psychophysiological interaction. The connectivity analysis is based on the reasoning that if a region is implicated in the integration process, the region must (i) contain information about the overall subjective value and (ii) have enhanced effective connectivity at the time of valuation with regions encoding each of the constitutive nutritive attributes of a food. The psychophysiological interaction analysis tested whether the value-related OFC subregions, identified in the searchlight MVPA (Fig. 2.2b), had increased task-related connectivity at the time of valuation with the lateral OFC subregions encoding each of the four subjective nutrient attributes (Fig. 2.3c). We found evidence for a significant increase in effective connectivity at the time of valuation between the value-related medial OFC subregion and the lateral OFC subregions representing the nutrient attributes (P<0.05, conjunction test; Fig. 2.5a). This result was further validated by a nonparametric bootstrap test (Efron and Tibshirani, 1994) (P<0.05; Methods), which is known to be relatively robust against potential outliers. On the other hand, we found no significant increase in the effective connectivity at the time of bidding or feedback (P>0.05, conjunction test; Supplementary Fig. 6a,b), although a subset of the individual attributes did show a connectivity effect. Also, we did not find robust evidence for a significant integration of nutrient attribute signals in the value-related lateral OFC region (P>0.05, conjunction test; Fig. 2.5b). These results indicate that the medial OFC satisfies both of the above two criteria for a brain region implicated in information integration, consistent with the notion that representations about elemental nutritive attributes of food in the lateral OFC are primarily integrated at the time of valuation in the medial OFC to compute subjective values.

Representation of value and nutrient factors in other brain regions

Previous studies demonstrated that value or reward signals are ubiquitously encoded not only in the OFC but also in other cortical regions, as well as in the amygdala (Vickery, Chun, and Lee, 2011; Kahnt, Park, et al., 2014; Gottfried, O'Doherty, and Dolan, 2003). As post hoc investigations beyond our original hypotheses, we tested for encoding of value and subjective nutrient factors in the following six ROIs: dorsomedial PFC (including anterior cingulate cortex), dorsolateral PFC, ventrolateral PFC, posterior parietal cortex (PPC), insula and amygdala (see Supplementary Fig. 7a for information about the ROIs). Consistent with previous findings, all these ROIs were found to significantly encode information about subjective value ($P<0.05$ for all; Supplementary Fig. 7b). On the other hand, information about the four subjective nutrient factors identified as value predictors could be significantly decoded only in the PPC ($P<0.05$, conjunction test; Supplementary Fig. 7c), while ventrolateral PFC and dorsolateral PFC represented only one and two factors, respectively (Supplementary Fig. 7c). However, when we applied a correction for multiple comparisons across these post hoc ROIs, the PPC ceased to be significant ($P>0.05$, conjunction test with Bonferroni correction). We also implemented a whole-brain searchlight analysis, which revealed that V1 also contained information about the four subjective nutrient factors ($P<0.05$ cluster-level FWE correction with the cluster-forming threshold $P=0.001$, conjunction test; Supplementary Fig. 8). These results are consistent with the possibility that not only lateral OFC but also V1 and potentially PPC represent information about the nutrient factors.

To further characterize the functional roles of the three regions, we performed the following additional analyses. First, we assessed whether those regions contain information about low-level visual features of the food images. The analyses revealed that information about low-level visual features could be decoded from both PPC and V1 ($P<0.01$; Supplementary Fig. 7d), consistent with the previous findings that PPC and V1 are major parts of a visual pathway (Mishkin, Ungerleider, and Macko, 1983) (i.e., the 'dorsal stream'). However, we note that this is not the case in the lateral OFC, where basic visual features were not represented (Supplementary Fig. 7d; see also Fig. 2.4a). Second, we compared the decoding accuracies of the low-level visual features with those of the subjective nutrient factors. Accuracy for the visual features was found to be significantly higher only in V1 (P<0.01; Supplementary Fig. 7d), while we found the opposite pattern in PPC and the lateral OFC (P<0.05; Supplementary Fig. 7d). These results together may suggest a functional gradient from V1 to PPC and lateral OFC: V1 predominantly represents the visual information, lateral OFC predominantly represent the nutritive information, and PPC is the intermediate locus. However, because the PPC result did not survive correction for multiple comparisons across ROIs, this result should be treated with caution until it can be independently replicated.

Discussion

This study elucidates the constituent nutritive attributes underlying valuation of food rewards. Behaviorally, we demonstrated that the subjective value of a food was best predicted by beliefs about the content of fat, carbohydrates, protein, and vitamins. This result suggests that food value is computed at least in part through integrating information about elemental nutritive attributes.

We then uncovered how information about the constituent attributes is represented and integrated in the brain. MVPA of fMRI data revealed that, while both lateral and medial parts of the OFC represented value signals, only lateral OFC represented information about the subjective nutrient factors. Furthermore, we found evidence for effective connectivity between the value-related medial OFC subregion and the lateral OFC subregions representing each of the individual nutrient attributes. Recent human neuroimaging studies have demonstrated that medial OFC and adjacent regions of medial PFC encode value information independently of the category of goods as a ‘common currency’ (McNamee, Rangel, and John P O’Doherty, 2013; Howard, Gottfried, et al., 2015), while lateral OFC encodes value in an identity-specific manner (Howard, Gottfried, et al., 2015; Klein-Flügge et al., 2013), and that identity-specific value representations are modulated by selective devaluation (Howard and Kahnt, 2017). Our findings go beyond these previous studies, in that we elucidate which constituent attributes underlie the construction of food value and how the constituent attributes are represented in the OFC.

There has been substantial debate in the literature about the distinct roles of the lateral and medial OFC in value-based decision-making (Noonan et al., 2010). Based on cytoarchitectonic structures and patterns of connectivity, neuroanatomical studies have identified a broad distinction between the medial part of the OFC, including adjacent vmPFC and the lateral part of the OFC (Öngür and Price, 2000). It has also been suggested that lateral OFC is involved in the initial assignment or representation of value (Padoa-Schioppa and Assad, 2006; Noonan et al., 2010) while medial OFC is more involved in a value comparison necessary for decision-making (Noonan et al., 2010). The present study, together with these previous findings, could lead to the conjecture that information about the elemental nutritive attributes of food is first represented in the lateral OFC and then subsequently integrated in the medial OFC to guide behavior. Our finding that the initial integration of food attributes needed to compute subjective value occurs in the medial OFC, alongside a lack of evidence that such integration occurs in the lateral OFC, raises the question of how the subjective value signal located in the lateral OFC is generated. One possibility is that this signal is a secondary representation elicited via reciprocal inputs from value signals in the medial OFC. However, further work will be necessary to investigate the nature of the local circuits within OFC in more detail in order to test this possibility.

In our experiment, participants were asked to report subjective values of food items (WTP task), and then rate the quantities of six nutrient factors contained in the same foods, as well as to provide an estimate of the overall caloric content (attribute-rating task). Due to the experimental design, one might argue that ratings about nutrient factors could be biased, in that the participants justified their subjective value ratings a posteriori. We believe this was not the case in our experiment. It is unlikely that participants were able to solve the complex multidimensional inverse problem: that is, to remember and artificially manipulate their ratings about the six nutrient factors to ensure consistency with the prior ratings of subjective value. Furthermore, an easier way to ensure the consistency would be to manipulate ratings about overall caloric content, but we found that the subjective caloric content was a poor predictor of the subjective value. Taken together, we conclude that the participants were unlikely to manipulate their ratings about nutrient factors to justify the subjective value ratings post hoc.

It is important to note that, while we have shown here that the value of a food reward can in part be predicted from beliefs about its subjective nutrient qualities, the overall value of a stimulus such as food is unlikely to depend exclusively on beliefs about nutritive composition. Instead, an individual's history of past experience with that food, including the amount of past exposure to the food and the past pairing of that food with other positive and negative experiences, are also likely to play critical roles in determining overall value. Moreover, we have left open the possibility that the overall value of a food is driven by some nonlinear combinations of the constitutive nutritive attributes corresponding to hidden superordinate properties of the food. It is also worth noting that the overall value of a food can be affected by cultural factors (Rozin and Vollmecke, 1986). While we recruited participants from the general population in the greater Los Angeles area (California, USA), people living in other regions such as Asia and Africa might potentially have different preferences for food. Yet while such overall preferences might vary based on culture, and this might lead to differences in the weightings given to different nutritive attributes in computing subjective value across cultures, there is no reason to expect that the fundamental aspects of the neuronal organization of the computation of food value from its elemental attribute representations would differ across cultures. This notwithstanding, a fruitful research agenda will involve quantifying all of the additional elemental and cultural factors that influence valuation, determining the neural representation of those variables and establishing how those various signals get integrated in order to compute an overall value.

To conclude, in this study, we provide substantial insights into how a value signal for a food reward can be constructed from its constituent nutritive attributes in the brain. Given that dysfunctional food-valuation processes may play a large role in the development of obesity and anorexia (Foerde et al., 2015; Carnell et al., 2012), our findings have implications for understanding neural and psychological mechanisms underlying eating disorders, which is an important step toward the goal of developing novel treatments for such disorders.

Methods

Participants

We recruited 24 healthy participants from the general population as part of the recruitment pool for the NIMH Caltech Conte center for social decision-making. Data from one participant were excluded due to technical problems with the fMRI scan. We therefore used the data from the remaining 23 participants (8 females; age 30.7±4.12 years, mean±s.d.; and BMI 23.51±4.00, mean±s.d.). All the participants were preassessed to exclude those with any previous history of neurological/psychiatric illness. We also confirmed that the participants were not on a diet or seeking to lose weight for any reason. They gave their informed written consent and received monetary and food rewards depending on their performance in the WTP task (see below) in addition to the participation fee of $50. No statistical methods were used to predetermine the sample size, but our sample size was motivated by those used in previous studies (McNamee, Rangel, and John P O'Doherty, 2013; Chib et al., 2009). The study protocol was approved by the Institutional Review Board of the California Institute of Technology.

Stimuli

In our experiment, we used 56 food items (for example, snacks, fruits, salads, etc.; some were selected from the previous study (Hare, Malmaud, and Rangel, 2011); Supplementary Table 2). These items were highly familiar and available at local stores. Indeed, during the attribute-rating task (see below), on average only 1.43±2.84 (mean±s.d.) food items of the 56 were rated as 'not familiar at all'. Information about the objective nutrient factors of the items was obtained from the package label or from an online calorie counter.

All the items were presented to participants as high-resolution color images. Information about the low-level visual features of the images (luminance, contrast, red intensity, green intensity, blue intensity, hue, saturation and brightness) was extracted using the Image Processing Toolbox included with Matlab. For each of the images, red, green and blue intensities in each pixel were extracted using the toolbox, and then luminance was computed as the weighted sum of the intensities ($0.2126 \times red + 0.7152 \times green + 0.0722 \times blue$). We also computed hue, saturation, and brightness in each pixel using Matlab's rgb2hsv function. For each whole image, each low-level visual feature was defined as the averaged values across all of the pixels. Finally, the (local) contrast of each image was defined as the s.d. of the pixel luminance values (Chikazoe et al., 2014).

Experimental Tasks

Participants performed the WTP task inside the MRI scanner, and then subsequently performed the attribute-rating task outside the scanner. To enhance participants' motivation for the foods, we asked them to refrain from eating or drinking any liquids, besides water, for 3hr before the experiment. Compliance was confirmed by self-reports, and the participants' hunger rating was on average 4.13 0.92 (mean ± s.d.; scaled from 1, 'not at all hungry,' to 6, 'very hungry'). Furthermore, participants were asked to stay at the laboratory for 30 min after the experiment, during which time the only thing they were able to eat was the food obtained in the experiment.

WTP task (inside the MRI scanner)

Following the procedure used in previous studies from our laboratory (McNamee, Rangel, and John P O'Doherty, 2013; Chib et al., 2009), we employed a modified version of the BDM auction task (Becker, DeGroot, and Marschak, 1964) to measure participants' willingness to pay (i.e., subjective value) for food items (Fig. 2.1a). In each trial of this task, a participant was endowed with \$3 and made a bid (\$0, \$1, \$2 or \$3) for one of the 56 items. At the end of the experiment, the computer randomly selected one of the trials to be implemented. For the selected trial, a random counter-bid was drawn from \$0, \$1, \$2, \$3 with equal probability. If the participant's bid was equal to or greater than the counter-bid, he or she paid the counter-bid and received the food item. Otherwise, the participant kept the initial endowment \$3 and received no food. The auction mechanism is incentive-compatible in the sense that the optimal strategy for the participants is to always bid the number closest to their true willingness to pay for obtaining that item (Becker, DeGroot, and Marschak, 1964). Participants were explicitly instructed in the optimal strategy, and using a questionnaire, we confirmed that they correctly understood the experimental mechanism. Furthermore, to control for effects of retail price, we instructed the participants that the amount of each food item was determined so that the retail price is around \$4.

This task consisted of four fMRI runs of the 56 trials. In each of the runs 1 and 3, a randomly selected 28 of the 56 items were presented twice in random order (i.e., 56 trials per run). The other 28 items were presented twice in each of the runs 2 and 4. In total, participants made a bid four times for each food item. We refer to the averaged amount of bid (i.e., willingness to pay) over the four trials as the 'subjective value' of the item.

At the beginning of each trial, a participant was shown one food item (valuation phase, 3s; Fig. 2.1a). In the next phase, the participant made a bid for that item by pressing the key on a numeric keypad that corresponded to the bid dollar amount (bid phase, within 2.5s). Here to dissociate the bid amount from the spatial information, mappings between keys and bid amounts were randomized across trials. The bid the participant made was immediately presented (feedback phase, 0.5s), followed by a jittered intertrial interval (ITI phase, 2–12s). During this task, participants failed to make a response only in 1.55 ± 2.94% of trials (mean ± s.d.), and the missed trials were modeled as a nuisance regressor in the fMRI analysis (see below).

Attribute-rating task (outside the MRI scanner)

Participants rated the subjective nutrient factors of the 56 food items (Fig. 2.1b). Notably, the instructions for the attribute-rating task was given after completing the WTP task, and thus the participants were not aware during the WTP task that they would be subsequently required to rate nutrient factors of the items. This helped us exclude an explanation for the behavioral and fMRI results in terms of somehow biasing or artificially inducing participants to focus on such food attributes at the time of valuation.

This task consisted of eight sessions. In each session, participants were asked to answer one of the following eight questions for the 56 items about the six nutrient factors as well as the overall calorie content and the familiarity:

(1) how high is the item in fat?

(2) how high is the item in carbohydrates?

(3) how high is the item in protein?

(4) how high is the item in vitamins?

(5) how high is the item in sugar?

(6) how high is the item in sodium (salt)?

(7) how high is the item in calories?

(8) how familiar is the item?

The order of the eight questions was randomized across participants. Notably, in this task the participants were asked to rate the ‘density’ of the nutrient factors in each food item. In the instruction sheets, we explicitly told participants, “please indicate your guess about the density of the nutrient, that is, the amount of the nutrient contained per unit of weight (for example, 10 oz. of the item).”

On each trial, the participant answered a question for one item on a continuous scale from ‘not at all’ to ‘very much’ by moving a red pointer, with no time constraint (Fig. 2.1b). The initial position of the pointer was randomized on each trial, and the pointer moved toward the right (or left) by pressing the key [1] (or [2]) on a numeric keypad. The answer was finally registered by pressing the key [3].

Behavioral analyses

Regression analysis with subjective nutrient factors

To examine which combination of the six subjective nutrient factors provided the best prediction of the subjective value, we ran the following linear regression analysis. For each participant, we regressed the value of each item against the subjective nutrient factors. Comparing all the possible $2^6 = 64$ models including none, some, or all of the six factors, we found that a combination of the four factors, fat, carbohydrates, protein, and vitamins provided the best prediction performance (Supplementary Table 1).

Here the prediction performance of each model (i.e., combination of the nutrient factors) was assessed by a leave-one-item-out cross-validation. That is, for each participant and each model, (i) we ran the regression analysis, leaving out one of the 56 items; (ii) computed the predicted value of the left-out item on the basis of the obtained regression coefficients; (iii) repeated the above procedure for each of the 56 items; and (iv) computed the correlation between the predicted and the actual values. The overall performance of the model was obtained by averaging the correlation across participants.

As a robustness check, we also conducted a logistic regression with the categorical predicted values (low and high subjective values split for each participant by the median value). The procedure was the same as for the linear regression, except that we used ‘accuracy’ as a measure of performance instead of the correlation between the predicted and actual values. The analysis demonstrated that, consistent with the linear regression results, the combination of the four factors fat, carbohydrates, protein, and vitamins provided the best prediction (Supplementary Table 1).

Regression analysis with objective nutrient factors

We ran the same linear and logistic regression analyses, using the objective nutrient factors as explanatory variables (Fig. 2.1d,e).

Regression analysis with the overall calorie content

We also ran the same analyses using subjective or objective estimates of total calorie content (Fig. 2.1d,e).

fMRI data acquisition

We collected the fMRI images using a 3 T Siemens (Erlangen) Trio scanner located at the Caltech Brain Imaging Center (Pasadena, CA) with a 32-channel radio frequency coil. The BOLD signal was measured using a one-shot T2*-weighted echo planar imaging sequence (Volume TR = 2,780 ms, TE = 30 ms, FA = 80°). We acquired 44 oblique slices (thickness = 3.0 mm, gap = 0 mm, FOV = 192 × 192 mm, matrix = 64 × 64) per volume. The slices were aligned 30° to the AC–PC plane to reduce signal dropout in the orbitofrontal area Deichmann et al., 2003. After the four functional runs, high-resolution (1 mm3) anatomical images were acquired using a standard MPRAGE pulse sequence (TR = 1,500 ms, TE = 2.63 ms, FA = 10°). The fMRI data were analyzed using SPM8 in Matlab R2013b on a MacBook Pro (Retina, 15-inch, mid-2015; Mac OS X 10.11.6). Data collection and analysis were not performed blind to the conditions of the experiments.

fMRI data preprocessing

fMRI images for each participant were preprocessed using the standard procedure in SPM8: after slice-timing correction, the images were realigned to the first volume to correct for participants' motion, spatially normalized and temporally filtered (using a high-pass filter width of 128s). Spatial smoothing with an 8-mm FWHM Gaussian kernel was applied to the fMRI images only for psychophysiological interaction analysis (see below) but not for MVPA or representational similarity analysis (RSA). For searchlight MVPA and RSA, smoothing was applied to the accuracy and the correlation maps, respectively, but not to the fMRI images (see below).

Multivoxel pattern analysis (MVPA)

To examine whether information about subjective value can be decoded from patterns of fMRI response, we conducted a classification analysis, multivoxel pattern analysis (see below). Also, the same procedure was applied to the classification analyses for nutrient factors and low-level visual features.

Classification samples

We extracted voxel-wise fMRI responses to each food item as classification samples. For each participant and each run, we designed a general linear model (GLM). The GLM contained 28 regressors indicating the valuation phases (duration = 3s) of the 28 different food items, as well as four regressors indicating the bid phases (duration = reaction time), feedback phases (duration = 0.5s), timing of the key press (duration = 0s) and missed trials (valuation phase, duration = 3s). All the regressors were convolved with a canonical hemodynamic response function. In addition, six motion-correction parameters and the linear trend were included as regressors of no interest to account for motion-related artifacts. For each voxel, the parameter estimates of the first 28 regressors corresponded to the fMRI responses to each of the 28 food items in each run. The fMRI responses to each food item were then entered into the classification analysis as classification samples.

Classification algorithm

We employed a linear support vector machine with a cost parameter C = 1 as a classifier. We performed the classification analysis using The Decoding Toolbox (TDT) (Hebart, Görgen, and Haynes, 2015). Classification accuracy was estimated using a leave-one-run-out cross-validation: for each of the four runs, a classifier was trained on the other three runs and tested on the remaining focal run; and the procedure was repeated for the four runs (accuracy scores were averaged).

More specifically, to avoid label imbalance bias in each run (see the “Classification label” section, above), we performed a bootstrap sampling procedure repeated 1,000 times (Hebart, Görgen, and Haynes, 2015). That is, we randomly removed some samples (without replacement) to ensure that the number of samples in each label was equalized for each run; the above classification analysis was then performed on the balanced data; and the procedure was repeated 1,000 times resulting in an average classification accuracy.

ROI analysis

We anatomically defined regions of interest (ROIs), lateral OFC, medial OFC and other areas based on the AAL database (Tzourio-Mazoyer et al., 2002). See Supplementary Figures 2a and 7a for details. fMRI responses in each of the ROIs were entered into the above classification analysis. We then examined, for each ROI, whether the mean accuracy across participants was greater than 50% (chance, given the binary label) using one-sampled t tests (one-tailed). A two-tailed test was employed only for the cross-decoding analysis (see main text) to examine whether the mean accuracy was greater or less than 50%. We also employed a permutation test (permuting the classification labels within each participant 1,000 times; one-tailed) to check whether the mean accuracy was significantly greater than chance. See Allefeld, Görgon, & Haynes for advanced issues pertaining to population-level inferences in MVPA studies.

Searchlight analysis

We also conducted a searchlight decoding analysis (Kriegeskorte, Goebel, and Bandettini, 2006) with a radius of 3 voxels (i.e., 9 mm), as in our previous studies (McNamee, Rangel, and John P O'Doherty, 2013; McNamee, Liljeholm, et al., 2015), within the entire OFC ROI (i.e., summation of the lateral and medial OFC). In this analysis, each participant's accuracy map was spatially smoothed with an 8-mm FWHM Gaussian kernel and entered into the second-level analysis performed by SPM8. The statistical significance was assessed by t test vs. 50% with a voxel-level FWE small-volume correction within the lateral and medial anatomical OFC ROIs. For the whole-brain analysis, we employed a cluster-level FWE correction for multiple comparisons (cluster-forming threshold, $P = 0.001$).

Conjunction test

In the conjunction test (Nichols et al., 2005), if all of the individual factors are significantly decoded ($P < 0.05$), we reject the null hypothesis that at least one of the factors was not represented; such result thus supports the alternative hypothesis that all of the factors were represented. In this study, we mainly employed conjunction analyses using t tests, while for some key results we also performed conjunction tests based on a permutation test (Fig. 2.3a).

Additional analysis (regressing out the effects of value)

In this analysis, we regressed out the effect of value from both the ratings about nutrient factors (i.e., classification labels) and the fMRI responses to each food item (i.e., classification samples), and then tested whether each of the nutrient factors could be still decoded. For the ratings data, in each participant we regressed values of food items against the ratings about each of the nutrient factors and took the residuals. We also regressed values against the fMRI responses to food items and then obtained the residuals (note: this procedure was performed for each participant and each run).

Representational similarity analysis (RSA)

To further examine the manner in which subjective nutritive information is represented in the OFC, we performed a representational similarity analysis (RSA) (Kriegeskorte and Kievit, 2013; Kriegeskorte, Mur, and Bandettini, 2008).

Voxel-wise representational dissimilar matrix (RDM)

As in the case of MVPA (see the "Classification samples" section, above), we extracted voxel-wise fMRI responses to each food item for each participant and each run. Averaging the fMRI responses over the runs, we estimated each voxel's response to each item for each participant (Supplementary Fig. 5a). We then created an RDM based on the correlation distance (i.e., 1 – Pearson's correlation coefficient across voxels) for each pair of the 56 items (Supplementary Fig. 5a).

Behavioral RDM

A behavioral RDM was created based on the correlation distance for each item pair in bundles of the four subjective nutrient factors (fat, carbohydrates, protein, and vitamins; and for each nutrient factor, rating values were z-normalized across the items). Note that the correlation distance in bundles reflects the dissimilarity between two items in terms of the relative contents of the four nutrient factors (Supplementary Fig. 5b).

Comparison of voxel-wise and behavioral RDMs

We computed the Spearman's rank correlation between upper triangular portions of the voxel-wise and the behavioral RDMs. The Fisher z-transformed correlation coefficient for each participant was then entered into the population-level inference.

ROI analysis

For the lateral and medial OFC ROIs (Supplementary Fig. 2a), we performed the above analysis and then examined whether the mean correlation coefficient was greater than 0 using one-sampled t tests (one-tailed).

Searchlight analysis

We also conducted a searchlight analysis (Kriegeskorte, Goebel, and Bandettini, 2006) with a radius of 3 voxels (i.e., 9 mm), as in the MVPA, within the entire OFC ROI (i.e., summation of the lateral and medial OFC ROIs). In this analysis, each participant's correlation map was spatially smoothed with an 8-mm FWHM Gaussian kernel and entered into the second-level random-effect analysis performed by SPM8. The statistical significance was assessed by performing a t test vs. 0 with a voxel-level FWE small-volume correction within the lateral and medial anatomical OFC ROIs.

Psychophysiological interaction (PPI) analysis

Following the standard procedure in SPM8, we performed a PPI analysis on the spatially smoothed fMRI images, as follows.

Extraction of BOLD signals

We first constructed a GLM for the extraction of BOLD signals. The GLM contained regressors indicating the valuation phase (duration = 3s), bid phase (duration = reaction time), feedback phase (duration = 0.5s), timing of the key press (duration = 0s), missed trials (valuation phase, duration = 3s), six motion-correction parameters and the linear trend, as well as parametric modulators of the valuation phase regressor depicting the subjective value and the four subjective nutrient factors (z-normalized across items). Based on the GLM, we extracted BOLD signals (eigenvariates adjusted for the valuation phase) from the lateral and medial OFC ROIs identified as encoding value information by the searchlight MVPA (Fig. 2.2b; spheres with a radius of 3 voxels centered at the respective peak voxels).

PPI model specification and estimation

We then constructed another GLM for the PPI analysis including the following regressors: (i) a physiological factor, the BOLD signal from lOFC; (ii) a physiological factor, the BOLD signal from mOFC; (iii) a psychological factor, the boxcar regressor indicating the valuation phase (duration = 3s; we call this regressor VAL); (iv) a psychophysiological interaction (PPI) factor, an interaction of the deconvolved lOFC BOLD signal and the psychological factor (VAL); and (v) a PPI factor, an interaction of the deconvolved mOFC BOLD signal and the psychological factor (VAL):

$$\beta_1\mathrm{lOFC} + \beta_2\mathrm{mOFC} + \beta_3\mathrm{VAL} + \beta_4\mathrm{OFCxVAL} + \beta_5\mathrm{mOFCxVAL} + \chi\beta + \epsilon$$

where Y denotes a BOLD signal in the target ROI, χ is a set of the other regressors (see below), β values indicate regression coefficients, and ϵ represents the residual. Note that in the PPI analysis, the mOFC BOLD signal, the lOFC BOLD signal and the corresponding PPI factors were included in the same GLM. To control for nuisance effects, we included four regressors indicating bid phases (duration = reaction time), feedback phases (duration = 0.5s), timing of the key press (duration = 0s) and missed trials (valuation phase, duration = 3s), as well as parametric modulators of the valuation-phase regressor representing the subjective value and the four subjective nutrient factors of the presented food item. All of the regressors except for the physiological factors were convolved with a canonical HRF. In addition, six motion-correction parameters were included as regressors of no interest to account for motion-related artifacts. For each participant, regression coefficients of the PPI factors were estimated at the lateral OFC ROIs identified in the searchlight MVPA as representing each of the four subjective nutrient factors (Fig. 2.3c; spheres with a radius of 3 voxels centered at the respective peak voxels).

Statistical test of the PPI effect

We then examined for each of the four ROIs to determine whether the mean regression coefficient across participants was greater than 0 using one-sampled t tests (one-tailed). To further support the examination, we also employed a bootstrap test (Efron and Tibshirani, 1994), which is known to be relatively robust against potential outliers. In the bootstrap test, we obtained 100,000 bootstrap datasets of the same size as the original sample size by resampling from the original data with replacement; then obtained the distribution of their mean values; finally we tested whether the 5% quintile of the distribution was greater than 0.

Overview of the statistical tests used in the present study

Parametric tests were used with the assumption of normality (the normality of the data was not formally tested). This approach is typical in the analysis approaches used for neuroimaging (McNamee, Rangel, and John P O'Doherty, 2013; McNamee, Liljeholm, et al., 2015; Penny et al., 2011). It is worth noting that, for some key results, we also conducted permutation tests and bootstrap tests, which do not require normality assumptions about the data. We employed one-tailed tests unless otherwise noted, as the tests examined whether the decoding accuracy is greater than chance. A two-tailed test was employed for the cross-decoding analysis (see the main text) to examine whether the mean accuracy was greater or less than 50%. In searchlight analyses, the statistical significance was assessed with a voxel-level FWE small-volume correction for the ROI analyses and a cluster-level FWE correction (cluster-forming threshold, P = 0.001) for the whole-brain analysis. A Life Sciences Reporting Summary is available.

Data and code availability

The data and code that support the findings of this study are available from the corresponding author upon reasonable request. The MRI data will also be posted to the NDARS data repository at `https://ndar.nih.gov/edit_collection.html?id=2417afterpublication.`

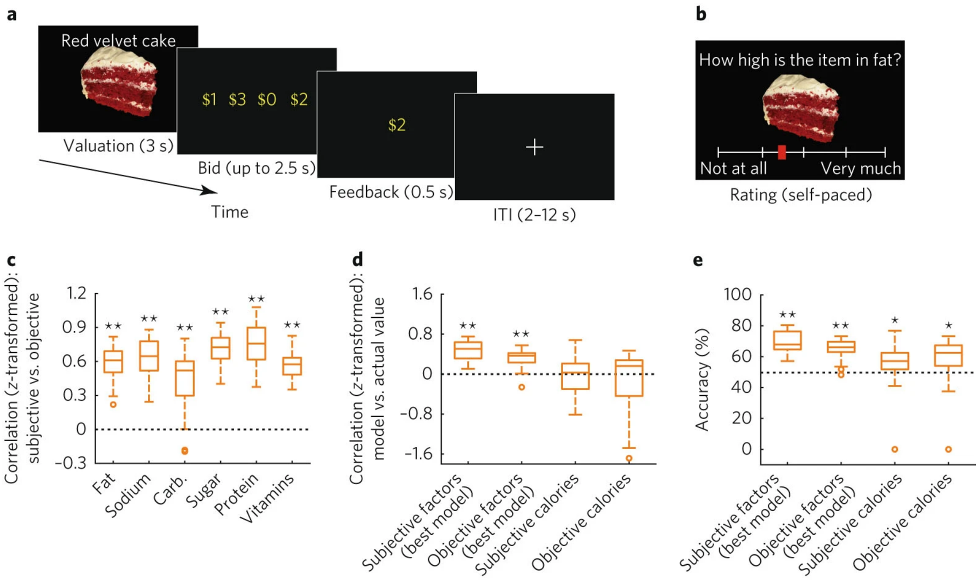


Figure 2.1: **Experimental task and behavior.**

a. Timeline of one trial in the WTP task. On each trial, participants reported their willingness to pay (i.e., subjective value) for one food item. Note that in the bid phase, mappings between keys and dollar amounts were randomized across trials.

b. Timeline of one trial in the attribute-rating task. On each trial, participants answered one question (for example, ‘How high is the item in fat?’) for one item on a continuous scale from ‘not at all’ to ‘very much’ by moving a red pointer, with no time constraint.

c. Correlations between the subjective and the objective nutrient factors (n = 23 participants). In each box and whisker plot, the central line denotes the median, and the bottom and top edges of the box indicate the 25th and 75th percentiles (q_{25} and q_{75}, respectively). The ends of the whiskers represent the maximum and minimum data points not considered outliers. Data points are considered outliers (open circles) if they are greater than $q_{75} + 1.5 \times (q_{75} - q_{25})$ or less than $q_{25} - 1.5 \times (q_{75} - q_{25})$. **P < 0.01, t test (fat: $t_{22} = 17.73$, $P < 0.001$; sodium: $t_{22} = 18.38$, $P < 0.001$; carbohydrates (carb.): $t_{22} = 7.71$, $P < 0.001$; sugar: $t_{22} = 26.34$, $P < 0.001$; protein: $t_{22} = 18.64$, $P < 0.001$; vitamins: $t_{22} = 23.70$, $P < 0.001$).

d. Prediction performance of the subjective value in each regression model (n = 23 participants). Performance was assessed by the cross-validated correlation between the predicted and actual values. Box and whisker plots are as in c. **P < 0.01, t test (subjective factors: $t_{22} = 12.36$, $P < 0.001$; objective factors: $t_{22} = 8.34$, $P < 0.001$; subjective calories: $t_{22} = –0.26$, $P = 0.607$; objective calories: $t_{22} = –1.18$, $P = 0.875$).

e. Prediction performance of the subjective value in each logistic regression model (n = 23 participants). Performance was assessed by cross-validated accuracy. Box and whisker plots are as in c. **P < 0.01 and *P < 0.05, t test vs. 50% (subjective factors: $t_{22} = 13.61$, $P < 0.001$; objective factors: $t_{22} = 9.98$, $P < 0.001$; subjective calories: $t_{22} = 2.05$, $P = 0.026$; objective calories: $t_{22} = 2.43$, $P = 0.012$).

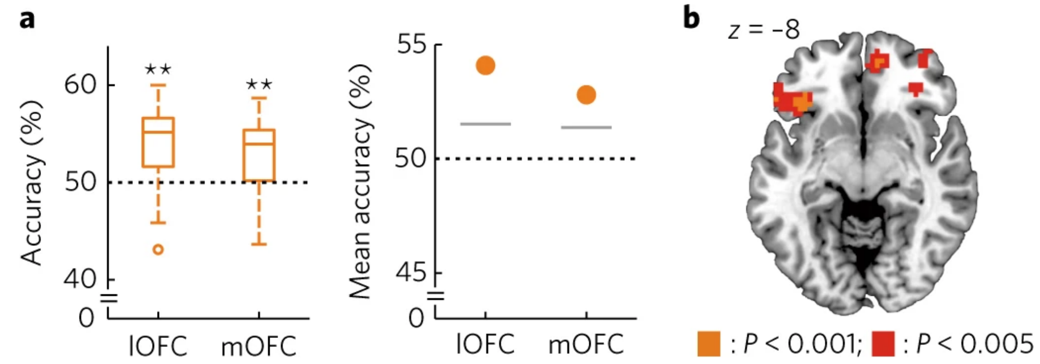


Figure 2.2: **Neural representation of subjective value.**
a. Subjective value signals can be decoded in both lateral and medial OFC (lOFC and mOFC, respectively). Decoding accuracy is plotted for the lOFC and the mOFC ROIs (n = 23 participants). Left: box and whisker plots are as in Fig. 2.1c. **P < 0.01, t test vs. 50% (lOFC: t_{22} = 4.75, P < 0.001; mOFC: t_{22} = 3.57, P < 0.001). Right: each point denotes the mean accuracy across participants. Gray horizontal lines indicate the 95th percentiles of the null distributions obtained from the permutation test procedure (lOFC: P < 0.001; mOFC: P < 0.001).
b. Subregions of the OFC encoding subjective value. The decoding accuracy map obtained from the searchlight analysis is thresholded at P < 0.005 (uncorrected) for display purposes (n = 23 participants). Peak voxels: Montreal Neurological Institute coordinates (MNI): x, y, z = –36, 26, –11 and 12, 53, –8 (P < 0.05, small-volume corrected).

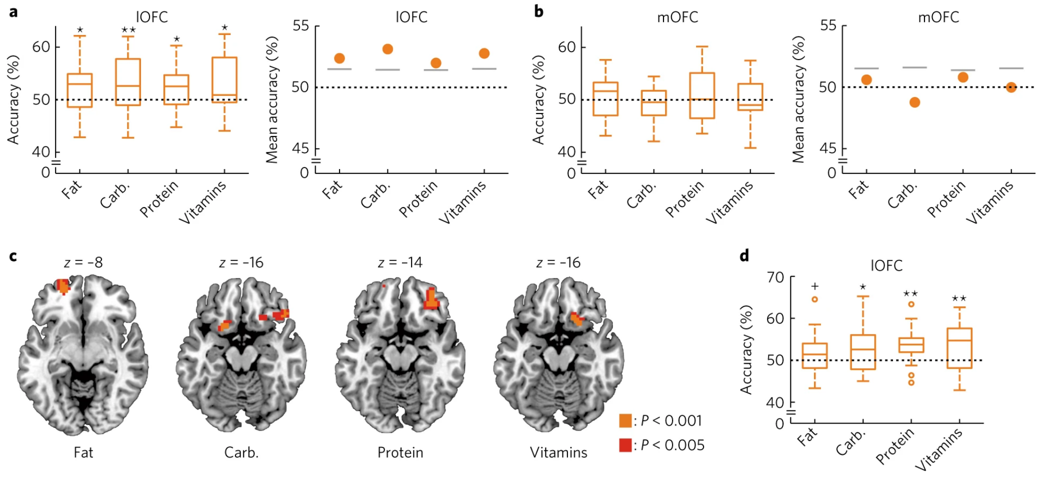


Figure 2.3: **Neural representation of subjective nutrient factors.**

a. Subjective nutrient factors can be significantly decoded from lOFC. Decoding accuracies are plotted for the lOFC ROI (n = 23 participants). Significant encoding was found for each of the nutrient factors, thereby indicating a significant conjunction effect18 at $P < 0.05$. Left: box and whisker plots are as in Fig. 2.1c. $^{*}P < 0.05$ and $^{**}P < 0.01$ for each factor, t test vs. 50% (fat: $t_{22} = 2.40$, $P = 0.013$; carb.: $t_{22} = 2.77$, $P = 0.006$; protein: $t_{22} = 2.31$, $P = 0.015$; vitamins: $t_{22} = 2.32$, $P = 0.015$). Right: as in Fig. 2.2a (right). Permutation test (fat: $P = 0.004$; carb.: $P < 0.001$; protein: $P = 0.013$; vitamins: $P = 0.001$).

b. As in a but for mOFC. Subjective nutrient factors were not significantly decodable above chance levels in mOFC (n = 23 participants). Left: t test vs. 50% (fat: $t_{22} = 0.68$, $P = 0.250$; carb.: $t_{22} = -1.74$, $P = 0.952$; protein: $t_{22} = 0.75$, $P = 0.230$; vitamins: $t_{22} = -0.02$, $P = 0.508$). Right: permutation test (fat: $P = 0.238$; carb.: $P = 0.923$; protein: $P = 0.159$; vitamins: $P = 0.519$).

c. Subregions of lOFC encoding each of the subjective nutrient factors (n = 23 participants). Decoding accuracy maps obtained from the searchlight analyses, thresholded at $P < 0.005$ (uncorrected) for display purpose. Peak voxels: MNI x, y, z = –21, 56, –8 for fat ($P < 0.05$, small-volume corrected); –15, 14, –17 for carbohydrates ($P < 0.05$, small-volume corrected); 33, 38, –14 for protein ($P < 0.05$, small-volume corrected); and 18, 17, –20 for vitamins ($P = 0.080$, small-volume corrected).

d. Decoding of subjective nutrient factors in lOFC after regressing out the effect of value (n = 23 participants). Box and whisker plots are as in a. $+P < 0.10$, $^{*}P < 0.05$ and $^{**}P < 0.01$ for each factor, t test vs. 50% (fat: $t_{22} = 1.53$, $P < 0.070$; carb.: $t_{22} = 2.20$, $P < 0.020$; protein: $t_{22} = 4.06$, $P < 0.001$; vitamins: $t_{22} = 2.90$, $P = 0.004$).

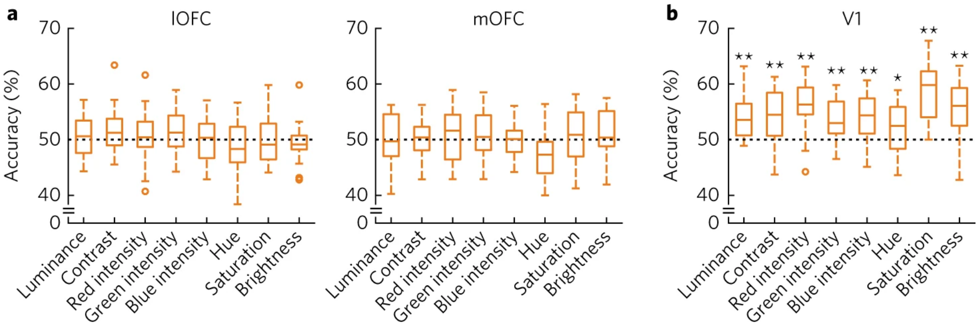


Figure 2.4: **Neural representation of low-level visual features.**

a. Low-level visual features could not be significantly decoded from lOFC or mOFC above chance levels. Decoding accuracies are plotted for lOFC and mOFC ROI (n = 23 participants). Box and whisker plots are as in Fig. 2.1c. Left: t test vs. 50% (luminance: t_{22} = 0.83, P = 0.208; contrast: t_{22} = 1.64, P = 0.058; red: t_{22} = 0.88, P = 0.195; green: t_{22} = 1.64, P = 0.057; blue: t_{22} = 0.29, P = 0.387; hue: t_{22} = –1.30, P = 0.896; saturation: t_{22} = 0.00, P = 0.520; brightness: t_{22} = –0.86, P = 0.800). Right: t test vs. 50% (luminance: t_{22} = –0.26, P = 0.603; contrast: t_{22} = 0.22, P = 0.414; red: t_{22} = 0.81, P = 0.212; green: t_{22} = 0.87, P = 0.200; blue: t_{22} = –0.15, P = 0.440; hue: t_{22} = –3.12, P = 0.998; saturation: t_{22} = 0.50, P = 0.313; brightness: t_{22} = 1.05, P = 0.153).

b. Low-level visual features could be robustly decoded from V1 (n = 23 participants). Box and whisker plots are as in Fig. 2.1c. *P < 0.05 and **P < 0.01 for each factor, t test vs. 50% (luminance: t_{22} = 5.07, P < 0.001; contrast: t_{22} = 4.60, P < 0.001; red: t_{22} = 6.30, P < 0.001; green: t_{22} = 5.03, P < 0.001; blue: t_{22} = 4.60, P < 0.001; hue: t_{22} = 2.20, P = 0.019; saturation: t_{22} = 8.32, P < 0.001; brightness: t_{22} = 5.38, P < 0.001).

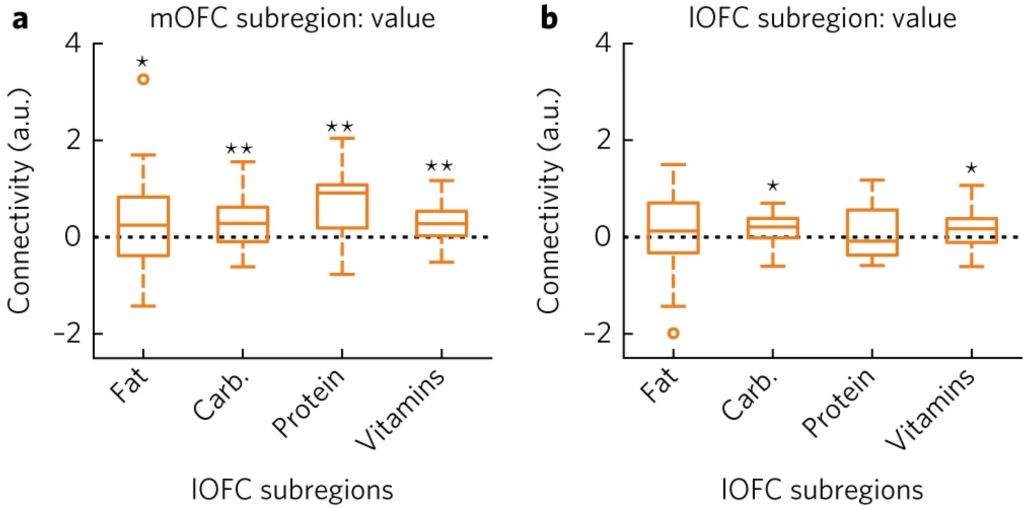


Figure 2.5: **Effective connectivity between OFC subregions at the time of valuation.**

a. Results of an effective connectivity analysis between the value-encoding mOFC subregion and the lOFC subregions encoding each of the four nutrient factors. A significant connectivity effect was found for each of the nutrient factors, thereby indicating a significant conjunction effect18 at $P < 0.05$. Effect sizes of the psychophysiological interaction (PPI) regressors are plotted ($n = 23$ participants). Box and whisker plots are as in Fig. 2.1c. **P < 0.01 and *P < 0.05 for each factor, t test (fat: $t_{22} = 1.74$, $P = 0.048$; carb.: $t_{22} = 2.85$, $P = 0.005$; protein: $t_{22} = 5.05$, $P < 0.001$; vitamins: $t_{22} = 3.25$, $P = 0.002$).

b. Results of an effective connectivity analysis between the value-encoding lOFC subregion and other lOFC subregions encoding each of the four nutrient factors. Box and whisker plots are as in a; t test (fat: $t_{22} = 0.62$, $P = 0.272$; carb.: $t_{22} = 1.78$, $P = 0.045$; protein: $t_{22} = 1.12$, $P = 0.137$; vitamins: $t_{22} = 1.78$, $P = 0.045$). While a significant connectivity effect was found for two of the factors (carb. and vitamins), the other two factors did not reach significance, and thus an overall significant conjunction effect was not found in lateral OFC. A.u., arbitrary units.

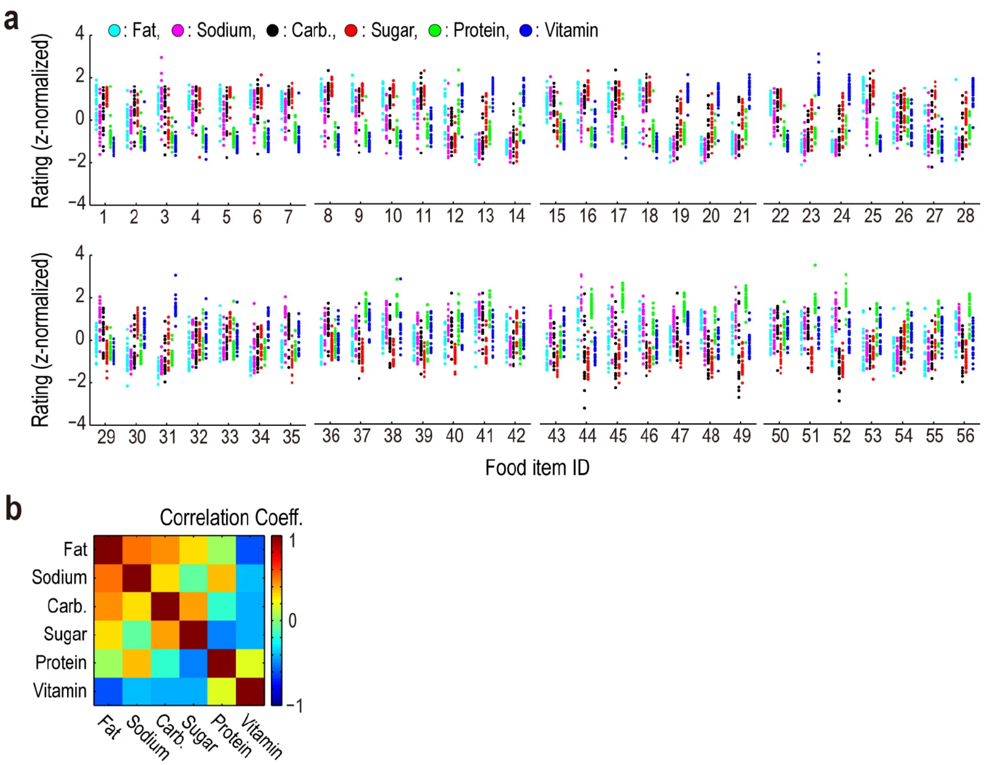


Figure 2.6: **Supplementary Figure 1: Ratings of nutrient factors.**
a. Subjective ratings about the 56 food items. For each item, we plot the participants' ratings about the six nutrient factors (cyan: fat; magenta: sodium; black: carbohydrate; red: sugar; green: protein; and blue: vitamin). See Table S2 for the item list. The rating data were z-normalized across the food items, within each participant and each nutrient factor. (b) Pair-wise correlations among subjective ratings of the nutrient factors (MEAN across participants; n = 23).

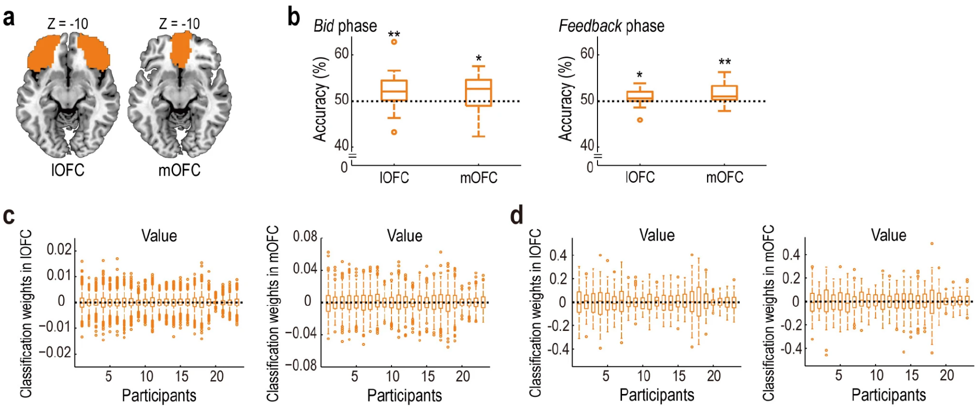


Figure 2.7: **Supplementary Figure 2: Supplementary results of the neural representation of subjective value.**

a. Anatomical OFC ROIs used in this study. The ROIs are defined based on the AAL database as follows: lOFC, bilateral MNI_Frontal_Mid_Orb + MNI_Frontal_Inf_Orb + MNI_Frontal_Sup_Orb; and mOFC, bilateral MNI_Frontal_Med_Orb. lOFC, lateral orbitofrontal cortex; and mOFC, medial orbitofrontal cortex.

b. Evidence for significant decoding of subjective value at the Bid phase (left) and at the Feedback phase (right) (n = 23 participants). The format is the same as in Fig. 2.2a (left). **P < 0.01 and *P < 0.05, t-test against 50% (Bid phase, lOFC: t_{22} = 2.78, P = 0.005; mOFC: t_{22} = 1.96, P = 0.031; and Feedback phase, lOFC: t_{22} = 2.23, P = 0.018; mOFC: t_{22} = 3.40, P = 0.001).

c. Weights of voxels in the value classifiers obtained from the ROI analyses (see Fig. 2.2a). We plot the weights of the voxels for each participant within the lOFC (left) and the mOFC (right) ROIs separately. Format of the box and whisker plots is the same as in Fig. 2.1c.

d. Weights of voxels in the value classifiers obtained from the searchlight analyses (see Fig. 2.2b). We plot the weights of the voxels within a radius of 3 voxels (i.e., 9 mm) around the peak voxels in lOFC (left) and mOFC (right). See Fig. 2.2b for information about the peak voxels. Format of the box and whisker plots is the same as in Fig. 2.1c.

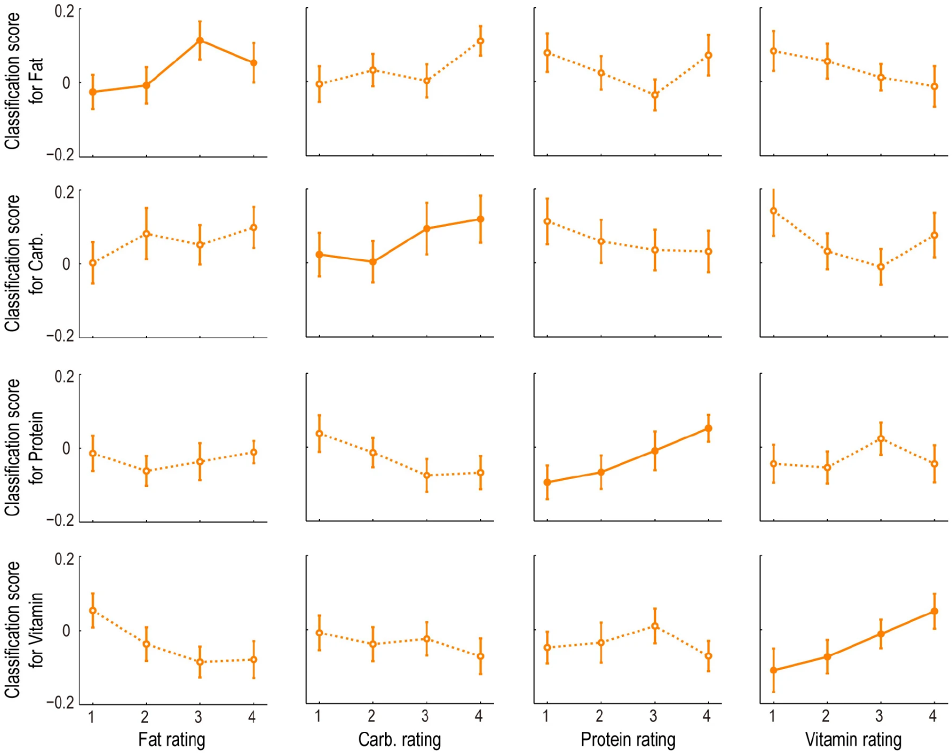


Figure 2.8: **Supplementary Figure 3: Classification scores in the decoding analysis for subjective nutrient factors.** We plot the classification scores in the lOFC ROI obtained by the classifier trained on fat, carb., protein, and vitamin respectively, as functions of the subjective nutrititive ratings (MEAN±SEM across participants; n = 23; see Fig. 2.3a). Note that ratings for each nutrient factors were binned based on the rank order; that each classifier is trained to discriminate high vs. low ratings (i.e., 1 & 2 vs. 3 & 4); and that the classification weights of each voxel were estimated on a subset of the data and the classification scores were computed on the other subset of the data (i.e., leave-one-run-out cross-validation; see Methods for details). lOFC, lateral orbitofrontal cortex; and Carb., carbohydrade.

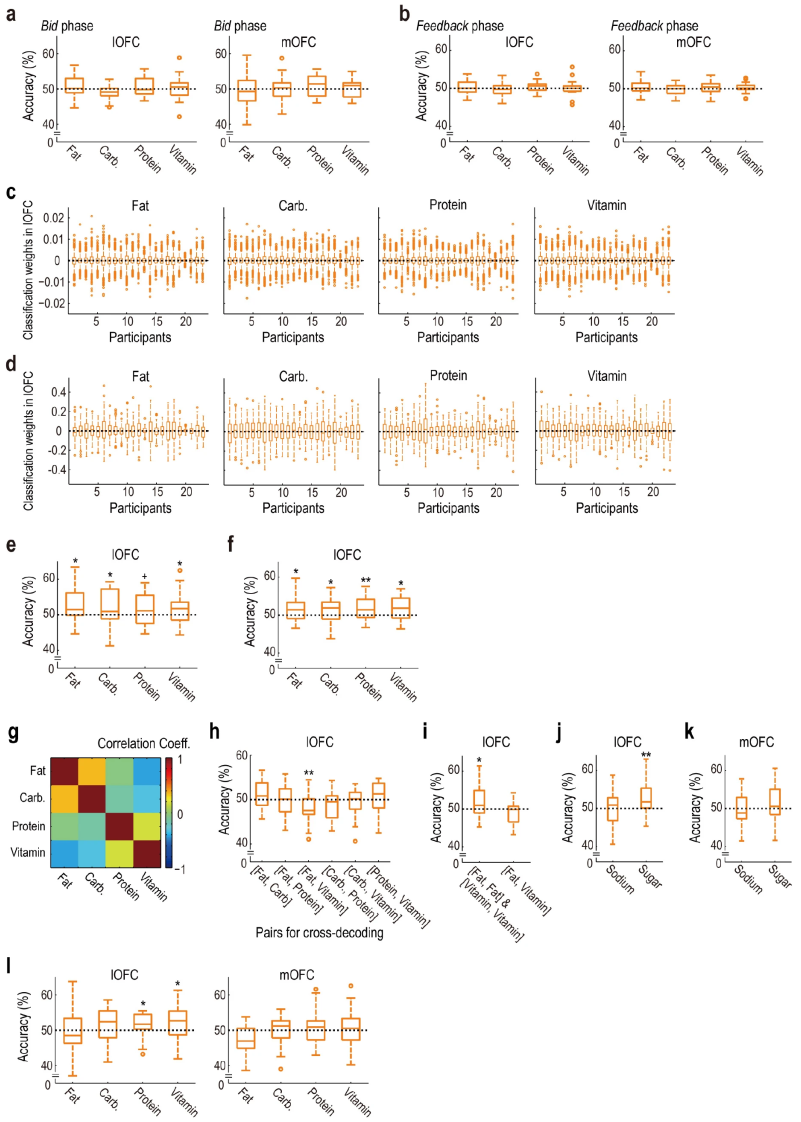


Figure 2.9: **Supplementary Figure 4: Supplementary results of the neural representation of subjective nutrient factors.**

a. Decoding accuracies of subjective nutrient factors at the time of bidding revealing a lack of significant decoding at this time-point (n = 23 participants). The format is the same as in Fig. 2.3a (left). Left. t-test against 50% (fat: t_{22} = 1.38, P = 0.091; carb.: t_{22} = -2.47, P = 0.989; protein: t_{22} = 1.11, P = 0.139; and vitamin: t_{22} = 0.48, P = 0.320). Right. t-test (fat: t_{22} = -0.42, P = 0.660; carb.: t_{22} = 0.14, P = 0.444; protein: t_{22} = 1.26, P = 0.110; and vitamin: t_{22} = 0.42, P = 0.339). lOFC, lateral orbitofrontal cortex; mOFC, medial orbitofrontal cortex; and Carb., carbohydrate.

b. Decoding accuracies of subjective nutrient factors at the time of feedback revealing little evidence for significant decoding at this time-point (n = 23 participants). The format is the same as in Fig. 2.3a (left). Left. t-test against 50% (fat: t_{22} = 0.72, P = 0.239; carb.: t_{22} = -0.43, P = 0.664; protein: t_{22} = 1.38, P = 0.090; and vitamin: t_{22} = 0.18, P = 0.431). Right. t-test (fat: t_{22} = 0.90, P = 0.190; carb.: t_{22} = -0.80, P = 0.783; protein: t_{22} = 1.07, P = 0.149; and vitamin: t_{22} = 1.01, P = 0.162).

c. Weights of voxels in the classifiers for each of the subjective nutrient factors obtained from the ROI analyses (see Fig. 2.3a). We plot the weights of the voxels for each participant within the lOFC ROI. Format of the box and whisker plots is the same as in Fig. 2.1c.

d. Weights of voxels in the classifiers for each of the subjective nutrient factors obtained from the search analyses (see Fig. 2.3c). We plot the weights of the voxels within a radius of 3 voxels (i.e., 9 mm) around the peak voxels in lOFC. See Fig. 2.3c for information about the peak voxels. Format of the box and whisker plots is the same as in Fig. 2.1c.

e. Decoding accuracies of the subjective nutrient factors for novel food items (n = 23 participants). The format is the same as in Fig. 2.3a (left). + $P < 0.10$, *$P < 0.05$ and **$P < 0.01$ for each factor, t-test against 50% (fat: $t_{22} = 2.42$, $P = 0.012$; carb.: $t_{22} = 1.90$, $P = 0.035$; protein: $t_{22} = 1.41$, $P = 0.087$; and vitamin: $t_{22} = 1.84$, $P = 0.039$).

f. Decoding accuracies of the subjective nutrient factors in the reduced lOFC ROIs (n = 23 participants). In this analysis, (i) we randomly re-sampled adjacent 533 voxels from the lOFC ROI (i.e., forming a continuous cluster consisting of the 553 voxels); then (ii) we tested if information about the subjective nutrient factors could be decoded from the re-sampled voxels; and (iii) the above procedure was repeated 100 times (the decoding accuracies were averaged). The format is the same as in Fig. 2.3a (left). *$P < 0.05$ and **$P < 0.01$ for each factor, t-test against 50% (fat: $t_{22} = 2.45$, $P = 0.011$; carb.: $t_{22} = 1.81$, $P = 0.042$; protein: $t_{22} = 2.59$, $P = 0.008$; and vitamin: $t_{22} = 2.46$, $P = 0.011$).

g. Pair-wise correlations among the classifiers' weights for the four nutrient factors (MEAN across participants; n = 23). For each pair of the nutrient factors, we obtained the correlation coefficient in the classification weights of the voxels within the lOFC ROI.

h. Decoding accuracies in the cross-decoding analyses (n = 23 participants). Format of the box and whisker plots is the same as in Fig. 2.1c. Two nutrient factors in each parenthesis denote the pair used for the cross-decoding. **$P < 0.01$, two-tailed t-test against 50% ([fat, carb.]: $t_{22} = 1.59$, $P = 0.127$; [fat, protein]: $t_{22} = -0.35$, $P = 0.729$; [fat, vitamin]: $t_{22} = -3.18$, $P = 0.004$; [carb., protein]: $t_{22} = -1.96$, $P = 0.062$; [carb., vitamin]: $t_{22} = -1.06$, $P = 0.299$; and [protein, vitamin]: $t_{22} = 0.84$, $P = 0.408$).

i. Decoding accuracies on the re-sampled food items (see the main text; n = 23 participants). Format of the box and whisker plots is the same as in Fig. 2.1c. Left, accuracy of fat and vitamin (one classifier was trained and tested on fat; the other classifier was on vitamin; and the accuracy scores were averaged). Right, accuracy in the cross-decoding analysis between fat and vitamin. Two nutrient factors in the parenthesis denote the pair used for the cross-decoding. That is, we trained a classifier on one factor and tested it on the other factor (and the reverse; and the decoding accuracy was assessed by the average across both directions). *P < 0.05, two-tailed t-test against 50% ([fat, fat] & [vitamin, vitamin]: t_{22} = 2.10, P = 0.048; and [fat, vitamin]: t_{22} = -1.10, P = 0.282).

j. Significant decoding of sugar but not sodium content in lOFC (n = 23 participants). The accuracies are plotted for the lOFC ROI. Format of the box and whisker plots is the same as in Fig. 2.1c. **P < 0.01, t-test (Sodium: t_{22} = 0.06, P = 0.474; and Sugar: t_{22} = 2.67, P = 0.007). (k) Neither sodium nor sugar content was significantly decodable in mOFC (n = 23 participants). t-test against 50% (Sodium: t_{22} = -0.15, P = 0.557; and Sugar: t_{22} = 1.34, P = 0.100). mOFC, medial orbitofrontal cortex. Format of the box and whisker plots is the same as in Fig. 2.1c. (l) Decoding accuracies of objective nutrient factors at the time of valuation, demonstrating relatively weak effects of objective nutrient factors (n = 23 participants). The format is the same as in Fig. 2.3a (left). Left. *P < 0.05, t-test against 50% (fat: t_{22} = -0.33, P = 0.626; carb.: t_{22} = 1.58, P = 0.064; protein: t_{22} = 2.05, P = 0.026; and vitamin: t_{22} = 2.26, P = 0.017). Right. t-test (fat: t_{22} = -3.10, P = 0.997; carb.: t_{22} = 0.29, P = 0.387; protein: t_{22} = 0.78, P = 0.222; and vitamin: t_{22} = 0.08, P = 0.469).

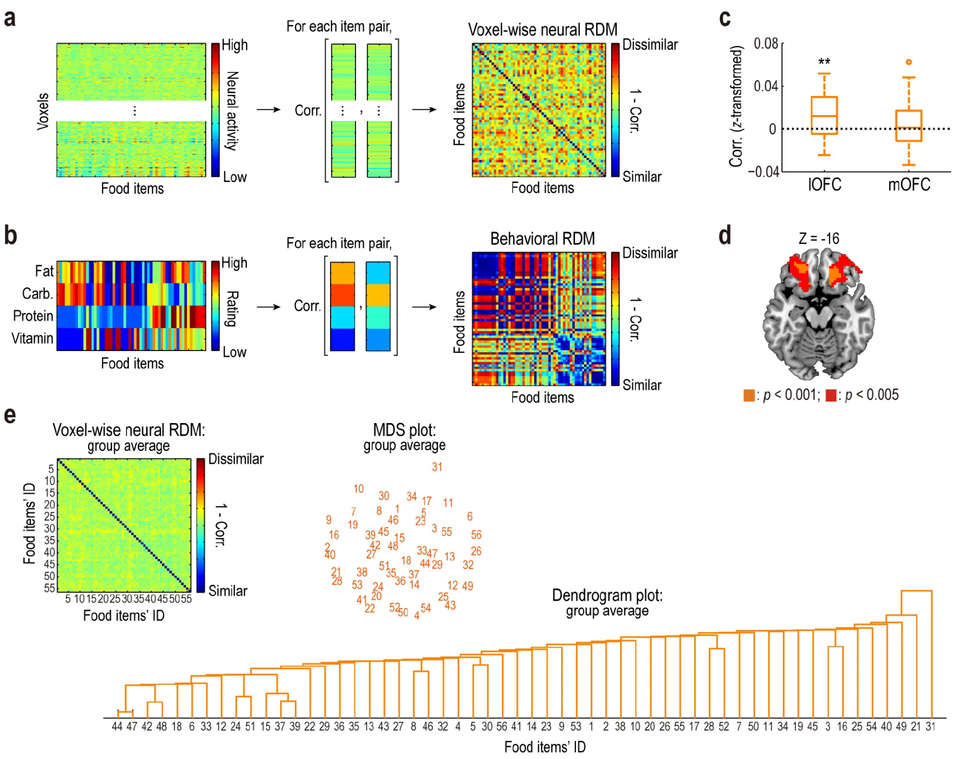


Figure 2.10: **Supplementary Figure 5: Procedure and results of the representational similarity analysis (RSA).**

a. Procedure for construction of the voxel-wise Representational Dissimilarity Matrix (RDM). The voxel-wise RDM is created based on the correlation across the voxels' activities for each pair of the items. See Methods for details. Corr., Pearson's correlation coefficient.

b. Procedure for construction of the behavioral RDM. The behavioral RDM is created based on the correlation in bundles of the four subjective nutrient factors for each item pair. See Methods for details.

c. Results of the ROI analyses. Spearman's rank correlation (z-transformed) between the voxel-wise neural and the behavioral RDMs is plotted for the lOFC and the mOFC ROIs (n = 23). Format of the box and whisker plots is the same as in Fig. 2.1c. **P < 0.01, t-test (lOFC: t_{22} = 2.85, P = 0.005; and mOFC: t_{22} = 1.13, P = 0.135). lOFC, lateral orbitofrontal cortex; and mOFC, medial orbitofrontal cortex.

d. Results of the searchlight analysis. The RSA correlation map is thresholded at P < 0.005 (uncorrected) for display purposes, generated by performing a t-test (n = 23 participants). Peak voxels, [MNI: x, y, z = 12, 23, -23] and [-21 38 -23] (P < 0.05 small-volume corrected) for right and left OFC, respectively. OFC, orbitofrontal cortex.

e. Pattern of fMRI response to each of the 56 food items in a space of the pair-wise correlation across voxels' activities in the lOFC. We plot the voxel-wise neural RDM averaged over the participants (top left, n = 23). To visualize the approximate geometric structure, we also show the same data as a two-dimensional MDS plot (top center) and a dendrogram plot obtained by an agglomerative hierarchical clustering (bottom right). In the MDS plot, the digits depict the food items' ID. See Table S2 for detailed information about the food items. MDS, multi dimensional scaling.

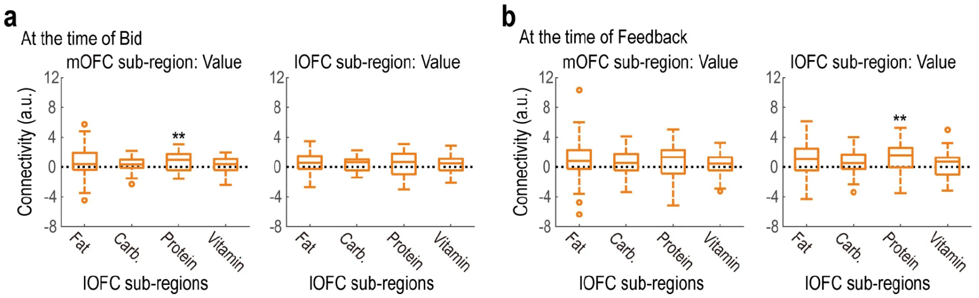


Figure 2.11: **Supplementary Figure 6: Effective connectivity between OFC subregions at the time of bidding and the time of feedback.**

a. Results of an effective connectivity analysis at the time of bidding. Effect sizes of the PPI regressors are plotted (n = 23 participants). The format is the same as in Fig. 2.5ab. $^{**}P < 0.01$ for each factor. Left. t-test (fat: $t_{22} = 1.37$, $P = 0.092$; carb.: $t_{22} = 1.28$, $P = 0.108$; protein: $t_{22} = 3.20$, $P = 0.002$; and vitamin: $t_{22} = 1.40$, $P = 0.088$). Right. t-test (fat: $t_{22} = 1.59$, $P = 0.063$; carb.: $t_{22} = 1.61$, $P = 0.060$; protein: $t_{22} = 1.54$, $P = 0.068$; and vitamin: $t_{22} = 1.33$, $P = 0.099$). lOFC, lateral orbitofrontal cortex; mOFC, medial orbitofrontal cortex; Carb., carbohydrate; and PPI, psychophysiological interaction.

b. Results of an effective connectivity analysis at the time of feedback. Effect sizes of the PPI regressors are plotted (n = 23 participants). The format is the same as in Fig. 2.5ab. $^{**}P < 0.01$ for each factor. Left. t-test (fat: $t_{22} = 1.30$, $P = 0.104$; carb.: $t_{22} = 1.67$, $P = 0.055$; protein: $t_{22} = 1.22$, $P = 0.118$; and vitamin: $t_{22} = 1.02$, $P = 0.159$). Right. t-test (fat: $t_{22} = 1.64$, $P = 0.058$; carb.: $t_{22} = 1.32$, $P = 0.100$; protein: $t_{22} = 2.80$, $P = 0.005$; and vitamin: $t_{22} = 1.01$, $P = 0.161$).

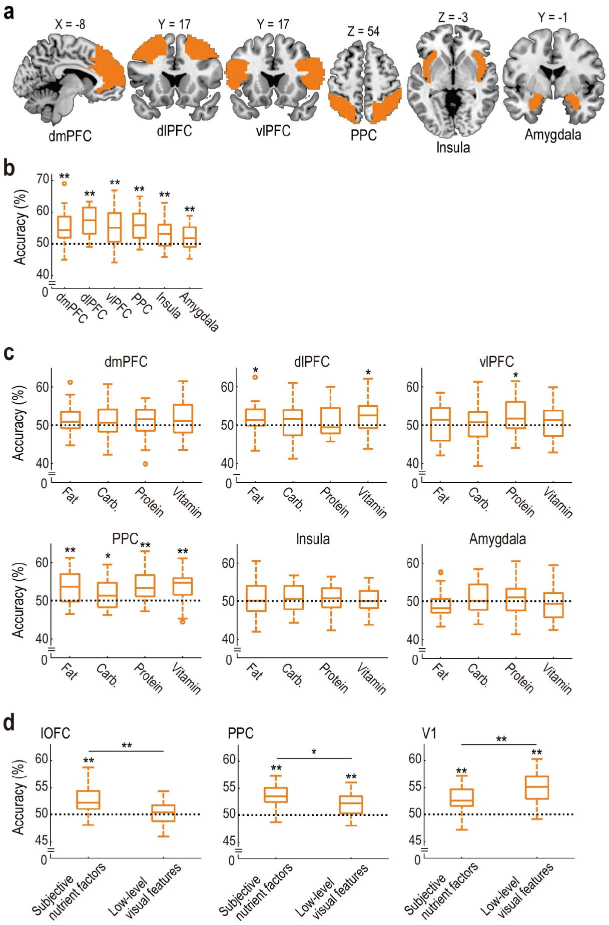


Figure 2.12: **Supplementary Figure 7: Decoding of subjective value and nutrient factors in other brain regions.**

a. Anatomical ROIs used in the additional post hoc analyses. The ROIs are defined based on the AAL database42 as follows: dmPFC, bilateral MNI_ Frontal_ Sup_ Medial + MNI_ Cingulum_ Ant; dlPFC, bilateral MNI_ Frontal_ Mid + MNI_ Frontal_ Sup; vlPFC, bilateral MNI_ Frontal_ Inf_ Oper + MNI_ Frontal_ Inf_ Tri; PPC, bilateral MNI_ Parietal_ Inf + MNI_ Parietal_ Sup; Insula, bilateral MNI_ Insula; and Amygdala, bilateral MNI_ Amygdala. dmPFC, dorsomedial prefrontal cortex; dlPFC, dorsolateral prefrontal cortex; vlPFC, ventrolateral prefrontal cortex; and PPC, posterior parietal cortex.

b. Decoding accuracies of subjective value across the ROIs (n = 23 participants). The format is the same as in Fig. 2.2a (left). **P < 0.01 for each region, t-test against 50% (dmPFC: t_{22} = 4.55, P < 0.001; dlPFC: t_{22} = 7.30, P < 0.001; vlPFC: t_{22} = 4.39, P < 0.001; PPC: t_{22} = 6.52, P < 0.001; Insula: t_{22} = 3.48, P = 0.001; and Amygdala: t_{22} = 2.92, P = 0.004).

c. Decoding accuracies of subjective nutrient factors (n = 23 participants). The format is the same as in Fig. 2.3a (left). *P < 0.05 and **P < 0.01 for each factor. Top left. t-test against 50% (fat: t_{22} = 1.57, P = 0.066; carb.: t_{22} = 1.12, P = 0.137; protein: t_{22} = 0.98, P = 0.168; and vitamin: t_{22} = 1.71, P = 0.050).

Top middle. t-test (fat: $t_{22} = 2.25$, $P = 0.018$; carb.: $t_{22} = 1.26$, $P = 0.111$; protein: $t_{22} = 1.52$, $P = 0.071$; and vitamin: $t_{22} = 2.57$, $P = 0.009$). Top right. t-test (fat: $t_{22} = 0.51$, $P = 0.307$; carb.: $t_{22} = 0.52$, $P = 0.305$; protein: $t_{22} = 2.44$, $P = 0.012$; and vitamin: $t_{22} = 0.90$, $P = 0.189$). Bottom left. t-test (fat: $t_{22} = 4.11$, $P < 0.001$; carb.: $t_{22} = 2.37$, $P = 0.014$; protein: $t_{22} = 4.46$, $P < 0.001$; and vitamin: $t_{22} = 4.50$, $P < 0.001$). Bottom middle. t-test (fat: $t_{22} = 0.65$, $P = 0.261$; carb.: $t_{22} = 0.83$, $P = 0.209$; protein: $t_{22} = 0.83$, $P = 0.208$; and vitamin: $t_{22} = 0.05$, $P = 0.481$). Bottom right. t-test (fat: $t_{22} = -0.79$, $P = 0.780$; carb.: $t_{22} = 0.73$, $P = 0.236$; protein: $t_{22} = 0.50$, $P = 0.312$; and vitamin: $t_{22} = -0.23$, $P = 0.592$).

d. Decoding accuracies of low-level visual features and comparison with subjective nutrient factors in lOFC, PPC and V1. The decoding accuracies of the low-level visual features (averaged over the eight features) and the subjective nutrient factors (averaged over the four factors identified as value predictors) are plotted for the lOFC, the PPC and the V1 (BA17) anatomical ROIs (n = 23 participants). Format of the box and whisker plots is the same as in Fig. 2.1c. * and ** on each plot, respectively, denote $P < 0.05$ and $P < 0.01$, for each factor, t-test against 50%. * and ** on the horizontal lines denote significant differences between the indicated pairs of data at $P < 0.05$ and $P < 0.01$ respectively, two-tailed paired t-test. Left. t-test (subjective nutrient factors: $t_{22} = 4.72$, $P < 0.001$; low-level visual features: $t_{22} = 0.70$, $P = 0.247$; and subjective nutrient factors vs. low-level visual features: $t_{22} = 3.18$, $P = 0.004$). Middle. t-test (subjective nutrient factors: $t_{22} = 7.04$, $P < 0.001$; low-level visual features: $t_{22} = 4.01$, $P < 0.001$; and subjective nutrient factors vs. low-level visual features: $t_{22} = 2.62$, $P = 0.0157$). Right. t-test (subjective nutrient factors: $t_{22} = 5.85$, $P < 0.001$; low-level visual features: $t_{22} = 8.34$, $P < 0.001$; and subjective nutrient factors vs. low-level visual features: $t_{22} = -3.15$, $P = 0.005$). lOFC, lateral orbitofrontal cortex; PPC, posterior parietal cortex; V1, primary visual cortex; and BA17, Brodmann area 17.

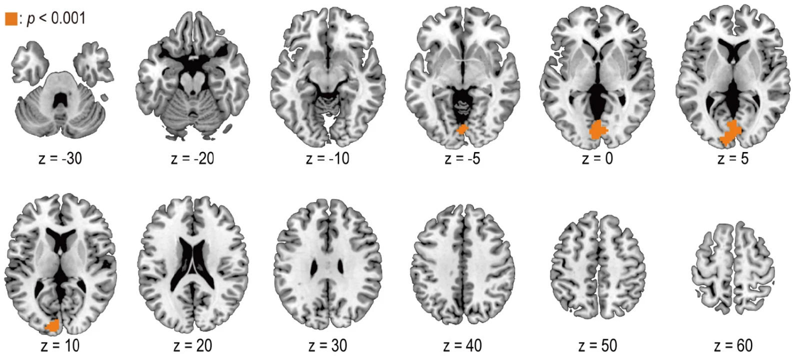


Figure 2.13: **Supplementary Figure 8: A region of V1 in which all of the four subjective nutrient factors can be decoded.**
The decoding accuracy map obtained from the whole-brain searchlight analysis is thresholded at P < 0.05 (cluster-level FWE correction with the cluster-forming threshold P = 0.001; n = 23 participants), conjunction-test. Peak voxel, [MNI: x, y, z = -9, -94, 7]. V1, primary visual cortex.

	Rank	Explanatory variable	Performance
Linear regression	**1**	**fat, carbohydrate, protein, vitamin**	**0.481**
	2	fat, sodium, carbohydrate, protein, vitamin	0.478
	3	fat, sodium, protein, vitamin	0.474
	4	fat, protein, vitamin	0.465
	5	sodium, carbohydrate, protein, vitamin	0.463
	6	fat, sodium, sugar, protein, vitamin	0.455
	7	fat, carbohydrate, sugar, protein, vitamin	0.455
	8	carbohydrate, protein, vitamin	0.453
	9	fat, sodium, carbohydrate, sugar, protein, vitamin	0.452
	10	fat, sugar, protein, vitamin	0.440
Logistic regression	**1**	**fat, carbohydrate, protein, vitamin**	**69.95**%
	2	fat, sodium, carbohydrate, protein, vitamin	69.64%
	3	fat, sodium, protein, vitamin	69.26%
	4	fat, protein, vitamin	69.10%
	5	sodium, carbohydrate, protein, vitamin	68.94%
	6	fat, sodium, sugar, protein, vitamin	68.79%
	7	fat, carbohydrate, sugar, protein, vitamin	68.79%
	8	carbohydrate, protein, vitamin	68.79%
	9	fat, sodium, carbohydrate, sugar, protein, vitamin	68.48%
	10	fat, sugar, protein, vitamin	68.25%

Table 2.1: **Prediction performances of the subjective value.**
Prediction performances of the best 10 models are shown for linear and logistic regression analyses (the best model in each analysis is shown in bold). Performance, z-transformed correlation between the predicted and the actual values for the linear regression analysis, and prediction accuracy for the logistic regression analysis. See Methods for details.

Food items used

1. 3 Musketeers[d]
2. Barnum’s Animal Crackers[d]
3. Doritos Nacho Cheese[e]
4. Chips Ahoy![d]
5. Kit Kat[d]
6. Pop-Tarts Brown Sugar Cinnamon[d]
7. Pop-Tarts Brown Sugar Strawberry[d]
8. Hostess Powdered Donettes[d]
9. Twix Cookie Bars[d]
10. Hershey’s Whatchamacallit Candy[d]
11. Apple Pie[d]
12. Avocado[a]
13. Blackberries[a]
14. Cauliflower[a]
15. Ritz Crackers’n Cheese Dip[e]
16. Cherry Pie[d]
17. Chocolate Muffins[d]
18. Hostess Donuts[d]
19. Granny Smith Apple[a]
20. Green Grapes[a]
21. Mango[a]
22. Milano Cookies[d]
23. Orange[a]
24. Raspberries[a]
25. Red Velvet Cake[d]
26. Quaker Chewy Granola Bar[d]
27. Starburst Candy[d]
28. Strawberry[a]
29. Sun Chips[e]
30. Dole Mixed Fruit[a]
31. Grapefruit[a]
32. Banana Chips[d]
33. Dark Chocoloate Bananas[d]
34. Crispy Apple[e]
35. Vegetable Chips[e]
36. Sweet Potato Chips[e]
37. Chopped Salad Chicken[c]
38. Mexicali Salad[c]
39. Caesar Salad[c]
40. Veggie Wrap[c]
41. Super Burrito[c]
42. Chocolate and Berry[d]
43. Green Beans Chips[e]
44. Salami[b]
45. Smoked Turkey[b]
46. American Cheese[b]
47. Chicken and Roasted Beet[c]
48. Mozzarella Cheese[b]
49. Roast Beef[b]
50. Caprese Sandwich[c]
51. Tuna Salad Wrap[c]
52. Smoked Salmon[b]
53. Plain Yogurt[b]
54. Strawberry Yogurt[b]
55. Blueberry Yogurt[b]
56. Deviled Eggs[b]

Table 2.2: [a] fresh vegetables and fruits (e.g., Orange and Apple), [b] meet and dairy products (e.g., salami and yogurt), [c] cooked products (e.g., salad and wrap), [d] sweet snacks (e.g., chocolate bar and cake), and [e] salty snacks (e.g., chips and crackers).

Chapter 3

INVESTIGATING THE NEURAL MECHANISMS OF BUNDLE VALUATION

Abstract

The computational mechanisms of multiattribute decision-making have begun to be uncovered at the behavioral and neural level. The results of Chapter 2 suggest that the brain decomposes stimuli into their constituent attributes and computes value as a weighted integration of these attributes. Here we test this framework as human participants evaluated bundles of multiple items. An extension of the multiattribute integration model explains significant variance in how bundles are valued, as the constituent items themselves become the features to be integrated. This integration occurs with a subadditive function for which bundle values are systematically discounted versus the sum of the individual item values. A distributed network throughout the prefrontal cortex (PFC) computes the value of a bundle with the same value code used to evaluate individual items, suggesting that these general value regions contextually adapt within this hierarchy. Additionally, the value representation moves across levels in this hierarchy with a normalization process that rescales the code to the distribution of values in the current context, as opposed to an absolute value code.

Introduction

In daily life, humans have to evaluate options that contain multiple isolated components, such as a meal, a vacation package, bundles of channels on cable TV, or an investment fund. Valuing bundles such as these must require a hierarchical process of valuing the bundle's constituent items and then integrating them to value the bundle as a whole. This process is analogous to attribute integration theories of value, which propose that the value of a stimulus is constructed by assigning values to the stimulus's attributes and integrating them (Bettman, Luce, and Payne, 1998; Suzuki, Cross, and O'Doherty, 2017). Recent human neuroimaging studies have identified encoding of stimulus attributes in cortical areas specialized in the processing of these features (Suzuki, Cross, and O'Doherty, 2017; Lim, John P. O'Doherty, and Rangel, 2013). For example, when evaluating food items, the elemental nutritive attributes of the food are encoded in the lateral orbitofrontal cortex (OFC). Notably, the areas encoding stimulus features exhibit functional connectivity to medial OFC and the nearby ventromedial prefrontal cortex (vmPFC) (Suzuki, Cross, and O'Doherty, 2017; Lim, John P. O'Doherty, and Rangel, 2013). Given that value signals have been identified in mOFC and vmPFC in association to a wide variety or rewards (Clithero and Rangel, 2014; McNamee, Rangel, and John P O'Doherty, 2013; Kable and Glimcher, 2007; Gross et al., 2014; Chib et al., 2009), this evidence suggests that attribute values are integrated in mOFC/vmPFC to compute the value of the item being assessed. In the present study, we generalize this attribute integration theory of value to the process of valuing bundles of items, for which the components that need to be integrated operate on a different level of abstraction: they are items themselves. To our knowledge, no studies have formally examined how bundles of items are evaluated in the human brain, as most neuroeconomics studies have solely investigated how individual items are valued.

Therefore, in this study participants valued food items, noncomestible consumer goods, and bundles of these items. Then participants made choices with these items and bundles while being scanned with functional MRI (fMRI). Our first item was to characterize how the values of bundles are constructed from the values of their constituent items. We hypothesized that participants would discount bundles in relation to the sum of the values of the bundle's items, and therefore bundle value could be modeled as a subadditive function of the individual item values.

At the neural level, there is conflicting evidence in the literature about whether the brain uses the same neural code to represent the value of stimulus independently of the category of the stimulus or the task context. There is evidence from neuroimaging and electrophysiology studies that OFC/vmPFC regions encode value with a 'common currency' across different types of goods and stimuli (Padoa-Schioppa and Assad, 2008; Chib et al., 2009; Lin, Adolphs, and Rangel, 2012; Levy and Glimcher, 2011). In contrast, another study demonstrated that by using a more sensitive fMRI analysis method called multivariate pattern analysis (MVPA), category-dependent distributed value representations could be identified in spatially distinct areas of OFC/vmPFC for valuing food items and consumer items separately (McNamee, Rangel, and John P O'Doherty, 2013). Additionally, this study characterized a topographical gradient of value representations in vmPFC, with more anterior regions encoding the value of more abstract rewards and more dorsal regions implementing a 'common currency.' Therefore, we tested whether spatially distinct regions separately compute the values of single items and bundles, or whether the same regions utilize the same value representation across these two conditions in accordance with a general value code.

A related question concerns how the neural value code represents bundle values in relation to single item values. Recent evidence suggests that decision-making regions in the brain implement a relative value code by adapting to the temporal and spatial context of the choice set (Louie, Glimcher, and Webb, 2015). As bundle values are higher on average than single item values, we hypothesize that value regions switch between levels in this hierarchy by rescaling the value code according to the distribution of values in the current context, rather than encoding the absolute value of an option.

Results

To investigate the neural mechanisms involved in evaluating a bundle of items, we recruited participants in a 3 day experiment. On each day, a participant first reported how much they would be willing to pay (WTP) for various foods and non-comestible consumer goods (trinkets) and pairs of these items (bundles) (Figure 3.1a). These WTP ratings represent an item's/bundle's subjective value and are recorded with an incentive compatible procedure used in previous studies (see Methods for details) (Becker, DeGroot, and Marschak, 1964; Chib et al., 2009; McNamee, Rangel, and John P O'Doherty, 2013; Suzuki, Cross, and O'Doherty, 2017). Participants were allocated a budget of \$0-\$20 and submitted a wide distribution of WTP bids across categories, albeit with most items and bundles valued at small dollar amounts (Figure 3.2a; mean WTP value for each category: individual item food = \$3.34 ± 3.30 s.d., individual item trinket = \$3.30 ± 3.50 s.d., food bundle = \$5.96 ± 4.44 s.d., trinket bundle = \$4.92 ± 4.81 s.d., mixed bundle \$5.394 ± 4.65 s.d.).

After the WTP task on each day, participants were scanned using functional MRI (fMRI) while they performed a choice task with the same items and bundles. Three participants were scanned with a high-resolution fMRI protocol (voxel size = 1.5mm isotropic) designed to record from medial prefrontal cortex (mPFC) regions with high fidelity (Figure S1). The remaining eleven participants were scanned with a standard wholebrain protocol (2.5 mm isotropic). On each trial in the choice task, a participant made a choice between an item (or a bundle) vs a reference monetary amount equal to their median WTP bid on that category (Figure 3.1b). This ensured that participants would value the item/bundle in isolation and choose the item/bundle about half the time and the reference monetary amount about half the time, which was seen empirically in the data other than a slight bias to choosing the money on individual item trials (Figure S2).

Bundle value is a subadditive function of the individual item values

Behaviorally, we first examined how a bundle's value is constructed from the values of its constituent items. This relationship can be qualitatively visualized in the density plot in Figure 3.2b; as the sum of the individual item values increases, bundle value tends to also monotonically increase as expected. However, if bundle value was equal to a simple addition of the values of the individual items, the data would lie along the diagonal. Instead, there is more density below the Y=X diagonal, suggesting that bundle value is a subadditive function of the individual item values.

To test this hypothesis quantitatively, we modeled bundle value as a function of the constituent item values with regression analyses. The simplest model involved predicting the value of a bundle as a linear combination of the individual item values (Bundle Value = $\beta_0 + \beta_1$ * Item1 Value + β_2 * Item2 Value), and this model explained a large amount of the variance in bundle WTP ratings (R^2=0.777). Both parameter estimates for the individual items were significantly less than 1 (β_1=0.726, $P < 10^{-10}$, T=6.24; β_2=0.734, $P < 10^{-11}$, T=6.72; Intercept β_0=0.832), consistent with the hypothesis that bundle value is a subadditive function of the individual item values (Figure 3.2b). We additionally tested three nonlinear models, a polynomial model, a power model, and a logarithmic model (see Methods). The power model and logarithmic marginally fit the data better than the linear model, even when controlling for the number of parameters (Table 1). Additionally, these two nonlinear models are concave (Figure 3.2b), thus participants discount their subjective value of a bundle more as the individual item values increase.

We next analyzed how this utility function varies depending on the category of the bundle. The linear model fit to the data of each type of bundle is shown in Figure 3.2c. The slopes of this model are smallest for bundles that are duplicates of the same item (β_1 = 0.615±0.013sem, β_2 = 0.615±0.013sem, Intercept β_0 = 0.987±0.126sem), which is consistent with the economic principle that the utility of an item decreases with each additional unit (diminishing marginal utility). Moreover, food and mixed bundles have larger slopes than same item bundles and trinket bundles (Food: β_1 = 0.742±0.039sem, β_2 = 0.773±0.041sem, Intercept β_0 = 1.073±0.261sem; Mixed: β_{Food} = 0.708±0.047sem, $\beta_{Trinket}$ = 0.710±0.059sem, Intercept β_0 = 0.772±0.255sem; Trinket: β_1 = 0.671±0.062sem, β_2 = 0.633±0.060sem, Intercept β_0 = 0.700±0.191sem).

Altogether, these behavioral results show that the value of a bundle is computed as a subadditive combination of the individual item values. This process is analogous to attribute integration theories of value (citations), which state that the value of a good is computed by integrating the value of the good's component attributes. In this scenario where bundles are evaluated, the attributes that are integrated are the values of the tangible goods that constitute a bundle rather than the values of abstract features that constitute a good.

Neural representation of subjective value and choice

We next investigated how subjective value and choice are encoded in the brain during the choice task. To do so, we constructed a univariate general linear model (GLM) that included regressors for a stimulus's WTP rating time locked to the onset of the trial, and a regressor for the choice made on that trial time locked to the time of choice (in addition to a regressor for trial type and other covariates of no interest; see Methods). Several regions were more active when participants choose an item or bundle vs. when they choose the reference monetary amount, including clusters in dmPFC areas such as superior frontal gyrus (SFG), anterior cingulate gyrus (ACC), vmPFC, and the angular gyrus (P < 0.001 FDR corrected cluster-level, Figure 3.3a). By including the regressor for the WTP of the item/bundle shown in a trial in the same GLM, we were also able to isolate subjective value signals independent of choice. After cluster-level false discovery rate correction (P < 0.001, FDR), one large cluster in the anterior portion of vmPFC and frontal pole showed a positive correlation to subjective value (Figure 3.3b). These results are consistent with previous results (Clithero and Rangel, 2014; Bartra, McGuire, and Kable, 2013), and set the stage for further probing how the representations in these regions are modulated trial type.

A central question about value computations in the brain is whether the same neural resources code for the value of a stimulus independently of the category of the stimulus or the broader context. In the context of this study, this would imply that the same regions compute the value of individual items and the value of bundles. An alternative hypothesis is that additional regions are recruited when computing the value of bundles, since this involves a hierarchical process of valuing the individual items and then integrating them to evaluate the bundle. A previous study identified a spatial topography of category-dependent value codes in which food value was represented in posterior mOFC and consumer good values (trinkets) were represented in anterior mOFC (McNamee, Rangel, and John P O'Doherty, 2013). This result is also consistent with a meta-analysis that reported an anterior versus posterior gradient according to the abstractness or complexity of the reward, with more abstract reinforcers encoded in anterior regions (Kringelbach and Rolls, 2004). The study additionally identified a ventral-dorsal gradient in vmPFC, with ventral OFC regions encoding category-dependent value codes and more dorsal regions encoding a category-independent common currency (McNamee, Rangel, and John P O'Doherty, 2013).

In order to examine whether the brain recruits additional circuitry when evaluating a bundle versus valuing a single item, we tested for the interaction of value and trial type in the previously described GLM (contrasts: bundle value > single item value and single item value > bundle value). At the group level, no clusters survived correction for either comparison, thereby providing no evidence for a topography of separable single item value and bundle value codes at the univariate level in the current dataset (Figure S3).

However, as previously reported neural activity can scale with value similarly across categories while also containing distinct and distributed category-dependent codes in the same region (McNamee, Rangel, and John P O'Doherty, 2013). Since multivariate pattern analysis (MVPA) is more sensitive to picking up these distributed codes, we implemented decoding analyses across conditions to test for the existence of distinct bundle value codes, single item value codes, or general value regions. A ridge regression decoder was trained on distributed voxel patterns in several regions of interest (ROIs) in PFC. The decoder was trained on samples from one trial type (ie. single item trials) for 14/15 runs and tested on the both trial types separately on the held out run (and cross validated with the leave-one-run-out method). Decoder performance was assessed with the Pearson correlation between the predicted values and the true values in the test set. To test for condition-independent general value regions, we analyzed if the decoder could predict the value of samples drawn from the opposite condition from what it was trained on. This would identify general value regions that utilize the same distributed code when computing the value of both single items and bundles. To test for condition-dependent value regions, we analyzed if the decoder could only predict samples within the condition it was trained on, while failing to predict samples from the opposite condition. This would identify distinct single item value regions or bundle value regions.

Value could be predicted above chance on all four types of train/test splits ('train and test on single items,' 'train and test on bundles,' 'train on single items/test on bundles,' 'train on bundles/test on single items') in vlPFC, dmPFC, dlPFC, MFG, and IFG (Figure 3.3c, two-sided one-sample Wilcoxon signed rank test $P < 0.05$ and FDR-corrected for multiple comparisons $q = 0.05$). By being able to decode value across conditions in these regions, this analysis suggests that they possess distributed general value codes that are independent of condition. Additionally, value was decoded above chance for the 'train on single items/test on bundles' partition in rACC, dACC, vmPFC, anterior OFC, lateral OFC, and medial OFC, providing some evidence for generalized value codes in these regions as well (although prediction accuracies were not significant for the reverse partition 'train on bundles/test on single items'). To identify a condition-dependent value region, we assessed if a region's decoder only had significant prediction accuracy when trained and tested on a single condition (ie. only significant for 'train and test on single items' but not for the other three partitions). No region met this criteria. Additionally, there were no significant differences in a decoder's prediction accuracy between conditions for any ROI (two-sided two-sample Wilcoxon signed rank test $P < 0.05$ and FDR-corrected for multiple comparisons $q = 0.05$). For example, although there is a 0.06 difference in pearson correlation between 'train on bundles/test on single items' and 'train and test on bundles' for dACC, this difference is not significant after correcting for multiple comparisons ($P = 0.426$). Thus, the cross decoding analyses do not yield any evidence for the existence of a distributed condition-dependent value code in any ROI.

Normalization of the value code

Thus far we have demonstrated that the brain's value regions represent value in a fashion that generalizes across evaluating individual item values and bundles of those items. A subsequent question emerges from this pattern: does the value code adaptively normalize to trial type? The common currency hypothesis suggests that options are encoded with the same value scale so they can be compared (Padoa-Schioppa and Assad, 2008; Chib et al., 2009; Lin, Adolphs, and Rangel, 2012; Levy and Glimcher, 2011), thereby value regions may code for items and bundles on the same range that scales with their WTP ratings. However, given the biological constraints of neurons, in order to evaluate decisions about which house to buy or which entree to buy with the same valuation system, the brain must also adaptively normalize the value code to the distribution of values available in a decision-making context (Louie, Glimcher, and Webb, 2015). Since the distribution of bundle values is systematically larger than the distribution of single item values (Figure 3.2a), the neural code may normalize to the context of the current trial. A value region could therefore encode a $4 rated item very differently than a $4 rated bundle if options are appraised based on their utility relative to the other options within a condition (Figure 3.4a). Relative value coding has been observed empirically in experiments where a different distribution of values is presented in each block (Padoa-Schioppa, 2009; Louie, Grattan, and Glimcher, 2011), but crucially in this experiment, bundle trials and single item trials are randomly intermixed within a block.

To test whether the neural representation codes for value in absolute or relative terms, we constructed GLMs with different value regressors. One GLM included a regressor for value according to the WTP bid in dollars for the item or bundle displayed in that trial, while in a separate GLM value was normalized (with a z-score) by trial type. Value contrasts were then compared at the second level (normalized value > absolute value and absolute value > normalized value). A significant cluster in vmPFC emerged for the normalized value > absolute value comparison (peak voxel: x=-2, y=62, z=-7; t_{13}=5.75; $P < 0.001$ uncorrected cluster forming threshold followed by $P < 0.001$ FDR cluster-level correction), suggesting that the neural representation of value in vmPFC is normalized. No significant clusters emerged after correction for the opposite analysis (absolute value > normalized value).

An alternative explanation for this result is that vmPFC is computing the difference in value between the item/bundle presented in a trial and the reference monetary amount it is choosing against. This computation would also produce a relative value code, but is a simpler form of adaptation that does not put the single item value and bundle value distributions on the same scale. Thus, another GLM was built with a modified value regressor equal to the WTP bid minus the reference monetary amount, and the resulting value contrast was tested against the fully normalized value contrast. Again, a significant cluster in vmPFC resulted from the normalized value > value difference contrast (peak voxel: x=0, y=52, z=-14; t_{13}=4.83; $P < 0.001$ uncorrected cluster forming threshold followed by $P < 0.005$ FDR cluster-level correction), and the opposite contrast yielded no significant clusters. These results demonstrate that at least at the univariate level, vmPFC normalizes the value code by condition.

To test the same hypothesis at the multivariate level, we completed another cross decoding analysis. With ridge regression, decoders were trained on 14/15 runs with one condition (single item trials or bundle trials) and tested on samples from both conditions in the held out run (with leave-one-run-out cross validation). The held out predictions are then compared (with pearson correlation) to the true value labels in their absolute value WTP < amount and their relative value after normalizing by condition. If the decoder systematically predicts that the bundle trial samples in the test set are higher value than the single item samples in the test set, the distribution of these held out predictions will look similar to an absolute value code as in the left histogram of Figure 3.4a. The held out predictions would thus have a higher correlation to the absolute value representation. In contrast, if the region tested normalizes value, the decoder's predictions will place the bundle samples and single item samples in the same distribution (Figure 3.4a right), and these held out predictions will have a higher correlation to the relative value code. The results demonstrate that the normalized value representation better matches the neural data, as predictions were significantly more correlated to the relative value code in every ROI tested (two-sided Wilcoxon signed rank tests, $P < 0.05$ and FDR-corrected for multiple comparisons ($q = 0.05$).

To ensure that these results were not just due to the parametric nature of the ridge regression decoder, we additionally performed representational similarity analysis (RSA) to assess whether the representational geometry in these regions was more reflective of an absolute value or relative value code. RSA provides a more data-driven method to examine the structure of neural representations by constructing dissimilarity matrices (DSMs) according to how the multidimensional voxel space changes from trial to trial or from condition to condition (Kriegeskorte, Mur, and Bandettini, 2008). These data-driven DSMs can then be compared to model DSMs that encode how the features of the task, such as value, evolve from trial to trial. Here, trial by trial DSMs were built for every ROI and correlated to model DSMs for absolute value and relative/normalized value. Similarly to the decoding analysis, the relative value DSM had significantly higher correlations to the neural DSMs in every ROI (two-sided Wilcoxon signed rank tests, $P < 0.05$ and FDR-corrected for multiple comparisons ($q = 0.05$). Altogether these analyses show that these PFC regions encode value in a normalized fashion, both at the univariate level and distributed multivariate level.

Discussion

In real-world decision-making, consumers often have to make choices between options that are each made up of multiple goods. However, it is unknown how the human brain constructs the value of bundles of multiple items. To investigate this question, we used a BDM auction procedure to elicit subjective values for food items, noncomestible consumer goods, and bundles of these items. Then participants made choices with these items and bundles during fMRI scanning.

WTP behavior could be predicted by modeling bundle value with an weighted integration model that uses the individual item values as attributes. It was found that bundle value is computed with a subadditive function of the individual item values. With parameter estimates of a linear model that are significantly less than 1, bundles are systematically discounted in relationship to the sum of the individual item values. Concave nonlinear models additionally capture this behavior, and this concavity suggests that participants discount their subjective value of a bundle more as the individual item values increase.

These results are are compatible with the expectation most consumers have that bundles of multiple goods are usually discounted in comparison to the total price of the constituent products purchased separately (Thorne, 2004; Nagle and Holden, 1987). Value meals at fast food restaurants, snack variety packs, vacation packages, and season tickets for sports teams all represent discounted bundles in the real-world marketplace. However, the incentives of buyers and sellers in the marketplace are usually different than the one in our experimental setup. Sellers offer bundles at a discount to increase the probability of a buyer purchasing an additional good they would not have purchased otherwise, similarly to discounting the purchase of the same item in bulk. For example, a McDonald's customer always has the optionality to buy a hamburger and fries separately, so a bundle of the two needs to be discounted in order to become an attractive package. In our experiment, each item and bundle is bid on separately and a BDM auction trial is selected at random, so when evaluating a bundle, participants do not have the optionality usually afforded to them in the marketplace to purchase them separately. Therefore, the results of our experiment suggests that bundle discounting is a more general phenomenon of multi-item valuation and not just the result of market equilibrium prices. This pattern is analogous to the law of diminishing marginal utility, which describes how one receives a reduction in utility per unit for each additional unit consumed of a good. In our experiment, bundles of the same item were discounted slightly more than the other bundles, which suggests that bundles of different items are discounted similarly to bulk quantities of the same item but not as much.

Microeconomic theory also accounts for the possibility that bundles of goods are substitutes or complements (Varian, 2014). Substitutes have interchangeable functions, and therefore are often discounted when bundled together. For example, tea and coffee could be viewed as substitutes as they both over a caffeine boost. A consumer typically does not buy both at cafe because of this, and their utility bundled together is likely less than the sum of their independent utilities to the average consumer. Complements are products which are typically bought and used together, such as pasta and pasta sauce. Other examples, like left and right shoes, can exhibit superadditivity, where the value of the left shoe by itself is much lower than half of the value of both shoes together. We did not optimize our experiment to include substitutes or complements, but it would be interesting to pick an item set that probes how substitutes and complements are evaluated in future work.

At the neural level, we tested for condition-dependent and condition-independent value signals at the univariate and multivariate levels. Previous research identified a spatial topography in OFC where food value was represented in posterior mOFC and value codes for consumer goods were represented in anterior mOFC (McNamee, Rangel, and John P O'Doherty, 2013). However, we found no evidence for separate single item value regions or bundle value regions or a topography of value complexity. An anterior portion of vmPFC showed a correlation to subjective value across both individual item trials and bundle trials. In a cross decoding analysis similar to methods previously used to identify category-dependent and category-independent value codes (McNamee, Rangel, and John P O'Doherty, 2013), a decoder was trained on samples from one trial type and tested to predict value for the other trial type. Distributed condition-independent value codes were revealed throughout PFC, including in dmPFC, vlPFC, and MFG. Although these regions encode value in a general way across trial types, they may not be implementing a common currency as previously suggested (Levy and Glimcher, 2012). A true common currency would map all stimuli to the same shared scale. Our results show that value is represented by the same distributed voxel patterns across all trial types, but this value code normalizes to the context and does not encode absolute value through a common currency representation. The possibility remains that general value regions encode a common currency for all stimuli within a context (Padoa-Schioppa and Assad, 2008), but context-dependent modulation of value has also been observed when choices are made between stimuli of the same type (Louie, Grattan, and Glimcher, 2011).

However, this project can only provide a coarse grained snapshot into the mechanism of normalization operating in these regions due to the limited spatial and temporal resolution of fMRI. Value can be encoded by a relative code with many different types of normalization (Miller and MacKay, 1994; Louie, Khaw, and Glimcher, 2013). Since vmPFC activity was more correlated to the normalized value regressor (via z-score) than the value difference regressor, our results are more consistent with divisive normalization models than subtractive normalization models. Electrophysiological studies of bundle valuation are needed to shed more light on the computations neural populations implement to rescale the value code, similarly to what has already been done in other neural circuits. (Carandini and Heeger, 2012).

Methods

Participants

Participants (N=14) were recruited from the general population through the Caltech Brain Research Participant System (7 females, 7 males, 24.9 ± 3.74 years, mean ± s.d.). They did not have any food allergies and were not dieting at the time of the experiment. They were given a participation fee of $40 ($20 per hour), in addition to receiving monetary, food, and non-comestible consumer goods as rewards depending on their choices in the experiment. Each subject gave their informed consent, and the study was approved by the Institutional Review Board of the California Institute of Technology.

Stimuli

Across participants, 70 food items and 40 non-comestible consumer goods were used as stimuli in the experiment. Food items included fruits, snacks, and mains (including microwaveable meals) that are available at local grocery stores. Consumer goods included a diverse array of items under $40 in price, including cell phone chargers, kitchen items, Caltech memorabilia, and books. Many of these items have been used in previous studies (Suzuki, Cross, and O'Doherty, 2017). The full list of items can be found in the Supplementary Table 1.

Experimental tasks

Participants performed the experiment in three sessions on three separate days in order to maximize the amount of fMRI data within-subject. On each day, participants first performed a willingness-to-pay (WTP) task outside the scanner, then performed the choice task in the scanner where they were asked to choose between an item or bundle of two items versus a reference monetary amount. Participants were asked to refrain from eating 4 hours before the experiment in order to ensure that food items were valuable to them. Compliance was confirmed by self-reports.

Each day, a participant performed these tasks with 20 items, 10 food items and 10 non-food items. 10 of these items were presented in all three days, and 10 new items were introduced every day, adding up to a total of 40 unique items presented throughout the experiment. To construct bundles, every item was paired with each other, including pairs of the same items. Thus, 210 bundles were included each day (20 choose 2 = 190 + 20 pairs of the same item). On the final day of the experiment after the choice task, outside the scanner participants rated how familiar they were with each of the 40 items. For each item, the participant indicated their familiarity with the item on a continuous scale from 'not at all' to 'very much' by moving a red pointer, with no time constraint, as done in the previous experiment presented in Chapter 1 (Suzuki, Cross, and O'Doherty, 2017).

WTP Task (outside the MRI scanner)

Participants completed an untimed BDM auction task to measure their willingness-to-pay for items and bundles, with a procedure similar to previous studies, including the task presented in Chapter 1 (Chib et al., 2009; McNamee, Rangel, and John P O'Doherty, 2013; Suzuki, Cross, and O'Doherty, 2017). The BDM auction is a reliable incentive-compatible method to elicit subjective values for items (Becker, DeGroot, and Marschak, 1964). Participants were endowed with a $20 budget in cash, and instructed that they can use this cash to purchase items from our laboratory store (and keep the money they do not use). In each trial, an item or bundle was shown and the participant was asked to type in how much they would be willing to pay from $0-$20 for that item/bundle (Figure 3.1a). Participants first bid on the individual items (20 each day) and then bid on the bundles (210 each day).

Each trial was to be treated independently, as a random trial from the entire experiment was selected at the end of the experiment. If the selected trial was from the WTP task, the participants' bid on that trial was then compared against a randomly generated price (uniform probability from $0-$20), and if their bid was greater than or equal to that price, they received the item(s) and paid the corresponding price with their $20 budget. If their bid was less than the price, they did not receive the item(s), and they did not have to pay anything. Participants were explicitly instructed about this auction procedure, and about how the optimal strategy is to bid their true subjective value for a given item/bundle. With a questionnaire, we confirmed that participants understood the mechanism of the auction.

Choice Task (inside the MRI scanner)

In the scanner, participants made choices involving the items and bundles previously bid on in the WTP task. Each trial involved a binary choice between an item or bundle and a reference monetary amount (Figure 3.1b). The reference monetary amount was equal to the participant's median WTP bid for that category (individual items were chosen against the median bid on individual items and bundles were chosen against the median bid on bundles). This ensured that the participants would choose the reference monetary amount about half the time and choose the item or bundle half the time. For a trial, the stimulus appeared in the middle of the screen, and the word 'ITEM' appeared on the bottom left or right with equal probability while the reference monetary amount in '$X' appeared on the other side. Participants selected their choices with a button box, with the leftmost button indicating choosing the left option and vice versa for the rightmost button. Participants had 5 seconds to make a decision, after which their choice was presented on the screen for 0.5 seconds and followed by a jittered intertrial interval (ITI phase, 2–7s).

On each of the three days, participants were scanned for five runs. Each run included 62 trials, where each of the 20 individual items was presented once per run, and each of the 210 bundles in a day was presented once on that day. On day 1, anatomical scans were collected after the choice task.

fMRI Data Acquisition

The fMRI data was acquired on a Siemens Prisma 3T scanner at the Caltech Brain Imaging Center (Pasadena, CA) with a 32-channel radio frequency coil. At the end of the first day of scanning, T1 and T2 weighted anatomical high-resolution scans were collected with 0.9mm isotropic resolution.

High resolution data

A high-resolution partial volume slab was collected in three participants with a 1.5mm isotropic voxel size (Figure S1) and the following parameters: multiband acceleration = 4, 64 slices, TR = 1100ms, TE = 26ms, flip angle = 63°, FOV= 192mm x 192mm, in-plane GRAPPA (R = 2), echo spacing = 0.68ms. The protocol is optimized to view mPFC in high-resolution and therefore the partial volume slab cuts off portions of the motor cortex and parietal lobe. EPI-based fieldmaps of positive and negative polarity were also collected before each run with similar parameters as the sequence.

Standard resolution data

A wholebrain multiband echo-planar imaging (EPI) protocol was collected in eleven participants with a 2.5 mm isotropic voxel size and the following parameters: multiband acceleration = 4, 72 slices, TR = 1120ms, TE = 30ms, A-P phase encoding, -30 degrees slice orientation from AC-PC line, flip angle = 54°, FOV= 192mm x 192mm. EPI-based fieldmaps of positive and negative polarity were also collected before each run with similar parameters as the sequence.

fMRI Preprocessing

Data was preprocessed using a standard pipeline for preprocessing of multiband data. Using FSL (Smith et al., 2004), images were brain extracted, realigned, high-pass filtered (100s threshold), and unwarped. Images were denoised by ICA component removal. Components were extracted using FSL's Melodic, classified into signal or noise with a classifier trained on separate datasets or manually classified for the high-resolution dataset. T2 images were aligned to T1 images with FSL FLIRT, and then both were normalized to standard space using ANTs (using CIT168 high resolution T1 and T2 templates (Avants et al., 2009; Tyszka and Pauli, 2016)). The functional data was first co-registered to anatomical images using FSL's FLIRT, then registered to the normalized T2 using ANTs. For univariate analyses, data was spatially smoothed in FSL with a 5-mm FWHM Gaussian kernel. For multivariate analyses, data was spatially smoothed with a 2-mm FWHM Gaussian kernel.

Behavioral Analyses

Linear and nonlinear regression analyses were performed to model how bundle value is computed as a function of the values of the constituent items in the bundle ($BundleValue = f(ItemValues)$). The linear model represents bundle value as a linear combination of the individual item values:

$$BundleValue = \beta_0 + \beta_1 * Item1Value + \beta_2 * Item2Value$$

Item 1 and item 2 simply correspond to the item shown on the left and right, respectively, during stimulus presentation. A mixed-effects model was estimated across all subjects, with subject-specific random effects terms for intercept and slope. Nonlinear mixed effects models were also constructed:

Polynomial: $BundleValue =$
$$\beta_0 + \beta_1 * ItemValue1^2 + \beta_2 * ItemValue1^3 + \beta_3 * ItemValue2^2 + \beta_4 * ItemValue2^3$$

Power: $BundleValue = \beta_0 + \beta_1 * ItemValue1^{\beta_2} + \beta_3 * ItemValue2^{\beta_4}$

Logarithmic:
$$BundleValue = \beta_0 + \beta_1 * log(\beta_3 + ItemValue1) + \beta_4 * log(\beta_5 + ItemValue2).$$

All models and model statistics were estimated with Matlab, with the fitlme and nlmefit functions, and model fits were evaluated with BIC and R^2 (Table 1). To plot the fitted line/curve for each model in Figure 3.2b, bundle value was computed with the fitted function parameters with ItemValue1 set equal to ItemValue2 for all values from 0-10 (with a 0.01 step size). The sum of these values (double of the value set both for ItemValue1 and ItemValue2) is represented by the x-axis.

The linear models were also separately fit on data from each type of bundle, food, trinket, mixed, and duplicates of the same item. Due to a small amount of data per subject, random effects terms could not be properly estimated for bundles of the same item, and therefore only a fixed effects model was estimated.

Regions of interest

Regions of interest (ROIs) were defined using the AAL database (Tzourio-Mazoyer et al., 2002). The labels used in the paper are mapped to the original ROI names as follows: ACC_pre: rACC, ACC_sup: dACC, Frontal_Inf_Orb_2: vlPFC, Frontal_Med_Orb: vmPFC,OFCant: OFCant, OFClat: OFClat,
OFCmed: OFCmed, OFCpost: OFCpost , Frontal_Sup_Medial: dmPFC,
Frontal_Sup_2: dlPFC, Frontal_Mid_2: MFG, Frontal_Inf_Tri: IFG.

Univariate Analyses

Univariate analyses were conducted in SPM12. General linear models (GLMs) were constructed to examine how subjective value and choice are encoded in the brain during the choice task. This GLM included a regressor for value time locked to the onset of the trial as a parametric modulator, which was modified in separate GLMs to test hypotheses about value representation. The stimulus onset regressor additionally had another parametric modulator for trial type (-1s for single item trials and 1s for bundle trials). Thus, a contrast corresponding to the representation of bundle value or single item value could be computed as the interaction between value and trial type. Reaction times were additionally included as a third parametric modulator. Regressor for the choice made on a trial was time locked to the time of choice, with separate regressors for choosing and item/bundle and choosing the reference monetary amount. Regressors of no interest were included: left and right button presses (duration=0), motion regressors, and run. Missed trials were not modeled. Data across all three days in a subject were included in the same model, with different days and separate regressors per day entered as different sessions.

Different representations of value were included in the value regressor. The value contrast in Figure 3.3b used a normalized version of value, where value was z-scored by trial type (item or bundle). The absolute value model uses the raw WTP $ amount that the subject rated an item/bundle outside the scanner. WTP minus reference (value difference) used WTP minus the reference monetary amount on that trial.

Multivariate Analyses (MVPA)

To examine the nature of the bundle value code as the fine-grained distributed level, we implemented MVPA and RSA analyses. All analyses used the PyMVPA toolbox (Hanke et al., 2009), Scikit-learn functions, and custom Python code.

MVPA samples

To prepare for MVPA and RSA, we extracted trial by trial voxel-wise fMRI responses. A GLM was designed for each participant that modeled the onset and duration of each trial separately to extract the voxel responses that were unique to each trial. Other regressors of no interest modeled the other events and were not separated by trial (one regressor across an entire day): outcome phase (onset and duration), left and right button presses (duration=0). As in the univarate analyses, data across all three days in a subject were included in the same model, with different days and separate regressors per day entered as different sessions. After estimating the models, the parameter estimate maps (beta maps) for each trial were concatenated into a 4D dataset, with length equal to the number of trials a subject completed across the experiment.

Cross Decoding Analysis

To test for distinct bundle value codes, single item value codes, and general value regions, we implemented a cross decoding analysis similar to previously used methods (McNamee, Rangel, and John P O'Doherty, 2013) at the ROI level. The 4D dataset of beta maps for every trial were loaded with PyMVPA functions. Value was z-scored by trial type (item or bundle) and included as the targets to predict in the PyMVPA dataset. Then the voxels for each ROI were entered as features for the cross decoding procedure. Ridge regression was used as the decoder in all analyses (Scikit-learn's, linear_model.Ridge), with alpha = 10^3 (this parameter was optimized with sweeping). Each run was used as a separate cross-validation fold. Decoders were trained on the training samples from one trial type at a time and tested on samples from the held out run (leave-one-run-out cross-validation). Decoders were tested on the samples from both trial types in the test set, even though they were trained on one type. However, decoder predictions were quantified separately for individual item samples in the test set and bundle samples in the test set. This ensured that performance could be compared across trial type separately from within trial type. Prediction accuracy was quantified by the Pearson correlation between the predictions and true value labels and averaged across cross-validation folds. This resulted in four decoder prediction accuracy metrics for each ROI and subject according to the four train/test splits: 'train and test on single items,' 'train and test on bundles,' 'train on single items/test on bundles,' 'train on bundles/test on single items'. The average prediction accuracy across participants for each of these train/test splits is plotted in Figure 3.3c. Significance was assessed for each split vs. chance level (r=0) with two-sided one-sample Wilcoxon signed rank tests at $P < 0.05$ and FDR-corrected for multiple comparisons (ROIs) at $q = 0.05$. To assess if a decoder's prediction accuracy was significantly different between conditions in the test set, we used two-sided two-sample Wilcoxon signed rank tests at $P < 0.05$ and FDR-corrected for multiple comparisons (ROIs) at $q = 0.05$.

Cross Decoding — Normalization of the value code

A similar cross decoding analysis was performed to test whether the neural representation codes for value in absolute or relative terms at the multivariate level. Ridge regression decoders were trained on 14/15 runs in one condition as described above. The only difference in the decoder procedure is that absolute value codes were used as targets rather than the relative/normalized values used in the first cross decoding analysis. The decoder is then tested on the samples from both trial types in the test set. In this analysis the decoder's is tested on both the individual item samples and bundle samples at the same time, in order to test if the decoder ranks the samples across trial types in an absolute or relative fashion. This procedure thus outputs predictions on all trials in the held out run, and these predictions are correlated to 1. the true labels in absolute value (WTP $ amount) and 2. the true labels normalized by condition (relative value). These prediction accuracy metrics were then averaged across participants and plotted in Figure 3.4d. Differences in prediction accuracy between absolute value and relative value were tested against chance with a nonparametric version of the Paired T-test, the two-sided two-sample Wilcoxon signed rank test $P < 0.05$ and FDR-corrected for multiple comparisons (ROIs) ($q = 0.05$).

Representational Similarity Analysis (RSA)

To examine whether the representational geometry of the regions of interest correlated to an absolute value or relative value code, we conducted representational similarity analysis (RSA). The 4D dataset of beta maps for every trial were loaded with PyMVPA functions as in the cross decoding analysis. Trial by trial neural dissimilarity matrices (DSMs) were constructed for every ROI by computing pairwise comparisons of the beta map across trials with PyMVPA's PDist function. Euclidean distance was used as the distance metric. Two model DSMs were constructed: 1. trial by trial pairwise distances according to the difference in absolute $ value of the stimuli between trials 2. trial by trial pairwise distances according to the difference in normalized value of the stimuli between trials (where value was z-scored by trial type (item or bundle)). For all DSMs, within day comparisons were removed to avoid potential confounds due to similarity being driven by patterns being in the same run or day. Neural DSMs and model DSMs were then compared with Pearson correlations. Differences in the correlations between absolute value and relative value were tested against chance with a nonparametric version of the Paired T-test, the two-sided two-sample Wilcoxon signed rank test $P < 0.05$ and FDR-corrected for multiple comparisons (ROIs) ($q = 0.05$).

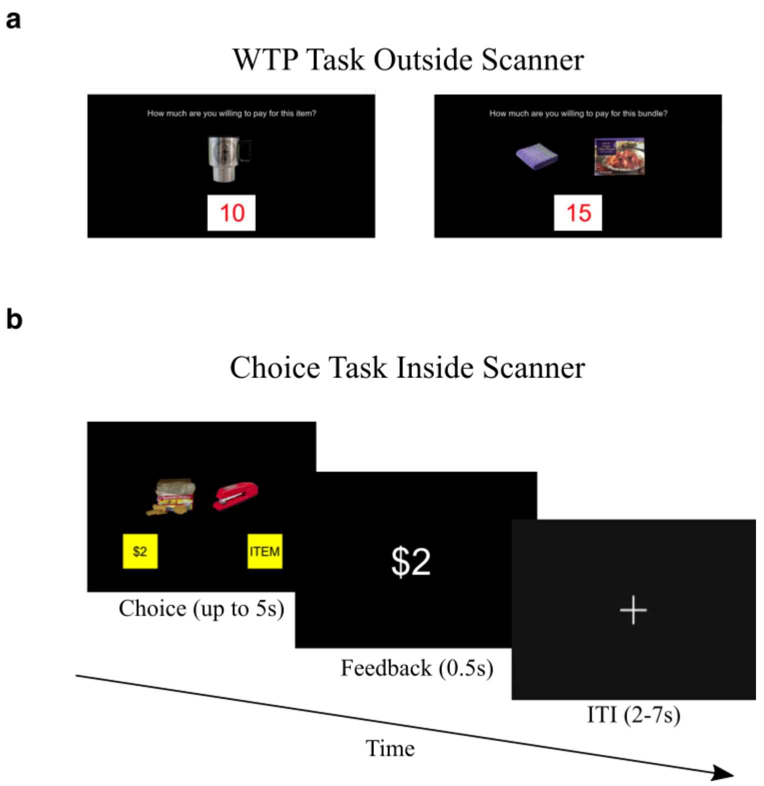

a
WTP Task Outside Scanner
How much are you willing to pay for this item?
10
How much are you willing to pay for this bundle?
15
b
Choice Task Inside Scanner
$2
ITEM
Choice (up to 5s)
$2
Feedback (0.5s)
ITI (2-7s)
Time

Figure 3.1: **Experimental Design**
a. WTP task. Participants reported how much they would be willing to pay for items and bundles of two items in a BDM auction. This was untimed outside the scanner.
b. Choice task. Inside the scanner, participants made choices with the items and bundles. During single item trials, a choice was made between an item and a reference monetary amount equal to the median bid of single item trials in the WTP task. Similarly during bundle trials, a choice was made between a bundle and a reference monetary amount equal to the median bid of WTP bundle trials. Participants had up to 5s to make a choice indicated by a right or left button press. The experiment involved 3 days of scanning 5 runs of the task for a total of 15 runs.

Model	R^2	BIC
Linear	0.7771	39151
Polynomial	0.7722	39510
Power	0.7797	39049
Logarithmic	0.7777	39097

Table 1. **Behavioral models of bundle value**

R^2 and BIC scores reflect fit across participants. Each model fits random effects parameters per participant. See Methods for model details.

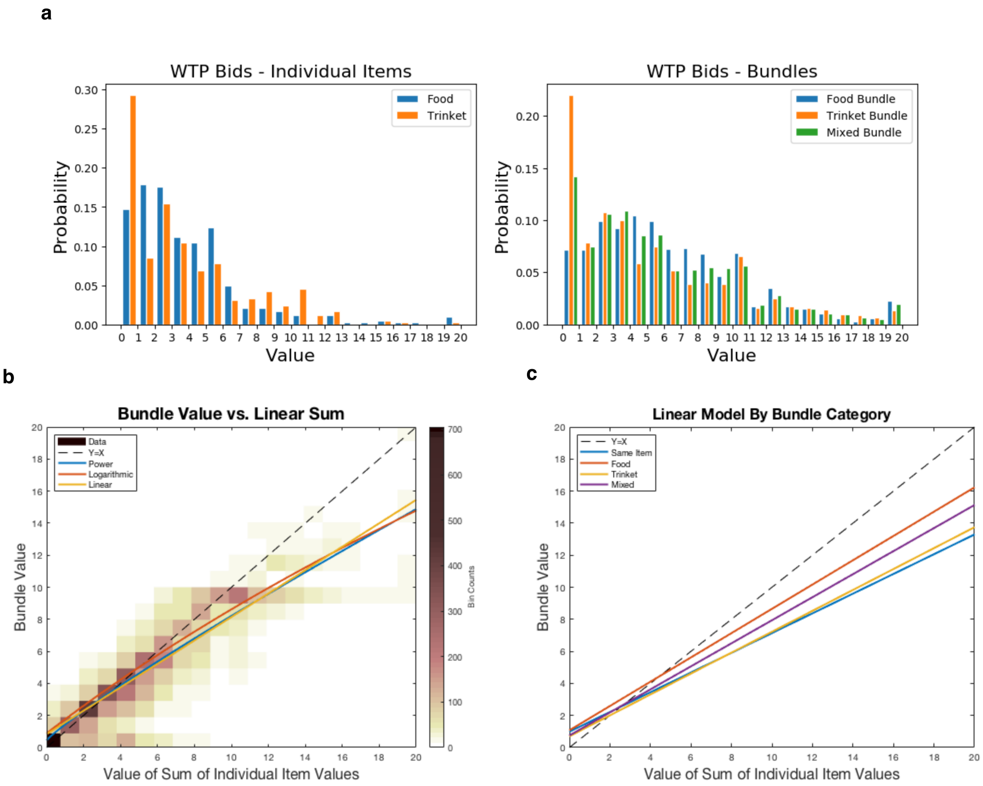


Figure 3.2: **WTP Behavior**

a. Histograms of WTP bids for individual items and bundles across all subjects.

b. Density plot of the WTP value of a bundle vs the sum of the values of the constituent items in a bundle in all subjects. If bundle value was equal to a linear addition of the constituent item values, bundle values would lie along the diagonal Y=X. Three models were constructed to predict the value of a bundle as a function of the individual item values: a linear model, a nonlinear power model, and a nonlinear logarithmic model. The fitted curves for each model are plotted along the density plot. All three models display that bundle value is a subadditive function of the individual item values, as they extend below the Y=X line.

c. The fitted linear model stratified by bundle type.

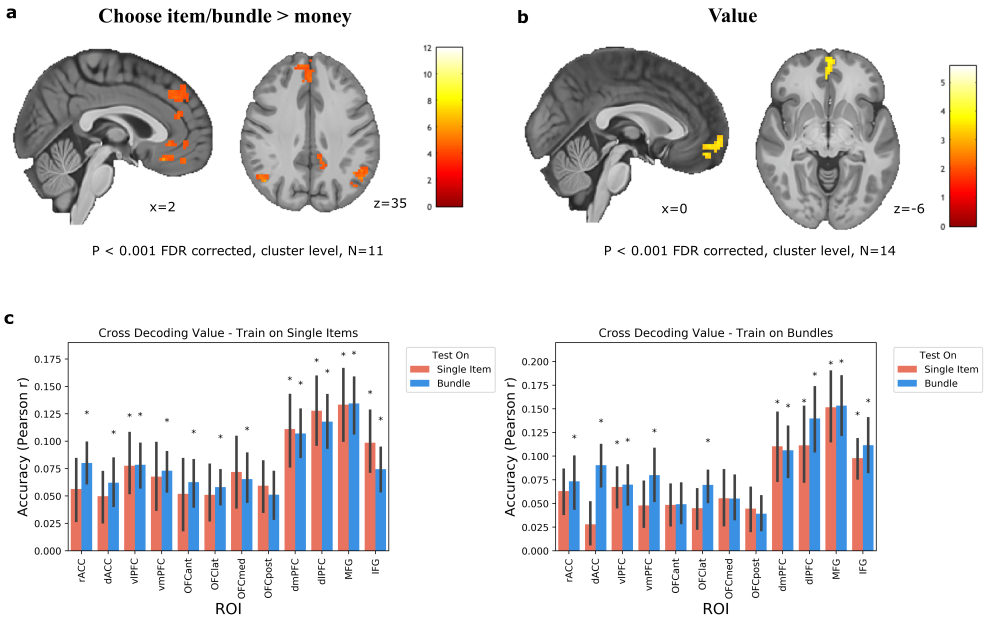


Figure 3.3: **Neural representation of subjective value and choice.**

a. Areas more active when the item or bundle was chosen than when the reference monetary amount was chosen. Significant clusters in dmPFC/SFG, ACC, vmPFC, and angular gyrus (N=11). Clusters are defined by a $P < 0.001$ uncorrected cluster forming threshold followed by $P < 0.001$ FDR cluster-level correction. Second level result shown without high-resolution subjects due to the partial volume high resolution scan's limited coverage in dorsal frontal areas such as SFG.

b. Neural correlates of the value of the item/bundle presented. Significant cluster in anterior vmPFC (N=14). Clusters are defined by a $P < 0.001$ uncorrected cluster forming threshold followed by $P < 0.001$ FDR cluster-level correction.

c. MVPA cross decoding analysis results. Ridge regression decoders were trained on samples from one condition and tested on samples from a held out run in both conditions. Left: decoders trained on trials of single items. Right: decoders trained on bundle trials. Asterisks * represent significant prediction accuracies on a test partition (two-sided one-sample Wilcoxon signed rank test $P < 0.05$ and FDR-corrected for multiple comparisons $q = 0.05$). There were no significant paired differences in prediction accuracies between test conditions for any ROI (two-sided two-sample Wilcoxon signed rank test $P < 0.05$ and FDR-corrected for multiple comparisons $q = 0.05$). Error bars reflect SE across participants.

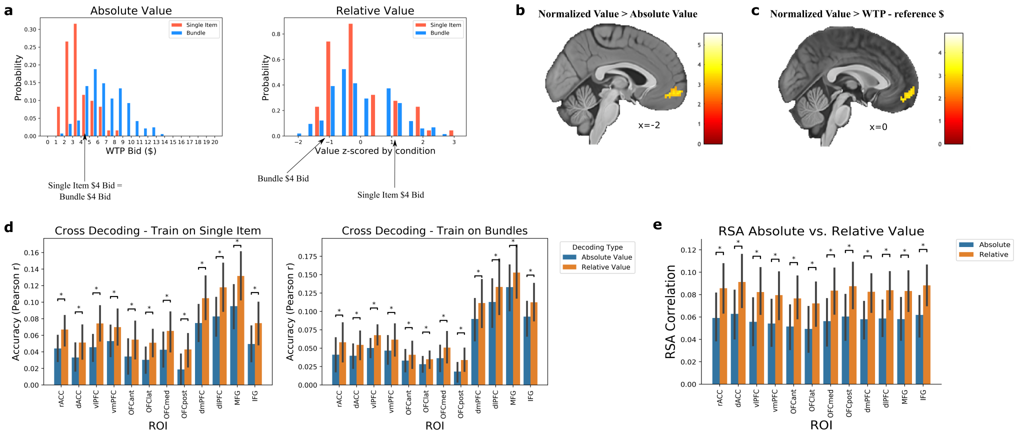


Figure 3.4: **Normalization of the value code.**

a. A depiction of the distributions of value by condition with absolute value and relative value codes. An absolute value code (left) represents value according to the participants' WTP bids, an incentive compatible measure of the subjective value of an item or bundle. A relative value code normalizes value by condition, plotted here after z-scoring individual item values and bundle values separately (left). A relative code adapts to the context of the task and puts values from different distributions on the same scale.

b. Contrast normalized value > absolute value. A cluster in vmPFC is better fit by the normalized value model, indicating that the value representation in this region is normalized. Clusters are defined by a P < 0.001 uncorrected cluster forming threshold followed by P < 0.001 FDR cluster-level correction.

c. Contrast normalized value > value difference, with a cluster in vmPFC emerging similarly to B. Value difference is the WTP bid $ amount - the reference $ amount. P < 0.005 FDR cluster-level corrected (P < 0.001 uncorrected cluster forming threshold).

d. Cross decoding analysis results with absolute and relative value. Results depicted for 12 ROIs in PFC. Ridge regression decoder is trained on one condition (single items: left; bundles: right) and tested on both conditions in a held-out run with leave-one-out cross validation. Depending on how the decoder ranks the held out samples in both conditions, predictions will yield a higher correlation to the absolute or relative value code. Predictions were more correlated to a relative value code for all ROIs, significant two-sided Wilcoxon signed rank test P < 0.05 and FDR-corrected for multiple comparisons (q = 0.05). Error bars reflect SE across participants.

e. Representational Similarity Analysis (RSA) results comparing absolute value and relative value DSMs to neural DSMs in each ROI. Neural DSMs show a significantly higher correlation to the relative value DSM in each ROI with two-sided Wilcoxon signed rank tests P < 0.05 and FDR-corrected for multiple comparisons (q = 0.05). Error bars reflect SE across participants.

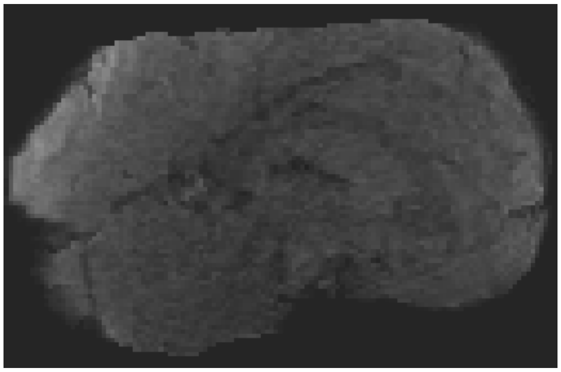


Figure 3.5: **Supplementary Figure 1. High Resolution Sequence**
High-resolution partial volume scans were collected in three participants with a 1.5mm isotropic voxel size.

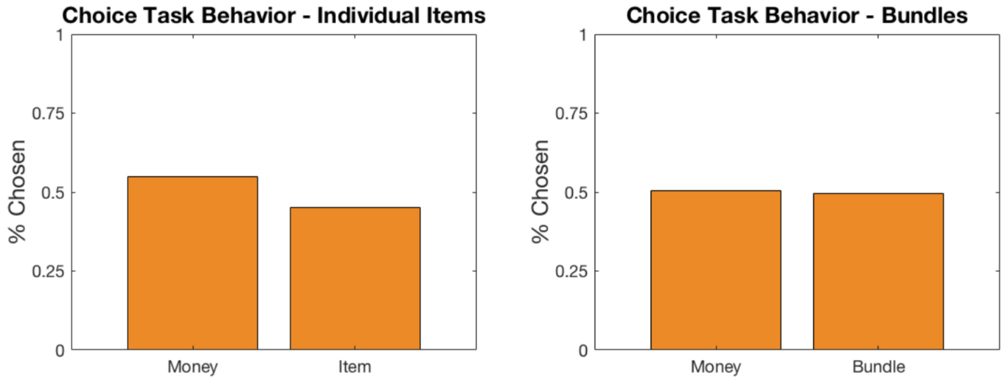
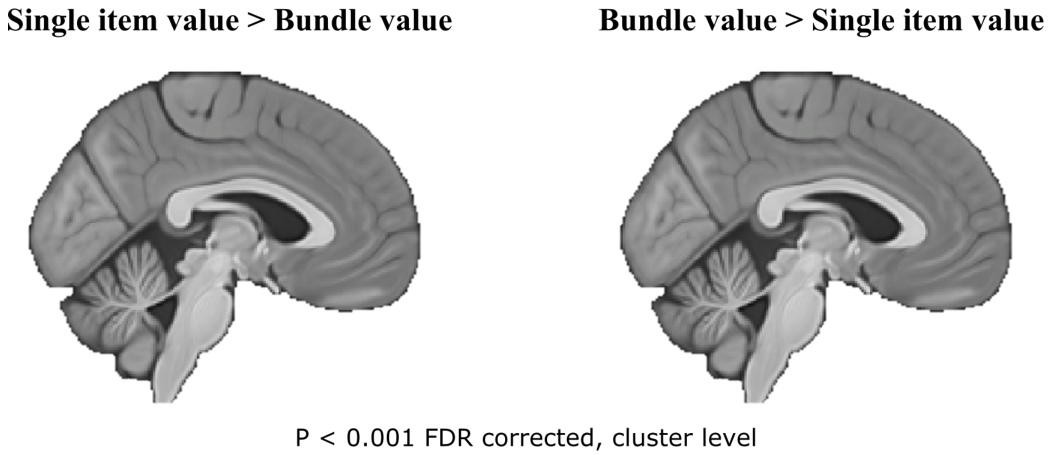


Figure 3.6: **Supplementary Figure 2. Behavior on the choice task.**
Percentage of trials in which the item or bundle was chosen vs. the reference monetary amount.



Single item value > Bundle value

Bundle value > Single item value

P < 0.001 FDR corrected, cluster level

Figure 3.7: **Supplementary Figure 3. Bundle Value vs Single Item Value.** Univariate contrasts testing the interaction of value and trial type. No clusters survived in either comparison after multiple comparisons correction.

Food items used	
1. 3 Musketeers	36. Sun Chips
2. Barnum’s Animal Crackers	37. Dole Mixed Fruit
3. Doritos Nacho Cheese	38. Grapefruit
4. Chips Ahoy!	39. Banana Chips
5. Kit Kat	40. Dark Chocoloate Bananas
6. Pop-Tarts Brown Sugar Cinnamon	41. Crispy Apple
7. Pop-Tarts Brown Sugar Strawberry	42. Vegetable Chips
8. Ghiradelli Chocolates	43. Sweet Potato Chips
9. Twix Cookie Bars	44. Chopped Salad Chicken
10. Hershey’s Whatchamacallit Candy	45. Mexicali Salad
11. Apple Pie	46. Caesar Salad
12. Avocado	47. Veggie Wrap
13. Blackberries	48. Super Burrito
14. Cauliflower	49. Chocolate and Berry
15. Ritz Crackers’n Cheese Dip	50. Green Beans Chips
16. Cherry Pie	51. Salami
17. Chocolate Muffins	52. Smoked Turkey
18. Powdered Donuts	53. American Cheese
19. Granny Smith Apple	54. Chicken and Roasted Beet
20. Green Grapes	55. Mozzarella Cheese
21. Mango	56. Roast Beef
22. Milano Cookies	57. Caprese Sandwich
23. Orange	58. Tuna Salad Wrap
24. Raspberries	59. Smoked Salmon
25. Red Velvet Cake	60. Plain Yogurt
26. Quaker Chewy Granola Bar	61. Strawberry Yogurt
27. Starburst	62. Blueberry Yogurt
28. Strawberry	63. Deviled Eggs
29. Crunchy Donuts	64. Smore’s Chewy Bars
30. Chicken Tikka Masala	65. Gnocci
31. Lamb Vindaloo	66. Magherita Pizza
32. Pollo Asado Burrito	67. Macarons
33. Bean and Cheese Burrito	68. Blueberry Crisp Clif Bars
34. Chocolate Chip Clif Bars	69. Yogurt Pretzels
35. Ferrero Chocolates	70. Chocolate Pretzels
Consumer goods used	
1. A Brief History of Time book	21. Lock
2. Freakonomics book	22. Notebook
3. 1984 book	23. Bathroom scale
4. Water bottle	24. Playing cards
5. Wireless mouse	25. Honey clementine candle
6. Yoga mat	26. Roses candle
7. Hitchhikers book	27. Umbrella
8. Lord of the Rings book	28. Android charger
9. Caltech backpack	29. iPhone charger
10. Caltech hat	30. Clothes hangers
11. Caltech banner	31. Beach towel
12. Caltech keychain	32. Cooking supplies
13. USB stick 16GB	33. Kitchen utensils
14. Caltech mug	34. Pens
15. Caltech drawstring bag	35. Plates
16. Desk lamp	36. Portable charger
17. Stapler	37. Portable speaker
18. Over the ear headphones	38. Screwdrivers
19. Head backpack	39. Sunglasses
20. Batteries	40. Surge Protector

Table 3.1: Items used in experiment.

Chapter 4

USING DEEP REINFORCEMENT LEARNING TO REVEAL HOW THE BRAIN ENCODES ABSTRACT STATE-SPACE REPRESENTATIONS IN HIGH-DIMENSIONAL ENVIRONMENTS

Abstract

Humans possess an exceptional aptitude to efficiently make decisions from high-dimensional and noisy sensory observations in the real-world. However, it is largely unknown how the brain compactly represents the current state of the environment to guide this decision-making process. Deep reinforcement learning algorithms, such as the deep Q-network (DQN) (Mnih et al., 2015) solve this problem by using a deep neural network to capture highly nonlinear mappings from multivariate inputs to the values of potential actions. We deployed DQN as an end-to-end model of brain activity and behavior in participants playing three Atari video games during fMRI scanning. We found that stimulus features in hidden layers of the DQN agent exhibit a striking resemblance to voxel activity patterns in a distributed sensorimotor network, extending throughout the dorsal visual pathway into posterior parietal cortex (PPC). By comparing various feature sets to fMRI activity, we found that neural state-space representations emerge from nonlinear transformations of the pixel space that bridge perception to action and reward. Furthermore, we show that these transformations reshape axes to reflect relevant high-level features and strip away information about task-irrelevant sensory features. Taken together, our findings shed light on how the brain solves the thorny computational problem of identifying and encoding the relevant states of the world needed to drive decision-making in real-world situations.

Introduction

The human brain is adapted to effortlessly act on the world and can do so effectively even in environments it has not seen before. The decision-making system is interlinked with a hierarchical perceptual system that rapidly infers low-dimensional and abstract structure from the high-dimensional information the nervous system directly senses. The development of the reinforcement learning framework has led to an unprecedented collaboration between researchers in artificial intelligence, neuroscience, and psychology that has yielded substantive insights into how a decision-making system should learn from feedback in the environment (Niv and Langdon, 2016). This theory characterizes the decision-making process as one where an agent interacts with the environment in a series of state-action-reward loops. Two decades of research has identified efficient algorithmic strategies for learning which actions to take in a given environmental state (Sutton and Barto, 2018; Watkins, 1992), and revealed neural substrates of these processes (John P. O'Doherty, Dayan, J. Schultz, et al., 2004; W. Schultz, Dayan, and Montague, 1997; W. Schultz, 1998; Steinberg et al., 2013).

However, much of this work has focused on the mechanisms of learning signals and value representations, divorced from the perceptual systems that are actively coupled to these mechanisms in the real-world. Additionally, in typical reinforcement learning experiments performed in neuroscience, the state-spaces are low-dimensional and discrete, which is often characterized by a small set of distinctive stimuli and actions to learn about. In more naturalistic environments however, the brain faces a continuous stream of high-dimensional inputs, and the brain has to identify and extract the relevant states by constructing a lower dimensional state-space representation internally (Botvinick, Wang, et al., 2020; Niv, 2019). The brain is then able to efficiently select actions with even completely novel sensory inputs by using this state-space to generalize from past experience given what previously worked well in similar states.

This computational problem is so challenging that it proved to be a major barrier to progress in artificial intelligence, until the recent emergence of deep reinforcement learning. For example, the deep Q-network (DQN) is capable of learning high-dimensional tasks like Atari video games with human level performance (Mnih et al., 2015). The marriage of reinforcement learning and deep learning provides a framework for solving the task representation problem by leveraging the outstanding ability of deep neural networks to extract useful features from naturalistic raw input. In addition, deep reinforcement learning algorithms represent end-to-end models for how sensory processing systems can be linked to action evaluation and selection mechanisms. Thus, the human brain may utilize similar computational principles in dynamic decision-making environments, which motivates our current study.

To uncover the mechanisms the brain uses to solve state-space representation problems, we scanned human participants with fMRI while they played three different classic Atari video games: Pong, Enduro, and Space Invaders. These Atari games are highly complex and unstructured compared to standard trial-based tasks, and posed severe challenges to computational reinforcement learning approaches in artificial intelligence before the advent of deep reinforcement learning (Mnih et al., 2015). Thus, we used DQN as a model for how the brain might solve the dimensionality reduction, state representation, and action evaluation problems humans face when mapping high-dimensional pixel inputs to actions.

We first tested whether human behavior during Atari gameplay could be predicted using the features in the hidden layers of a DQN that was independently trained on the same games. This allowed us to establish whether the DQN ended up converging on a similar behavioral policy to that used by human participants during gameplay. We next examined the relationship between the features encoded in the hidden layers of the DQN and patterns of activity in the human brain while human participants played the Atari games. This enabled us to test whether the human brain utilizes similar mechanisms for encoding state space representations as the DQN.

Additionally, comparing the neural predictivity of various control models and different features within DQN helped reveal which computational principles the brain uses to encode a compact state-space representation and how this representation changes between regions. We reasoned that abstract state-space representations should only encode sensory information that is relevant for gameplay behavior, by encoding the most important high-level features such as the position of the Pong ball, while ignoring low-level features that are irrelevant. The richness of two out of three of the Atari games (Space Invaders and Enduro) enabled us to determine that abstract features which generalize across perceptually different inputs are mapped to posterior parietal cortex areas.

Results

We used three Atari tasks of varied complexity (Pong, Enduro, and Space Invaders, Figure 4.1A). The relatively simple game of Pong involves getting the ball past your opponent's paddle while avoiding being scored against. Enduro is a driving game where a player needs to drive as fast as possible while avoiding other cars, and Space Invaders is a fixed shooter game where a player shoots enemy spaceships. The trained DQN reaches human-level performance on all three games (Mnih et al., 2015) (Table S1). Therefore, we hypothesized that the DQN agent could be utilized as an end-to-end model for how the brain maps high-dimensional inputs to actions, and that its hidden layers could serve as a model for state-space representation (Figure 4.1B).

We acquired fMRI data from 6 participants who each completed 4.5 hours of gameplay (1.5 hours on each game). This in-depth fMRI approach has been utilized in previous studies utilizing similar encoding analyses (Güçlü and Gerven, 2015; Kay et al., 2008). Rather than testing a large group of participants for a short period of time as is typical in group fMRI studies, here we obtained sufficiently large amounts of data in a small set of participants to enable us to robustly establish in each participant the relationship between that individual's gameplay and DQN's representations. For our analyses, we ran the frames from the human gameplay data through DQN models that were trained independently from any human data. This produced Q-value outputs and a large set of nonlinear stimulus features represented by the activations in the four hidden layers (three convolutional, one fully connected) for every time point the participants experienced.

DQN state-space representations resemble human state-space representations

Since DQN training was done completely independently of human data, it is unclear whether the state-space representations learned by the DQN agent would resemble the state-space used by human participants or if DQN would develop a policy that resembled that of human players at all. For example, the DQN agent may have not sampled and learned about the states the human participants visit during their gameplay trajectories or may have developed strategies beyond or tangential to that of humans.

The distribution of human actions appeared to diverge from the DQN's when fed human gameplay frames (Figure S1A). However, these differences are largely trivial due to an increased propensity for humans to take NOOP actions (meaning no action) and a reduced tendency for action combinations. This is expected, since unlike DQN, humans encounter a metabolic cost for taking actions and physical constraints limit rapid switching from one action to another. Consequently, we focused on DQN action values when human participants take a "move left" or "move right" action (or any combination with fire or brake). Across all games, DQN action values were significantly higher for the corresponding human action (Figure 4.2A). For example, when a human participant moves left to avoid hitting a car in Enduro, DQN also values moving left more than right. This suggests that the DQN mirrors human policies at these crucial decision points.

The DQN's state-space is not explicitly represented by the output action value layer; rather, it is encoded in the four hidden layers preceding this output layer, as the Q-values used for action selection are linearly computed from the last hidden layer. Therefore, to investigate whether these internal representations could similarly map to human policies, we tested whether the hidden layer activations could be used to predict human behavior. Using a linear decoder, human actions (move left vs. move right) in each of our 6 participants could be reliably predicted from the hidden representations in all three games, demonstrating that DQN encodes stimulus features about the state-space that can be used to model human actions (average accuracy Enduro=84.3, Pong=75.0%, Space Invaders=67.9%; cross-validated by run; chance level accuracy=50%; $P < 0.001$, block permutation test; Figure 4.2B). We were also able to isolate contributions from different features and different layers by averaging the absolute value of the coefficients across a layer (Figure S2). For Enduro and Space Invaders, features from the last two layers, the last convolutional layer and the fully connected layer after it, were the most useful for predicting actions. This suggests that more nonlinear transformations of the sensory input are needed to construct the features humans use to evaluate actions. For Pong, the simplest game of the three, layers 1 and 2 contributed more, and the contribution of each layer was more varied across participants.

Encoding model reveals a distributed network representing a state-space

After validating the use of DQN hidden layers as a model for human state-space representation that could predict behavior, we wanted to localize brain regions involved in encoding this state-space. In order to isolate brain responses with similar representations to DQN hidden layers, we employed an encoding model analysis to create a linear mapping of neural network activations to voxel responses, as done in previous studies that utilized deep neural networks for object recognition (Güçlü and Gerven, 2015; Yamins, Hong, et al., 2014). After reducing the dimensionality of the four hidden layers to 100 features each with PCA, the neural network activations from all hidden layers were then used to model and predict the response of individual voxels with ridge regression (Figure 4.3A).

Across the three games, we found that the DQN model could significantly predict voxel responses throughout the dorsal visual stream and posterior parietal cortex (PPC) (cross-validated by run; $P < 0.001$, FDR corrected; block permutation tests; Figure 4.3B-E, Figure S2). prediction accuracies were significantly higher in the dorsal visual stream regions extending into the parietal cortex, in comparison to ventral stream regions extending into temporal cortex, suggesting a specific role for the dorsal visual pathway in state-space representation for naturalistic visuomotor tasks like video games (two-sample T-test, $P < 1e\text{-}10$, Figure S3A).The encoding model also captured fMRI responses in motor and premotor cortex, SMA, and superior frontal gyrus in all three games. Outside of primary sensory and motor areas, many additional regions of PPC were mapped to DQN hidden layers, including the superior parietal lobule, supramarginal gyrus, and precuneus. Previous studies of object recognition have found evidence for a gradient in neural activity in the ventral visual stream such that later neural network layers better explain neural activity in higher order visual regions, while early layers better explain activity in the early ventral visual pathway (Güçlü and Gerven, 2015). To determine whether a similar pattern exists in our analysis, we examined the coefficients in the encoding models. For example, to evaluate whether early visual regions were more selective of early DQN layers, we averaged the coefficient magnitudes (absolute valued to account for negative coefficients) across the 100 regressors in a layer to see if the first two layers had higher coefficient magnitudes. However, no clear gradient was identified for Enduro and Pong (Figure S3B). For Space Invaders only, the coefficients for the first two layers were lower in PPC, motor, and frontal regions than early visual regions. For all games, every region had very high magnitude coefficients for the last convolutional layer (hidden layer 3). One possible explanation for the apparent lack of evidence supporting a regional selectivity gradient across layers could be that different subsets of features within a layer are mapped to different regions rather than entire layers being mapped to different regions. We investigate this possibility later in this chapter.

Control analyses

An alternative explanation for the encoding model results is that they reflect basic visual features and not information related to reward or action evaluation. To test this, we performed control analyses with feature representations of variable complexity. We used motor regressors as a basic motor control and principal components (PCs) of the pixel space to control for low-level visual properties. We also included two deep neural network (DNN) controls: a DQN agent trained on a separate game, and a variational autoencoder (VAE) (Kingma and Welling, 2013), an unsupervised representation learning method used previously to extract state representations (Ha and Schmidhuber, 2018; Higgins, Pal, et al., 2017; Watter et al., 2015) (see Methods and Figure S4A for examples of VAE outputs). Since the VAE does not encode value or action information, this allows us to test whether this information is needed to reach the prediction accuracies of the DQN encoding model.

DQN outperformed all control models ($p < 1e\text{-}10$, paired t test across voxels) across games except in one participant (Figures 4A and S4B). Furthermore, DQN was best in all regions of interest (ROIs) (except in one participant), especially in PPC (Figures 4B and S5A). The relative performance of different feature sets reveals the computational principles accounting for DQN's ability to explain neural activity. Nonlinear feature representations outperformed linear ones, as both the DQN trained on another game and the VAE consistently showed higher prediction accuracies than a linear principal-component analysis (PCA) model. Additionally, the original DQN surpasses the other two DNN models by linking perception to action and reward.

We next examined whether neural to DQN feature correlations are maintained when all models are included in the same analysis to compete for variance. This reveals whether DQN offers unique predictive information even after controlling for basic visual and motor activity and alternative sensory models. For this, we constructed a general linear model with the first 10 PCs of the most relevant models (DQN layers 1–4, VAE, and PCA) and other regressors of no interest such as game events.

We found that many voxels within each ROI are significantly modulated by unique variance in each model, particularly DQN layers 3 and 4 ($p < 0.001$ family-wise error rate [FWER] corrected, cluster level, F-test; Figure 4.4C). In Figure 4.4C, the results show the proportion of voxels per ROI correlating with a given model above and beyond variance explained by every other model. After controlling for both VAE and PCA, all DQN layers still explain significant variance in a substantial proportion of the voxels per ROI. Additionally, VAE and PCA models explain significant variance after controlling for the effects of the DQN layers. Since early visual and motor regions encode features in DQN layers 3 and 4 when controlling for the other models, this suggests that even these primary sensory regions process more complex sensorimotor features than in conventional visual and motor models.

Representational geometry of DQN's internal representations

The highly distributed representation and numerous parameters within a DNN make its representation rather opaque. To shed light on what DQN is encoding, we utilized representational similarity analysis (RSA). RSA allows comparison of the representational space of many different data types and models of varying dimensionality (e.g., deep network, fMRI patterns, and hand-drawn features), helping to illustrate how a model's representation changes throughout a task as well as aiding comparison across models (Haxby, Connolly, and Guntupalli, 2014; Kriegeskorte, Mur, and Bandettini, 2008).

We first examined Pong, which can be fully characterized with a few high-level features that we manually annotated frame by frame: the positions of the two paddles, the ball position (X and Y), and the ball's velocity (X and Y). A useful and compact state-space should encode this information in some form. An exemplar dissimilarity matrix (DSM; see Methods) for these hand-drawn features is illustrated in Figure 4.5A alongside the DSM of the last convolutional layer in DQN (layer 3) for the same game frames. Similarity is high between two time points when feature vectors in those time points are close in a distance metric (i.e., Euclidean). The representational geometry of DQN resembles the hand-drawn feature DSM, suggesting that it may encode these game-relevant features directly.

To quantify similarities among different DQN layers, hand-drawn features, and other models, we correlated the model DSMs with each other. In Pong, the internal representations in DQN start to become highly similar to hand-drawn features in layers 3 and 4 (Figure 4.5B; Spearman ρ = 0.53, 0.55, respectively), suggesting that DQN constructs a compact state-space representation by realigning its axes to code for these high-level features in later layers. Although this object information is present in the input pixels, they share a relatively low correlation with the pixel space (ρ = 0.058), suggesting some form of nonlinear transformation is required to disentangle this information from the input (DiCarlo and Cox, 2007; Higgins, Amos, et al., 2018). Additionally, the first layer of DQN in Pong is highly similar to the pixel space and PCA model (ρ = 0.9; ρ = 0.78), suggesting that the input data are not yet highly compressed in the first layer of DQN. In contrast, the later layers become increasingly dissimilar to the pixel and PCA representation as they start encoding a lower-dimensional subspace for game-relevant features. A similar pattern is seen in Space Invaders, where the first DQN layer is highly correlated to the pixel space and PCA model (ρ = 0.91; ρ = 0.69), but the last layer is highly dissimilar (ρ = 0.16; ρ = 0.04). In Enduro, representations in all four layers are highly similar to each other, suggesting that differences between them might be more subtle, raising the possibility that there may be more interesting variance within a layer rather than between layers. In all games, the VAE representations are moderately similar to the DQN's, especially for the first three DQN layers.

The brain's state-space representation in Pong encodes the spatial information about objects

Next, we tested whether the brain similarly encodes the spatial positions of the objects in Pong by computing DSMs from voxel activity and correlating these DSMs with a hand-drawn feature DSM (downsampled to fMRI resolution). For all subjects, the hand-drawn feature DSM was significantly correlated to all brain areas in the sensorimotor pathway previously identified in the encoding model analyses (Figures 5C and S6 for individual subjects; block permutation tests, $p < 0.01$, FWER corrected for multiple comparisons). This suggests that similarly to DQN, the brain's state-space representation in Pong involves coding for high-level features tracking the spatial positions of the relevant objects.

Additionally, brain DSMs are significantly correlated to DQN layers 3 and 4 for all subjects in early visual, PPC, and motor/frontal ROIs (and to DQN layer 2 for early visual regions). Representations in early visual areas are already highly correlated to hand-drawn features, which may explain why these regions prefer DQN layers 3 and 4 rather than earlier layers.

Action values encoded in motor and premotor areas

DQN hidden layers encode a state-space to compute Q-values in the output of the network for action evaluation. To identify whether similar action value computations occur in the brain, we implemented a computational model-based general linear model (GLM) analysis (John P. O'Doherty, Hampton, and Kim, 2007) using the DQN output as the computational model.

The action value regressor identifies regions encoding continuous values for the chosen DQN action as a function of the state the participant sees (action advantages used; see Methods). Significant encoding of action values was found in premotor, SMA, and primary visual and motor cortex in all games (Figures 6B and S7). Significant clusters at $p < 0.001$ (FWER corrected, cluster level) are located in motor or SMA/premotor regions for all participants in Enduro, five out of six participants in Pong (six out of six at uncorrected $p < 0.001$), and three out of six participants in Space Invaders. These results indicate that action values are computed in SMA and premotor cortex during Atari gameplay.

Convolutional filter analyses

Thus far, we have shown that a brain-like representation emerges most notably in DQN layers 3 and 4. We see that all ROIs, even early visual regions, prefer these last two DQN layers, suggesting multiple nonlinear transformations of the input pixels are necessary to derive features most predictive of cortical responses during Atari gameplay. However, even though the last two layers best predict voxels across the brain, different regions might prefer different artificial neurons or features within these layers. If so, could we leverage this variability to further shed insight into the features the brain is encoding and how the brain's internal representations transform from one region to another?

We test this by retraining the encoding model on each convolutional filter in the last convolutional layer separately (layer 3, 64 filters; DQN architecture illustrated in Figure 4.1B). The convolutional filter of a convolutional neural network (CNN) represents a feature the network is looking to detect in the input, and this feature can be somewhat visualized with guided backpropagation/deconvolution (Springenberg et al., 2014; Zeiler and Fergus, 2014) (Figure 4.7E). For example, early layers in a typical CNN encode low-level features such as edges and contours.

We then estimated how well each filter predicted voxel responses by averaging prediction accuracies across voxels in our ROIs, a metric we term "neural predictivity." This quantifies how well each filter explains neural responses in general and enables us to test whether neural predictivity changes across different ROIs.

The RSA results in Pong suggested that the shared representation between the brain and DQN in Pong corresponds to a mutual encoding of the spatial positions of objects. We tested this explicitly with our neural predictivity metric, as convolutional filters containing more information about high-level features may better explain brain responses. To quantify this, we calculate the degree to which a layer 3 filter encodes the Pong hand-drawn features with a mutual information metric.

We found that filters with higher neural predictivity encode more information about the hand-drawn features. These correlations are significant for ball position, ball velocity, and paddle positions in every participant ($p < 0.0001$; Figure 4.7A), indicating that the nature of the DQN to brain mapping in Pong lies at the representation of the high-level features.

Filter neural predictivity across regions

To estimate whether different regions prefer different filters, we averaged prediction accuracies for each filter across each ROI. We then computed correlations between the 64 filter scores across regions. For Pong, high correlations between filter scores were found across all regions, suggesting that the same filters are useful for explaining responses uniformly across the brain (Figure 4.7B).

However, in Enduro and Space Invaders, different ROIs only have partially overlapping sets of filters mapped to them, suggesting a more heterogeneous representation across regions (Figure 4.7B). We found visual, parietal, and motor clusters of filter encoding with high correlations within cluster and moderate correlations between cluster. These patterns may differ from the more homogeneous filter selectivity in Pong because of the increased complexity of these games.

Neurally predictive filters generalize across participants and can predict behavior

To investigate if all our participants converge on similar useful representations for solving the task, we correlated each filter's neural predictivity score across participants. We observed high correlations between all participants in all games (Figure 4.7C), meaning the same filters were mapped to the brain across participants.

This result also suggests that some filters in the network are universally useful for explaining neural responses and some are universally useless. Enduro layer 3 filter 40 was one of the best-fitting filters for explaining brain activity in every participant.

Through guided backpropagation (Springenberg et al., 2014), we could see that the filter detects cars and the sides of the road, which are useful features for acting in the game (Figure 4.7E). By contrast, Enduro layer 3 filter 56 was one of the worst-fitting filters for explaining brain activity in five out of six participants. This filter detects the score at the bottom of the screen, which is correlative of reward, since the score board changes when reward is received, but not causally related to reward.

A sample of filter deconvolutions for five random filters in each game is also plotted in Figure S8A.

Next, we evaluated how well each filter modeled human behavior by retraining the decoding human behavior model (Figure 4.2B) on every filter in layer 3 separately. Similar to the neural predictivity analysis, this allows us to probe how useful every layer 3 filter is for predicting human actions. We found correlations between how well a filter explains voxel activity (the neural predictivity score) and how well a filter explains human behavior (Figure 4.7D). This correlation was most pronounced for Enduro and Pong ($p < 0.05$ in six out of six participants in Enduro and six out of six participants in Pong, but only two out of six participants in Space Invaders). Thus, the brain encodes the features most relevant for behavior, and DQN encodes features that not only are brain-like in a universal way across participants, but also predict human actions.

State-space representations are nuisance invariant in PPC

An abstract state-space representation should ideally be pruned of sensory features not necessary for learning or behavior. For Pong, this involves encoding high-level features about the relevant objects in the game. However, the other two games are more complex and involve a large number of features that are difficult to hand label. Thus, rather than isolating relevant high-level features in these games, we next identify irrelevant features that an abstract state-space should ignore.

We wanted to find brain regions where the state-space encoding is insensitive to sensory information irrelevant for task performance, a pattern known as nuisance invariance (Lenc and Vedaldi, 2015). For Enduro, one nuisance variable is the weather and time of day. Driving gameplay starts off during the day and gradually becomes nighttime with various weather patterns. The colors of the pixels and visual input dramatically change, while the overall gameplay remains mostly the same.

Formally, this weather variable had no relationship with the participant's actions in an information-theoretic sense (see Methods). A good state-space representation should localize objects independently of colors in the game. Thus, it should often project inputs that are very far away in the pixel space to similar regions of the latent state-space if an agent should act similarly across them (illustrated in Figure 4.8A). In contrast, even small changes in pixel space may necessitate opposite actions. For example, in Figure 4.8A, an agent should move left or right depending on the location of the car in front of it, even though the two pairs of frames are perceptually similar.

For Space Invaders, the number of on-screen invaders explains a lot of variance in the pixel space but has a marginal effect on what actions participants take (see Methods). This is because as an agent kills more invaders, the screen becomes more and more black. This information does not heavily impact which actions an agent should take, because the relative positions of the invaders above an agent matter the most.

To estimate whether ROI representations are nuisance invariant, we quantified the mutual information between a filter and the nuisances identified for Enduro and Space Invaders, giving each filter a metric for how insensitive it was to the nuisances (see Methods). We computed the correlation between each filter's nuisance invariance, and its neural predictivity in a ROI, which we define as a nuisance invariance score for each region (normalized across voxels; see Methods). Simply put, this score estimates how each region prefers the filters that are nuisance invariant.

Regions in PPC and in the late dorsal visual stream (i.e., lateral occipital cortex [LOC]) were more insensitive to nuisances than early visual cortex regions V1–V4 (Figures 8B, 8C, S8B, and S8C) in both games. Early visual cortex regions exhibited the lowest nuisance invariance scores in both games, suggesting that filters mapped to these regions still encoded the low-level nuisance variables. Additionally, LOC, which is later in the dorsal visual pathway, had a higher nuisance invariance score than these earlier visual regions. For Enduro, a PPC region exhibited the highest or second highest score of any region in every participant. In five out of six participants in Space Invaders, premotor/prefrontal cortex regions also exhibited high nuisance invariance scores.

These results suggest that irrelevant visual input is stripped from the neural code as information passes through the dorsal visual stream to the PPC. This leads to a lower-dimensional, compressed, and abstract representation that projects similar game situations to the same part of the state-space as depicted in Figure 4.8A.

Discussion

One of the major unresolved questions in decision-making neuroscience is how relevant features from the environment are identified, extracted, and relationally structured to be used for action evaluation and selection in real-world environments. Here we aimed to address this question by using a highly complex set of tasks, whereby humans were asked to play three different classic Atari games (Pong, Enduro, and Space Invaders) while undergoing fMRI scans. Taking our cue from recent advances in machine-learning and artificial intelligence (Mnih et al., 2015), we utilized a computational modeling approach where a deep neural network, which has classically been applied to high-dimensional categorization problems, is married to a reinforcement learning system. Using the deep reinforcement learning approach, we were able to demonstrate that representations in DQN show a remarkable similarity to representations used by humans. Features in DQN hidden layers could predict human actions and fMRI activity in a distributed sensorimotor network extending from the dorsal visual stream and posterior parietal cortex to premotor areas. Not only did the DQN model significantly outperform control models of varying levels of complexity, but DQN features also explained unique variance in these ROIs when controlling for the other models. These results suggest that these regions do not simply encode low-level sensory information, but produce a state representation that links sensory information to reward and action selection. Further validating this approach of using DQN as an end-to-end model of how the brain maps pixel inputs to actions, we found an encoding of the action value output of the DQN agent in the supplementary motor area, along with primary visual and motor cortex. In alignment with previous results from a more traditional trial-based study (Wunderlich, Rangel, and John P. O'Doherty, 2009), our results support a role for SMA in action valuation and generalize this neural process to an environment with fast moving and high-dimensional state dynamics. When taken together, these results help unveil the nature of the shared representation between DQN and the human brain during Atari gameplay.

The present findings build on a growing catalogue of intriguing similarities shared by deep neural networks and the human brain (Eickenberg et al., 2017; Güçlü and Gerven, 2015; Khaligh-Razavi and Kriegeskorte, 2014; Wen et al., 2018; Yamins, Hong, et al., 2014; Yamins and DiCarlo, 2016; Wang et al., 2018; Iigaya et al., 2020). Neural activity across regions of the ventral visual system have been found to resemble activation patterns in different layers of deep learning networks trained to recognize objects or even to predict the value of visual art (Iigaya et al., 2020). In a converging line of research, we also found strong representational similarities between a model network that can perform complex tasks in an intelligent manner and activity in the brain while human participants performed those same tasks. Nevertheless, to our knowledge this is the first study directly applying a deep RL model to neuroscientific data in a naturalistic task, as research to connect these two fields is still in its early stages (Botvinick, Wang, et al., 2020).

Unlike the passive visual feature encoding used for object recognition or aesthetic valuation, we did not find evidence for a gradient of abstraction in the mapping of early layers to early visual regions and later layers to brain regions further along in the processing pathway. All regions of interest consistently preferred DQN layers 3 and 4 over the first two layers of the network. By examining the representational geometry of different DQN layers and other models, we were able to identify computational principles that could account for this pattern. For Pong and Space Invaders especially, the internal representation is not highly dissimilar to the pixel space until DQN layers 3 and 4, whereas the information reaching early visual cortex may already be heavily compressed. Prior research suggests that a considerable amount of compression and nonlinear processing of visual input before the cortex, with a combination of retina, LGN processing, eye movements, and recurrence/feedback connections (Gollisch and Meister, 2010; Hayhoe and Ballard, 2005; Hosoya, Baccus, and Meister, 2005; Kietzmann et al., 2019). The relative roles of each of these components in shaping the representation in the visual cortex necessitates future research. For Pong, the RSA and convolutional filter analyses also suggest that both the brain and later layers of DQN exhibit a state-space representation that encodes high-level features about the spatial positions of the ball and paddles. This may account for why early visual regions have more similarity to layers 3 and 4, since DQN only begins to disentangle these features from the pixel space in these later layers. Additionally, many of the DNNs used in visual neuroscience have 8 or more layers, with layers 2-4 often constituting the most similarity to early visual cortex, rather than layer 1 (Khaligh-Razavi and Kriegeskorte, 2014; Seeliger et al., 2018; Wen et al., 2018). Thus, if the network had more layers, it is possible that gradient of representation at the layer level would emerge, with early visual regions still preferring layers 3 and 4, but more anterior regions mapping to even deeper layers. In addition to more layers, future studies could also utilize architectures with a wide number of more biologically plausible characteristics, such as attention processes and recurrent connectivity. However, for our purposes DQN provides a satisfactory account of both behavior and neural data, and the most interesting variance for explaining cortical activity is packed into layers 3 and 4. Therefore, we analyzed how the different features within layer 3 (the last convolutional layer) explain activity across the brain.

To do so, we retrained separate encoding models on stimulus features from the convolutional filters in DQN layer 3. This analysis showed that filters most predictive of voxel activity are also predictive of human behavior, suggesting that these features are used by the brain to guide behavior. Filter selectivity is highly correlated between participants, indicating a common task representation across individuals. For Pong, the filter analysis provided more evidence that this common state-space represents high-level features such as the spatial positions of the relevant objects. This is in line with a recent proposal that the dorsal stream and PPC encode spatial positions of objects by projecting high-dimensional inputs onto a low-dimensional manifold of physical space (Summerfield, Luyckx, and Sheahan, 2020).

For Enduro and Space Invaders, the mapping from DQN features to neural responses was more heterogeneous between regions, suggesting that different regions preferred different underlying features in the network. We found that the posterior parietal cortex (PPC) encodes features that are more generalizable and nuisance invariant than early visual regions. Thus, the representation in PPC is able to ignore and abstract away information from the sensory stream that is not relevant for behavioral performance, such as the changing colors and backgrounds in Enduro. These findings directly support a recently proposed theory that PPC regions act as a central interface for isolating behaviorally relevant stimuli by integrating visual, cognitive, and motor information (Freedman and Ibos, 2018). These ideas can also be synthesized with a substantial literature in motor neuroscience that suggests PPC regions are involved in sensorimotor transformations, linking perception to decision-making and action (Andersen and Buneo, 2002; Andersen and Cui, 2009;

Gold and Shadlen, 2007). This function is highly compatible with the role we ascribe to PPC regions in this paper, as a state-space representation that links perception to action in a reinforcement-learning system. Past studies have also associated the parietal cortex with updating the state-space representations needed for reinforcement-learning, with a specific role in encoding the relationship between states, actions and subsequent states (Gläscher et al., 2010). The present work suggests that these past findings and proposed theories can be integrated into a broader conceptualization of the posterior parietal cortex as encoding abstract state-space features that link perception to learning and action selection.

Overall, our results point toward key properties fostering an effective state-space for tasks of real-world complexity. Initially, compression to a lower-dimensional space takes place to avoid the curse of dimensionality, where learning complexity scales exponentially with the number of states to learn about. However, exploiting the raw statistical properties of the input data, as in unsupervised learning techniques, is not enough; it must also disentangle a purely sensory manifold into appropriate axes linked to rewards and the actions that deliver them (DiCarlo and Cox, 2007; Higgins, Amos, et al., 2018). For Pong, these axes code for relevant data-generating factors, the spatial positions of the ball and paddles. In addition, a state-space would likely benefit from being invariant to nuisances irrelevant for task performance (Lenc and Vedaldi, 2015). This property further reduces state-space dimensionality by only transmitting useful signals through an information bottleneck (Achille and Soatto, 2018; Shwartz-Ziv and Tishby, 2017). This added compression helps protect against overfitting by shaping an abstract task representation orthogonal to low-level sensory properties that can change in future settings. Humans are clearly equipped with abstract representations with this property (Behrens et al., 2018), as they can seamlessly adapt to novel circumstances, such as driving on new roads without having to relearn the driving process.

It should be noted that the DQN objective function and architecture itself does not explicitly promote the learning of nuisance invariant representations, and most filters still retain information about the nuisances we highlighted (weather/time of day in Enduro, number of invaders left in Space Invaders). Additionally, DQN performance is not robust to even moderate visual changes such as the contrast of the image space during the testing of the algorithm if the change was not in the training distribution. Most deep reinforcement learning algorithms are not explicitly trained to learn a representation (unlike representation learning algorithms), but are only trained to approximate value-based and policy-based functions with a deep neural network and thereby learn a task representation as a side-effect. These approaches are plagued with sample efficiency and generalization issues (Kaiser et al., 2019; Lake et al., 2017). Therefore, we suggest that the sample efficiency and generalization performance of deep reinforcement learning algorithms would greatly benefit from explicitly learning a representation with the principles we previously outlined and other characteristics in line with the inductive biases humans possess about the structure of the world (Botvinick, Ritter, et al., 2019). Promising work has started to develop methods for accomplishing this goal in the emerging field of state representation learning (Anand et al., 2019; Botvinick, Ritter, et al., 2019; Ha and Schmidhuber, 2018; Higgins, Pal, et al., 2017; Jaderberg et al., 2016; Van den Oord, Li, and Vinyals, 2018; Lesort et al., 2018; Srinivas, Laskin, and Abbeel, 2020; Zhang et al., 2020). We also hope that our work will promote more cross-talk between decision neuroscientists and artificial intelligence researchers at the level of representations for a reinforcement learning system, whereas thus far most of the interaction between these fields has occurred at the level of learning signals (Botvinick, Wang, et al., 2020; Dabney et al., 2020; Niv and Langdon, 2016).

The present findings suggest that even with notable architectural differences between the human brain and deep RL models, DQN still does remarkably well in capturing variance in both human behavior and brain activity throughout the dorsal visual stream and the parietal and premotor cortices in high-dimensional decision-making contexts. These findings further help to establish the deep and sustained relationship between progress in artificial intelligence and in computational neuroscience. Our results suggest that this interdisciplinary interplay is continuing to evolve and that in particular, a synergy between deep RL and decision neuroscience offers the continuing prospect to yield rich insights about the internal representations of intelligent systems.

Methods

Participants

We recruited six healthy participants from the Caltech and Pasadena community (4 male and 2 females, age 26 ± 3.4). All participants performed the tasks over the course of four separate days and received a participation fee of $40 a day. The

Caltech Institutional Review Board approved the protocol, and all participants gave their informed consent on each day of the experiment.

Experimental Paradigm/Atari Gameplay

Across the four days of the experiment, each participant went through 33 runs of gameplay. The runs were 10 minutes in duration, with 8 minutes of gameplay in between a minute of rest and a fixation cross before and after gameplay. Eyetracking was recorded, but not analyzed for this paper. Each participant played the games Space Invaders, Pong, and Enduro 11 times each. On day 1, each game was played twice, in random order with the one constraint of never playing the same game twice in a row. The six runs were then followed by anatomical scans on day 1. On days 2-4, each game was played three times, in random order with the same constraint of never playing the same game twice in a row. Before scanning on the first day, each participant went through a training session to become familiar with each game by playing each game for 5 minutes on a laptop.

The Atari games were presented through the Arcade Learning Environment (Bellemare et al., 2013), with modified code to log actions, rewards, MRI pulses, and frames with proper timestamps. A button box with four buttons was used as an Atari controller (Figure 4.1A). Participants held the button box with two hands, using their left thumb to press the 1 and 2 buttons corresponding to move left and move right, respectively, and using their right thumb to press the 3 and 4 buttons to hit brake and fire, respectively. Brake is only used in Enduro, and fire is only used in Enduro and Space Invaders.

In Enduro, participants control a race car that must move as fast as possible while avoiding other cars on the road. Participants get a reward of 1 for every car they pass, and the main objective is to pass a certain number of cars before the end of the day (200 cars in level 1 and 300 cars in level 2). The sky and weather patterns change throughout the gameplay to simulate the passing of time in the day ('sunny,' 'snow,' 'blue dusk,' 'red dusk,' 'night,' ‘fog,' 'sunrise'), with the sky eventually becoming black and the sun beginning to rise before time runs out after 13312 frames.

In Pong, points are awarded to a player when the white ball moves past their opponent's paddle. Participants control the green paddle on the right side of the screen and try to defend their goal and score on their opponent's goal by moving their paddle up and down in the white ball's path.

In Space Invaders, participants control a green ship that can move from left to right at the bottom of the screen. The objective is to destroy enemy ships to get reward and avoid being hit by missiles from the enemy ships while having 3 lives before the game ends.

fMRI data acquisition

We collected two datasets on two separate scanners at the Caltech Brain Imaging Center (Pasadena, CA). The first dataset included two participants and was collected using a 3T Siemens Magneto TrioTim scanner. After an upgrade to a Siemens Prisma, a second dataset was collected with four participants. Both datasets used a 32-channel radio frequency coil. These parameters were shared across the two sequences: whole-brain BOLD signal acquired using multiband acceleration of 4, 56 slices, voxel size = 2.5mm isotropic, TR = 1,000ms, TE = 30 ms, FA = 60°, FOV = 200mm x 200mm. At the end of the first day of scanning, T1 and T2 weighted anatomical high-resolution scans were collected with 0.9mm isotropic resolution.

fMRI preprocessing

Data was preprocessed using a standard pipeline for preprocessing of multiband data. Using FSL (Smith et al., 2004), images were brain extracted, realigned, high-pass filtered (100 s threshold), and unwarped. Images were denoised by ICA component removal. Components were extracted using FSL's Melodic, classified into signal or noise with a classifier trained on separate datasets for the first dataset, and manually classified for the second dataset since the scanner was different from the one used in the classifier training set. T2 images were aligned to T1 images with FSL FLIRT, and then both were normalized to standard space using ANTs (using CIT168 high resolution T1 and T2 templates (Avants, Tustison, Song, et al., 2009; Tyszka and Pauli, 2016). The functional data was first co-registered to anatomical images using FSL's FLIRT, then registered to the normalized T2 using ANTs. For GLMs in SPM 12 (Penny et al., 2011) (encoding model control GLM and action value analysis) the data was spatially smoothed in FSL with a 5-mm FWHM Gaussian kernel. Smoothing was not initially applied to the fMRI images for the voxelwise encoding model analyses to preserve fine-grained detail at the voxel level but was applied with a 5-mm kernel for visualization.

Deep Q-Network training

Deep Q-networks were trained separately for each of the three games using the Neon deep learning library, by making modifications to open source code `https://github.com/tambetm/simple_dqn`. As in the original paper (Mnih et al., 2015), DQN takes a tensor of four input frames as input, has three convolutional layers (Layer 1: 32 filters of 8x8 with stride of 4; Layer 2: 64 filters of 4x4 with stride of 2; Layer 3: 64 filters of 3x3 with stride of 1) followed by one fully connected layer (512 units), and outputs Q-values for every available action. DQN takes the action with the highest Q-value. Convolutional layers are locally connected with each neuron having a receptive field. Convolutional filters learn visual features which are then convolved across the input to detect the presence of that feature. Fully connected layers do not have this local connectivity as every neuron is connected to every neuron in the previous layer.

The Arcade Learning Environment was used as the Atari environment during training (Bellemare et al., 2013). The training consisted of 100 epochs of 250,000 steps in each epoch for each game. One modification was made for Pong by restricting the action set to noop, up, and down, since the default available action set for this game includes redundant actions up/right, and down/left.

To output Q-values and hidden unit activations that are used for all analyses, the human gameplay frames were run through the trained network. Since the input to DQN is a tensor of four consecutive images, a frame from the human data is concatenated with its three preceding frames. Thus, the fourth frame in a run is the first one put through DQN. In Enduro, each level is won after passing 200 cars in the first level and 300 cars in the second level, signified by flags appearing on the scoreboard. When this happens, the game engine no longer gives reward until the day ends/clock stops even though the participant is still tasked with controlling the car and trying to avoid other cars. Thus, the network would detect the flags and predict 0 reward when this happened, resulting in meaningless Q-value traces. This would happen occasionally in a participant's run and would last a couple of minutes. To ensure that the activations and Q-values we extracted from the network were useful, we altered the images from Enduro human gameplay before they were put through DQN so that the scoreboard would never change. Specifically, the scoreboard from a reference image midway through a run was copied into every frame.

Human actions analyses

To analyze the human state-space in relation to the DQN's state space, we analyzed the actions participants took and how these compare to the actions DQN selects when fed the human gameplay data. This initially involved plotting the distributions for the actions executed by DQN and human participants (Figure S1A). To analyze DQN's action values with respect to human actions, the Q-values for every participant were included after downsampling by 10 and removing the first 100 frames in a run. Action values were computed with an action advantage function by subtracting the average Q-value as an action-independent baseline (Sutton and Barto, 2018). This allows us to isolate action related variance from state value related variance. Action values/advantages were then LOWESS smoothed across frames (using the Statsmodels Python package with the "frac" parameter = 0.005) and normalized with Scikit-learn's StandardScaler. All the frames involving a "move-left" or "move-right" human action were selected, including combination actions (ie. "fire left"). Then average action values for the corresponding frames are computed across a human action category. For Enduro and Space Invaders that have combination actions, the maximum Q-value for a "move" action was taken (ie. for the "move left" Q-value in a frame in Enduro, we take the max between "move left," "brake left," and "fire left"). To test for significance, we test the interaction term in the linear model Action Value C(DQN Action Value) + C(Human Action) + C(DQN Action Value):C(Human Action) with Statsmodels.

For decoding human actions, we model human actions with the hidden layers of DQN using LASSO logistic regression (L1 regularization) using Scikit-learn functions and custom Python code. Each hidden layer was projected to a dimensionality of 100 using PCA, giving a concatenated feature set of 400. Time points were downsampled by a factor of 10 to ease computation. The PCA transformation matrices were estimated using the frames for Sub001. These transformation matrices were used in every participant, to ensure that the PCs of every participant would be in the same space. LASSO logistic regression classifiers were then trained to predict left versus right actions, after frames where no action or other actions occurred were removed. The time points when other actions were selected in combination with left versus right were also included. Decoding accuracy was determined by cross-validating across runs. Optimal regularization parameters were found through grid search and were fixed across participants per game. Decoding accuracies were tested against a null distribution created from permutation tests of 1000 permutations. To maintain the autocorrelation of action trajectories, the cross validated data was shuffled in blocks of 40 time points (Wen et al., 2018). The predicted responses from the model were then compared against these shuffled datasets. The accuracy of every model in every participant exceeded the accuracy of the maximum value in the null distributions. To determine which layers were most useful for decoding actions, the model was trained on all runs (no cross-validation) and coefficients were absolute valued and averaged by layer.

Encoding model

To map hidden representations in DQN to voxels in the brain, we performed deep learning based encoding model analyses (Güçlü and Gerven, 2015). All analyses were run in custom Python code using functions from PyMVPA (Hanke et al., 2009) and Scikit-learn. First, image frames from the participant's gameplay data were run through the trained DQNs in order to generate neural network activations in every layer at every time point. As done in the decoding human actions analysis, PCA is used to reduce the dimensionality to 400 (100 PCs per layer). To downsample from the video game framerate to the TR of 1 Hz, each feature's values are averaged over a second. Then, copied time courses are shifted by both 5 s and 6 s to account for the hemodynamic delay of the fMRI signal. These two shifted time courses are concatenated into a feature set of 800. Next, voxelwise ridge regression (L2 regularization) is performed to predict each voxel's responses as a linear combination of this feature set. Optimal regularization parameters were found using grid search. Voxels are preprocessed as described above without spatial smoothing. Each voxel's response is z-scored to ensure every voxel is on the same scale. Accuracy is estimated using cross-validation across runs and calculating the Pearson correlation between predicted and actual time courses.

Statistical significance was quantified through permutation tests (since fMRI data may not be normally distributed) methods similar to previous approaches where 100,000 permutation tests are performed on 14 random voxels (Eickenberg et al., 2017). In each permutation, the time course of the held-out validation set was shuffled in a blockwise manner of blocks of 40 TRs to keep autocorrelation intact (Wen et al., 2018). The Pearson correlation between the shuffled time course and the predicted responses from the model were then computed. These permuted distributions are then concatenated, and voxel accuracy scores are compared to this concatenated null distribution to obtain one-sided p values for every voxel. Rather than selecting 14 completely random voxels to estimate a global null hypothesis for all brain voxels, we took a more conservative approach and selected 14 random voxels who were in the 90th percentile or above of scores in the encoding model analysis. This condition ensured that voxels with strong signal were selected. Voxels were then multiple comparisons corrected using FDR and plotted at the corrected threshold as indicated. Maps are transformed to standard space and spatially smoothed (5mm kernel) for visualization. To estimate layer selectivity, the coefficients from the models were absolute valued, averaged across layer, and then averaged across region. Average coefficients across participants are shown in Figure S3B.

Regions of interest and atlases

To define regions of interests for visualization and further analyses, we used the Harvard-Oxford Atlas. To distinguish V1, V2, V3, and V4 in the visual cortex, we used the Juelich Histological Atlas. Both atlases were accessed with FSLview. The early visual ROI consists of V1-V4; PPC includes LOC superior, superior parietal lobule, supramarginal gyrus, precuneus; Motor/Frontal includes motor and premotor cortex, SMA, and superior frontal gyrus.

Encoding model control analyses

Various control models were tested in the encoding model to help identify what computational principles play a role in the DQN model explaining neural responses. In an identical pipeline as the DQN encoding model analysis, these control feature sets were downsampled and time shifted by 5 s and 6 s (other than motor regressors where this preprocessing has already taken place) before cross-validated ridge regression was performed to compute prediction accuracies for every voxel.

Motor

Two motor regressors corresponding to making responses with the left and right hands were used. These regressors were taken directly from the GLMs in SPM for action value that are described below.

PCA

To construct a control model for basic visual features that represented the statistical structure of the images, the 84x84x4 pixel tensor was linearly projected to dimensionality 100 with principal component analysis using Scikit-learn. Although the DQN encoding model includes 400 features and we match this dimensionality with the cross game DQN and VAE control models, using 100 principal components outperformed using 400. This linear projection of the input uncovers features that explain the low-level statistical structure in the input that vary the most during gameplay without any representation of reward. Similar approaches have been used to explain neural responses to a remarkable degree throughout the visual pathway (Chang and Tsao, 2017; Olshausen and Field, 1996) (Chang and Tsao, 2017; Olshausen and Field, 1996). Additionally, since we perform PCA on the tensor of 4 consecutive frames that are input into DQN, the principal components uncover statistical properties of motion and change detection that are appropriate to model the dorsal visual pathway. As with the other PCA analyses, transformation matrices were estimated using sub001's data and used across participants to project every all data to the same space.

These principal components were also used to estimate their representation of the nuisance variables. Scikit-learn's 'mutual_info_classif' function was used to calculate the mutual information between the first principal component and the nuisance variables.

Cross game DQN

We also compared our encoding model results with a DQN trained on a different game. The Space Invaders network was used as this control for Enduro, Enduro for Pong, and Pong for Space Invaders. Other than shifting the networks, the regressors were constructed identically to the original encoding model.

VAE

To compare DQN with another state of the art method for state representation learning using a deep neural network (Higgins, Pal, et al., 2017; Mohamed and Jimenez Rezende, 2015; Watter et al., 2015), we trained variational autoencoders in Tensorflow for each game by modifying an existing template `https://github.com/tensorflow/docs/blob/master/site/en/tutorials/generative/cvae.ipynb`. The architecture we used was designed to be as similar as possible to DQN.

This consisted of an encoder of three convolutional layers (Layer 1: 32 filters of 8x8 with stride of 4; Layer 2: 64 filters of 4x4 with stride of 2; Layer 3: 64 filters of 3x3 with stride of 1), followed by a fully connected layer to output the set of mean and log-variance parameters for the latent representation of dimensionality 400. The decoder architecture consisted of a fully connected layer followed by four convolution transpose layers (Layer 1: 64 filters of 4x4 with stride of 1; Layer 2: 64 filters of 4x4 with stride of 2; Layer 3: 32 filters of 8x8 with stride of 2; Layer 4: 1 filter of 8x8 with stride of 1). All activation functions are rectified linear units (ReLU). The network was trained on each game separately by maximizing the evidence lower bound (ELBO) on the marginal log-likelihood of the training data. Data frames of the first 8 runs of the first participant were used as the training set, and frames from the last 3 runs were used as the test set for tracking generalization (training sets and test sets were downsampled by 5 to ease computation). Training included 1000 epochs over the entire training set, but converged well before that for every game (training loss for first 500 epochs plotted in Figure 4.4A). After training, performance on the test set was nearly equivalent to performance on training set.

The human frames from every participant were then run through the trained encoder to map them to the latent distribution, which outputs 400 means and log-variances for the latent dimensions. The means were then used as a 400 dimensional stimulus feature set for the control encoding model, and preprocessed with downsampling and time lags identically to the other feature sets used for encoding models.

General linear model (GLM) control analysis

In order to test for whether brain responses could still be predicted by DQN when controlling for the other models and game events, we constructed GLMs in SPM12 similarly to previous approaches (Iigaya et al., 2020). The first 10 principal components for each DQN layer, VAE, and PCA models were added as parametric modulators to the same onset at the temporal resolution of the 1 Hz TR after averaging across volumes. Orthogonalization was turned off. Other regressors of no interest included all of the regressors described in the computational model-based GLM section below, including regressors for motor responses, reward/punishment, and action values. To quantify a voxel's correlation to the unique variance in each of the six models (four DQN layers, VAE, PCA) F-tests were computed on the betas for the 10 PCs in each model, which tests whether a voxel is significantly modulated by at least one principal component in a model. The percentage of significant voxels in a region of interest for each model is reported in Figure 4.4C.

Control region analysis

To rule out the possibility that our analyses are picking up on artifacts such as head motion that affect the entire properties of the fMRI images, we completed a control region analysis with one subject (sub001). A control region was represented by two spheres of air drawn directly in front of the brain. The encoding model was then run on every voxel within those spheres and the distribution of prediction accuracies were plotted alongside comparison ROIs (V1 and superior parietal lobule) in Figure S5B. No voxels in these spheres had significant prediction accuracies and the whole distribution of scores were very close to zero.

Representational similarity analysis

We performed representational similarity analyses (RSA) to examine how the representations transform throughout the DQN layers. Dissimilarity matrices (DSMs) were constructed at the frame level for DQN layers 1-4, VAE, PCA, the pixel space, and the hand drawn features for Pong. Each model was first downsampled by 20 and data was concatenated across runs within subject. DSMs were constructed by computing pairwise comparisons across frames for each model with pyMVPA. Within day comparisons were removed to avoid potential confounds due to similarity being driven by patterns being in the same run or day. For the pixel space, the 84x84x4 tensor of images that are fed to DQN were reshaped into a 28224 dimensional response vector. For the PCA model, weights fit to the data of sub001 were again used to transform the pixel space into a 100 dimensional space. In Pong, each hand drawn feature (the positions of the two paddles, the ball position X and Y, and the ball's velocity X and Y) was z-scored and input into one response vector. Euclidean distance was used as the distance metric for the Pong hand drawn features, and correlation distance was used for every other model. Every DSM was rank-ordered to compare model DSMs without assuming a linear relationship between models. Models were then compared with Spearman correlations (the Pearson correlation on the rank-ordered DSMs).

For comparing model DSMs and fMRI DSMs in Pong, each DSM was created at the TR level. This involved using the same feature sets that were used in the encoding model, where responses were averaged across volumes to downsample to TR resolution (1 Hz) and shifted by 6 s to account for hemodynamic delay. Again, correlation distance was used for every model except the hand drawn features (Euclidean) and DSMs were rank-ordered.

For fMRI data, DSMs for three brain areas were constructed, early visual, posterior parietal cortex (PPC), and motor/frontal. Early visual regions included all visual cortex ROIs. PPC included superior lateral occipital cortex, superior parietal lobule, supramarginal gyrus, and precuneus. Motor/frontal included motor and premotor cortex, SMA, and superior frontal gyrus.

To test for significance we performed block permutation tests for every model, since the data may not be normally distributed. Similarly to the encoding model permutation tests, fMRI data volumes were shuffled blockwise in blocks of 40 TRs to keep autocorrelation intact (Wen et al., 2018), then DSMs were reconstructed and correlated to the non shuffled model DSMs. Then, to test if the correlation in a model was significantly different than zero, the correlation score had to be greater than the maximum correlation in the permutation test distribution (one-sided). To test if the differences between models were significant, this difference was tested (two-sided) against a distribution based on computing the differences between the models in every permutation. All scores were corrected for multiple comparisons.

Computational model-based GLMs

To localize the neural correlates of action value computations, we conducted computational model-based generalized linear model (GLM) analyses (John P. O'Doherty, Hampton, and Kim, 2007). This novel analysis differs from previous approaches in two ways: a deep neural network is used to approximate the value function that is used to construct regressors, and the model is trained independently of any human behavioral data.

All univariate GLMs were conducted using SPM12 software. Initially the image frames from the human gameplay data were run through the trained DQN to output Q-values at every frame in a run as described above. Next, the Q-values were decomposed into action advantages/values to separate action related variance from reward related variance. Taking inspiration from actor critic approaches to isolate action advantages (Sutton and Barto, 2018), we define state value (V(s), s = state) as the average of all Q-values, and action advantages (A(s,a), s = state, a = action) as the difference between an actions Q-value and the state value.

$$A(s,a) = Q(s,a) - V(s)$$

$$V(s) = \frac{1}{|A|} \sum_{a'} Q(s,a)$$

Similar to the analysis in a previous study (Wunderlich, Rangel, and John P. O'Doherty, 2009), the action value regressor here is computed as the chosen value (the maximum) between move left value and move right value. Chosen value is then LOWESS smoothed across frames, and downsampled to 10 Hz for every volume (TR = 1 s). The regressor is then z-scored and entered into a GLM where it is convolved with a hemodynamic response function. Across the games, other covariates included left and right hand motor responses, parametric regressors for both positive reward and negative reward, game presentation (8 minutes of gameplay per run with one minute of rest before and after), run, and day. Losing a life was included as the negative reward regressor in Space Invaders, although the Atari engine does not explicitly deliver negative reward for loss of life, and the negative ramifications are reflected in the opportunity cost of gaining more points. Additional regressors for Space Invaders also included fire action value and the number of invaders left on the screen. For Enduro, the action value for the brake action was also included (which simultaneously approximates the anti-correlated fire action value, thus the fire action value was not also included).

Filter analyses

To further interpret the encoding model results, we wanted to identify which filters were useful for modeling neural responses, and whether this varied between regions of interest. To do this, we retrained the encoding model on each filter in layer 3 (the last convolutional layer) on each voxel that was significant in the encoding model analyses. This layer had 64 filters of 7x7 receptive field size. We use cross-validated prediction accuracy of a voxel response using a convolutional filter's explanatory features to quantify that filter's Neural Predictivity. This Neural Predictivity score was averaged across a region to estimate how well that filter predicted responses in a region. With 64 Neural Predictivity scores per region, correlations of these scores across regions were computed to evaluate the variability of filter selectivity between regions and to construct a similarity matrix (Figure 4.7B). This similarity matrix reflects the average similarity matrix across the six participants. A similar procedure was used to compute correlations of Neural Predictivity across participants, where in this case filter scores were averaged across all voxels in all ROIs in a participant rather than by region (Figure 4.7C). For computing a filter score for decoding human actions (Figure 4.7D), we similarly retrain the model from "Decoding Human Actions" on each filter separately in layer 3. These filter scores were then rescaled with min-max normalization for subsequent correlation analyses and visualizations. Thus, the best filter has a score of 1 and the worst filter has a score of 0.

To visualize the features encoded by the filters (Figure 4.7E), we use Neon's deconvolution visualization function and modified code from https://github.com/tambetm/simple_dqn. This procedure finds frames from the human gameplay data that activate a filter the most (which is depicted on the right side), then uses guided backpropagation to identify parts of the image that led to this activation (left side). The colors reflect changes and motion across the image tensor of three frames, meaning the filter detects motion in this location.

We annotate six high-level features in Pong using custom Python code that localizes the corresponding objects in the pixel space: ball X position, ball Y position, ball X velocity, ball Y velocity, left paddle position, and right paddle position. To assess how much each filter encodes each feature, we use Scikit-learn's 'mutual_info_regression' function to calculate the mutual information between a filter and these continuous variables. The mutual information scores were averaged across ball X and Y positions to get one score for ball position. We similarly averaged across the ball X and Y velocity and the left and right paddle position to get scores for ball velocity and the paddle positions, respectively. This outputs a MI score for each of the 7x7 receptive fields in a filter, which were then averaged to get one metric per filter for each high-level feature. These metrics are then correlated with each filter's Neural Predictivity across the whole brain in Pong (Figure 4.7A).

Nuisance invariance scores

We completed additional analyses to identify how the regions of interest encode sensory information that is irrelevant for task performance. To uncover this, we utilized a concept from the machine-learning sub-field of representation learning: nuisance invariance (Lenc and Vedaldi, 2015). A nuisance variable is any variable in the input that is irrelevant to the task, and is mathematically defined as any variable where the mutual information between it and the task output is zero (I(y;n) = 0), where y is a task label and n is a nuisance variable). Common examples include translation and illumination invariance in object recognition, as the location of an object on an image and the overall brightness of a picture are usually unrelated to classifying it correctly. Thus, nuisance invariance in neural networks suggests that a compressed and abstract representation has been learned.

The game Enduro has a unique feature that we leveraged to study nuisance invariance in the gameplay environment. The colors on the screen constantly change as the weather and time of day in the game frequently changes. These stages include sunny, snowy, foggy, dusk, and night-time. Therefore, the pixel space changes dramatically while the overall gameplay dynamics are stable. In fact, we calculated that the mutual information between human left and right actions and the weather/time of day variable equals zero using Scikit-learn's 'mutual_info_regression' function (I(time of day; actions) = 0), which indicates that weather/time of day is a nuisance variable. This metric can only be zero if and only if two random variables are independent. To put this in perspective, the mutual information between weather/time of day and the first principal component of the pixel space is 1.70 (I(time of day; PC 1) = 1.70), and the mutual information of weather/time of day with itself is 1.81. This shows that a large amount of variance in the pixel space is due to these changing weather patterns as the first principal component codes for these conditions.

Although there was no factor that is as obviously a nuisance variable in Space Invaders as the changing colors on the screen was in Enduro, the total number of invader ships on the screen explains a lot of variance in the visual pixel space, and has a high mutual information with the first principal component of the pixel space (I(num. invaders; PC 1) = 1.52). However, in this game the relative positions of the invaders above an agent matter more than their absolute position and the global features, as the invaders above an agent will be in the agent's line of fire and the agent will be in the invader's line of fire. One exception is when there is one invader left and it starts to speed up faster than usual. To quantify this pattern, we calculated that the mutual information between the number of invaders on the screen and left and right actions is relatively low (I(num. invaders; action) = 0.07).

To compute a nuisance invariance score for each filter, we again use sklearn's 'mutual_info_regression' function to calculate the mutual information between a filter and a nuisance variable (weather/time of day for Enduro, number of invaders on the screen for Space Invaders — calculated with downsampled data from sub001 to ease computation).This function outputs a MI score for each of the 7x7 receptive fields, thus these scores were averaged to get a single score per filter. This score was multiplied by -1 to get the inverse of this MI metric, to denote insensitivity of the nuisance rather than encoding of the nuisance. Next, the 64 filter nuisance invariance scores are Pearson correlated with the 64 Neural Predictivity scores in a region. Intuitively, this analysis estimates whether a region prefers filters that are more insensitive to the nuisances (positive correlation) or filters that code for the nuisance (negative correlation). To increase interpretability and enhance the variability across regions that we are most interested in assessing, we z-score this metric across voxels in a participant. Thus, a nuisance invariance score of 0 is average with respect to the other voxels in a participant and the magnitude of the score reflects how many standard deviations it is from the mean.

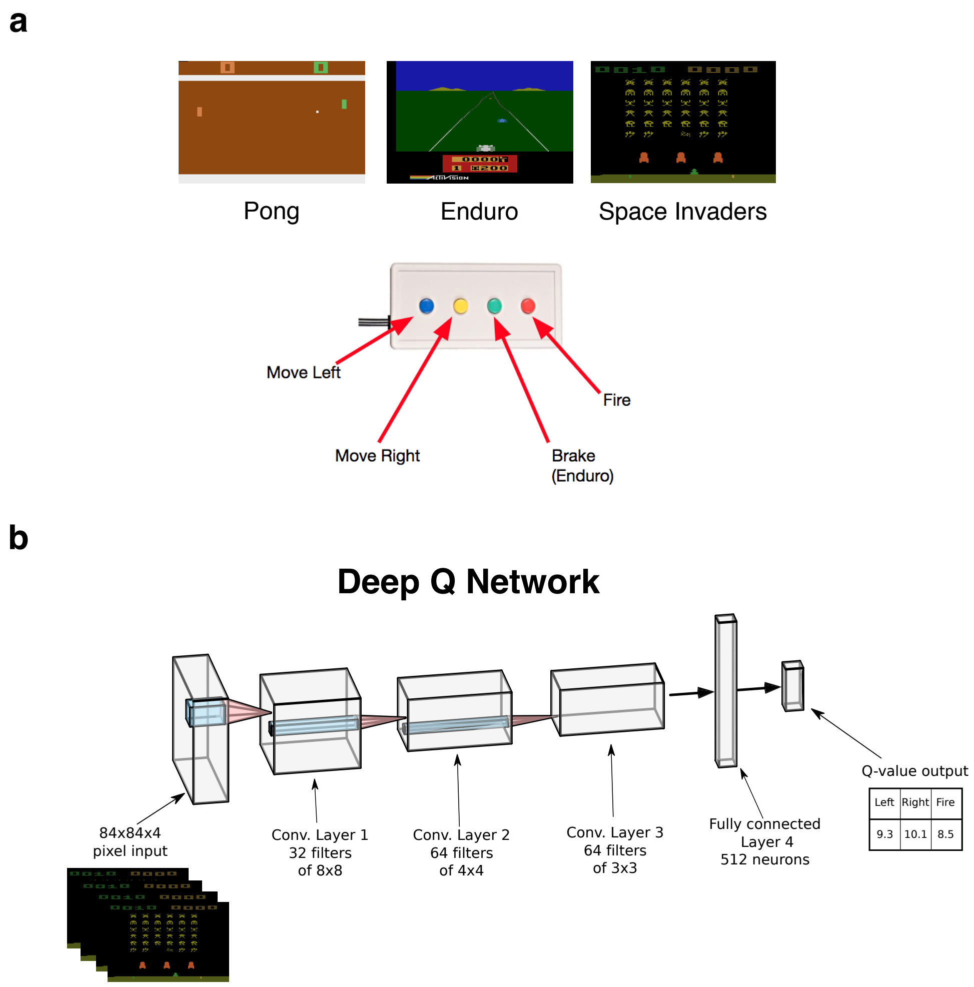


Figure 4.1: **Atari game setup and DQN**

a. Participants played Atari games in the fMRI scanner (Pong, Enduro, and Space Invaders). A button box was used as a controller.

b. DQN is used as a model for how the brain maps high-dimensional inputs to actions. See (Mnih et al., 2015) and Methods for more details.

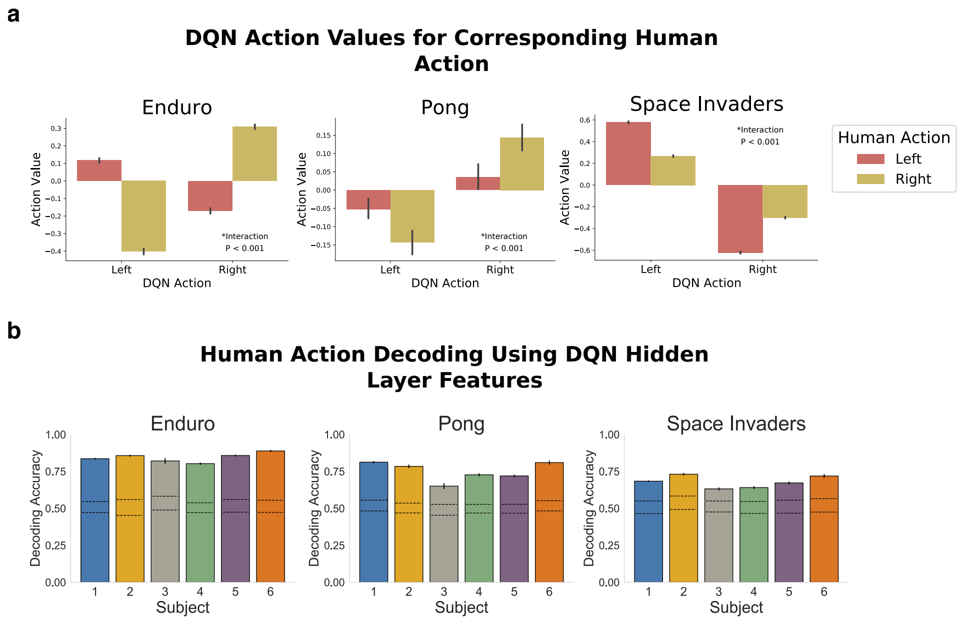


Figure 4.2: **Predicting human behavior using DQN hidden layers.**

a. DQN action values are higher for actions that participants chose. DQN action values depicted for "left" and "right" actions for frames where human participants took either a "left" or "right" action of any combination with fire or brake. Action values correspond to normalized action advantages (see Methods).

b. Human actions are linearly decodeable from the features in DQN hidden layers. Logistic regression models were trained to predict left versus right actions in all games. Features in the model included 100 principal components (PCs) of each DQN layer. Graphs depict cross-validated classification accuracy. Error bars depict SE across 11 cross-validation folds. Dashed lines correspond to the max and min accuracies of null distributions computed with block permutation tests of 1,000 shuffles.

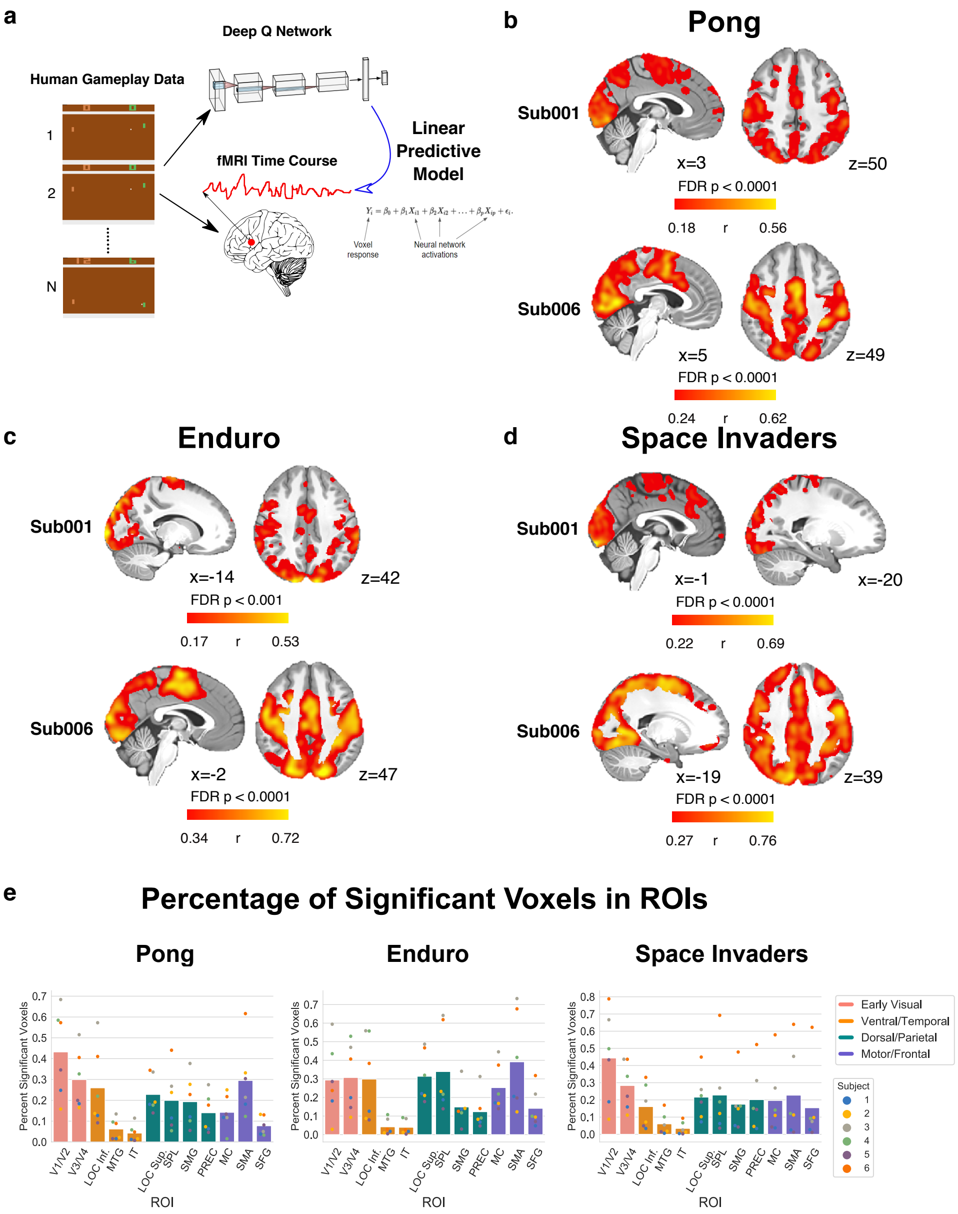


Figure 4.3: **Encoding model: DQN hidden layers mapped to distributed network across the brain, including dorsal stream.**

a. Visualization of encoding model analysis. Human gameplay frames were run through a trained DQN to extract neural network activations in the hidden layers at every time point in an fMRI run. Voxel responses were modeled using ridge regression. The explanatory features included the first 100 PCs from each DQN hidden layer.

b. Voxels mapped to hidden layers for Pong. Cross-validated prediction accuracy uses Pearson correlation between the predicted and actual voxel responses. Whole-brain threshold at $p < 0.001$ or $p < 0.0001$ FDR corrected. Thresholds are determined via cross-validated prediction accuracy against the null distribution using block permutation testing on a subset of voxels. Data are from two participants; others are shown in Figure S2A.

c. Same as in (B), but for Enduro.

d. Same as in (B), but for Space Invaders.

e. Percentage of voxels in a region of interest that are significant in the respective thresholds in (B)–(D). ROIs are noted as V1/V2, V3/V4, LOC Inf. (inferior lateral occipital cortex), MTG (middle temporal gyrus), IT (inferior temporal lobe), LOC Sup. (superior lateral occipital cortex), SPL (superior parietal lobule), SMG (supramarginal gyrus), PREC (precuneus), MC (motor cortex), SMA (supplementary motor area), and SFG (superior frontal gyrus). Plots for individual participants are shown in Figure S2B.

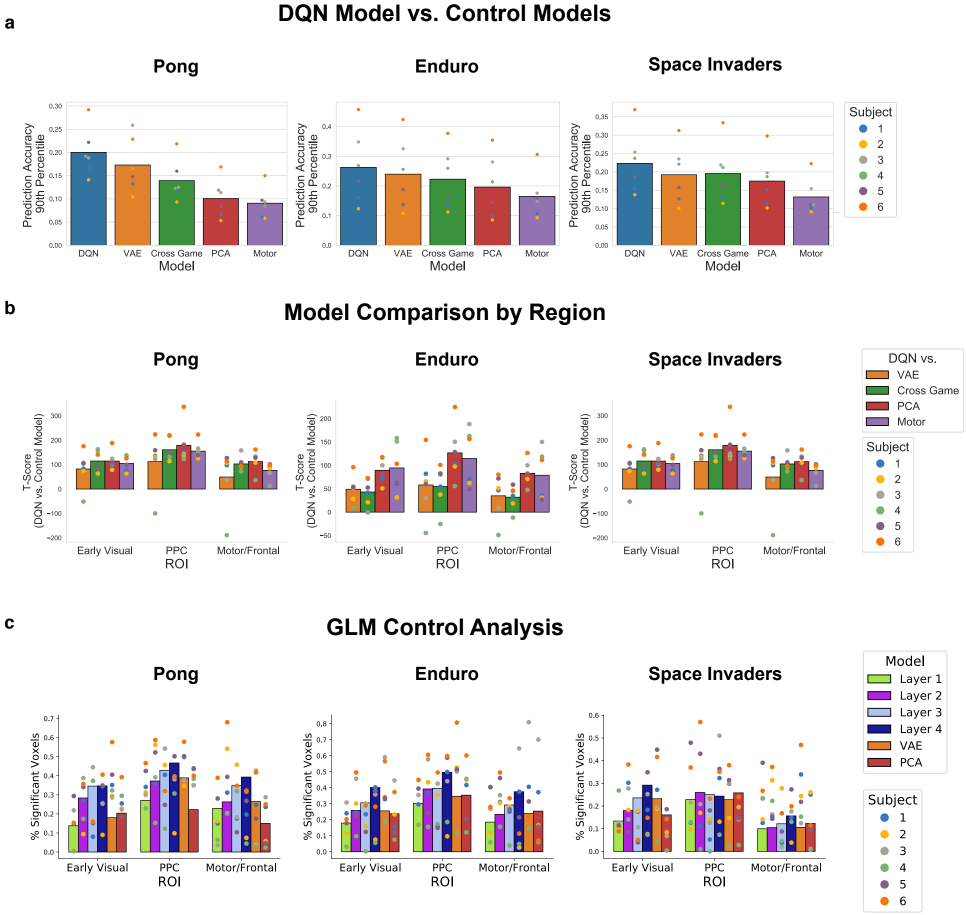


Figure 4.4: **Control models.**

a. Encoding analysis control models: motor regressors, PCA on the input pixels, DQN trained on one of the other games, and a VAE. Bar plots show prediction accuracies for the 90th percentile of prediction accuracies across the whole brain for each model (averaged across six participants with each participant's values shown). Boxplots for distribution of scores in the upper 20th percentile for each model and participant shown in Figure S4B.

b. T-scores by region of interest comparing DQN prediction accuracies to prediction accuracies from control models. T-values reflect average T-scores across participants with each participant's T-scores shown. Plots for individual participants are depicted in Figure S5A.

c. Percentage of significant voxels for each ROI in a GLM where all DQN layers, VAE model, and PCA model compete for variance ($p < 0.001$ FWER corrected, cluster level, F-test across 10 PCs representing a model's regressors).

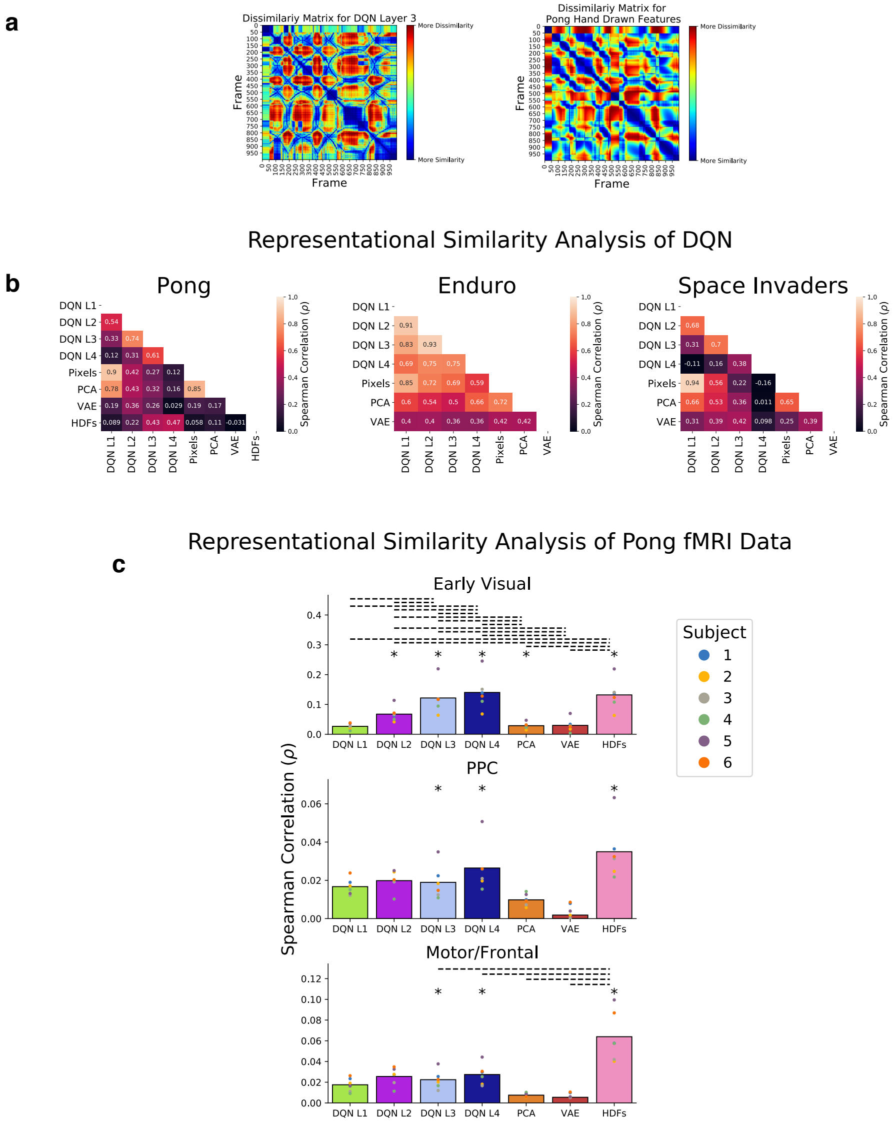


Figure 4.5: **Representational similarity analysis.**
a. Illustrations of what dissimilarity matrices (DSMs) look like for Pong. DSMs represent pairwise comparisons of model representations across time, depicted here for the first 1,000 frames in an example Pong run. The DSM on the left represents the DSM for DQN layer 3 and the DSM on the right represents the DSM for the hand-drawn features in Pong: the positions of the two paddles, the ball position, and the ball's velocity.
b. Representational similarity analysis on DQN. Correlations of all the model DSMs for all games and also the hand-drawn features (HDFs) for Pong. The internal representations in Pong become more dissimilar to the pixel space and PCA model and more similar to the hand-drawn features from DQN layers 1–4. DQN representations in later layers also become more dissimilar to the input space in Space Invaders.

c. Representational similarity analysis on fMRI data for Pong. fMRI DSMs for three ROIs were correlated with model DSMs including HDFs, each layer of DQN, PCA, and VAE. Asterisks ($*$) above bars indicate significance in six out of six subjects (block permutation tests, $p < 0.01$, FWER corrected for multiple comparisons). Dotted lines above bars indicate significant differences between models in six out of six subjects (block permutation tests, $p < 0.01$, FWER corrected for multiple comparisons). All brain areas in all subjects were significantly correlated to the HDF DSM and DQN layers 3 and 4. See Figure S6 for individual subject plots.

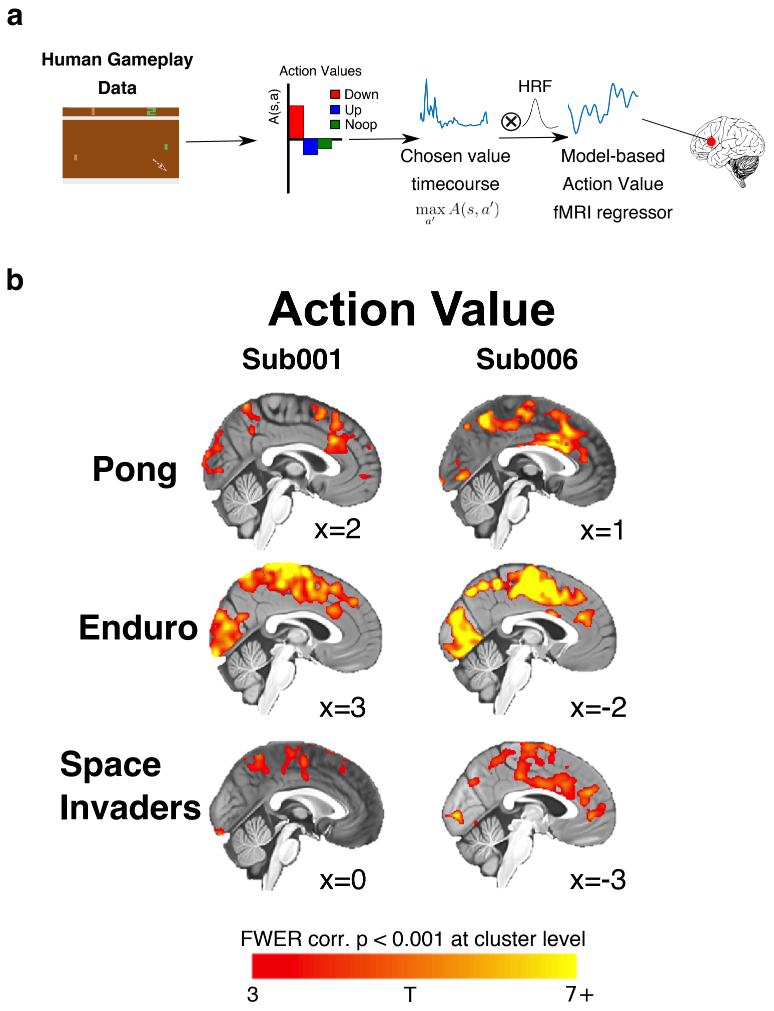
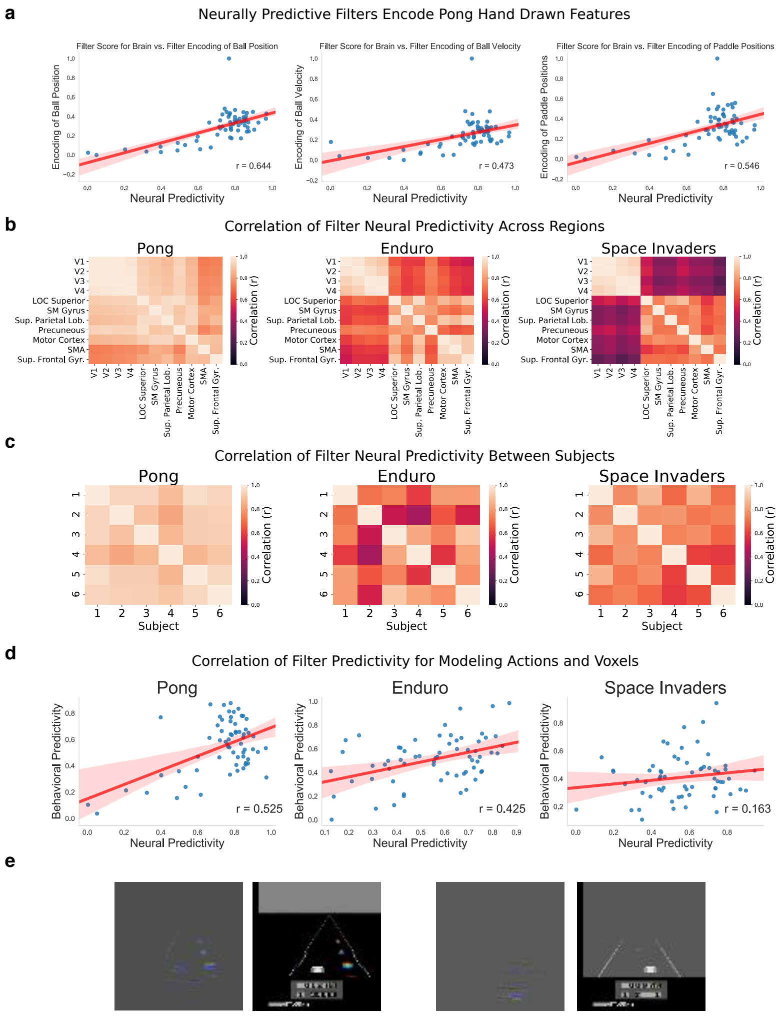


Figure 4.6: **Action value results.**

a. Depiction of action value GLMs. Human gameplay frames were run through DQN to evaluate action/chosen values. Traces were then downsampled to 10 Hz and convolved with a hemodynamic response function to reveal GLM regressors for action values.

b. Neural encoding of action value in premotor/SMA areas. Whole-brain maps were thresholded at p < 0.001 (FWER corrected, cluster level). Significant representation of action value was also found in primary visual and motor cortex. Other participants are shown in Figure S7.



Good filter example:
Layer 3 Filter 40 in Enduro
detects side of the road and cars
Neural Predictivity Rank: 5

Bad filter example:
Layer 3 Filter 56 in Enduro
detects score on bottom of screen
Neural Predictivity Rank: 56

Figure 4.7: **Filter analyses on brain activity.**

a. Neurally predictive filters in Pong encode the spatial positions of objects. Encoding models were run separately on each layer 3 filter to estimate filter neural predictivity. Separately, each filter was assessed on how much it encoded the hand labeled features in Pong. Significant correlations were found between the filter neural predictivity scores and the metric about how much information the filters encoded about the hand-labeled features in every participant ($p < 0.0001$). The average scores and correlations across participants are plotted.

b. Correlations in filter neural predictivity scores across regions. Neural predictivity scores were correlated across regions to estimate whether the same filters are useful for predicting all neural responses or whether the mapping is more heterogeneous. In both Enduro and Space Invaders, more clustering occurs separating visual, parietal, and motor networks.

c. Filter scores are correlated across participants. There are high correlations across all participants and nearly perfect correlations for Pong.

d. Correlations between neural predictivity and behavioral predictivity. Axes represent normalized scores with worst filter at 0 and best filter at 1. Data aggregated across participants are depicted.

e. Visualization of two example filters using guided backpropagation in Neon. Images to the right of each example represent an image from the human gameplay data that activate the filter the most. Gray images to the left of each example represent which parts of the pixel space affect the activation of the filter the most from this input image. Red, green, and blue reflect pixels that changed across the frames in the input. Five randomly selected filters for each game are also visualized in Figure S8A.

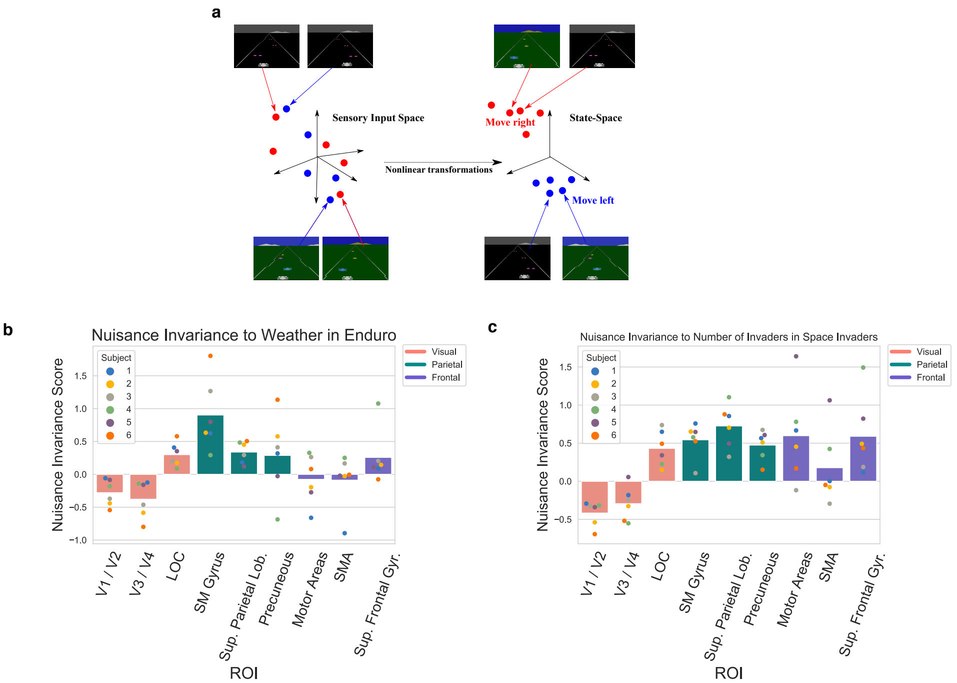


Figure 4.8: **Representations become more insensitive to nuisances in posterior parietal cortex.**
a. Illustration of what a useful representation would do in Enduro. The sky color changes frequently, but these changes have no effect on human actions. The input space on the left depicts how situations are clustered by perceptual features such as color in the pixel space. Within each night/day cluster, there are examples of a car in front of an agent on both the right and left. Therefore, one must take opposite actions in each scenario to avoid a collision. A good state-space localizes the positions of the relevant objects independently of visual nuisances. The resulting state-space representation on the right clusters together perceptually dissimilar situations if they share the same underlying semantic meaning for the policy.
b. Nuisance invariance to weather/time of day in Enduro. We calculate a nuisance invariance score in every region. This score is defined as the correlation of a filter's neural predictivity in a region and that filter's nuisance invariance to weather. The motor area ROI includes both the primary motor cortex and premotor cortex.
c. Nuisance invariance to number of invaders on the screen in Space Invaders. We similarly calculate a nuisance invariance score for every region as defined in (B). For the game Space Invaders, the proxy nuisance variable was the number of invaders on the screen.

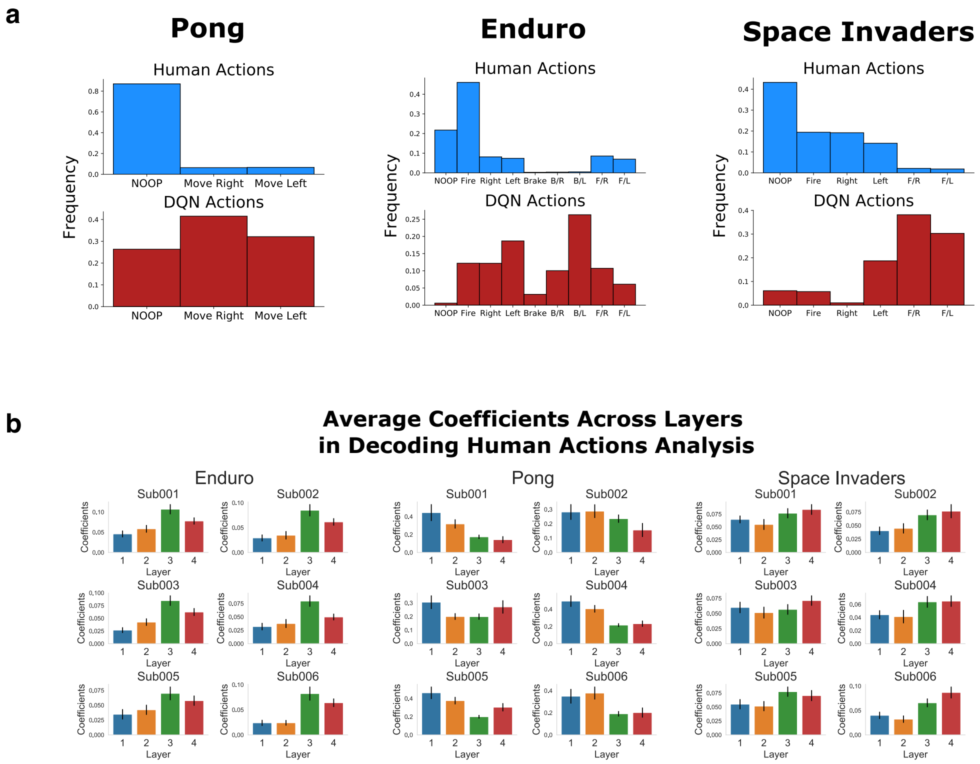


Figure 4.9: **Supplementary Figure 1**

a. Distribution of actions for human participants and DQN. For Enduro and Space Invaders, F/R and F/L correspond to the pairwise combination of fire and move right or move left respectively. For Enduro, B/R and B/L similarly correspond to the pairwise combination of brake and move right or move left.

b. Absolute value of the coefficients in the decoding human actions logistic regression model averaged across layers. For Enduro and Space Invaders, layers 3 and 4 were the most useful for predicting human actions in every participant. For Pong, the contributions of each layer were more heterogeneous across participants with layers 1 and 2 having larger coefficients. Error bars depict SE across neurons in a layer.

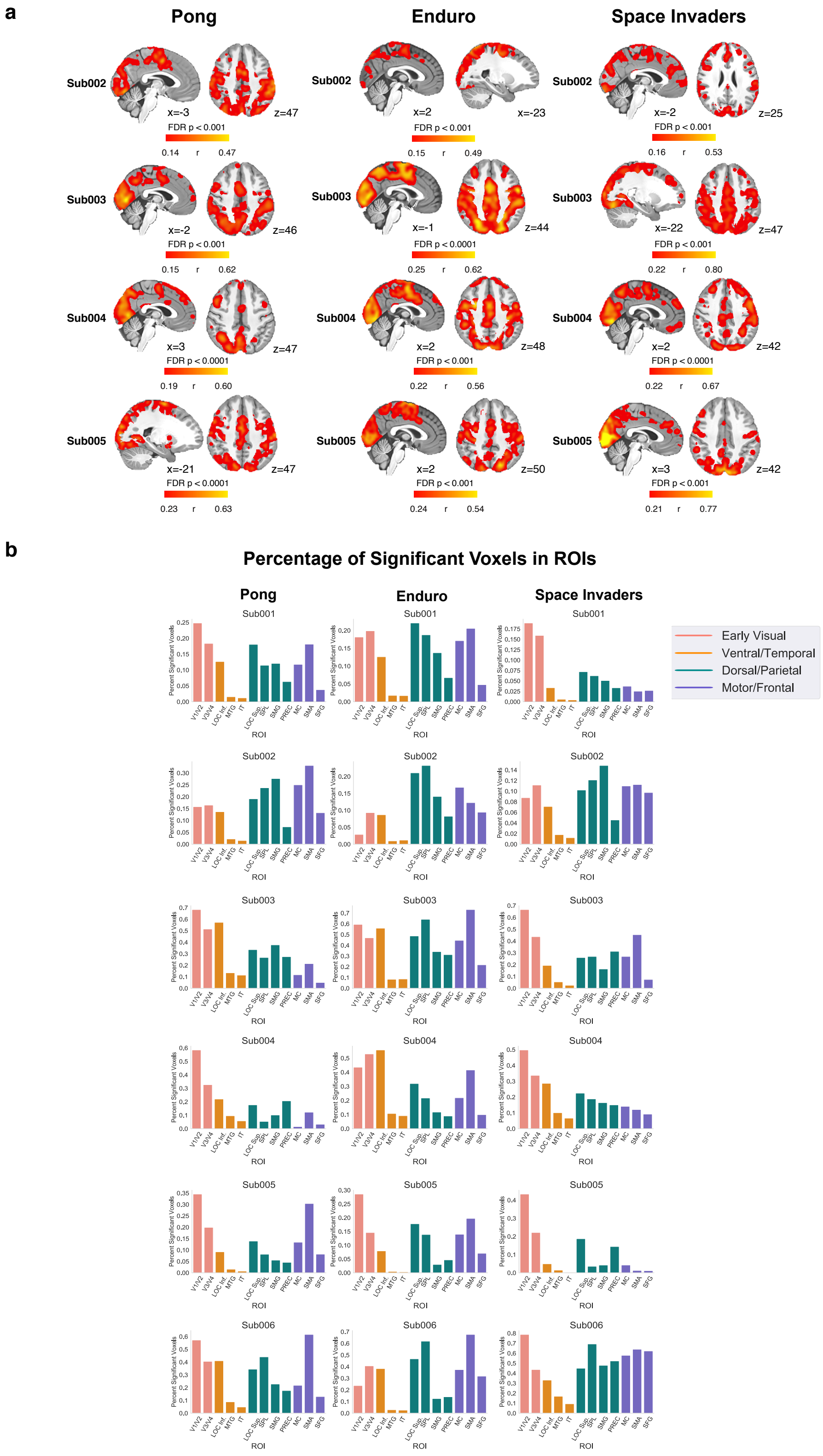


Figure 4.10: **Supplementary Figure 2**

a. Encoding model results for other participants. As in Figure 4.3, whole brain maps are thresholded as noted.

b. Encoding model results by ROI for individual participants. As in Figure 4.3e.

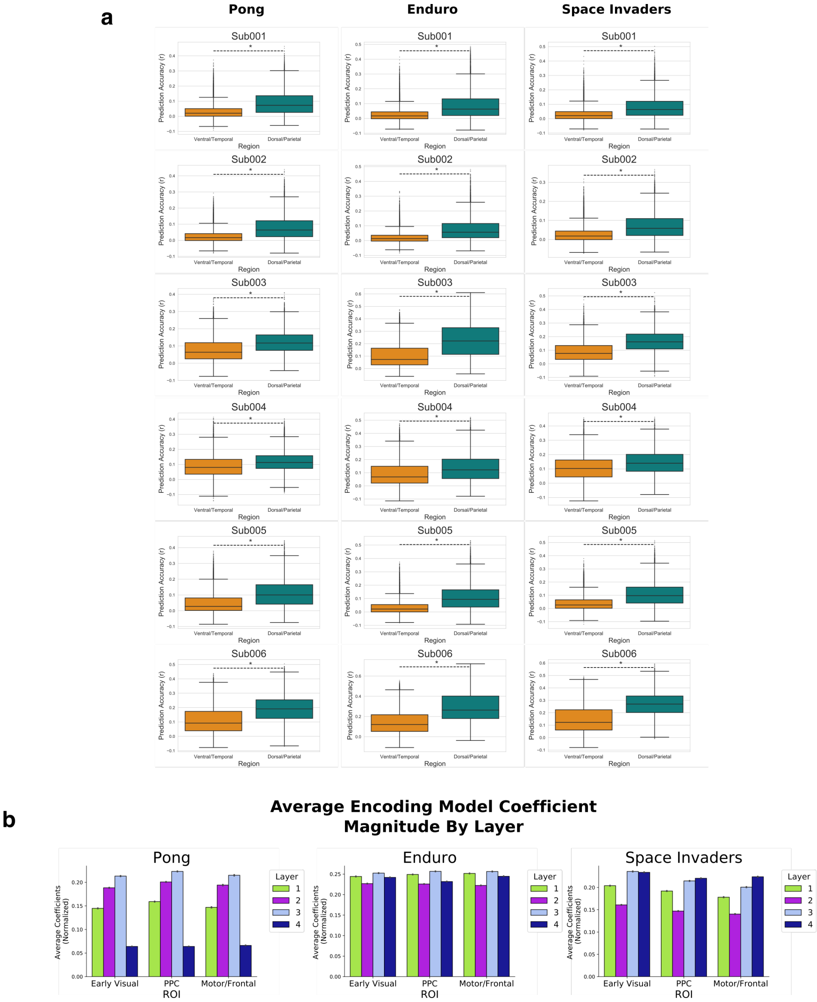


Figure 4.11: **Supplementary Figure 3**

a. Encoding model prediction accuracies are higher in dorsal visual stream than in ventral visual stream. Prediction accuracies (Pearson r) for regions of interest in the dorsal visual stream and/or parietal lobe (superior lateral occipital cortex, superior parietal lobule, supramarginal gyrus, and precuneus) and in the ventral visual stream and/or temporal lobe (inferior lateral occipital cortex, middle temporal gyrus, inferior temporal lobe). Prediction accuracies are significantly higher in dorsal stream/parietal lobe ROIs for all subjects and all games (two-sample T-test, P < 1e-10, signified by *).

b. Average coefficient magnitude by layer. Absolute value of the coefficients by layer in the encoding model analysis averaged across participants. Each layer has 100 coefficients corresponding to 100 principal components of that layer. Error bars reflect SEM across voxels in all participants. Early visual ROI includes V1-V4; PPC includes LOC superior, superior parietal lobule, supramarginal gyrus, precuneus; Motor/Frontal includes motor and premotor cortex, SMA, and superior frontal gyrus.

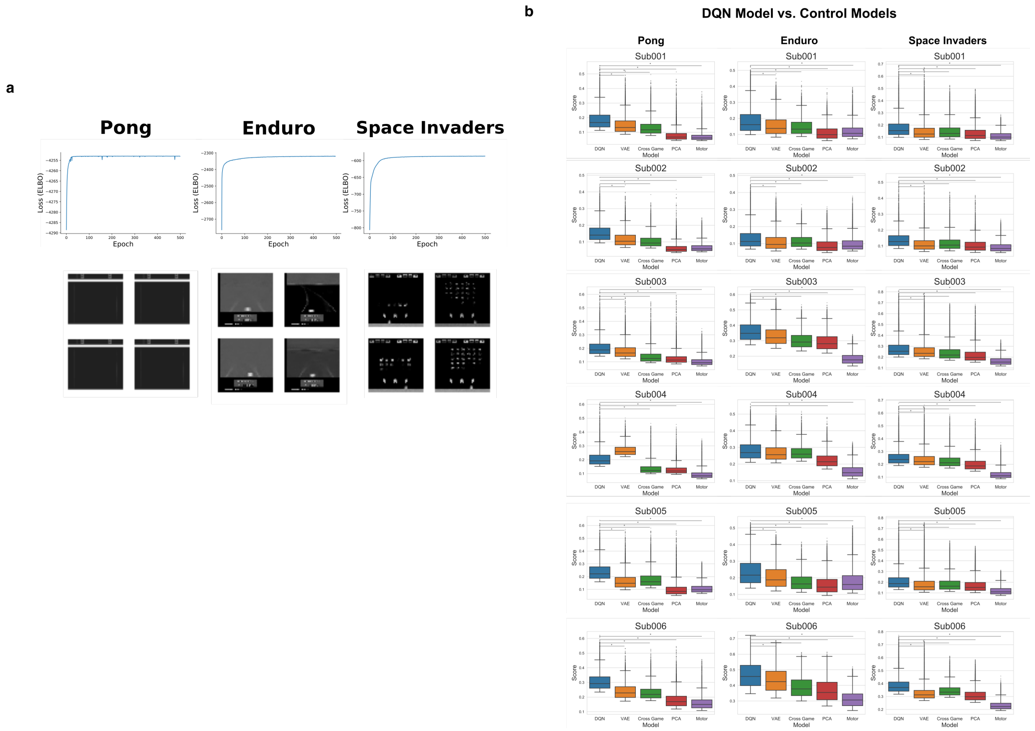


Figure 4.12: **Supplementary Figure 4**

a. VAE Control Model. Top row shows the training performance for the first 500 training epochs. Models were trained to maximize the evidence lower bound (ELBO) on the log-likelihood of the data. Bottom row: examples of VAE generated outputs. Images are generated by sampling latent vectors from a Gaussian distribution and inputting these samples into the VAE decoder.

b. Control models prediction accuracy distributions. Boxplots show distributions of the prediction accuracies in the upper 20th percentile for each model. Outlier points represent the voxels with the highest prediction accuracies in the model.The DQN outperforms all other models in every game and participant other than sub004 (shown P < 1e-10 with * symbols).

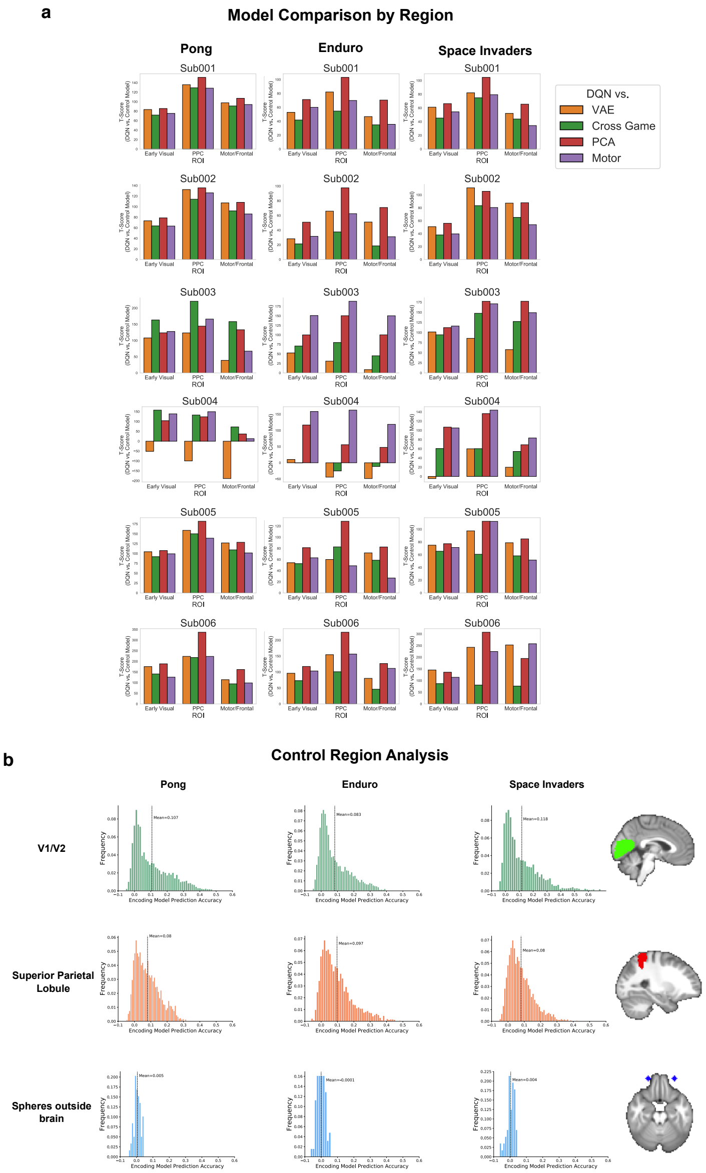


Figure 4.13: **Supplementary Figure 5**
a. DQN vs. control models by region for individual participants. As in Figure 4.4B. **b**. Control Region Analysis. To rule out the possibility that the encoding model analysis is picking up on motion related artifacts or other nuisances that affect the whole fMRI image, we ran the encoding model pipeline on two spheres of air directly outside of the brain (anterior) for sub001. The distribution of scores were around zero for every game and no voxels had significant prediction accuracies. The distribution of scores for V1/V2 and the superior parietal lobule are shown for comparison.

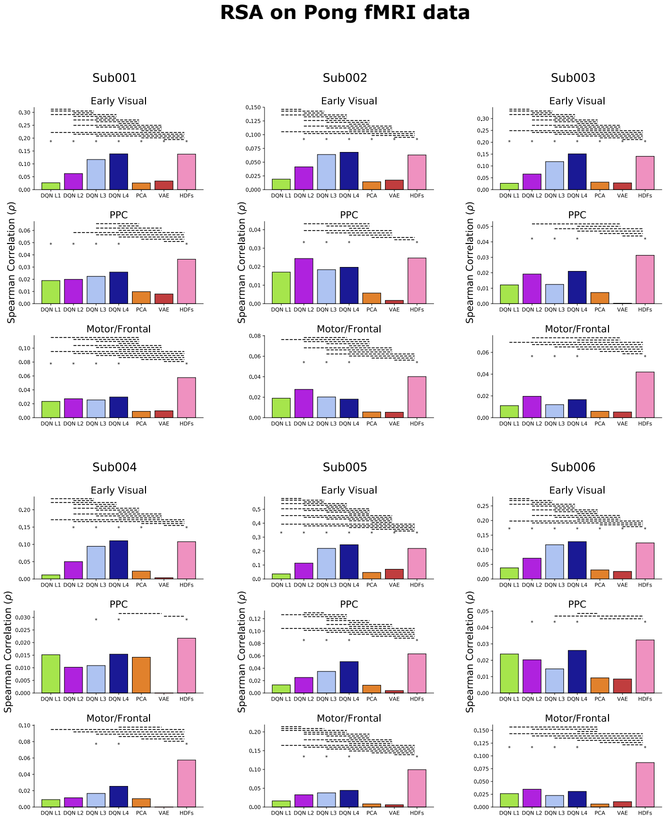


Figure 4.14: **Supplementary Figure 6** Representational Similarity Analysis on fMRI data for Pong for individual subjects. As in Figure 4.5C. Asterisks (*) above bars indicate significance (block permutation tests, P < 0.01, FWER corrected for multiple comparisons). Dotted lines above bars indicate significant differences between models (block permutation tests, P < 0.01, FWER corrected for multiple comparisons).

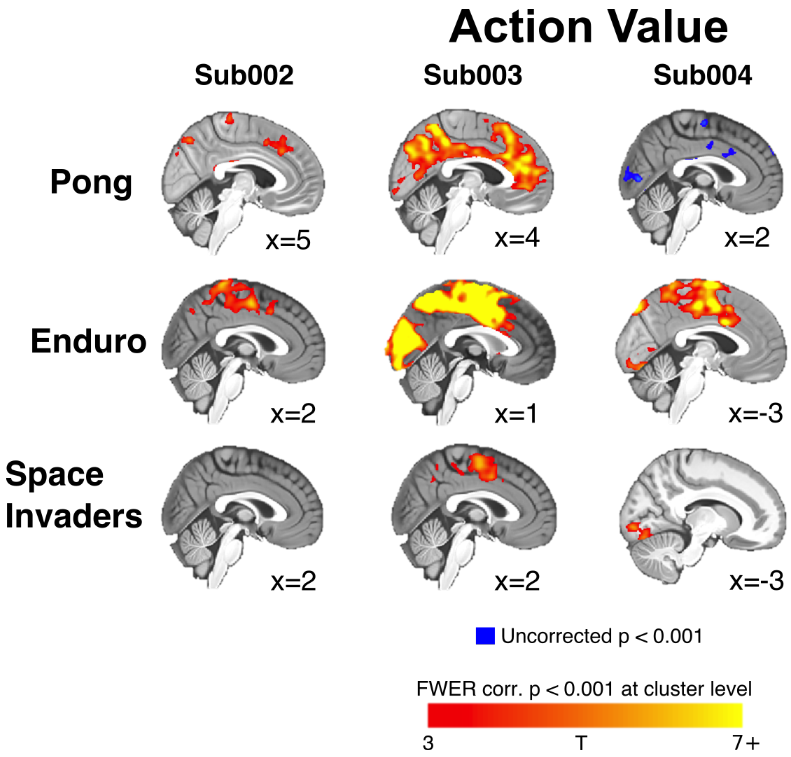


Figure 4.15: **Supplementary Figure 7**
Representational Similarity Analysis on fMRI data for Pong for individual subjects. As in Figure 4.5C. Asterisks (*) above bars indicate significance (block permutation tests, P < 0.01, FWER corrected for multiple comparisons). Dotted lines above bars indicate significant differences between models (block permutation tests, P < 0.01, FWER corrected for multiple comparisons).

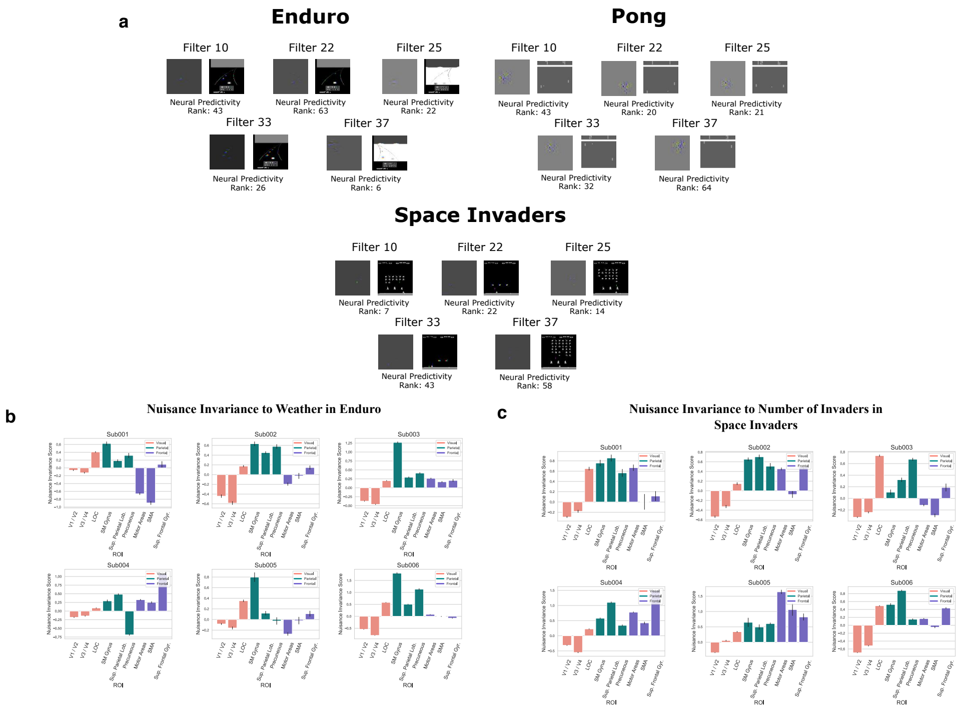


Figure 4.16: **Supplementary Figure 8**

a. Visualization of example DQN layer 3 filters using guided backpropagation. Five randomly selected filters for each game are visualized for each game (as in Figure 4.7E), along with their Neural Predictivity rank. Images to the right of each example represent one image from the human gameplay data that activates the filter very highly. Gray images to the left of each example represent which parts of the pixel space affect the activation of the filter the most from this input image. Red, green, and blue colors reflect pixels that changed across the frames in the input.

b. Nuisance invariance to weather/time of day in Enduro in individual participants. Results depicted as in Figure 4.8B. The Motor Areas ROI includes both the primary motor cortex and premotor cortex.

c. Nuisance invariance to number of invaders on the Screen in Space Invaders in individual participants. Results depicted as in Figure 4.8C.

	DQN	DQN (Mnih 2015)	Sub 001	Sub 002	Sub 003	Sub 004	Sub 005	Sub 006	Rand. Agent
Pong	17.2 ±2.7	18.9 ±1.3	-0.6 ±3.4	0.7 ±2.6	-1.9 ±4.3	1.5 ±3.2	-3.1 ±2.6	-0.1 ±2.2	-20.6 ±0.9
Enduro	348.0 ±108	301.8 ±25	230.0 ±37	201.7 ±5.1	202.2 ±3.7	198.3 ±7.4	201.6 ±1.8	199.5 ±4.1	0
Space Invaders	910 ±335	1976 ±893	949 ±514	890 ±601	1399 ±827	1242 ±632	778 ±280	664 ±336	158 ±135

Table 4.1: **Game performance for all participants and the DQN agent we trained.**

DQN from Mnih 2015 included for reference along with a random agent. ± refers to standard deviation of scores. For evaluation, DQN tested with ϵ-greedy policy (ϵ = 0.05) to minimize risk of evaluating an overfit policy. Subject scores were often limited by game cutting off prematurely due to time constraint of 8 minute runs.

DISCUSSION

Summary

Pioneering work in the tangential fields of decision neuroscience, neuroeconomics, and reinforcement learning has begun to uncover the neurobiology of value-based decision making (Rangel, Camerer, and Montague, 2008; Glimcher and Fehr, 2013). Each option or action up for consideration when making a decision is assigned a value related to the expected utility of the decision outcomes. Options can then be compared in this abstract value space. A multitude of neuroimaging studies in humans have isolated subjective value representations in the ventromedial prefrontal cortex (vmPFC) and orbitofrontal cortex (OFC) (Levy and Glimcher, 2012; Clithero and Rangel, 2014). An important open question emerges from this research: how is this value code constructed?

This thesis outlines a theory for how value is constructed in the brain: stimuli are decomposed into their constituent attributes which are then integrated to compute an abstract value signal that is used to compare options and select actions. The nature of the attributes that are integrated is highly dependent on what features are useful within the context of the task. In Chapter 2, we demonstrated that food items are evaluated by a weighted combination of their constituent nutrient attributes, such as protein, fat, carbohydrates and vitamin content. Items themselves can also become features when valuing bundles of multiple items. This is described in detail in Chapter 3. By using a connectionist deep reinforcement learning model, we generalize this feature integration approach to an environment with more real-world complexity, Atari video games. In naturalistic and high-dimensional environments such as these games, features are processed in a hierarchical topography, as multistep nonlinear transformations of the input reshape a sensory space into a task representation that encodes the high-level features that are relevant for maximizing reward. The value of states and actions can then be constructed as a nonlinear weighted combination of these high-level features.

By decomposing stimuli into features and integrating across them, even the value of novel stimuli that have never been encountered can be actively constructed. Humans flexibly do this in daily life when they explore novel dishes at a new restaurant or drive in neighborhoods they have never been in before by representing stimuli in a multidimensional feature space. When you read about a new dish on a restaurant menu, you use this information to infer its nutrient makeup and flavor profile and can compute how rewarding its expected to be as a function of these features (and potentially other higher-level features, discussed below). In the latter example, you do not need to relearn the driving process from scratch when you drive on a new street, you can generalize from your past experience in similar states. Moreover, this generalization process can operate at the level of high-level features that may be divorced the from high-dimensional stream of sensory input. The concept of 'turn left' should implement similar motor actions irrespective of the colors of the cars on the road or the amount of clouds in the sky even if these irrelevant variables dramatically change the image directly projected on the retina.

Additionally, the values of stimuli can rapidly change depending on the external or internal context, and decomposing a stimulus into features allows them to be reweighted between contexts (O'Doherty, Rutishauser, and Iigaya, 2021). For example, fatty foods can become less desirable after consuming a double cheeseburger or more valuable after a 24 hour fast. Therefore, this contextual information needs to be integrated as well with the other appetitive and aversive properties of a stimulus. Further research is needed however to uncover the exact mechanisms of how interoceptive information and other contextual information is combined with the low-level and high-level features of a stimulus. Similarly, the broader implications of each of the projects presented in this thesis bring forth exciting new research questions. These implications and further directions will be outlined below.

Broader Implications and Future Directions

In Chapter 2, we apply this weighted feature integration theory of value construction to the evaluation of food items. The value of a food was best predicted by the beliefs about its density of carbohydrates, protein, fat, and vitamins. Crucially, the subjective ratings participants gave about the nutrient factors was more predictive of value than the objective quantities of nutrients. This demonstrates that beliefs about a stimulus's features are also actively constructed and inferred rather than recovered with perfect fidelity from the objective world, and these beliefs are used downstream to compute value. This process may account for framing effects and other context-dependent modulations of subjective value (Tversky and Kahneman, 1985), as a person's beliefs about a stimulus's attributes will be influenced by how that stimulus is presented. Marketers understand how to highlight the positive features of a good: would you rather have 80% fat free yogurt or yogurt with 20% fat?

Additionally, humans likely use other features other than the nutrient attributes when computing the value of a food item. Before an item is consumed for the first time, its visual properties probably play a large role computing a value. An interesting future direction would involve including the visual properties of a stimulus in combination with the nutrient factors in a model predicting food preference, similarly to what has been done with visual art (Iigaya et al., 2020). High-level features, such as brand, food category (ie. snack, lunch, dessert, cuisine type), healthiness (Hare, Camerer, and Rangel, 2009), personal history, and cultural factors (Rozin and Vollmecke, 1986) also likely influence a persons food preferences. There may also be nonlinear interactions between the attributes, even between the sensory features and high-level features. For example, a salty dessert may be less desirable than a sweet one. The attribute integration framework presented in this thesis is amenable to all of these caveats. Our model may be expanded upon with the inclusion of many types of features at multiple levels of abstraction and by adding nonlinear interactions between features.

In the neural data, we used multivariate pattern analysis (MVPA) to localize information about the nutrient factors in lateral OFC, while value signals were found in both lateral and medial OFC. Since the lateral OFC contains the secondary taste cortex (Rolls, 1996), it is unclear whether stimulus features are always encoded in lOFC, or if this computation is specific to food valuation. On the other hand, lOFC also receives input from the visual, auditory, olfactory, and somatosensory systems (Rolls, 1996; Grabenhorst and Rolls, 2011), which suggests that its role in attribute integration may generalize to other types of stimuli beyond food. Our analyses uncovered no evidence of mOFC or lOFC encoding information about low-level visual features, but this may potentially be because these features are not utilized in the computation of value in our experiment. A clarifying experiment could perform the same analysis when participants evaluate non-comestible consumer goods and examine whether stimulus attributes can be decoded from lOFC. A related question concerns where high-level features are encoded when computing the value of a good. When evaluating a piece of art, there is a hierarchical organization of feature encoding, such that early visual regions in the ventral stream encode low-level features and higher-order visual areas in the temporal cortex represent higher-level features (Iigaya et al., 2020). There may similarly be a topographical gradient of organization of feature encoding when one values consumer goods, within OFC or distributed throughout other cortical areas.

More work is also needed to characterize how the regions that contain information about stimulus features interact with each other and vmPFC/OFC value regions during the value integration process. Our research sheds some light on this topic at the level of functional connectivity, as effective connectivity between mOFC and lOFC increased at the time of valuation. However, due to the limited temporal and spatial resolution of fMRI, these analyses only provide a coarse grained picture about the circuit level mechanism. Detailed electrophysiological and causal perturbation studies (with tools like optogenetics) need to be performed to delineate exactly how attribute information is projected to vmPFC/mOFC. It is also likely that recurrent connections and other top-down modulatory processes influence the coding of stimulus features. Given the context, some features are more important than others and therefore attention mechanisms and feedback connections may modify the weights of the integration process accordingly.

The features used to compute the value of an option are also not always sensory properties or high-level stimulus attributes; they can be items themselves. In Chapter 3, we describe how bundles of items are valued as a subadditive function of the values of the constituent items. Therefore, the value construction process can operate over higher-level abstractions if the decision-making context affords decomposing the stimuli in this way.

In future work, we wish to examine how feature representations change depending on the context. Are substimulus attributes such as nutrient factors encoded by the same brain areas when valuing an individual item and valuing a bundle? Or does a region like lOFC only encode the features relevant to the current level of analysis required by the context, such that it encodes the substimulus attributes when items are valued in isolation and only encodes representations of items and their values when bundles are valued? Relatedly, another open question concerns how this hiearchical multistep valuation process happens at the circuit level. Are values of the constituent items evaluated first and then integrated subsequently? Or is this all happening in pallelel? Methods with higher temporal resolution are needed to test these hypotheses.

As discussed in Chapter 3, the valuation of bundles of goods can be modulated by their relationship to each other. Substitutes are alternative goods that provide the same function, such as coffee and tea. Complements are goods with synergistic effects when used together, such as pasta and pasta sauce. We posit that these interactive effects happen at the level of stimulus attributes. A subsequent experiment could select an item set that systematically varies how the constituent items of a bundle fit together, and model how stimulus features interact to modulate value.

Our results also show that the value representation in PFC rescales to the distribution of values in a context as one switches between levels of the valuation hierarchy. It is unknown whether similar normalization processes operate over stimulus attributes. Recent work suggests that range adaptation and hierarchical normalization processes occur at the attribute level (Louie, Glimcher, and Webb, 2015; Soltani, De Martino, and Camerer, 2012; Hunt, Raymond J Dolan, and Behrens, 2014). Therefore, this canonical neural computation may be implemented at multiple steps of the decision-making process (Carandini and Heeger, 2012). Additionally, this computation may be ubiquitous because it efficiently encodes and transmits information, in accordance with the efficient coding hypothesis (Louie and Glimcher, 2012).

The experiments presented in Chapters 2 and 3 precisely control the stimuli participants evaluate and use event-related designs that afford segregation of various decision-related signals. However, decision-making in the real-world often does not embody this simplicity. The brain continually processes a stream of high-dimensional exteroceptive and interoceptive input and must extract the features that are relevant from the noise. In chapter 4, we tackle the question of how the brain makes decisions in more ecologically valid environments by scanning participants as they played Atari video games. The model we used to analyze behavior and neural data during Atari gameplay is a high-dimensional extension of the weighted integration models described in chapters 2 and 3. This model, the deep Q-network (DQN), combines the deep learning approach with reinforcement learning, which allows it to learn to play dozens of games at superhuman levels (Mnih et al., 2015). Deep neural networks in general have an outstanding ability to extract useful features from naturalistic raw input (LeCun, Bengio, and Hinton, 2015). They do this by processing input features with connection weights that are integrated into activations of artificial neurons at the next layer. Another step of weighted integration occurs to nonlinearly combine the features represented by neural activations at this layer. This process repeats again and again to construct a hierarchical set of features that are useful for the objectives of the task. Deep learning is therefore a useful tool in computational neuroscience for multiple reasons.

First, deep learning provides methods for extracting stimulus features in a data-driven way, in contrast to hand labeling the features the brain might represent. In environments as complex as the Atari games, it is difficult to hand label relevant features or even identify what features are important. By training DQN, a large set of features at multiple levels of complexity are encoded in its intermediate layers. Using these features, we could predict neural responses throughout the dorsal visual pathway and posterior parietal cortex (PPC). Additionally, this feature set outperformed control models that similarly computed features by exploiting the structure of the input data (a variational autoencoder and principal component analysis). We believe that DQN outperforms these models because it incorporates reward information and links perception to action and reward.

Secondly, as artificial networks are directly inspired by the brain, deep learning is also useful as a direct model of how neural systems solve tasks. Even though the deep learning framework is only a coarse grained analogue of a neural circuit, and a lot of the important biological structure is stripped away, research in the past decade has identified many similarities between deep neural networks and the brain at the representational level (Eickenberg et al., 2017; Güçlü and Gerven, 2015; Khaligh-Razavi and Kriegeskorte, 2014; Wen et al., 2018; Yamins, Hong, et al., 2014; Yamins and DiCarlo, 2016; Wang et al., 2018; Iigaya et al., 2020). Therefore, the weighted integration of attributes may be the canonical computation performed by the brain, and may explain not only value-based decision-making but other aspects of cognition such as object recognition (Yamins and DiCarlo, 2016), audition (Kell et al., 2018), and language (Schwartz and Mitchell, 2019).

Our results suggest that DQN and the human brain converge to a similar state-space representation that is useful for solving the task. Training separate encoding models with stimulus features from different convolutional filters demonstrated that the filters that are most predictive of voxel activity and also predictive of behavior. Additionally, this filter selectivity is highly correlated between participants. These analyses suggest that different people and DQN utilize a shared set of features that are useful for evaluating states and actions. For Pong, this common space represents high-level features such as the spatial positions of the relevant objects in the game. In Enduro and Space Invaders, PPC encodes abstract features that are more invariant to task-irrelevant features like color.

These findings point to several key takeaways about what makes a good task representation in environments of real-world complexity. High-dimensional input is projected to a lower-dimensional space that has disentangled a sensory manifold into an abstract space that represents behaviorally relevant features (DiCarlo and Cox, 2007; Higgins et al., 2018a). These features are representative of the data generating factors and elements of the environment that can acted on or controlled. Thus, they may represent affordance-like features that are codetermined by the agent and the environment rather than objective properties of a stimulus (**rosch1991embodied**; Gibson, 1977). The relevant data generating factors that are encoded in Pong's abstract state-space include the spatial positions of the ball and paddles. This information needs to be carefully extracted from the input, since the geometry of the pixel space is more reflective of low-level visual properties than these object features. For Enduro, the optimal state-space should be abstracted away from the sensory input even more, and become invariant to nuisance features like the changing colors in the background. The latter property is crucial for an agent to generalize beyond the exact data distribution it is trained on. For example, an agent only trained in the sunny day context in Enduro, would break when tested on the snow or night-time context if its internal representation were not divorced from the sensory particularities of the sunny day context. In contrast, if it possesses an internal representation of higher-level concepts, such as the positions of the cars on the road in Enduro and the ball and paddles in Pong, it is more robust to dramatic changes in low-level visual properties (which happens all the time for humans even due to something as simple as a change in ambient light levels). This concept connects to the idea of factorized representations, which isolate structural representations of the world from the raw sensory information they are associated with. Factorized representations have been linked to the interaction between the entorhinal cortex and hippocampus in more cognitive tasks (Behrens et al., 2018; Manns and Eichenbaum, 2006). Here, we show that the posterior parietal cortex plays a fundamental role in encoding an analogous abstract representation during visuomotor tasks like Atari video games.

Our findings highlight the contribution of the dorsal visual stream, the parietal cortex in particular, as encoding abstract state-space representations during Atari gameplay. In contrast, the orbitofrontal cortex (OFC) has also been implicated in encoding of stimulus features in the work described in Chapter 2. Additionally, other studies have described how OFC plays a role in the representation of task state-space (Niv, 2019; Schuck et al., 2016; Wilson et al., 2014). The OFC has been found to contribute in situations where the relevant states have to be inferred on the basis of partially observable information (Niv, 2019). In the present study we did not find evidence for the involvement of the OFC, which may pertain to the fact that the Atari games we used here involve fully observable states, thereby not relying on the hidden state inference process attributed to the OFC. Moreover, in order to perform the Atari game tasks, it is necessary for individuals to rapidly select actions based on fast-moving, multivariate sensory information. By contrast, previous tasks implicating the OFC relied instead on trial-based tasks with low dimensional (Schuck et al., 2016; Wilson et al., 2014). Thus, the involvement of the parietal cortex may be especially pertinent under conditions where rapid visuomotor integration is required for task performance.

In future work, we hope to uncover not only what properties make up a useful state representation, but how state representations change throughout the learning process and across human players of varying skill levels. How does the brain reshape its representation of a task on the path to expertise? One possibility is that DQN and other deep RL algorithms become the most predictive of behavior and neural data once a human reaches a certain performance level or within certain high performing blocks of gameplay. Or alternatively, human experts may encode unique task representations that branch away from the local optima that deep RL algorithms tend to converge to. Additionally, it would be informative to project the neural patterns within different humans and artificial agents of the same skill level to a common representational space (Chen et al., 2015; Haxby, Guntupalli, et al., 2011; Lu et al., 2018). This would afford us a way to examine exactly how different substrates and different instantiations of a network in the same substrate can converge to similar state-space representations.

BIBLIOGRAPHY

Abler, Birgit et al. (2006). “Prediction error as a linear function of reward probability is coded in human nucleus accumbens”. In: *Neuroimage* 31.2, pp. 790–795.

Achille, Alessandro and Stefano Soatto (2018). “Emergence of invariance and disentanglement in deep representations”. In: *The Journal of Machine Learning Research* 19.1, pp. 1947–1980.

Anand, Ankesh et al. (2019). “Unsupervised state representation learning in Atari”. In: *Advances in Neural Information Processing Systems* 32.

Andersen, Richard A. and Christopher A. Buneo (2002). “Intentional maps in posterior parietal cortex”. In: *Annual Review of Neuroscience* 25.1, pp. 189–220.

Andersen, Richard A. and He Cui (2009). “Intention, action planning, and decision making in parietal-frontal circuits”. In: *Neuron* 63.5, pp. 568–583.

Ariely, Dan and Simon Jones (2008). *Predictably irrational*. HarperCollins New York.

Avants, Brian B., Nick Tustison, Gang Song, et al. (2009). “Advanced normalization tools (ANTS)”. In: *Insight Journal* 2.365, pp. 1–35.

Barron, Helen C., Raymond J. Dolan, and Timothy E.J. Behrens (2013). “Online evaluation of novel choices by simultaneous representation of multiple memories”. In: *Nature Neuroscience* 16.10, pp. 1492–1498.

Bartra, Oscar, Joseph T. McGuire, and Joseph W. Kable (2013). “The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value”. In: *Neuroimage* 76, pp. 412–427.

Bechara, Antoine et al. (1994). “Insensitivity to future consequences following damage to human prefrontal cortex”. In: *Cognition* 50.1-3, pp. 7–15.

Becker, Gordon M., Morris H. DeGroot, and Jacob Marschak (1964). “Measuring utility by a single-response sequential method”. In: *Behavioral science* 9.3, pp. 226–232.

Behrens, Timothy E.J. et al. (2018). “What is a cognitive map? Organizing knowledge for flexible behavior”. In: *Neuron* 100.2, pp. 490–509.

Bellemare, Marc G. et al. (2013). “The arcade learning environment: An evaluation platform for general agents”. In: *Journal of Artificial Intelligence Research* 47, pp. 253–279.

Bettman, James R., Mary Frances Luce, and John W. Payne (1998). “Constructive consumer choice processes”. In: *Journal of Consumer Research* 25.3, pp. 187–217.

Botvinick, Matthew, Sam Ritter, et al. (2019). “Reinforcement learning, fast and slow”. In: *Trends in Cognitive Sciences* 23.5, pp. 408–422.

Botvinick, Matthew, Jane X. Wang, et al. (2020). “Deep reinforcement learning and its neuroscientific implications”. In: *Neuron* 107.4, pp. 603–616.

Carandini, Matteo and David J. Heeger (2012). “Normalization as a canonical neural computation”. In: *Nature Reviews Neuroscience* 13.1, pp. 51–62.

Carnell, Susan et al. (2012). “Neuroimaging and obesity: Current knowledge and future directions”. In: *Obesity Reviews* 13.1, pp. 43–56.

Chang, Le and Doris Y. Tsao (2017). “The code for facial identity in the primate brain”. In: *Cell* 169.6, pp. 1013–1028.

Chen, Po-Hsuan Cameron et al. (2015). “A reduced-dimension fMRI shared response model”. In: *Advances in Neural Information Processing Systems* 28.

Chib, Vikram S. et al. (2009). “Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex”. In: *Journal of Neuroscience* 29.39, pp. 12315–12320.

Chikazoe, Junichi et al. (2014). “Population coding of affect across stimuli, modalities and individuals”. In: *Nature Neuroscience* 17.8, pp. 1114–1122.

Clithero, John A. and Antonio Rangel (2014). “Informatic parcellation of the network involved in the computation of subjective value”. In: *Social Cognitive and Affective Neuroscience* 9.9, pp. 1289–1302.

Cuayáhuitl, Heriberto, Simon Keizer, and Oliver Lemon (2015). “Strategic dialogue management via deep reinforcement learning”. In: *arXiv preprint arXiv:1511.08099*.

Dabney, Will et al. (2020). “A distributional code for value in dopamine-based reinforcement learning”. In: *Nature* 577.7792, pp. 671–675.

Daw, Nathaniel D. and John P. O’Doherty (2014). “Multiple systems for value learning”. In: *Neuroeconomics*. Elsevier, pp. 393–410.

Daw, Nathaniel D. and Philippe N. Tobler (2014). “Value learning through reinforcement: The basics of dopamine and reinforcement learning”. In: *Neuroeconomics*. Elsevier, pp. 283–298.

Day, Jeremy J. et al. (2007). “Associative learning mediates dynamic shifts in dopamine signaling in the nucleus accumbens”. In: *Nature Neuroscience* 10.8, pp. 1020–1028.

De Araujo, Ivan E. et al. (2008). “Food reward in the absence of taste receptor signaling”. In: *Neuron* 57.6, pp. 930–941.

De Martino, Benedetto et al. (2009). “The neurobiology of reference-dependent value computation”. In: *Journal of Neuroscience* 29.12, pp. 3833–3842.

Deichmann, Ralf et al. (2003). “Optimized EPI for fMRI studies of the orbitofrontal cortex”. In: *Neuroimage* 19.2, pp. 430–441.

DiCarlo, James J. and David D. Cox (2007). “Untangling invariant object recognition”. In: *Trends in Cognitive Sciences* 11.8, pp. 333–341.

Efron, Bradley and Robert J. Tibshirani (1994). *An introduction to the bootstrap*. CRC press.

Eickenberg, Michael et al. (2017). “Seeing it all: Convolutional network layers map the function of the human visual system”. In: *NeuroImage* 152, pp. 184–194.

Espeholt, Lasse et al. (2018). “Impala: Scalable distributed deep-rl with importance weighted actor-learner architectures”. In: *International Conference on Machine Learning*. PMLR, pp. 1407–1416.

Foerde, Karin et al. (2015). “Neural mechanisms supporting maladaptive food choices in anorexia nervosa”. In: *Nature Neuroscience* 18.11, pp. 1571–1573.

Freedman, David J. and Guilhem Ibos (2018). “An integrative framework for sensory, motor, and cognitive functions of the posterior parietal cortex”. In: *Neuron* 97.6, pp. 1219–1234.

Gibson, James J. (1977). “The theory of affordances”. In: *Hilldale, USA* 1.2, pp. 67–82.

Gigerenzer, Gerd and Reinhard Selten (2002). *Bounded rationality: The adaptive toolbox*. MIT press.

Gläscher, Jan et al. (2010). “States versus rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning”. In: *Neuron* 66.4, pp. 585–595.

Glimcher, Paul W. and Ernst Fehr (2013). *Neuroeconomics: Decision making and the brain*. Academic Press.

Gold, Joshua I. and Michael N. Shadlen (2007). “The neural basis of decision making”. In: *Annual Review of Neuroscience* 30, pp. 535–574.

Gollisch, Tim and Markus Meister (2010). “Eye smarter than scientists believed: Neural computations in circuits of the retina”. In: *Neuron* 65.2, pp. 150–164.

Grabenhorst, Fabian and Edmund T. Rolls (2011). “Value, pleasure and choice in the ventral prefrontal cortex”. In: *Trends in Cognitive Sciences* 15.2, pp. 56–67.

Gross, Jörg et al. (2014). “Value signals in the prefrontal cortex predict individual preferences across reward categories”. In: *Journal of Neuroscience* 34.22, pp. 7580–7586.

Güçlü, Umut and Marcel A.J. van Gerven (2015). “Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream”. In: *Journal of Neuroscience* 35.27, pp. 10005–10014.

Ha, David and Jürgen Schmidhuber (2018). “World models”. In: *arXiv preprint arXiv:1803.10122*.

Haarnoja, Tuomas et al. (2018). “Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor”. In: *International Conference on Machine Learning*. PMLR, pp. 1861–1870.

Hanke, Michael et al. (2009). “PyMVPA: A Python toolbox for multivariate pattern analysis of fMRI data”. In: *Neuroinformatics* 7.1, pp. 37–53.

Hare, Todd A., Colin Camerer, Daniel T. Knoepfle, et al. (2010). “Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition”. In: *Journal of Neuroscience* 30.2, pp. 583–590.

Hare, Todd A., Colin Camerer, and Antonio Rangel (2009). “Self-control in decision-making involves modulation of the vmPFC valuation system”. In: *Science* 324.5927, pp. 646–648.

Hare, Todd A., Jonathan Malmaud, and Antonio Rangel (2011). “Focusing attention on the health aspects of foods changes value signals in vmPFC and improves dietary choice”. In: *Journal of Neuroscience* 31.30, pp. 11077–11087.

Harrison, Stephenie A. and Frank Tong (2009). “Decoding reveals the contents of visual working memory in early visual areas”. In: *Nature* 458.7238, pp. 632–635.

Haxby, James V., Andrew C. Connolly, and J. Swaroop Guntupalli (2014). “Decoding neural representational spaces using multivariate pattern analysis”. In: *Annual Review of Neuroscience* 37, pp. 435–456.

Haxby, James V., M. Ida Gobbini, et al. (2001). “Distributed and overlapping representations of faces and objects in ventral temporal cortex”. In: *Science* 293.5539, pp. 2425–2430.

Haxby, James V., J. Swaroop Guntupalli, et al. (2011). “A common, high-dimensional model of the representational space in human ventral temporal cortex”. In: *Neuron* 72.2, pp. 404–416.

Hayhoe, Mary and Dana Ballard (2005). “Eye movements in natural behavior”. In: *Trends in Cognitive Sciences* 9.4, pp. 188–194.

Haynes, John-Dylan (2015). “A primer on pattern-based approaches to fMRI: Principles, pitfalls, and perspectives”. In: *Neuron* 87.2, pp. 257–270.

Hebart, Martin N., Kai Görgen, and John-Dylan Haynes (2015). “The Decoding Toolbox (TDT): A versatile software package for multivariate analyses of functional imaging data”. In: *Frontiers in neuroinformatics* 8, p. 88.

Higgins, Irina, David Amos, et al. (2018). “Towards a definition of disentangled representations”. In: *arXiv preprint arXiv:1812.02230*.

Higgins, Irina, Arka Pal, et al. (2017). “Darla: Improving zero-shot transfer in reinforcement learning”. In: *International Conference on Machine Learning*. PMLR, pp. 1480–1490.

Hosoya, Toshihiko, Stephen A Baccus, and Markus Meister (2005). “Dynamic predictive coding by the retina”. In: *Nature* 436.7047, pp. 71–77.

Howard, James D., Jay A. Gottfried, et al. (2015). “Identity-specific coding of future rewards in the human orbitofrontal cortex”. In: *Proceedings of the National Academy of Sciences* 112.16, pp. 5195–5200.

Howard, James D. and Thorsten Kahnt (2017). “Identity-specific reward representations in orbitofrontal cortex are modulated by selective devaluation”. In: *Journal of Neuroscience* 37.10, pp. 2627–2638.

Hunt, Laurence T., Raymond J Dolan, and Timothy E.J. Behrens (2014). “Hierarchical competitions subserving multi-attribute choice”. In: *Nature Neuroscience* 17.11, pp. 1613–1622.

Iigaya, Kiyohito et al. (2020). “Aesthetic preference for art emerges from a weighted integration over hierarchically structured visual features in the brain”. In.

Jaderberg, Max et al. (2016). “Reinforcement learning with unsupervised auxiliary tasks”. In: *arXiv preprint arXiv:1611.05397*.

Kable, Joseph W. and Paul W. Glimcher (2007). “The neural correlates of subjective value during intertemporal choice”. In: *Nature Neuroscience* 10.12, pp. 1625–1633.

Kahneman, Daniel and Amos Tversky (2013). “Prospect theory: An analysis of decision under risk”. In: *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, pp. 99–127.

Kahnt, Thorsten et al. (2010). “The neural code of reward anticipation in human orbitofrontal cortex”. In: *Proceedings of the National Academy of Sciences* 107.13, pp. 6010–6015.

Kaiser, Lukasz et al. (2019). “Model-based reinforcement learning for Atari”. In: *arXiv preprint arXiv:1903.00374*.

Kay, Kendrick N. et al. (2008). “Identifying natural images from human brain activity”. In: *Nature* 452.7185, pp. 352–355.

Kell, Alexander J.E. et al. (2018). “A task-optimized neural network replicates human auditory behavior, predicts brain responses, and reveals a cortical processing hierarchy”. In: *Neuron* 98.3, pp. 630–644.

Khaligh-Razavi, Seyed-Mahdi and Nikolaus Kriegeskorte (2014). “Deep supervised, but not unsupervised, models may explain IT cortical representation”. In: *PLoS Computational Biology* 10.11, e1003915.

Kietzmann, Tim C et al. (2019). “Recurrence is required to capture the representational dynamics of the human visual system”. In: *Proceedings of the National Academy of Sciences* 116.43, pp. 21854–21863.

Kingma, Diederik P. and Max Welling (2013). “Auto-encoding variational Bayes”. In: *arXiv preprint arXiv:1312.6114*.

Klein-Flügge, Miriam Cornelia et al. (2013). “Segregated encoding of reward–identity and stimulus–reward associations in human orbitofrontal cortex”. In: *Journal of Neuroscience* 33.7, pp. 3202–3211.

Kriegeskorte, Nikolaus, Rainer Goebel, and Peter A. Bandettini (2006). “Information-based functional brain mapping”. In: *Proceedings of the National Academy of Sciences* 103.10, pp. 3863–3868.

Kriegeskorte, Nikolaus and Rogier A. Kievit (2013). “Representational geometry: Integrating cognition, computation, and the brain”. In: *Trends in Cognitive Sciences* 17.8, pp. 401–412.

Kriegeskorte, Nikolaus, Marieke Mur, and Peter A. Bandettini (2008). “Representational similarity analysis-connecting the branches of systems neuroscience”. In: *Frontiers in systems neuroscience* 2, p. 4.

Kriegeskorte, Nikolaus, Marieke Mur, Douglas A. Ruff, et al. (2008). “Matching categorical object representations in inferior temporal cortex of man and monkey”. In: *Neuron* 60.6, pp. 1126–1141.

Kringelbach, Morten L. and Edmund T. Rolls (2004). “The functional neuroanatomy of the human orbitofrontal cortex: Evidence from neuroimaging and neuropsychology”. In: *Progress in neurobiology* 72.5, pp. 341–372.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey Hinton (2012). “Imagenet classification with deep convolutional neural networks”. In: *Advances in Neural Information Processing Systems* 25.

Lake, Brenden M. et al. (2017). “Building machines that learn and think like people”. In: *Behavioral and brain sciences* 40.

Lebreton, Maël et al. (2009). “An automatic valuation system in the human brain: Evidence from functional neuroimaging”. In: *Neuron* 64.3, pp. 431–439.

LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton (2015). “Deep learning”. In: *Nature* 521.7553, pp. 436–444.

LeCun, Yann, Léon Bottou, et al. (1998). “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11, pp. 2278–2324.

Lenc, Karel and Andrea Vedaldi (2015). “Understanding image representations by measuring their equivariance and equivalence”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 991–999.

Lesort, Timothée et al. (2018). “State representation learning for control: An overview”. In: *Neural Networks* 108, pp. 379–392.

Levy, Dino J. and Paul W. Glimcher (2011). “Comparing apples and oranges: Using reward-specific and reward-general subjective value representation in the brain”. In: *Journal of Neuroscience* 31.41, pp. 14693–14707.

– (2012). “The root of all value: A neural common currency for choice”. In: *Current Opinion in Neurobiology* 22.6, pp. 1027–1038.

Lillicrap, Timothy P. et al. (2015). “Continuous control with deep reinforcement learning”. In: *arXiv preprint arXiv:1509.02971*.

Lim, Seung-Lark, John P. O’Doherty, and Antonio Rangel (2013). “Stimulus value signals in ventromedial PFC reflect the integration of attribute value signals computed in fusiform gyrus and posterior superior temporal gyrus”. In: *Journal of Neuroscience* 33.20, pp. 8729–8741.

Lin, Alice, Ralph Adolphs, and Antonio Rangel (2012). “Social and monetary reward learning engage overlapping neural substrates”. In: *Social Cognitive and Affective Neuroscience* 7.3, pp. 274–281.

Louie, Kenway and Paul W. Glimcher (2012). “Efficient coding and the neural representation of value”. In: *Annals of the New York Academy of Sciences* 1251.1, pp. 13–32.

Louie, Kenway, Paul W. Glimcher, and Ryan Webb (2015). “Adaptive neural coding: From biological to behavioral decision-making”. In: *Current Opinion in Behavioral Sciences* 5, pp. 91–99.

Louie, Kenway, Lauren E. Grattan, and Paul W. Glimcher (2011). “Reward value-based gain control: Divisive normalization in parietal cortex”. In: *Journal of Neuroscience* 31.29, pp. 10627–10639.

Louie, Kenway, Mel W. Khaw, and Paul W. Glimcher (2013). “Normalization is a general neural mechanism for context-dependent decision making”. In: *Proceedings of the National Academy of Sciences* 110.15, pp. 6139–6144.

Lu, Qihong et al. (2018). “Shared representational geometry across neural networks”. In: *arXiv preprint arXiv:1811.11684*.

Manns, Joseph R. and Howard Eichenbaum (2006). “Evolution of declarative memory”. In: *Hippocampus* 16.9, pp. 795–808.

McClure, Samuel M., Gregory S. Berns, and P. Read Montague (2003). “Temporal prediction errors in a passive learning task activate human striatum”. In: *Neuron* 38.2, pp. 339–346.

McNamee, Daniel, Mimi Liljeholm, et al. (2015). “Characterizing the associative content of brain structures involved in habitual and goal-directed actions in humans: A multivariate FMRI study”. In: *Journal of Neuroscience* 35.9, pp. 3764–3771.

McNamee, Daniel, Antonio Rangel, and John P O’Doherty (2013). “Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex”. In: *Nature Neuroscience* 16.4, pp. 479–485.

Miller, Kenneth D. and David J.C. MacKay (1994). “The role of constraints in Hebbian learning”. In: *Neural Computation* 6.1, pp. 100–126.

Mishkin, Mortimer, Leslie G. Ungerleider, and Kathleen A. Macko (1983). “Object vision and spatial vision: Two cortical pathways”. In: *Trends in Neurosciences* 6, pp. 414–417.

Mitchell, Tom M. et al. (2004). “Learning to decode cognitive states from brain images”. In: *Machine Learning* 57.1, pp. 145–175.

Mnih, Volodymyr et al. (2015). “Human-level control through deep reinforcement learning”. In: *Nature* 518.7540, pp. 529–533.

Mohamed, Shakir and Danilo Jimenez Rezende (2015). “Variational information maximisation for intrinsically motivated reinforcement learning”. In: *Advances in Neural Information Processing Systems* 28.

Morgenstern, Oskar and John Von Neumann (1944). *Theory of Games and Economic Behavior*. Princeton University Press.

Nagle, Thomas T. and Reed K. Holden (1987). *The strategy and tactics of pricing*. Vol. 3. Prentice Hall Englewood Cliffs, NJ.

Nichols, Thomas et al. (2005). “Valid conjunction inference with the minimum statistic”. In: *Neuroimage* 25.3, pp. 653–660.

Niv, Yael (2019). “Learning task-state representations”. In: *Nature Neuroscience* 22.10, pp. 1544–1553.

Niv, Yael and Angela Langdon (2016). “Reinforcement learning with Marr”. In: *Current Opinion in Behavioral Sciences* 11, pp. 67–73.

Noonan, MaryAnn et al. (2010). “Separate value comparison and learning mechanisms in macaque medial and lateral orbitofrontal cortex”. In: *Proceedings of the National Academy of Sciences* 107.47, pp. 20547–20552.

O’Doherty, John P., Peter Dayan, Karl Friston, et al. (2003). “Temporal difference models and reward-related learning in the human brain”. In: *Neuron* 38.2, pp. 329–337.

O’Doherty, John P., Peter Dayan, Johannes Schultz, et al. (2004). “Dissociable roles of ventral and dorsal striatum in instrumental conditioning”. In: *Science* 304.5669, pp. 452–454.

O’Doherty, John P., Alan Hampton, and Hackjin Kim (2007). “Model-based fMRI and its application to reward learning and decision making”. In: *Annals of the New York Academy of sciences* 1104.1, pp. 35–53.

O’Doherty, John P., Ueli Rutishauser, and Kiyohito Iigaya (2021). “The hierarchical construction of value”. In: *Current Opinion in Behavioral Sciences* 41, pp. 71–77.

Ogawa, Seiji et al. (1992). “Intrinsic signal changes accompanying sensory stimulation: Functional brain mapping with magnetic resonance imaging”. In: *Proceedings of the National Academy of Sciences* 89.13, pp. 5951–5955.

Olshausen, Bruno A. and David J. Field (1996). “Emergence of simple-cell receptive field properties by learning a sparse code for natural images”. In: *Nature* 381.6583, pp. 607–609.

Öngür, Dost and Joseph L Price (2000). “The organization of networks within the orbital and medial prefrontal cortex of rats, monkeys and humans”. In: *Cerebral Cortex* 10.3, pp. 206–219.

Padoa-Schioppa, Camillo (2009). “Range-adapting representation of economic value in the orbitofrontal cortex”. In: *Journal of Neuroscience* 29.44, pp. 14004–14014.

Padoa-Schioppa, Camillo and John A. Assad (2006). “Neurons in the orbitofrontal cortex encode economic value”. In: *Nature* 441.7090, pp. 223–226.

– (2008). “The representation of economic value in the orbitofrontal cortex is invariant for changes of menu”. In: *Nature Neuroscience* 11.1, pp. 95–102.

Penny, William D. et al. (2011). *Statistical parametric mapping: The analysis of functional brain images*. Elsevier.

Plassmann, Hilke, John P. O’Doherty, and Antonio Rangel (2007). “Orbitofrontal cortex encodes willingness to pay in everyday economic transactions”. In: *Journal of Neuroscience* 27.37, pp. 9984–9988.

Rangel, Antonio, Colin Camerer, and P. Read Montague (2008). “A framework for studying the neurobiology of value-based decision making”. In: *Nature Reviews Neuroscience* 9.7, pp. 545–556.

Rescorla, Robert A (1972). “A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement”. In: *Current Research and Theory*, pp. 64–99.

Rich, Erin L. and Jonathan D. Wallis (2016). “Decoding subjective decisions from orbitofrontal cortex”. In: *Nature Neuroscience* 19.7, pp. 973–980.

Rolls, Edmund T. (1996). “The orbitofrontal cortex”. In: *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 351.1346, pp. 1433–1444.

Rozin, Paul and Teresa A. Vollmecke (1986). “Food likes and dislikes”. In: *Annual Review of Nutrition* 6.1, pp. 433–456.

Rudebeck, Peter H. and Elisabeth A. Murray (2014). “The orbitofrontal oracle: Cortical mechanisms for the prediction and evaluation of specific behavioral outcomes”. In: *Neuron* 84.6, pp. 1143–1156.

Schuck, Nicolas W. et al. (2016). “Human orbitofrontal cortex represents a cognitive map of state space”. In: *Neuron* 91.6, pp. 1402–1412.

Schulman, John et al. (2017). “Proximal policy optimization algorithms”. In: *arXiv preprint arXiv:1707.06347*.

Schultz, Wolfram (1998). “Predictive reward signal of dopamine neurons”. In: *Journal of Neurophysiology* 80.1, pp. 1–27.

Schultz, Wolfram, Peter Dayan, and P. Read Montague (1997). “A neural substrate of prediction and reward”. In: *Science* 275.5306, pp. 1593–1599.

Schwartz, Dan and Tom M. Mitchell (2019). “Understanding language-elicited EEG data by predicting it from a fine-tuned language model”. In: *arXiv preprint arXiv:1904.01548*.

Seeliger, Katja et al. (2018). “Convolutional neural network-based encoding and decoding of visual object recognition in space and time”. In: *NeuroImage* 180, pp. 253–266.

Shwartz-Ziv, Ravid and Naftali Tishby (2017). “Opening the black box of deep neural networks via information”. In: *arXiv preprint arXiv:1703.00810*.

Silver, David et al. (2016). “Mastering the game of Go with deep neural networks and tree search”. In: *nature* 529.7587, pp. 484–489.

Simon, Herbert A (1957). *Models of man; social and rational.* Wiley.

Small, Dana M. et al. (2003). “Dissociation of neural representation of intensity and affective valuation in human gustation”. In: *Neuron* 39.4, pp. 701–711.

Soltani, Alireza, Benedetto De Martino, and Colin Camerer (2012). “A range-normalization model of context-dependent choice: A new model and evidence”. In: *PLoS Computational Biology* 8.7, e1002607.

Spicer, Julie et al. (2007). “Sensitivity of the nucleus accumbens to violations in expectation of reward”. In: *Neuroimage* 34.1, pp. 455–461.

Springenberg, Jost Tobias et al. (2014). “Striving for simplicity: The all convolutional net”. In: *arXiv preprint arXiv:1412.6806*.

Srinivas, Aravind, Michael Laskin, and Pieter Abbeel (2020). “Curl: Contrastive unsupervised representations for reinforcement learning”. In: *arXiv preprint arXiv:2004.04136*.

Stalnaker, Thomas A. et al. (2014). “Orbitofrontal neurons infer the value and identity of predicted outcomes”. In: *Nature Communications* 5.1, pp. 1–13.

Steinberg, Elizabeth E. et al. (2013). “A causal link between prediction errors, dopamine neurons and learning”. In: *Nature Neuroscience* 16.7, pp. 966–973.

Summerfield, Christopher, Fabrice Luyckx, and Hannah Sheahan (2020). “Structure learning and the posterior parietal cortex”. In: *Progress in Neurobiology* 184, p. 101717.

Sutton, Richard S. and Andrew G Barto (2018). *Reinforcement learning: An introduction*. MIT press.

Suzuki, Shinsuke, Ryo Adachi, et al. (2015). “Neural mechanisms underlying human consensus decision-making”. In: *Neuron* 86.2, pp. 591–602.

Suzuki, Shinsuke, Logan Cross, and John P. O'Doherty (2017). "Elucidating the underlying components of food valuation in the human orbitofrontal cortex". In: *Nature Neuroscience* 20.12, pp. 1780–1786.

Suzuki, Shinsuke, Norihiro Harasawa, et al. (2012). "Learning to simulate others' decisions". In: *Neuron* 74.6, pp. 1125–1137.

Tai, Lei et al. (2016). "A survey of deep network solutions for learning control in robotics: From reinforcement to imitation". In: *arXiv preprint arXiv:1612.07139*.

Tang, Deborah W., Lesley K. Fellows, and Alain Dagher (2014). "Behavioral and neural valuation of foods is driven by implicit knowledge of caloric content". In: *Psychological Science* 25.12, pp. 2168–2176.

Tellez, Luis A. et al. (2016). "Separate circuitries encode the hedonic and nutritional values of sugar". In: *Nature Neuroscience* 19.3, pp. 465–470.

Thorne, John (2004). "Discounted Bundling by Dominant Firms". In: *George Mason Law Review* 13, p. 339.

Tobler, Philippe N. et al. (2007). "Reward value coding distinct from risk attitude-related uncertainty coding in human reward systems". In: *Journal of Neurophysiology* 97.2, pp. 1621–1632.

Tom, Sabrina M. et al. (2007). "The neural basis of loss aversion in decision-making under risk". In: *Science* 315.5811, pp. 515–518.

Tversky, Amos and Daniel Kahneman (1985). "The framing of decisions and the psychology of choice". In: *Behavioral Decision Making*. Springer, pp. 25–41.

Tyszka, J. Michael and Wolfgang M. Pauli (2016). "In vivo delineation of subdivisions of the human amygdaloid complex in a high-resolution group template". In: *Human Brain Mapping* 37.11, pp. 3979–3998.

Tzourio-Mazoyer, Nathalie et al. (2002). "Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain". In: *Neuroimage* 15.1, pp. 273–289.

Van den Oord, Aaron, Yazhe Li, and Oriol Vinyals (2018). "Representation learning with contrastive predictive coding". In: *arXiv e-prints*, arXiv–1807.

Varian, Hal R (2014). *Intermediate microeconomics: A modern approach: Ninth international student edition*. WW Norton & Company.

Vinyals, Oriol et al. (2019). "Grandmaster level in StarCraft II using multi-agent reinforcement learning". In: *Nature* 575.7782, pp. 350–354.

Wang, Jane X. et al. (2018). "Prefrontal cortex as a meta-reinforcement learning system". In: *Nature Neuroscience* 21.6, pp. 860–868.

Watkins, Christopher J.C.H. (1992). "Daya. P: Technical Note: Q-Learning". In: *Machine learning* 8.3, pp. 279–292.

Watter, Manuel et al. (2015). “Embed to control: A locally linear latent dynamics model for control from raw images”. In: *Advances in Neural Information Processing Systems* 28.

Wen, Haiguang et al. (2018). “Neural encoding and decoding with deep learning for dynamic natural vision”. In: *Cerebral Cortex* 28.12, pp. 4136–4160.

Wilson, Robert C. et al. (2014). “Orbitofrontal cortex as a cognitive map of task space”. In: *Neuron* 81.2, pp. 267–279.

Wunderlich, Klaus, Antonio Rangel, and John P. O’Doherty (2009). “Neural computations underlying action-based decision making in the human brain”. In: *Proceedings of the National Academy of Sciences* 106.40, pp. 17199–17204.

Yamins, Daniel L.K. and James J. DiCarlo (2016). “Using goal-driven deep learning models to understand sensory cortex”. In: *Nature Neuroscience* 19.3, pp. 356–365.

Yamins, Daniel L.K., Ha Hong, et al. (2014). “Performance-optimized hierarchical models predict neural responses in higher visual cortex”. In: *Proceedings of the National Academy of Sciences* 111.23, pp. 8619–8624.

Zeiler, Matthew D. and Rob Fergus (2014). “Visualizing and understanding convolutional networks”. In: *European Conference on Computer Vision*. Springer, pp. 818–833.

Zhang, Amy et al. (2020). “Learning invariant representations for reinforcement learning without reconstruction”. In: *arXiv preprint arXiv:2006.10742*.

Zuker, Charles S. (2015). “Food for the brain”. In: *Cell* 161.1, pp. 9–11.