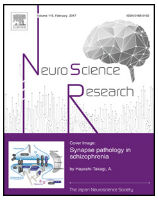
Contents lists available at ScienceDirect

Neuroscience Research

journal homepage: www.elsevier.com/locate/neures



Review article

Over-representation of fundamental decision variables in the prefrontal cortex underlies decision bias

Tomoya Ohnuki, Yuma Osako, Hiroyuki Manabe, Yoshio Sakurai, Junya Hirokawa*

Laboratory of Neural Information, Graduate School of Brain Science, Doshisha University, 1-3 Tatara Miyakodani, Kyotanabe, Kyoto, 610-0394, Japan

ARTICLE INFO

Article history:
Received 28 January 2021
Received in revised form 15 June 2021
Accepted 13 July 2021
Available online xxx

Keywords:
Decision making
Optogenetics
Prefrontal cortex
Decision bias
Psychiatric disorders
Mixed selectivity
Corticostriatal

ABSTRACT

The brain is organized into anatomically distinct structures consisting of a variety of projection neurons. While such evolutionarily conserved neural circuit organization underlies the innate ability of animals to swiftly adapt to environments, they can cause biased cognition and behavior. Although recent studies have begun to address the causal importance of projection-neuron types as distinct computational units, it remains unclear how projection types are functionally organized in encoding variables during cognitive tasks. This review focuses on the neural computation of decision making in the prefrontal cortex and discusses what decision variables are encoded by single neurons, neuronal populations, and projection type, alongside how specific projection types constrain decision making. We focus particularly on över-representationsöf distinct decision variables in the prefrontal cortex that reflect the biological and subjective significance of the variables for the decision makers. We suggest that task-specific over-representation in the prefrontal cortex involves the refinement of the given decision making, while generalized over-representation of fundamental decision variables is associated with suboptimal decision biases, including pathological ones such as those in patients with psychiatric disorders. Such over-representation of the fundamental decision variables in the prefrontal cortex appear to be tightly constrained by afferent and efferent connections that can be optogenetically intervened on. These ideas may provide critical insights into potential therapeutic targets for psychiatric disorders, including addiction and depression.

Contents

1. Introduction 00
2. Encoding of cognitive variables in neurons with mixed selectivity 00
3. Encoding of fundamental decision variables in PFC 00
4. Over-representation and flexible coding of task state 00
5. Biological constraints of cognition and behavior 00
6. Therapeutic targets for psychiatric disorders 00
7. Conclusion 00
Author contributions 00
Funding 00
Declaration of Competing Interest 00
Acknowledgments 00
References 00

* Corresponding author.
E-mail address: jhirokaw@mail.doshisha.ac.jp (J. Hirokawa).

https://doi.org/10.1016/j.neures.2021.07.002

1. Introduction

Despite the Japanese expression for a brain, n̈ou-miso,̈ which suggests that the brain looks like a homogenous substance made of miso paste (fermented soybean paste), decades of neuroscience research has demonstrated that the brain is organized into distinct functional and anatomical structures, including cortical areas, layers, and columns consisting of distinct neural circuits with distinct neuron types marked by specific gene expression patterns and connections (Chen et al., 2019; Harris et al., 2019; Kawaguchi and Kubota, 1997; Kepecs and Fishell, 2014; Luo et al., 2018; Rakic, 2009; Sakata and Harris, 2009; Takahata et al., 2014; Tasic et al., 2018; Watakabe et al., 2007; Watakabe and Hirokawa, 2018; Yamamori, 2011). These macro- and micro-structures in the brain are identifiable across different individuals and species (Adkins, 2020), demonstrating they are heavily influenced by evolutionary pressure. These common biological structures are considered to underlie behaviors dominated by innate components such as escape responses (avoidance movement from potential predation) and taxes (orientation movement to a stimulus or the presence of food) (Gorostiza, 2018) and to enable animals to more swiftly adapt to new environments (Okamoto et al., 2021). Conversely, the neural structure would be expected to limit freedom in behavior and cognitive repertories, which may, for instance, contribute to the formation of biases in decision making that deviate from a normative account (Kahneman, 2011; Santos and Rosati, 2015; Schacter and Buckner, 1998). However, there is still a considerable gap between our understandings of the computational mechanisms of decision making and biological neural circuits.

The main reason for the gap derives from the fact that it is still unclear what decision-related variables are represented by neural circuit components, particularly within the prefrontal cortex (PFC). The PFC is considered to be involved in various latent cognitive processes, including attention, working memory, planning, reward predictions, inhibitory control, metacognition, abstraction, and categorization (Funahashi, 2001; Miller and Cohen, 2001; Miyamoto et al., 2018; Schoenbaum et al., 2009). To estimate the task-relevant decision variables, one needs an appropriate model and behavioral readout. To this end, the conventional approach trained animals to conduct a cognitive task designed to reveal specific aspects of cognition, allowing researchers to attempt to find neural representations of decision variables. If a particular variable explains the observed neural activity better than other variables, it can be said that the neural activity “encodes” or “represents” that variable (deCharms and Zador, 2000; Kable and Glimcher, 2007; Padoa-Schioppa and Assad, 2006; Plassmann et al., 2007). For instance, Padoa-Schioppa and Assad showed that the orbitofrontal cortex (OFC) neurons encode subjective value variables in economic decision making (Padoa-Schioppa and Assad, 2006). The authors identified three groups of neurons in OFC that represent offered value, chosen value, and chosen juice, all of which are essential components that explain economic decision making in their task. In this line of single-unit studies, other research groups have attempted to fit neuronal activities to their models revealing that individual neurons in OFC encode key decision variables, including outcome expectancy (expectation of a reward), reward risk (outcome variance), decision confidence (estimated probability that a decision is correct), and chosen value (overall utility of choice) (Abe and Lee, 2011; Farovik et al., 2015; Feierstein et al., 2006; Furuyashiki et al., 2008; Kennerley et al., 2009, 2011; Kepecs et al., 2008; McGinty et al., 2016; Morrison and Salzman, 2009; O’Neill and Schultz, 2010; Ogawa et al., 2013; Padoa-Schioppa, 2013; Padoa-Schioppa and Assad, 2006; Roesch et al., 2006; Roitman and Roitman, 2010; Schoenbaum et al., 1998; Steiner and Redish, 2014; Sul et al., 2010; Thorpe et al., 1983; Tremblay and Schultz, 1999; Wallis and Miller, 2003).

While these studies shed light on the consistent role of OFC neurons in representing expected reward values or utility during decision making, several groups of researchers have pointed out the heterogeneity in the observed neuronal-response patterns, suggesting that individual neurons appear to be responsive to nearly all combinations of task-related variables with only quantitative differences between other PFC regions (Wallis and Kennerley, 2010). The property of neurons encoding different variables redundantly is referred to as m̈ixed selectivity(̈Raposo et al., 2014; Rigotti et al., 2013). This idea of mixed selectivity can be interpreted in two different scenarios. First, some individual neurons may be interpretable in response to a specific combination of multiple task variables. In this case, one may describe such complex neuronal activities as representing higher-order decision variables such as reward value, decision confidence, and risk. Such activities appeared to be randomly mixed when researchers do not have proper models to test the right entities (Fig. 1). However, such interpretation is known to suffer from confirmation bias and requires careful investigation, as discussed below. The second scenario interprets mixed selectivity as an inherent circuit feature of the brain, especially in the association cortex, enabling the neuronal population to flexibly realize various computational functions (Rigotti et al., 2013; Sakurai, 1999) (see (Fusi et al., 2016; Sakurai et al., 2018) for reviews). In this scenario, a neuronal population is a fundamental unit of computation, and thus, the activities of individual neurons are not necessarily interpretable by an explicit decision variable. Rather, some studies suggest that the individual neurons randomly mix information, termed *random mixed selectivity* (Mante et al., 2013; Raposo et al., 2014). Despite the computational benefits of neural encoding at the population level, it is still unclear whether these network functions are implemented in micro and macro-circuits of the brain because researchers have been blind to the identity and the connections of the analyzed neurons.

Are individual neuronal activities in PFC interpretable as decision variables without any specific assumptions? Are those neuronal representations structured or randomly distributed in a task-variable space? Moreover, how are such neural representations supported by biological structures, such as long-range projection neuron types (i.e., projection type) (Gabbott et al., 2005; Morishima et al., 2011)? In this review, we focus on the role of PFC, including OFC, in decision making. First, we discuss how neural representations of decision variables at the single-neuron level can be revealed while avoiding confirmation bias. Second, we discuss the nature of decision variables that can be encoded in single neurons of PFC. In particular, we discuss the fundamental decision variables that can be shared across different task demands. Third, we propose that over-exposure of animals to a specific environment creates over-representations of these fundamental variables in the relevant cortical areas, which are constrained by their input/output properties. Finally, we discuss the maladaptive over-representation of decision variables in biased decision making and the implications of projection-type-specific representations of particular decision variables, providing a novel insight for treating psychiatric disorders. We conclude that over-representation of the fundamental decision variables reflect circuit structures in PFC, which could be a promising therapeutic targets for psychiatric disorders.

2. Encoding of cognitive variables in neurons with mixed selectivity

It has been debated whether artificially defined cognitive variables (or psychological entities) are explicitly represented in the brain, particularly at the single-neuron level (Buzsáki, 2020). One line of studies revealed that broader information such as internal states might be coded in a form that is different from the firing rates

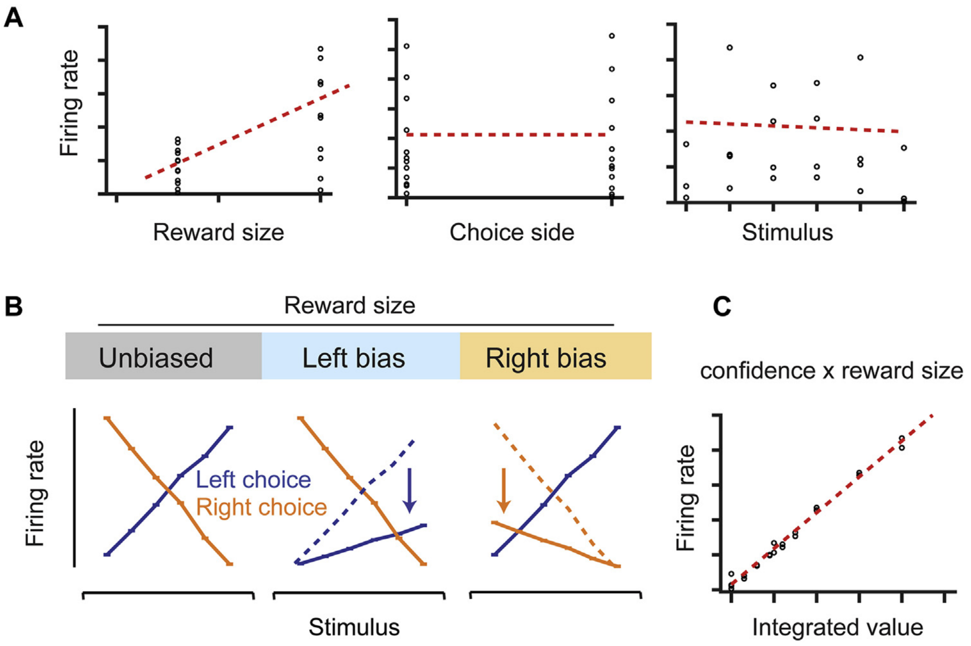


Fig. 1. Single neurons with mixed selectivity can represent higher-order variables.
A. An example simulated neuron that shows mixed selectivity to reward size, choice side and stimulus. Black dot indicates a firing rate for a particular condition in **B**. Red dotted lines indicate the linear regression with each variable. **B.** The same neuron shows a complex tuning pattern to the combination of three task variables when it is expanded to different conditions. **C.** The neuronal activity was parsimoniously explained using a single variable that coherently integrates three variables. The integrated value is calculated by multiplying the confidence and the chosen reward size.

of single neurons (Leavitt et al., 2017; Saez et al., 2015). Nevertheless, some cognitive variables can be explicitly represented at the single-neuron level, especially when a relevant behavioral readout is required for the task. From a reductionist view, if a set of single neurons encodes specific decision variables, those neurons are a promising target for investigating the computational mechanisms of the cognitive process linking those neurons with the related specific neural circuit components. However, determining neural representations of higher-order variables easily suffers from confirmation bias, as other untested variables may better explain the observed neuronal activity.

Several approaches have been proposed to compensate for confirmation bias. First, one can confirm that the neurons do not merely correlate with the task variables, which constitute decision variables, but fulfill other predictions independent of the task variables or behavioral readouts. For instance, if a neuron encodes the economic value of a chosen option, then its activity should reflect the subjective nature of the option's value, reflecting the motivational state (e.g., satiety) of the agent independent of the task variables. There is compelling evidence showing that neuronal responses and the behavioral measures of subjective value co-vary (Padoa-Schioppa and Assad, 2006; Raghuraman and Padoa-Schioppa, 2014; Tsutsui et al., 2016). Similarly, Kepecs et al. showed that neurons encode decision confidence also tracked the accuracy ofchoices regardless of stimulus difficulty, suggesting that such neurons encode abstract states or variables (Kepecs et al., 2008). Second, to demonstrate abstract coding, one needs to determine whether the representations are robust to different contexts, such as the modality of cue stimuli and behavioral outputs. For example, Masset et al. recently reported that OFC neuronal activities represent decision confidence and that this activity does not depend on the modality of sensory stimuli used in the task (Masset et al., 2020). In addition, this neuronal activity correlated with two different behavioral outputs: duration of time invested for rewards and choice patterns to previous rewards and stimuli, suggesting that representations of decision confidence in OFC reflect amodal information controlling different behaviors. Third, a data-driven approach to find neural correlates without an explicit assumption of particular models strongly mitigates the confirmation bias (Hirokawa et al., 2019; Rubin et al., 2019). This process is described below in detail. Finally, when targeted neurons reflect a specific representation of cognitive variables, revealing the behavioral impact of neuron-type-specific manipulation can provide strong evidence to support coding functions. However, this requires the identification of neuron types and their encoding, which has not been well established in PFC, as discussed below.

A model-free and data-driven approach to discover neural representations of cognitive variables has recently been applied by Hirokawa and Vaughan et al. to determine the neural correlates of decision variables in OFC (Hirokawa et al., 2019). In this study, the authors trained rats to conduct a two-alternative forced-choice perceptual decision making task with block-wise changes in reward size. The authors analyzed the neuronal distribution in the task parameter space consisting of 42 dimensions, including choice, reward size, odor stimuli, and past outcomes (Fig. 2), and found that ~90 % of the total recorded neurons were categorized as one of eight distinct clusters in the high-dimensional task-variable space. Next, the authors sought to determine the information that each of the neuronal clusters represented. The authors found that most of the observed neural representations could be interpreted as components of the normative model, despite the high dimensionality of the task-variable space. Each cluster represented basic decision variables such as choice direction, reward size, choice history, as well as highly integrative decision variables such as decision confidence and integrated value required to conduct the task optimally. Note that m̈odel-freeïn this study had several limitations, such as a limited number of task variables and training history of animals and the analyzed brain area. Nevertheless, the finding of eight distinct clusters of decision variables is not an inevitable consequence of a 42-dimensional clustering analysis. Rather, it can be interpreted as showing the categorical encodings of task-relevant decision variables, reflecting a specific internal process for decisions. Thus,

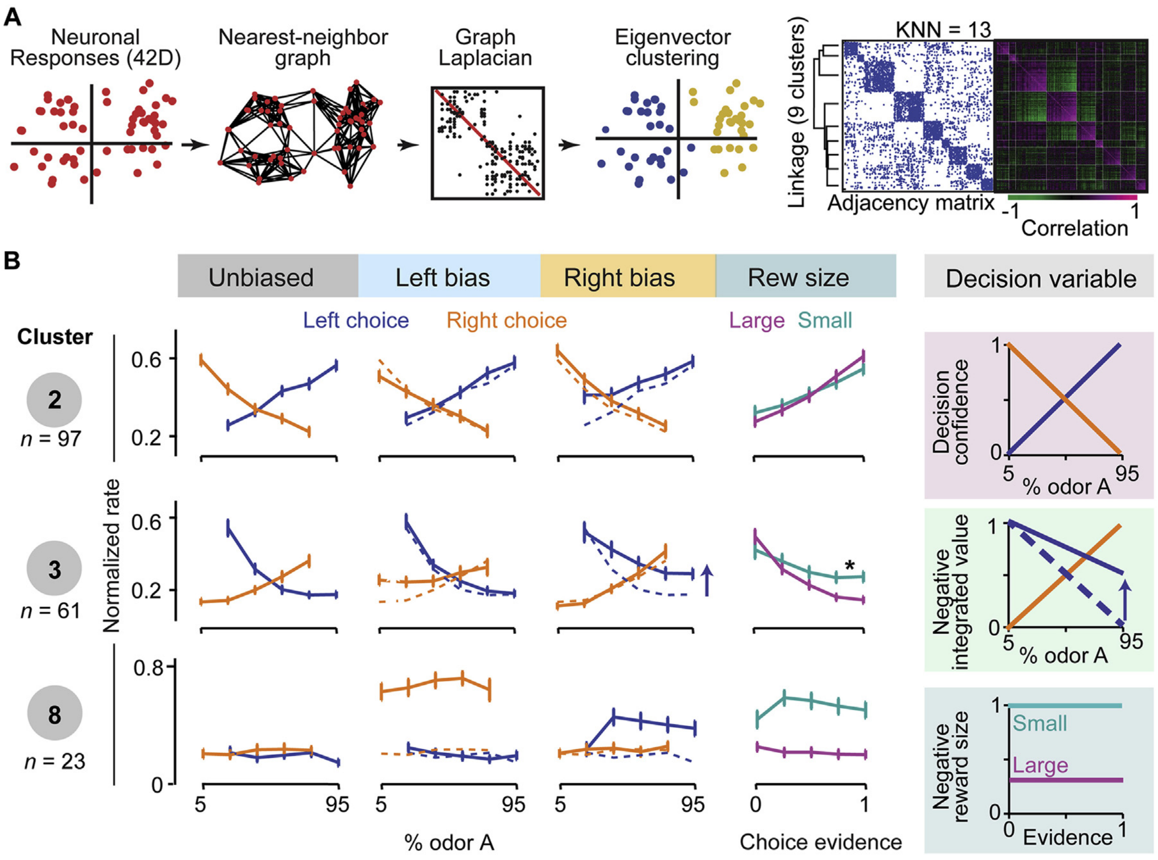


Fig. 2. Orbitofrontal cortex (OFC) neurons can be categorized into distinct groups of neurons representing specific decision variables.
A. Analysis pipeline for the clustering of neurons. Sorted adjacency and correlation matrices reveal substantial within-cluster similarity and between-cluster antagonism. **B.** Average response profiles for representative clusters. Each row represents a cluster, with average normalized firing rate plotted across behavioral conditions. The first three conditions represent combinations of stimulus and choice in different reward (rew) blocks (unbiased, left bias and right bias). The reward size column represents tuning as a function of choice evidence and reward size blocks (left and right bias). Tuning profiles suggest the specific decision variables represented. The figure was reproduced with a modification from Hirokawa et al., 2019.

mixed selectivity neurons (or abstract variable neurons) are not the artifactual product of the random mixed selectivity properties. Instead, this study points to the presence of the structured representations in OFC dominated by the fundamental decision variable representations. However, we note that the neurons within each cluster appear to hold random variability that will allow a flexible population-level mechanism to readout other task-specific information as necessary (Rigotti et al., 2013).

In addition to the categorical encoding of decision variables in OFC, Hirokawa et al. also revealed a relationship between the encoding and a specific projection pathway. They found that optogenetically identified OFC-striatum projection neurons homogeneously represented the integrated value of a chosen option during reward-anticipation and sustained this information during subsequent trials, suggesting the role of these neurons in updating reward values. This close relationship between specific projections and neuronal encodings has been reported before using classic electrophysiological techniques (Chen et al., 2013; Glickfeld et al., 2013; Movshon and Newsome, 1996). In addition, recent optogenetic manipulation studies have revealed that disrupting particular pathways can modify specific behavioral phenotypes (Kim et al., 2017; Li et al., 2015; Tye et al., 2011). For example, modifying corticostriatal projections leads to side-biased choice behavior (Friedman et al., 2015; Otis et al., 2017; Znamenskiy and Zador, 2013). Another line of studies suggests that certain subtypes of cortical projection neurons, intratelencephalic (IT) and pyramidal tract (PT) neurons, are highly differentiated at multiple levels, including morphology, gene expression, electrophysiological properties, and long-range and local connections, and that they are involved in various neurological and neuropsychiatric diseases in different ways (Shepherd, 2013). Indeed, a recent opto- and chemogenetic approach demonstrated that these neurons represent opposite valence information in response to drug experiences (Garcia et al., 2021). Such different representations of specific neuronal projection types can be also closely associated with the gene expression patterns of various receptors such as dopamine and serotonin, since the enriched receptor type affects specific representations (Leiser et al., 2015; Ranganath and Jacob, 2016).

The question of what types of information are encoded by specific pathways and gene expression profiles has just begun to be addressed (Bari et al., 2019; Ciocchi et al., 2015; Economo et al., 2018; Glickfeld et al., 2013; Kitanishi et al., 2021; Murugan et al., 2017), but it is likely that neurons with hard-wired long-range connections do not change the representation content dramatically, and that locally-connected neurons are relatively flexible. Therefore, mapping fundamental decision variables onto different projection types is a promising approach to test causal relationships between discovered abstract neural representations and behavior. Knowledge of these causal relationships may in turn help to precisely understand how animals exploit neural representations during the decision making process.

Taken together, we conclude that the mixed selectivity in single neurons in PFC is not as random as it has been presumed. Instead, the mixed selectivity can be interpreted as representing decision variables as far as the subject and the targeted brain area are engaged in a relevant task.

3. Encoding of fundamental decision variables in PFC

Abstract information coding by single neurons described above raises a question of how generalizable these single-neuron representations are across different conditions and contexts. Although PFC neurons have been shown to support high-dimensional encodings in highly demanding tasks with many variables (Asaad et al., 2000; Rigotti et al., 2013), recent studies suggest that task-related neural activities are often low-dimensional relative to the size of the recorded population (Bartolo et al., 2020; Cunningham and Yu, 2014; Gao and Ganguli, 2015). While low-dimensional representations in PFC are often considered to be attributed to low dimensionalities of the designed tasks, another important factor is the significance of the dimensions for the organism. Here, we discuss the possibility that decision making in animals is constrained by neural circuits tuned to exploit a small number of the coherent structures of the environment for the organisms that limits the dimensionality of the neural representations. In other words, neuronal representations in PFC may be intrinsically bounded so that animals can utilize a set of biologically significant common template variables for decision making, which are generalizable in natural situations.

The prefrontal representations are associated with goal-directed behavior (Miller and Cohen, 2001), and the goal depends on unique incentives that derive from various ethological factors for each species. Such unique incentives obviously limit behavioral repertories. For instance, in social animals such as mice, a subordinate male has fewer choices under the influence of social pressure from other dominant males (i.e., social value), even when multiple options are physically available (choices for mating partners, foods, and accessible areas) (Hollis and Kabbaj, 2014). Similarly, positive valance, negative valence, confidence, risk, efforts, and waiting time are subjective variables that sometimes prevent rational behavior and thus are a limiting factor for the dimensionality of the behavioral repertoire. Consistently, recent studies showed that human consumer decision making consisted of several principal components such as confidence, risk, and social value (Chapman et al., 2018; Stango and Zinman, 2020). We describe these variables as fundamental decision variables because they can constitute basic building blocks for decision making, and each of them appears to hold unique incentives for decision makers in natural environments.

These fundamental decision variables are predominantly represented in prefrontal areas (Kennerley and Wallis, 2009; Stalnaker et al., 2015) that apparently contribute to the low dimensionality of the neural representations in those areas. For instance, in a simple Pavlovian task with multiple odors as conditioned stimuli (Wang et al., 2020), neurons in the piriform cortex, but not OFC, represent the identity of individual odors. After conditioning, OFC neurons, but not piriform, represent their associated positive outcome (positive valence) without sensitivity to individual odor identities. After reversal learning, most of the OFC neurons changed their responses to track the same valence, supporting coherent encodings across different contexts. Note that reducing the dimensionality by encoding valence is not a necessary solution for this task. One can respond to reward by preparing neurons representing both identities of the odors and the associated outcome that maintains maximum dimensionality. Thus, OFC is designed to extract abstract value information projecting the odor and outcome information (i.e., high-dimensional sensory features) to a lower dimensional space (i.e., valence expectation).

The model-free approach is an effective way to discover such fundamental decision variables for each species. Recent studies suggest that representations of decision variables are generalized across different situations. For instance, Hirokawa and Vaughan et al. reported that their model-free clustering analysis yielded a cluster of neurons representing chosen reward size (Hirokawa et al., 2019). These neurons were not active when the reward size is equal between options (Fig. 2) but activated when the rats anticipate a reduced reward, indicating that those neurons are dedicated to conveying relative reward size during the task. Such "reserved neurons" that are only activated upon requirement have been observed during various decision making tasks in PFC (Bari et al., 2019; Tsutsui et al., 2016). These findings indicate that PFC neurons coherently represent one of the fundamental decision variables across different contexts.

It is possible that some fundamental variables can be represented in PFC, regardless of the direct task demands. Hirokawa and Vaughan et al. reported a large cluster of neurons (~30 %) that represented decision confidence even though the task did not require the subjects to report decision confidence (Fig. 2) (Hirokawa et al., 2019). Indeed, these variables, such as confidence and reward size, are useful variables in model-free learning that can cover most situations of decision making that animals would confront in their natural environment. For instance, it has been shown that predictions about the consequences of a behavior (i.e., confidence) are implicitly computed during perceptual decision tasks in different species, including humans (Lak et al., 2020). Therefore, variables such as confidence could be preferentially represented in PFC without explicit demands from the experimental design, suggesting that the low dimensionality of the neural representations is not always caused by the specific task demands.

Though these studies demonstrate that fundamental decision variables are represented by PFC neurons, they only analyzed the data after extensive training of the specific tasks. Therefore, it is unclear whether such representations are the product of the learning that exploits the specific organization of the reward structure in each task, or such representational organization existed from before the training reflecting the inputs organization into the area. For example, Hirokawa and Vaughan et al. found that neurons in OFC represent decision confidence and reward size coherently (i.e., in a positively correlated manner), implying that the neuronal population tends to represent the expected value of the outcome (Fig. 3A, B) (Hirokawa et al., 2019). It is likely that a larger proportion of neurons are recruited to represent the coherently integrated value as the animals learn the relevant task variables during training (Bartolo et al., 2020). However, this does not mean that such a low-dimensional solution for reward maximization is learned from scratch in every time. For instance, the coherent coding of reward probability (confidence) and magnitude can be "naturally trained" outside of the task because the expected value (magnitude x probability) is a general way to maximize reward in the natural environment (Glimcher et al., 2013; Yamada et al., 2021). The involved mechanism may be reminiscent of orientation-selective neurons within the primary visual cortex, which are naturally trained by vertical and horizontal lines in the natural environment (Blakemore and Cooper, 1970). Thus, some generalizable decision variables, such as expected values, could be implemented as intrinsic connections among the neurons through natural experiences afforded by specific neuronal organization of the prefrontal areas. Together, these studies suggest that clusters of neurons represent a fundamental set of variables exploited in various situations.

There are multiple advantages of constrained fundamental decision variables over a strategy that does not impose any limitations on such fundamental variable representations. First, when subjects face difficult decisions, representations of fundamental decision variables enable them to react quickly and accurately rather than having to discover potentially significant decision variables from scratch. Indeed, this is considered one of the most significant advantages of animal brains over artificial networks without any priors (Zador, 2019). Second, explicit representations of decision variables enforce the area to compute their combinations with

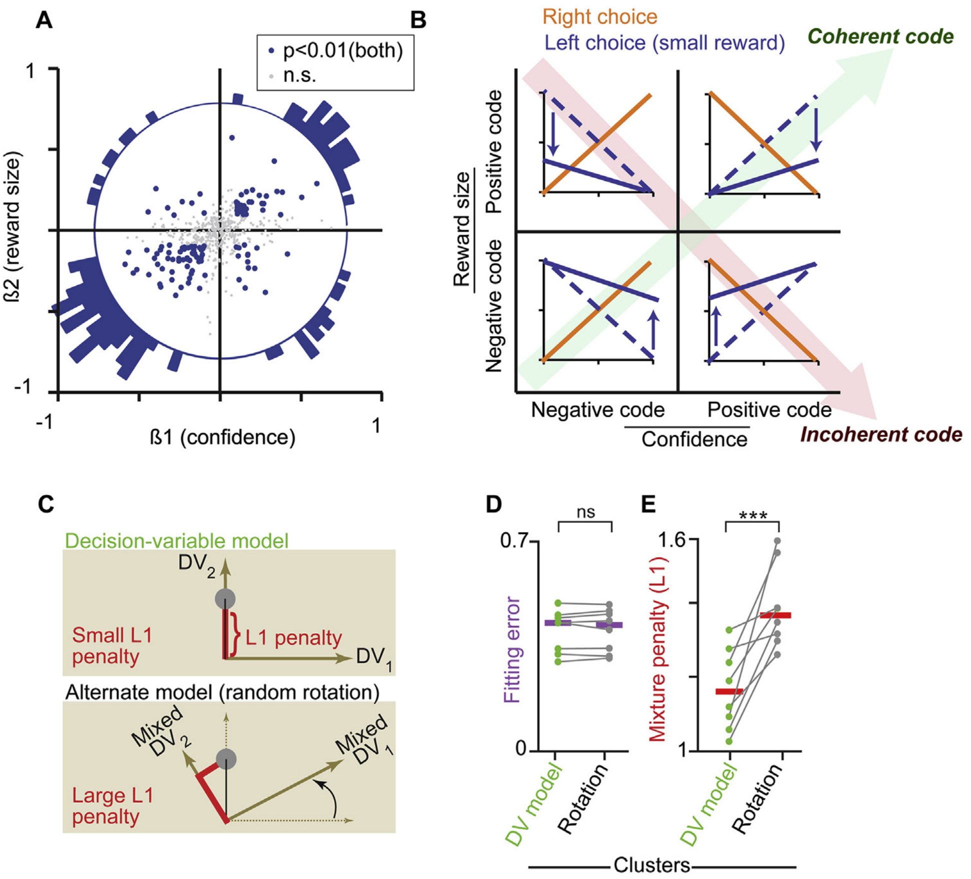


Fig. 3. Single neurons encode coherent combinations of decision variables.
A. Each neuron's response profile was fit to a two-parameter model representing confidence and reward size. For most neurons, regression coefficients (β) for each component share the same sign. Data are shown for all neurons (grey), and neurons with significant beta coefficients for both components are shown in blue ($P < 0.01$ threshold). The resulting polar histogram is significantly different from a uniform distribution ($P < 0.01$). **B.** Diagram explaining coherent (green) and incoherent (pink) integration of confidence and reward size. **C.** Least absolute shrinkage and selection operator (LASSO) regression analysis of individual clusters. **D, E.** Fits of canonical decision-variable model (DV model) to clusters required smaller mixture penalty (L1, $^{***}P < 0.001$, sign rank test) than rotated variables (that is, mixtures of decision variables), with similar fitting error. NS, not significant; horizontal lines show tested comparisons across groups. The figure was reproduced from Hirokawa et al., 2019 except **B.**

less computational effort. Thus, these representations can be basic building blocks to compute further abstract information (Hirokawa et al., 2019). Third, fundamental decision variable representations allow steady communications with the hard-wired downstream subcortical areas with relatively specific functions. For example, cortico-striatal projection neurons drive decisional and action biases (Amemori et al., 2018; Hirokawa et al., 2019; Tai et al., 2012; Znamenskiy and Zador, 2013), suggesting that such hard-wired projection neurons have a designated role.

The fact that fundamental decision variables constrain the neural representations seems to hinder flexible computations, which is a functional hallmark of higher-order cortical areas. However, constrained representations can support computational flexibility by allowing variability among individual neurons, while their averaged activity still cleanly represents distinct decision variables. Ohnuki et al. found that individual neurons in the perirhinal cortex coherently encode choice direction across distinct epochs in a sensory decision task, while these same individual neurons differently contribute to the population-level encodings between those epochs (Ohnuki et al., 2020). This result indicates that individual neurons representing abstract information are sensitive to ongoing neural computations at the population level, offering a good example of how single-neuron and population-level coding can be combined to simultaneously support abstract representations and capacity for flexible computations in cortical areas. Neurons showing weak or modest tuning to the fundamental variables might not be merely carrying "poor" representations, but instead are flexible computational units (Hardcastle et al., 2017; Osako et al., 2021) that carry information about other variables in the task besides the dedicated decision variables. Unstructured connections within and across areas may be the source of flexible representations (Bouchacourt and Buschman, 2019), and this flexibility is likely to be more evident in humans. Nevertheless, fundamental decision variables can still be strong constraints for human decisions.

In conclusion, the fundamental decision variable representations by single neurons play a significant role in expressing biologically rooted behavior and are constrained by afferent and efferent properties in the area. These categorical representations have a limited degree of freedom, but variable responses among individual neurons provide a potential mechanism to retain flexible cortical computations.

4. Over-representation and flexible coding of task state

Discrepancies in interpreting mixed selectivity (coherent single-cell coding vs. random coding) could arise in part from over-representation of decision variables due to the degree of subjects' long-term engagement in cognitive tasks. Conventional single-neuron studies, especially in non-human primate studies, tend to focus on animal behavior after extensive training of a task. These experiments involve the recording of a handful of neurons each day, repeatedly exposing the subjects to the same set of task variables and collecting a sufficient number of neurons for analysis over many days or even months. In this extreme situation, the relevant task variables become more evident to the subjects, and the neural representations may settle into a minimal number of dimensions for the task performance (Bartolo et al., 2020). Thus, individual neurons can tend to be well-tuned to the trained variables in the conventional single-unit studies. Although such over training of animals may be a key to explaining the different emphasis of encoding schemes between the single-neuron and population, insights gained from single-unit studies are not necessarily only applicable in such extreme situations. Here, we discuss this low-dimensional representational organization and provide insights for understanding the neural circuit basis of decision making.

Over-training of subjects tends to recruit a disproportionately large number of neurons to represent salient events or features (Bieszczad and Weinberger, 2012; Henschke et al., 2020; Polley et al., 2006; Reed et al., 2011; Rutkowski and Weinberger, 2005; Shiotani et al., 2020), which we refer to here as over-representation. Indeed, emphasized representations of the outside world and the body parts reflecting functional significance for animals are a fundamental feature of the brain. Over-representation is often associated with higher discriminability of external signals, which have been extensively documented in the sensory cortical areas (Bieszczad and Weinberger, 2010; Kim and Bao, 2013; Polley et al., 2006; Recanzone et al., 1993), hippocampal regions (Fischer et al., 2020; Igarashi et al., 2014; Kaufman et al., 2020; Robinson et al., 2020; Sato et al., 2020), and other subcortical areas (Burgess et al., 2016; Hafed and Chen, 2016; Portfors et al., 2009).

Unlike sensory, spatial, or time representations in the sensory cortical areas and hippocampus, it is not always straightforward to objectively evaluate the over-representation of decision variables in PFC, given their subjective nature. However, multiple studies have demonstrated that PFC neurons can disproportionately represent task-relevant decision variables in a manner corresponding to the utility of the variables (Barat et al., 2018; Hirokawa et al., 2019; Wikenheiser et al., 2017). Here, we defined such *disproportionate recruitment* of the neurons for given variables over other relevant variables as over-representations in PFC. The null hypothesis here is that all the relevant variables are *equally* represented. For instance, Hirokawa and Vaughan et al. found a disproportionately large number of neurons representing a coherently integrated decision variable (21.4 %) compared to those representing incoherent combinations (5%) (Fig. 3A, B), despite the fact that those variables are independently manipulated, demonstrating that OFC neurons over-represent coherent combinations. Note that the result can also be considered as “under-representation of incoherent combinations” compared to coherent ones. It is likely that the over-representation of certain variables in turn suppresses other variables through network mechanisms such as lateral inhibitions, resulting in under-representations, as reported in sensory cortex (Dragoi et al., 2000). The result is in line with the studies on working memory in non-human-primates, which have reported that the capability of individual PFC neurons to represent multiple variables in concurrent tasks determines working memory performance (Funahashi, 2017; Watanabe and Funahashi, 2014), suggesting that a larger number of recruited neurons with a particular memory can dominate working memory performance. Furthermore, different PFC areas preferentially over-represent the particular valence of the chosen value, and its extent is associated with approach/avoidance motivation (Amemori et al., 2015). Together, these studies suggest that over-representation of the decision variables in PFC is associated with the performance of conflicting decisions.

Over-representation contents are constrained by the input and output properties of each brain area. The PFC, especially the OFC, is considered to be a major hub for integrating internal variables from diverse brain areas (Cavada et al., 2000; Hoover and Vertes, 2007). These variables include sensory information from various sensory cortices, choice information from the parietal cortex (Hanks et al., 2015), reward information from the limbic areas (Burgess et al., 2016), and valence and salience information from the basolateral amygdala (Burgos-Robles et al., 2017; Klavir et al., 2017; Vogel et al., 2016). These varieties of inputs are thought to constitute the primary sources for decision variables. Additionally, neuromodulators such as cholinergic (Hangya et al., 2015; Hasselmo and Sarter, 2011; Major et al., 2018), monoaminergic (Bouret and Sara, 2004; Miyazaki et al., 2014; Uematsu et al., 2017) inputs may play an essential role in shaping over-representations. For instance, dopaminergic inputs into the medial PFC (mPFC) may enhance the over-representation of negative outcomes in this region by increasing the signal-to-noise ratio in the affected neuronal population (Vander Weele et al., 2018). Projection specificity from PFC can also constrain representational content, likely due to the distinct topographic arrangements of sensory and motor representations within the subcortical region (Gabbott et al., 2005; De Oca et al., 1998; Tekin and Cummings, 2002), as well as the functional specificity of the regions. Recent studies suggest that some of long-range projection neurons tend to encode specific information (Hirokawa et al., 2019; Lui et al., 2021; Namboodiri et al., 2019). This raises the possibility that such highly-tuned projection neurons can be a core source of information for the over-representation by modulating selectivity of neighboring neurons through local interactions (Peron et al., 2020; Silberberg and Markram, 2007) and that long-range projections to the neuromodulatory system may in turn affect the activities of many neurons in the relevant areas in specific contexts (Uematsu et al., 2017).

Functional studies suggest an area-specific over-representation of distinct decision variables. As described above, Hirokawa and Vaughan et al. demonstrated that the entire OFC population could be categorized into eight distinct clusters of neurons, each corresponding to specific decision variables. To demonstrate the degree of specificity of the representations, the authors used a regression approach with a canonical model consisting of relevant decision variables utilizing a lasso regularization that minimizes the number of regressors (Fig. 3C). The analysis revealed that the decision variable model requires a systematically smaller number of variables to explain the cluster activities compared with a model with their rotated versions of variable space (Fig. 3D), suggesting that individual decision variables, but not other combinations, provide the parsimonious explanation for the representations. One key finding of the study is that the authors did not find clusters of neurons with other task-relevant variables, such as lateralized variables (e.g., odor strength, action value), identities of odors, stimulus difficulty, and prediction error that can be found in other brain areas such as sensory, motor cortical areas, and other subcortical areas (Cohen et al., 2012; Hirokawa et al., 2011; Samejima et al., 2005; Shiotani et al., 2020; Tai et al., 2012). These results suggest that OFC over-represents specific higher-order variables such as decision confidence and integrated value, likely due to the distinct pattern of inputs into the area that constrains the generation of these representations. A recent study showed that odors associated with reward are over-represented when subjects learn new stimulus-outcome associations and that this over-representation shrinks after overtraining in OFC but not in mPFC (Wang et al., 2020). Shrinkage of the over-representation in OFC may be due to the simplicity of the task that does not require further discrimination of the stimuli. Another study also showed that neurons representing a particular decision variable (current value) and task-specific context information emerge after extensive training (Zhou et al., 2020). Therefore, the representational content in OFC can be flexible while still being constrained by innate requirements such as reward value.

This over-representation of variables in association areas has at least five potential functional advantages. First, the recruitment of many neurons in response to particular events could enhance the discriminability of the associated events at the population level (Morcos and Harvey, 2016). Second, recruiting many neurons for a particular variable could further contribute to integrating diverse information with the targeted variable (Ohnuki et al., 2020). Third, the over-representation in PFC may increase the salience of a given internal variable (Ogawa et al., 2013) and prioritize this variable over other variables, as discussed below. Such bias could enable animals to rapidly make decisions in uncertain situations. Fourth, related to the third, induction of over-representation in PFC may serve to generate distinct states in cognitive map (Kaufman et al., 2020; Király et al., 2020) that may help to keep track of the current state during a certain cognitive task (Wilson et al., 2014). Finally, simultaneous activation of many neurons aligned to specific events could increase the efficacy of information transmission to downstream regions. Along this line, it is interesting that particular decision variables can be mapped onto particular projection neurons in OFC (Groman et al., 2019; Hirokawa et al., 2019), indicating that over-representation also plays a role in information routing to downstream regions (Nakajima et al., 2019).

From these considerations, we propose that over-representation is a fundamental feature of PFC that aligns the neuronal population to prioritize biologically relevant variables and refine associated decision makings. In the next section, we discuss how such neuronal over-representations can be linked to biased decision making.

5. Biological constraints of cognition and behavior

Unlike the artificial network that requires a large number of data sets to learn the optimal solution, young animals, including humans, typically learn the essential behavior within a few trials (Santos and Rosati, 2015; Zador, 2019). It suggests that the brain strongly relies on a priori information to learn the structure of the environment rather than learning from scratch. Indeed, biologically significant stimuli are automatic and robust in drawing attention as compared to other events (Ohman and Mineka, 2001). Such prioritized attention and decisions become evident as heuristics or cognitive biases (Aue and Okon-Singer, 2020). Decision bias is generally considered a trend of decisions deviating from the normative account (Glimcher et al., 2013). The normative model of decision makings can define the list of potential choice options and the associated objective reward values that posit how efficient agents should behave for maximizing the reward.

Decision makings of human and other animals are profoundly biased by various factors such as availability, framing, immediacy, risk, and loss aversion, especially when the decision-maker faces uncertain conditions where available information is incomplete or overly complex (Kahneman, 2011). Even in a perceptual decision task, where the reward values for choice options are designed to be equal, subjects often show bias to a particular choice option influenced by specific aspects of the trial history, including stimulus, choice, outcome, and their combinations (Akaishi et al., 2014; Hirokawa et al., 2019; Hwang et al., 2019; Lak et al., 2020; Voss et al., 2008). From the experimenter's perspective who has complete knowledge of the relevant task variables, such choice patterns are sometimes seen as suboptimal for maximizing reward. However, from the subject's perspective, the expected values of the choice options vary in a trial-to-trial basis due to the insufficient experience and imperfect memory. The inherent mechanisms that create bias can be advantageous in certain environments with a high degree of uncertainty (Fan et al., 2018; Kacelnik, 2006; Lak et al., 2020; Mendonça et al., 2020; Pisupati et al., 2021). Thus, decision bias can sometimes be a mischaracterization of decisions due to the lack of consideration of the latent decision variables. However, specific choice patterns deviating from the apparent rationality provide significant insights for characterizing the biological features and rules of the cognitive process.

Several brain regions and mechanisms have been suggested to underlie various decision biases (Aue and Okon-Singer, 2020). In particular, PFC plays an important role in top-down attentional bias by actively maintaining relevant information in working memory (Funahashi et al., 1989; Goldman-Rakic, 1995; Miller et al., 2018). Distinct areas within PFC are distinguished primarily by the types of information maintained in working memory (Amemori et al., 2015; Monosov and Hikosaka, 2012). For instance, OFC preferentially represents reward values (Elliott et al., 2000; Schoenbaum and Roesch, 2005; Tremblay and Schultz, 1999) with the distinct preference of positive and negative outcomes of decisions between its medial and lateral parts, respectively (Gottfried et al., 2002; O'Doherty et al., 2001; Ursu and Carter, 2005). Attentional bias can be translated to spatial choice through subcortical regions such as the striatum (Lauwereyns et al., 2002; Nonomura et al., 2018) and superior colliculus (Felsen and Mainen, 2008; Hirokawa et al., 2011; Huda et al., 2020), which have topographic mapping of spatial and motor representations (Znamenskiy and Zador, 2013). These studies highlight that different subregions within PFC preferentially encode different outcome valance with distinct interactions with different subcortical regions that can mediate choice biases.

While these studies suggest a critical role of PFC and specific circuitries for decision biases, it is unclear what representational organization in PFC generates such subjective biases. Over-representation of a particular decision variable is one of the mechanisms that can lead to an imbalanced integration of decision variables. A number of studies suggest that the magnitude of neuronal firings in a set of PFC neurons with a goal representation correlates positively with the probability of biased choices, while other sets of neurons with the opposing representations negatively with the probability (Hirokawa et al., 2019; Maoz et al., 2013). It is likely that the emphasized activities of a particular set of neurons and the inhibition of opposing neurons through the lateral inhibition can lead to competitive interactions among neurons with distinct decision variable representations, resulting in over-weighted or under-weighted decision variables for the integration process. Furthermore, it has been shown that a part of ACC preferentially over-represents negative outcome in an approach-avoidance decision task and the activation causally induces choice avoidance (i.e., negative biases) relative to the control state (Amemori and Graybiel, 2012; Amemori et al., 2015). These studies implicate that the dominant representation in PFC is causally associated with decision biases.

Though these studies revealed the biological basis that generates specific trends of decisions, over-representations in PFC could reflect the significance of the variable as a consequence of the adaptation to the environment given the preference of the variable for the area. Thus, over-representation can be adaptive and is not always associated with suboptimal decision bias. On the other hand, there are pathological decision biases that might be associated with maladaptive over-representation. To assess such suboptimal decision biases, often observed in patients with psychiatric diseases, one needs to define "abnormality" relative to a proper control group of subjects. In this case, one needs to construct a normative model considering expected utility for control subjects rather than mathematical expected value. The normative expected utility model defines the average weights of the relevant variables that explain the decisions of normal subjects in a particular task. Then, this model can be used to evaluate the degrees of the bias and over-representations of the patients in the task. To determine the list of the decision variables for the model inputs, a data-driven approach adapting a decision making task with many task variables would be ideal. Such quantitative measures are crucial for revealing the causal role of the "maladaptive over-representation" for suboptimal decision biases, as discussed in the following section.

6. Therapeutic targets for psychiatric disorders

Chronic over-representation of fundamental decision variables may underly biased decision making, which is often observed in patients with psychiatric disorders (Cousijn et al., 2011; Raglan and Schulkin, 2014; Wada and Ide, 2016; Wiers et al., 2013; Zhou et al., 2012). For instance, a hallmark of drug addiction is an excessive attentional bias to drug-related cues (Harmer and Phillips, 1998; Hester et al., 2006; Lubman et al., 2000) and underestimation of negative outcomes (Groman et al., 2018; Parvaz et al., 2015). These observations suggest that the over-representation of reward-associated cues and under-representation of other potential variables underlie repeated risky choices in addiction, which is supported by both neuroimaging studies in humans (Dom et al., 2005; Goldstein and Volkow, 2011) and electrophysiological studies in rodents (Lucantonio et al., 2012). The over-representation of the drug over natural reward may enable animals to maximize their drug-taking efficacy, but such behavior can be interpreted as suboptimal according to a normative perspective.

Similarly, patients who suffer from major depression are occupied by an aberrant salience of negative outcomes of their decisions (Beck, 2008; Murrough et al., 2011) and/or an absence of positive bias. Functional imaging studies have shown that elevated brain responses to pessimistic keywords and negative cues in patients with depression compared with normal controls (Chamberlain and Sakakian, 2006), whereas diminished neural responses to positive stimuli are observed in depressed patients (Schaefer et al., 2006). In rodents, it is known that chronic stress causes the dendrite shrinkage in mPFC, whereas OFC shows the expansion of dendrites after the same chronic stress (Liston et al., 2006). The expansion of the dendrites in OFC after chronic stress may suggest over-representation of the negative expectation enhancing the saliency to the stressors (McEwen and Morrison, 2013), whereas the shrinkage of the dendrites in mPFC may involve in under-representation of other variables such as positive outcomes.

These findings support the notion that the PFC neurons are disproportionally recruited to represent cues related to the typical symptoms (i.e., type of bias) of their disorder in psychiatric patients. Post-traumatic stress disorder (PTSD) and generalized anxiety disorder (GAD) may also have over-represented negative expectations in PFC. Notably, there is an inverse relationship between the activity in the amygdala and mPFC during the expression of opposing symptoms (approach and avoidance of threat, respectively) within individual patients (Chiba et al., 2020), suggesting that this reciprocal circuit (Ishikawa et al., 2020; Senn et al., 2014) contributes to opposing over-representations and subsequent mental states. These studies point to causal relationships between psychiatric disorder symptoms and chronic over-representation of particular fundamental decision variables that are tightly regulated by specific neural circuits.

Recent studies have begun to address the contributions of PFC circuitry to psychiatric disorders through specific manipulations of targeted neural networks (Hare and Duman, 2020; Lüscher, 2016; Muir et al., 2019). Animal models of psychiatric disorders such as addiction, depression, and obsessive-compulsive disorder have demonstrated that some of the distinct behavioral aspects of these disorders could be regulated by manipulations of specific PFC circuitries (Burguiere et al., 2013). For instance, focal optogenetic stimulation of the lateral OFC-striatum projection was found to alleviate compulsive grooming in *Sapap3* mutant mice (Burguiere et al., 2013). In other studies, resistance to negative outcomes was found to segregate with activation in OFC, and optogenetic and chemogenetic manipulations of OFC affected resistance to a negative outcome in addiction model mice (Pascoli et al., 2015, 2018). Anxiety-like behavior can also be modulated using optogenetic manipulation of distinct regions within mPFC, suggesting the need to target discrete neuronal populations by input/output region (Hare and Duman, 2020). In addition, optogenetic stimulation of PFC neurons projecting to the dorsal periaqueductal gray (dPAG) induces defensive behavior in mice, while dopamine release in mPFC induces over-representation of negative expectations in dPAG projection neurons (Vander Weele et al., 2018). These studies suggest that the imbalanced representation of positive and negative outcomes results in maladaptive decisions despite negative outcomes. Therefore, manipulation of over-representation of fundamental decision variables is a promising target for relieving psychiatric symptoms.

Although previous studies have pointed to promising therapeutic targets for various psychiatric disorders, there is little information about the optimal choice for projection targets and stimulation protocol for each type of psychiatric disorder. It is crucial to monitor neuronal and neural circuit activities while optogenetic manipulation is applied because it has been reported that there are unintended effects in optogenetic stimulation (Li et al., 2019). Another significant challenge is the diversity of symptoms. It is essential to characterize what types of decision bias is observed in behavior and what decision variables are over-represented in individual patients. Such characterization of individual differences is useful to classifying subtypes of psychiatric disorders and predicting responses to treatments. Through the identification of such biomarkers, more effective circuitries can be chosen for targeting. Besides, the normalization of such aberrant circuit activities can be achieved by using alternative approaches such as electroconvulsive therapy (ECT) (Singh and Kar, 2017), repetitive transcranial magnetic stimulation (rTMS) (Dubin et al., 2017), deep brain stimulation (DBS) (Mayberg et al., 2005), and fMRI neurofeedback (Cortese et al., 2016). Although these alternative approaches are promising, the targeted structures can contain multiple subtypes of projection neurons that may have different functions (Friedman et al., 2015; Hooks et al., 2018). Therefore, further improvements in targeting accuracy can reduce the side effects of therapeutic manipulations. Identifying specific types of over-represented decision variables in PFC will provide promising therapeutic targets for the effective use of pathway-specific manipulations to treat various psychiatric disorders.

7. Conclusion

Although the clinical use of optogenetics is not currently available for psychiatric diseases (White et al., 2020) but see (Gauvain et al., 2021), it is crucial to establish the model of potential target neuron types for various psychiatric disorders for future clinical use. To achieve a greater understanding of decision making at the circuit level in PFC, a future challenge will be to employ model-free data-driven approaches to determine the neuronal representations of different projection types. This in turn will better establish the basic principles of prefrontal control of our cognition. Such an approach can provide significant insights for establishing therapeutic targets for various psychiatric disorders.

Author contributions

All the authors contributed to the writing of this manuscript.

Funding

This work was supported by the JSPS KAKENHI Grant Numbers 19J12634 (to T.O.), 19J21544 (to Y.O.), 20H00109, 20H05020 (to Y.S.), 19H05028, and 19K06948 (to J.H.).

Declaration of Competing Interest

The authors declare that there are no conflicts of interest.

Acknowledgments

We would like to thank past and present members of the Sakurai Laboratory and Dr. Adam Kepecs for many valuable discussions www.editage.com.

References

Abe, H., Lee, D., 2011. Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. Neuron 70, 731–741.

Adkins, R.S., et al., 2020. A multimodal cell census and atlas of the mammalian primary motor cortex. BioRXiv, http://dx.doi.org/10.1101/2020.10.19.343129.

Akaishi, R., Umeda, K., Nagase, A., Sakai, K., 2014. Autonomous mechanism of internal choice estimate underlies decision inertia. Neuron 81, 195–206.

Amemori, K.I., Graybiel, A.M., 2012. Localized microstimulation of primate pregenual cingulate cortex induces negative decision-making. Nat. Neurosci. 15, 776–785.

Amemori, K.I., Amemori, S., Graybiel, A.M., 2015. Motivation and affective judgments differentially recruit neurons in the primate dorsolateral prefrontal and anterior cingulate cortex. J. Neurosci. 35, 1939–1953.

Amemori, K., Amemori, S., Gibson, D.J., Graybiel, A.M., 2018. Striatal microstimulation induces persistent and repetitive negative decision-making predicted by Striatal beta-band oscillation. Neuron 99, 829-841.e6.

Asaad, W.F., Rainer, G., Miller, E.K., 2000. Task-specific neural activity in the primate prefrontal cortex. J. Neurophysiol. 84, 451–459.

Aue, T., Okon-Singer, H., 2020. Cognitive Biases in Health and Psychiatric Disorders. Elsevier Academic Press.

Barat, E., Wirth, S., Duhamel, J.R., 2018. Face cells in orbitofrontal cortex represent social categories. Proc. Natl. Acad. Sci. U. S. A. 115, E11158–E11167.

Bari, B.A., Grossman, C.D., Lubin, E.E., Rajagopalan, A.E., Cressy, J.I., Cohen, J.Y., 2019. Stable representations of decision variables for flexible behavior. Neuron 103, 922-933.e7.

Bartolo, R., Saunders, R.C., Mitz, A.R., Averbeck, B.B., Id, R.B., Saunders, R.C., Id, A.R.M., Id, B.B.A., Bartolo, R., Saunders, R.C., et al., 2020. Dimensionality, information and learning in prefrontal cortex. PLoS Comput. Biol. 16, 1–26.

Beck, A.T., 2008. The evolution of the cognitive model of depression and its neurobiological correlates. Am. J. Psychiatry 165, 969–977.

Bieszczad, K.M., Weinberger, N.M., 2010. Representational gain in cortical area underlies increase of memory strength. Proc. Natl. Acad. Sci. U. S. A. 107, 3793–3798.

Bieszczad, K.M., Weinberger, N.M., 2012. Extinction reveals that primary sensory cortex predicts reinforcement outcome. Eur. J. Neurosci. 35, 598–613.

Blakemore, C., Cooper, G.F., 1970. Development of the brain depends on the visual environment. Nature 228, 477–478.

Bouchacourt, F., Buschman, T.J., 2019. A flexible model of working memory. Neuron 103, 147-160.e8.

Bouret, S., Sara, S.J., 2004. Reward expectation, orientation of attention and locus coeruleus-medial frontal cortex interplay during learning. Eur. J. Neurosci. 20, 791–802.

Burgess, C.R., Ramesh, R.N., Sugden, A.U., Levandowski, K.M., Minnig, M.A., Fenselau, H., Lowell, B.B., Andermann, M.L., 2016. Hunger-dependent enhancement of food cue responses in mouse postrhinal cortex and lateral amygdala. Neuron 91, 1154–1169.

Burgos-Robles, A., Kimchi, E.Y., Izadmehr, E.M., Porzenheim, M.J., Ramos-Guasp, W.A., Nieh, E.H., Felix-Ortiz, A.C., Namburi, P., Leppla, C.A., Presbrey, K.N., et al., 2017. Amygdala inputs to prefrontal cortex guide behavior amid conflicting cues of reward and punishment. Nat. Neurosci. 20, 824–835.

Burguiere, E., Monteiro, P., Feng, G., Graybiel, A.M., Burguière, E., Monteiro, P., Feng, G., Graybiel, A.M., Burguiere, E., Monteiro, P., et al., 2013. Optogenetic stimulation of lateral orbitofronto-striatal pathway suppresses compulsive behaviors. Science 340, 1243–1246.

Buzsáki, G., 2020. The brain–cognitive behavior problem: a retrospective. eNeuro 7, 1–8.

Cavada, C., Compañy, T., Tejedor, J., Cruz-Rizzolo, R.J., Reinoso-Suárez, F., 2000. The anatomical connections of the macaque monkey orbitofrontal cortex. A review. Cereb. Cortex 10, 220–242.

Chamberlain, S.R., Sakakian, B.J., 2006. The neuropsychology of mood disorders. Curr. Psychiatry Rep. 8, 458–463.

Chapman, J., Dean, M., Ortoleva, P., Snowberg, E., Camerer, C., 2018. Econographics (No.w24931), Natl. Bur. Econ. Res http://pietroortoleva.com/papers/Econographics.pdf.

Chen, J.L., Carta, S., Soldado-Magraner, J., Schneider, B.L., Helmchen, F., 2013. Behaviour-dependent recruitment of long-range projection neurons in somatosensory cortex. Nature 499, 336–340.

Chen, X., Sun, Y.C., Zhan, H., Kebschull, J.M., Fischer, S., Matho, K., Huang, Z.J., Gillis, J., Zador, A.M., 2019. High-throughput mapping of long-range neuronal projection using in situ sequencing. Cell 179, 772-786.e19.

Chiba, T., Ide, K., Taylor, J.E., Boku, S., Toda, H., Kanazawa, T., Kato, S., Horiuchi, Y., Hishimoto, A., Maruyama, T., et al., 2020. A reciprocal inhibition model of alternations between under-/overemotional modulatory states in patients with PTSD. Mol. Psychiatry.

Ciocchi, S., Passecker, J., Malagon-Vina, H., Mikus, N., Klausberger, T., Malagon-Vina, H., Mikus, N., Klausberger, T., 2015. Selective information routing by ventral hippocampal CA1 projection neurons. Science 348 (80-), 560–563.

Cohen, J.Y., Haesler, S., Vong, L., Lowell, B.B., Uchida, N., 2012. Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature 482, 85–88.

Cortese, A., Amano, K., Koizumi, A., Kawato, M., Lau, H., 2016. Multivoxel neurofeedback selectively modulates confidence without changing perceptual performance. Nat. Commun. 7, 1–18.

Cousijn, J., Goudriaan, A.E., Wiers, R.W., 2011. Reaching out towards cannabis: approach-bias in heavy cannabis users predicts changes in cannabis use. Addiction 106, 1667–1674.

Cunningham, J.P., Yu, B.M., 2014. Dimensionality reduction for large-scale neural recordings. Nat. Neurosci. 17, 1500–1509.

De Oca, B.M., DeCola, J.P., Maren, S., Fanselow, M.S., 1998. Distinct regions of the periaqueductal gray are involved in the acquisition and expression of defensive responses. J. Neurosci. 18, 3426–3432.

deCharms, R.C., Zador, A., 2000. Neural representation and the cortical code. Annu. Rev. Neurosci. 23, 613–647.

Dom, G., Sabbe, B., Hulstijn, W., van den Brink, W., 2005. Substance use disorders and the orbitofrontal cortex: systematic review of behavioural decision-making and neuroimaging studies. Br. J. Psychiatry 187, 209–220.

Dragoi, V., Sharma, J., Sur, M., 2000. Adaptation-induced plasticity of orientation tuning in adult visual cortex. Neuron 28, 287–298.

Dubin, M.J., Liston, C., Avissar, M.A., Ilieva, I., Gunning, F.M., 2017. Network-guided transcranial magnetic stimulation for depression. Curr. Behav. Neurosci. Reports 4, 70–77.

Economo, M.N., Viswanathan, S., Tasic, B., Bas, E., Winnubst, J., Menon, V., Graybuck, L.T., Nguyen, T.N., Smith, K.A., Yao, Z., et al., 2018. Distinct descending motor cortex pathways and their roles in movement. Nature 563, 79–84.

Elliott, R., Dolan, R.J., Frith, C.D., 2000. Dissociable functions in the medial and lateral orbitofrontal cortex: evidence from human neuroimaging studies. Cereb. Cortex 10, 308–317.

Fan, Y., Gold, J.I., Ding, L., 2018. Ongoing, rational calibration of reward-driven perceptual biases. Elife 7, 1–26.

Farovik, X.A., Place, R.J., Mckenzie, S., Porter, X.B., Munro, C.E., Eichenbaum, H., 2015. Orbitofrontal cortex encodes memories within value-based schemas and represents contexts that guide memory retrieval. J. Neurosci. 35, 8333–8344.

Feierstein, C.E., Quirk, M.C., Uchida, N., Sosulski, D.L., Mainen, Z.F., 2006. Representation of spatial goals in rat orbitofrontal cortex. Neuron 51, 495–507.

Felsen, G., Mainen, Z.F., 2008. Neural substrates of sensory-guided locomotor decisions in the rat superior colliculus. Neuron 60, 137–148.

Fischer, L.F., Mojica Soto-Albors, R., Buck, F., Harnett, M.T., 2020. Representation of visual landmarks in retrosplenial cortex. Elife 9, 1–25.

Friedman, A., Homma, D., Gibb, L.G., Amemori, K.-I., Rubin, S.J., Hood, A.S., Riad, M.H., Graybiel, A.M., 2015. A corticostriatal path targeting striosomes controls decision-making under conflict. Cell 161, 1320–1333.

Funahashi, S., 2001. Neuronal mechanisms of executive control by the prefrontal cortex. Neurosci. Res. 39, 147–165.

Funahashi, S., 2017. Working memory in the prefrontal cortex. Brain Sci. 7.

Funahashi, S., Bruce, C.J., Goldman-Rakic, P.S., 1989. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. J. Neurophysiol. 61, 331–349.

Furuyashiki, T., Holland, P.C., Gallagher, M., 2008. Rat orbitofrontal cortex separately encodes response and outcome information during performance of goal-directed behavior. J. Neurosci. 28, 5127–5138.

Fusi, S., Miller, E.K., Rigotti, M., 2016. Why neurons mix: high dimensionality for higher cognition. Curr. Opin. Neurobiol. 37, 66–74.

Gabbott, P.L., Warner, T.A., Jays, P.R., Salway, P., Busby, S.J., 2005. Prefrontal cortex in the rat: projections to subcortical autonomic, motor, and limbic centers. J. Comp. Neurol. 492, 145–177.

Gao, P., Ganguli, S., 2015. On simplicity and complexity in the brave new world of large-scale neuroscience. Curr. Opin. Neurobiol. 32, 148–155.

Garcia, A.F., Crummy, E.A., Webb, I.G., Nooney, M.N., Ferguson, S.M., 2021. Distinct populations of cortical pyramidal neurons mediate drug reward and aversion. Nat. Commun. 12, 1–8.

Gauvain, G., Akolkar, H., Chaffiol, A., Arcizet, F., Khoei, M.A., Desrosiers, M., Jaillard, C., Caplette, R., Marre, O., Bertin, S., et al., 2021. Optogenetic therapy: high spatiotemporal resolution and pattern discrimination compatible with vision restoration in non-human primates. Commun. Biol. 4, 1–15.

Glickfeld, L.L., Andermann, M.L., Bonin, V., Reid, R.C., 2013. Cortico-cortical projections in mouse visual cortex are functionally target specific. Nat. Neurosci. 16, 219–226.

Glimcher, P.W., Camerer, C.F., Fehr, E., Poldrack, R.A., 2013. Neuroeconomics: Decision Making and the Brain. Academic Press, Cambridge.

Goldman-Rakic, P.S., 1995. Cellular basis of working memory. Neuron 14, 477–485.

Goldstein, R.Z., Volkow, N.D., 2011. Dysfunction of the prefrontal cortex in addiction: neuroimaging findings and clinical implications. Nat. Rev. Neurosci. 12, 652–669.

Gorostiza, E.A., 2018. Does cognition have a role in plasticity of "innate behavior"? A perspective from Drosophila. Front. Psychol. 9, 1502.

Gottfried, J.A., O'Doherty, J., Dolan, R.J., 2002. Appetitive and aversive olfactory learning in humans studied using event-related functional magnetic resonance imaging. J. Neurosci. 22, 10829–10837.

Groman, S.M., Rich, K.M., Smith, N.J., Lee, D., Taylor, J.R., 2018. Chronic exposure to methamphetamine disrupts reinforcement-based decision making in rats. Neuropsychopharmacology 43, 770–780.

Groman, S.M., Keistler, C., Keip, A.J., Hammarlund, E., DiLeone, R.J., Pittenger, C., Lee, D., Taylor, J.R., 2019. Orbitofrontal circuits control multiple reinforcement-learning processes. Neuron 103, 734-746.e3.

Hafed, Z.M., Chen, C.Y., 2016. Sharper, stronger, faster upper visual field representation in primate superior colliculus. Curr. Biol. 26, 1647–1658.

Hangya, B., Ranade, S.P., Lorenc, M., Kepecs, A., 2015. Central cholinergic neurons are rapidly recruited by reinforcement feedback. Cell 162, 1155–1168.

Hanks, T.D., Kopec, C.D., Brunton, B.W., Duan, C.A., Erlich, J.C., Brody, C.D., 2015. Distinct relationships of parietal and prefrontal cortices to evidence accumulation. Nature 520, 220–223.

Hardcastle, K., Maheswaranathan, N., Ganguli, S., Giocomo, L.M., 2017. A multiplexed, heterogeneous, and adaptive code for navigation in medial entorhinal cortex. Neuron 94, 375-387.e7.

Hare, B.D., Duman, R.S., 2020. Prefrontal cortex circuits in depression and anxiety: contribution of discrete neuronal populations and target regions. Mol. Psychiatry 25, 2742–2758.

Harmer, C.J., Phillips, G.D., 1998. Enhanced appetitive conditioning following repeated pretreatment with d-amphetamine. Behav. Pharmacol. 9, 299–308.

Harris, J.A., Mihalas, S., Hirokawa, K.E., Whitesell, J.D., Choi, H., Bernard, A., Bohn, P., Caldejon, S., Casal, L., Cho, A., et al., 2019. Hierarchical organization of cortical and thalamic connectivity. Nature 575, 195–202.

Hasselmo, M.E., Sarter, M., 2011. Modes and models of forebrain cholinergic neuromodulation of cognition. Neuropsychopharmacology 36, 52–73.

Henschke, J.U., Dylda, E., Katsanevaki, D., Dupuy, N., Currie, S.P., Amvrosiadis, T., Pakan, J.M.P., Rochefort, N.L., 2020. Reward association enhances stimulus-specific representations in primary visual cortex. Curr. Biol. 30, 1866-1880.e5.

Hester, R., Dixon, V., Garavan, H., 2006. A consistent attentional bias for drug-related material in active cocaine users across word and picture versions of the emotional Stroop task. Drug Alcohol Depend. 81, 251–257.

Hirokawa, J., Sadakane, O., Sakata, S., Bosch, M., Sakurai, Y., Yamamori, T., 2011. Multisensory information facilitates reaction speed by enlarging activity difference between superior colliculus hemispheres in rats. PLoS One 6, e25283.

Hirokawa, J., Vaughan, A., Masset, P., Ott, T., Kepecs, A., 2019. Frontal cortex neuron types categorically encode single decision variables. Nature 576, 446–451.

Hollis, F., Kabbaj, M., 2014. Social defeat as an animal model for depression. ILAR J. 55, 221–232.

Hooks, B.M., Papale, A.E., Paletzki, R.F., Feroze, M.W., Eastwood, B.S., Couey, J.J., Winnubst, J., Chandrashekar, J., Gerfen, C.R., 2018. Topographic precision in sensory and motor corticostriatal projections varies across cell type and cortical area. Nat. Commun. 9, 3549.

Hoover, W.B., Vertes, R.P., 2007. Anatomical analysis of afferent projections to the medial prefrontal cortex in the rat. Brain Struct. Funct. 212, 149–179.

Huda, R., Sipe, G.O., Breton-Provencher, V., Cruz, K.G., Pho, G.N., Adam, E., Gunter, L.M., Sullins, A., Wickersham, I.R., Sur, M., 2020. Distinct prefrontal top-down circuits differentially modulate sensorimotor behavior. Nat. Commun. 11, 1–17.

Hwang, E.J., Link, T.D., Hu, Y.Y., Lu, S., Wang, E.H.J., Lilascharoen, V., Aronson, S., O'Neil, K., Lim, B.K., Komiyama, T., 2019. Corticostriatal flow of action selection Bias. Neuron 104, 1126-1140.e6.

Igarashi, K.M., Lu, L., Colgin, L.L., Moser, M.B., Moser, E.I., 2014. Coordination of entorhinal-hippocampal ensemble activity during associative learning. Nature 510, 143–147.

Ishikawa, J., Sakurai, Y., Ishikawa, A., Mitsushima, D., 2020. Contribution of the prefrontal cortex and basolateral amygdala to behavioral decision-making under reward/punishment conflict. Psychopharmacology (Berl.) 237, 639–654.

Kable, J.W., Glimcher, P.W., 2007. The neural correlates of subjective value during intertemporal choice. Nat. Neurosci. 10, 1625–1633.

Kacelnik, A., 2006. Meanings of rationality. In: Rational Animals? Oxford University Press, Oxford.

Kahneman, D., 2011. Thinking, Fast and Slow. Farrar, Straus & Giroux, New York.

Kaufman, A.M., Geiller, T., Losonczy, A., 2020. A role for the locus coeruleus in hippocampal CA1 place cell reorganization during spatial reward learning. Neuron 105, 1018-1026.e4.

Kawaguchi, Y., Kubota, Y., 1997. GABAergic cell subtypes and their synaptic connections in rat frontal cortex. Cereb. Cortex 7, 476–486.

Kennerley, S.W., Wallis, J.D., 2009. Reward-dependent modulation of working memory in lateral prefrontal cortex. J. Neurosci. 29, 3259–3270.

Kennerley, S.W., Dahmubed, A.F., Lara, A.H., Wallis, J.D., 2009. Neurons in the frontal lobe encode the value of multiple decision variables. J. Cogn. Neurosci. 21, 1162–1178.

Kennerley, S.W., Behrens, T.E.J., Wallis, J.D., 2011. Double dissociation of value computations in orbitofrontal and anterior cingulate neurons. Nat. Neurosci. 14, 1581–1589.

Kepecs, A., Fishell, G., 2014. Interneuron cell types are fit to function. Nature 505, 318–326.

Kepecs, A., Uchida, N., Zariwala, H.A., Mainen, Z.F., 2008. Neural correlates, computation and behavioural impact of decision confidence. Nature 455, 227–231.

Kim, H., Bao, S., 2013. Experience-dependent overrepresentation of ultrasonic vocalization frequencies in the rat primary auditory cortex. J. Neurophysiol. 110, 1087–1096.

Kim, C.K., Ye, L., Jennings, J.H., Pichamoorthy, N., Tang, D.D., Yoo, A.C.W., Ramakrishnan, C., Deisseroth, K., 2017. Molecular and circuit-dynamical identification of top-down neural mechanisms for restraint of reward seeking. Cell 170, 1013-1027.e14.

Király, B., Hangya, B., Map, C., Coeruleus, L., Part, I., Király, B., Hangya, B., 2020. Cartographers of the cognitive map: locus coeruleus is part of the guild. Neuron 105, 951–953.

Kitanishi, T., Umaba, R., Mizuseki, K., 2021. Robust information routing by dorsal subiculum neurons. Sci. Adv. 7, http://dx.doi.org/10.1126/sciadv.abf1913.

Klavir, O., Prigge, M., Sarel, A., Paz, R., Yizhar, O., 2017. Manipulating fear associations via optogenetic modulation of amygdala inputs to prefrontal cortex. Nat. Neurosci. 20, 836–844.

Lak, A., Hueske, E., Hirokawa, J., Masset, P., Ott, T., Urai, A.E., Donner, T.H., Carandini, M., Tonegawa, S., Uchida, N., Kepecs, A., 2020. Reinforcement biases subsequent perceptual decisions when confidence is low: a widespread behavioral phenomenon. Elife 9.

Lauwereyns, J., Watanabe, K., Coe, B., Hikosaka, O., 2002. A neural correlate of response bias in monkey caudate nucleus. Nature 418, 413–417.

Leavitt, M.L., Pieper, F., Sachs, A.J., Martinez-Trujillo, J.C., 2017. Correlated variability modifies working memory fidelity in primate prefrontal neuronal ensembles. Proc. Natl. Acad. Sci. U. S. A. 114, E2494–E2503.

Leiser, S.C., Li, Y., Pehrson, A.L., Dale, E., Smagin, G., Sanchez, C., 2015. Serotonergic regulation of prefrontal cortical circuitries involved in cognitive processing: a review of individual 5-HT receptor mechanisms and concerted effects of 5-HT receptors exemplified by the multimodal antidepressant vortioxetine. ACS Chem. Neurosci. 6, 970–986.

Li, N., Chen, T.-W., Guo, Z.V., Gerfen, C.R., Svoboda, K., 2015. A motor cortex circuit for motor planning and movement. Nature 519, 51–56.

Li, N., Chen, S., Guo, Z.V., Chen, H., Huo, Y., Inagaki, H.K., Chen, G., Davis, C., Hansel, D., Guo, C., et al., 2019. Spatiotemporal constraints on optogenetic inactivation in cortical circuits. Elife 8, 642215.

Lubman, D.I., Peters, L.A., Mogg, K., Bradley, B.P., Deakin, J.F.W., 2000. Attentional bias for drug cues in opiate dependence. Psychol. Med. (Paris) 30, 169–175.

Lucantonio, F., Stalnaker, Ta, Shaham, Y., Niv, Y., Schoenbaum, G., 2012. The impact of orbitofrontal dysfunction on cocaine addiction. Nat. Neurosci. 15, 358–366.

Lui, J.H., Nguyen, N.D., Grutzner, S.M., Darmanis, S., Peixoto, D., Wagner, M.J., Allen, W.E., Kebschull, J.M., Richman, E.B., Ren, J., et al., 2021. Differential encoding in prefrontal cortex projection neuron classes across cognitive tasks. Cell 184, 489-506.e26.

Luo, L., Callaway, E.M., Svoboda, K., 2018. Genetic dissection of neural circuits: a decade of progress. Neuron 98, 256–281.

Lüscher, C., 2016. The emergence of a circuit model for addiction. Annu. Rev. Neurosci. 39, 257–276.

Major, A.J., Vijayraghavan, S., Everling, S., 2018. Cholinergic overstimulation attenuates rule selectivity in macaque prefrontal cortex. J. Neurosci. 38, 1137–1150.

Mante, V., Sussillo, D., Shenoy, K.V., Newsome, W.T., 2013. Context-dependent computation by recurrent dynamics in prefrontal cortex. Nature 503, 78–84.

Maoz, U., Rutishauser, U., Kim, S., Cai, X., Lee, D., Koch, C., 2013. Predeliberation activity in prefrontal cortex and striatum and the prediction of subsequent value judgment. Front. Neurosci. 7, 1–16.

Masset, P., Ott, T., Lak, A., Hirokawa, J., Kepecs, A., 2020. Behavior- and modality-general representation of confidence in orbitofrontal cortex. Cell 182, 112-126.e18.

Mayberg, H.S., Lozano, A.M., Voon, V., McNeely, H.E., Seminowicz, D., Hamani, C., Schwalb, J.M., Kennedy, S.H., 2005. Deep brain stimulation for treatment-resistant depression. Neuron 45, 651–660.

McEwen, B.S., Morrison, J.H., 2013. The brain on stress: vulnerability and plasticity of the prefrontal cortex over the life course. Neuron 79, 16–29.

McGinty, V.B., Rangel, A., Newsome, W.T., 2016. Orbitofrontal cortex value signals depend on fixation location during free viewing. Neuron 90, 1299–1311.

Mendonça, A.G., Drugowitsch, J., Vicente, M.I., DeWitt, E.E.J., Pouget, A., Mainen, Z.F., 2020. The impact of learning on perceptual decisions and its implication for speed-accuracy tradeoffs. Nat. Commun. 11, 2757.

Miller, E.K., Cohen, J.D., 2001. An integrative theory of prefrontal cortex function. Annu. Rev. Neurosci. 24, 167–202.

Miller, E.K., Lundqvist, M., Bastos, A.M., 2018. Working memory 2.0. Neuron 100, 463–475.

Miyamoto, K., Setsuie, R., Osada, T., Miyashita, Y., 2018. Reversible silencing of the frontopolar cortex selectively impairs metacognitive judgment on nonexperience in Primates. Neuron, 1–10.

Miyazaki, K.W., Miyazaki, K., Tanaka, K.F., Yamanaka, A., Takahashi, A., Tabuchi, S., Doya, K., 2014. Optogenetic activation of dorsal raphe serotonin neurons enhances patience for future rewards. Curr. Biol. 24, 2033–2040.

Monosov, I.E., Hikosaka, O., 2012. Regionally distinct processing of rewards and punishments by the primate ventromedial prefrontal cortex. J. Neurosci. 32, 10318–10330.

Morcos, A.S., Harvey, C.D., 2016. History-dependent variability in population dynamics during evidence accumulation in cortex. Nat. Neurosci. 19, 1672–1681.

Morishima, M., Morita, K., Kubota, Y., Kawaguchi, Y., 2011. Highly differentiated projection-specific cortical subnetworks. J. Neurosci. 31, 10380–10391.

Morrison, S.E., Salzman, C.D., 2009. The convergence of information about rewarding and aversive stimuli in single neurons. J. Neurosci. 29, 11471–11483.

Movshon, J.A., Newsome, W.T., 1996. Visual response properties of striate cortical neurons projecting to area MT in macaque monkeys. J. Neurosci. 16, 7733–7741.

Muir, J., Lopez, J., Bagot, R.C., 2019. Wiring the depressed brain: optogenetic and chemogenetic circuit interrogation in animal models of depression. Neuropsychopharmacology 44, 1013–1026.

Murrough, J.W., Iacoviello, B., Neumeister, A., Charney, D.S., Iosifescu, D.V., 2011. Cognitive dysfunction in depression: neurocircuitry and new therapeutic strategies. Neurobiol. Learn. Mem. 96, 553–563.

Murugan, M., Jang, H.J., Park, M., Miller, E.M., Cox, J., Taliaferro, J.P., Parker, N.F., Bhave, V., Hur, H., Liang, Y., et al., 2017. Combined social and spatial coding in a descending projection from the prefrontal cortex. Cell 171, 1663-1677.e16.

Nakajima, M., Schmitt, L.I., Halassa, M.M., 2019. Prefrontal cortex regulates sensory filtering through a basal ganglia-to-Thalamus pathway. Neuron 103, 445-458.e10.

Namboodiri, V.M.K., Otis, J.M., van Heeswijk, K., Voets, E.S., Alghorazi, R.A., Rodriguez-Romaguera, J., Mihalas, S., Stuber, G.D., 2019. Single-cell activity tracking reveals that orbitofrontal neurons acquire and maintain a long-term memory to guide behavioral adaptation. Nat. Neurosci. 22, 1110–1121.

Nonomura, S., Nishizawa, K., Sakai, Y., Kawaguchi, Y., Kato, S., Uchigashima, M., Watanabe, M., Yamanaka, K., Enomoto, K., Chiken, S., et al., 2018. Monitoring and updating of action selection for goal-directed behavior through the striatal direct and indirect pathways. Neuron 99, 1302-1314.e5.

O'Doherty, J., Kringelbach, M.L., Rolls, E.T., Hornak, J., Andrews, C., 2001. Abstract reward and punishment representations in the human orbitofrontal cortex. Nat. Neurosci. 4, 95–102.

O'Neill, M., Schultz, W., 2010. Coding of reward risk by orbitofrontal neurons is mostly distinct from coding of reward value. Neuron 68, 789–800.

Ogawa, M., van der Meer, M.A., Esber, G.R., Cerri, D.H., Stalnaker, T.A., Schoenbaum, G., 2013. Risk-responsive orbitofrontal neurons track acquired salience. Neuron 77, 251–258.

Ohman, A., Mineka, S., 2001. Fears, phobias, and preparedness: toward an evolved module of fear and fear learning. Psychol. Rev. 108, 483–522.

Ohnuki, T., Osako, Y., Manabe, H., Sakurai, Y., Hirokawa, J., 2020. Dynamic coordination of the perirhinal cortical neurons supports coherent representations between task epochs. Commun. Biol. 3, 406.

Okamoto, H., Cherng, B.W., Nakajo, H., Chou, M.Y., Kinoshita, M., 2021. Habenula as the experience-dependent controlling switchboard of behavior and attention in social conflict and learning. Curr. Opin. Neurobiol. 68, 36–43.

Osako, Y., Ohnuki, T., Tanisumi, Y., Shiotani, K., Manabe, H., Sakurai, Y., Hirokawa, J., 2021. Contribution of non-sensory neurons in visual cortical areas to visually guided decisions in the rat. Curr. Biol. 31 (13), 2757–2769.

Otis, J.M., Namboodiri, V.M.K., Matan, A.M., Voets, E.S., Mohorn, E.P., Kosyk, O., McHenry, J.A., Robinson, J.E., Resendez, S.L., Rossi, M.A., et al., 2017. Prefrontal cortex output circuits guide reward seeking through divergent cue encoding. Nature 543, 103–107.

Padoa-Schioppa, C., 2013. Neuronal origins of choice variability in economic decisions. Neuron 80, 1322–1336.

Padoa-Schioppa, C., Assad, J.A., 2006. Neurons in the orbitofrontal cortex encode economic value. Nature 441, 223–226.

Parvaz, M.A., Konova, A.B., Proudfit, G.H., Dunning, J.P., Malaker, P., Moeller, S.J., Maloney, T., Alia-Klein, N., Goldstein, R.Z., 2015. Impaired neural response to negative prediction errors in cocaine addiction. J. Neurosci. 35, 1872–1879.

Pascoli, V., Terrier, J., Hiver, A., Lüscher, C., 2015. Sufficiency of mesolimbic dopamine neuron stimulation for the progression to addiction. Neuron 88, 1054–1066.

Pascoli, V., Hiver, A., Van Zessen, R., Loureiro, M., Achargui, R., Harada, M., Flakowski, J., Lüscher, C., 2018. Stochastic synaptic plasticity underlying compulsion in a model of addiction. Nature 564, 366–371.

Peron, S., Pancholi, R., Voelcker, B., Wittenbach, J.D., Ólafsdóttir, H.F., Freeman, J., Svoboda, K., 2020. Recurrent interactions in local cortical circuits. Nature 579, 256–259.

Pisupati, S., Chartarifsky-Lynn, L., Khanal, A., Churchland, A.K., 2021. Lapses in perceptual decisions reflect exploration. Elife 10.

Plassmann, H., O'Doherty, J., Rangel, A., 2007. Orbitofrontal cortex encodes willingness to pay in everyday economic transactions. J. Neurosci. 27, 9984–9988.

Polley, D.B., Steinberg, E.E., Merzenich, M.M., 2006. Perceptual learning directs auditory cortical map reorganization through top-down influences. J. Neurosci. 26, 4970–4982.

Portfors, C.V., Roberts, P.D., Jonson, K., 2009. Over-representation of species-specific vocalizations in the awake mouse inferior colliculus. Neuroscience 162, 486–500.

Raghuraman, A.P., Padoa-Schioppa, C., 2014. Integration of multiple determinants in the neuronal computation of economic values. J. Neurosci. 34, 11583–11603.

Raglan, G., Schulkin, J., 2014. Decision making, mindfulness and mood: how mindfulness techniques can reduce the impact of biases and heuristics through improved decision making and positive affect. J. Depress. Anxiety 04, 1–8.

Rakic, P., 2009. Evolution of the neocortex: a perspective from developmental biology. Nat. Rev. Neurosci. 10, 724–735.

Ranganath, A., Jacob, S.N., 2016. Doping the mind: dopaminergic modulation of prefrontal cortical cognition. Neuroscientist 22, 593–603.

Raposo, D., Kaufman, M.T., Churchland, A.K., 2014. A category-free neural population supports evolving demands during decision-making. Nat. Neurosci. 17, 1784–1792.

Recanzone, G.H., Schreiner, C.E., Merzenich, M.M., 1993. Plasticity in the frequency representation of primary auditory cortex following discrimination training in adult owl monkeys. J. Neurosci. 13, 87–103.

Reed, A., Riley, J., Carraway, R., Carrasco, A., Perez, C., Jakkamsetti, V., Kilgard, M.P., 2011. Cortical map plasticity improves learning but is not necessary for improved performance. Neuron 70, 121–131.

Rigotti, M., Barak, O., Warden, M.R., Wang, X.-J., Daw, N.D., Miller, E.K., Fusi, S., 2013. The importance of mixed selectivity in complex cognitive tasks. Nature, 1–6.

Robinson, N.T.M., Descamps, L.A.L., Russell, L.E., Buchholz, M.O., Bicknell, B.A., Antonov, G.K., Lau, J.Y.N., Nutbrown, R., Schmidt-Hieber, C., Häusser, M., 2020. Targeted activation of hippocampal place cells drives memory-guided spatial behavior. Cell, 1–14.

Roesch, M.R., Taylor, A.R., Schoenbaum, G., 2006. Encoding of time-discounted rewards in orbitofrontal cortex is independent of value representation. Neuron 51, 509–520.

Roitman, J.D., Roitman, M.F., 2010. Risk-preference differentiates orbitofrontal cortex responses to freely chosen reward outcomes. Eur. J. Neurosci. 31, 1492–1500.

Rubin, A., Sheintuch, L., Brande-Eilat, N., Pinchasof, O., Rechavi, Y., Geva, N., Ziv, Y., 2019. Revealing neural correlates of behavior without behavioral measurements. Nat. Commun. 10.

Rutkowski, R.G., Weinberger, N.M., 2005. Encoding of learned importance of sound by magnitude of representational area in primary auditory cortex. Proc. Natl. Acad. Sci. U. S. A. 102, 13664–13669.

Saez, A., Rigotti, M., Ostojic, S., Fusi, S., Salzman, C.D., 2015. Abstract context representations in primate amygdala and prefrontal cortex. Neuron 87, 869–881.

Sakata, S., Harris, K.D., 2009. Laminar structure of spontaneous and sensory-evoked population activity in auditory cortex. Neuron 64, 404–418.

Sakurai, Y., 1999. How do cell assemblies encode information in the brain? Neurosci. Biobehav. Rev. 23, 785–796.

Sakurai, Y., Osako, Y., Tanisumi, Y., Ishihara, E., Hirokawa, J., Manabe, H., 2018. Multiple approaches to the investigation of cell assembly in memory research-present and future. Front. Syst. Neurosci., http://dx.doi.org/10.3389/fnsys.2018.00021.

Samejima, K., Ueda, Y., Doya, K., Kimura, M., 2005. Representation of action-specific reward values in the striatum. Science 310, 1337–1340.

Santos, L.R., Rosati, A.G., 2015. The evolutionary roots of human decision making. Annu. Rev. Psychol. 66, 321–347.

Sato, M., Mizuta, K., Islam, T., Kawano, M., Sekine, Y., Takekawa, T., Gomez-Dominguez, D., Schmidt, A., Wolf, F., Kim, K., et al., 2020. Distinct mechanisms of over-representation of landmarks and rewards in the Hippocampus. Cell Rep. 32, 107864.

Schacter, D.L., Buckner, R.L., 1998. Priming and the brain review. Cell 20, 185–195.

Schaefer, H.S., Putnam, K.M., Benca, R.M., Davidson, R.J., 2006. Event-related functional magnetic resonance imaging measures of neural activity to positive social stimuli in pre- and post-treatment depression. Biol. Psychiatry 60, 974–986.

Schoenbaum, G., Roesch, M., 2005. Orbitofrontal cortex, associative learning, and expectancies. Neuron 47, 633–636.

Schoenbaum, G., Chiba, aa, Gallagher, M., 1998. Orbitofrontal cortex and basolateral amygdala encode expected outcomes during learning. Nat. Neurosci. 1, 155–159.

Schoenbaum, G., Roesch, M.R., Stalnaker, T.A., Takahashi, Y.K., 2009. A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. Nat. Rev. Neurosci. 10, 885–892.

Senn, V., Wolff, S.B.E., Herry, C., Grenier, F., Ehrlich, I., Gründemann, J., Fadok, J.P., Müller, C., Letzkus, J.J., Lüthi, A., 2014. Long-range connectivity defines behavioral specificity of amygdala neurons. Neuron 81, 428–437.

Shepherd, G.M.G., 2013. Corticostriatal connectivity and its role in disease. Nat. Rev. Neurosci. 14, 278–291.

Shiotani, K., Tanisumi, Y., Murata, K., Hirokawa, J., Sakurai, Y., Manabe, H., 2020. Tuning of olfactory cortex ventral tenia tecta neurons to distinct task elements of goal-directed behavior. Elife 9, 1–24.

Silberberg, G., Markram, H., 2007. Disynaptic inhibition between neocortical pyramidal cells mediated by Martinotti cells. Neuron 53, 735–746.

Singh, A., Kar, S.K., 2017. How electroconvulsive therapy works?: understanding the neurobiological mechanisms. Clin. Psychopharmacol. Neurosci. 15, 210–221.

Stalnaker, T.A., Cooch, N.K., Schoenbaum, G., 2015. What the orbitofrontal cortex does not do. Nat. Neurosci. 18, 620–627.

Stango, V., Zinman, J., 2020. We are all behavioral, more or less: a taxonomy of consumer decision making. National bureau of economic research, Cambridge https://www.nber.org/papers/w28138.

Steiner, A.P., Redish, A.D., 2014. Behavioral and neurophysiological correlates of regret in rat decision-making on a neuroeconomic task. Nat. Neurosci. 17, 995–1002.

Sul, J.H., Kim, H., Huh, N., Lee, D., Jung, M.W., 2010. Distinct roles of rodent orbitofrontal and medial prefrontal cortex in decision making. Neuron 66, 449–460.

Tai, L.-H., Lee, a M., Benavidez, N., Bonci, A., Wilbrecht, L., 2012. Transient stimulation of distinct subpopulations of striatal neurons mimics changes in action value. Nat. Neurosci. 15, 1281–1289.

Takahata, T., Miyashita, M., Tanaka, S., Kaas, J.H., 2014. Identification of ocular dominance domains in New World owl monkeys by immediate-early gene expression. Proc. Natl. Acad. Sci. U. S. A. 111, 4297–4302.

Tasic, B., Yao, Z., Graybuck, L.T., Smith, K.A., Nguyen, T.N., Bertagnolli, D., Goldy, J., Garren, E., Economo, M.N., Viswanathan, S., et al., 2018. Shared and distinct transcriptomic cell types across neocortical areas. Nature 563, 72–78.

Tekin, S., Cummings, J.L., 2002. Frontal-subcortical neuronal circuits and clinical neuropsychiatry: an update. J. Psychosom. Res. 53, 647–654.

Thorpe, S.J., Rolls, E.T., Maddison, S., 1983. The orbitofrontal cortex: neuronal activity in the behaving monkey. Exp. Brain Res. 49, 93–115.

Tremblay, L., Schultz, W., 1999. Relative reward preference in primate orbitofrontal cortex. Nature 398, 704–708.

Tsutsui, K.I., Grabenhorst, F., Kobayashi, S., Schultz, W., 2016. A dynamic code for economic object valuation in prefrontal cortex neurons. Nat. Commun. 7.

Tye, K.M., Prakash, R., Kim, S., Fenno, L.E., Grosenick, L., Zarabi, H., Thompson, K.R., Gradinaru, V., Ramakrishnan, C., Deisseroth, K., 2011. Amygdala circuitry mediating reversible and bidirectional control of anxiety. Nature 471, 358–362.

Uematsu, A., Tan, B.Z., Ycu, E.A., Cuevas, J.S., Koivumaa, J., Junyent, F., Kremer, E.J., Witten, I.B., Deisseroth, K., Johansen, J.P., 2017. Modular organization of the brainstem noradrenaline system coordinates opposing learning states. Nat. Neurosci. 20, 1602–1611.

Ursu, S., Carter, C.S., 2005. Outcome representations, counterfactual comparisons and the human orbitofrontal cortex: implications for neuroimaging studies of decision-making. Cogn. Brain Res. 23, 51–60.

Vander Weele, C.M., Siciliano, C.A., Matthews, G.A., Namburi, P., Izadmehr, E.M., Espinel, I.C., Nieh, E.H., Schut, E.H.S., Padilla-Coreano, N., Burgos-Robles, A., et al., 2018. Dopamine enhances signal-to-noise ratio in cortical-brainstem encoding of aversive stimuli. Nature 563, 397–401.

Vogel, E., Krabbe, S., Gründemann, J., Wamsteeker Cusulin, J.I., Lüthi, A., 2016. Projection-specific dynamic regulation of inhibition in amygdala micro-circuits. Neuron 91, 644–651.

Voss, A., Rothermund, K., Brandtstädter, J., 2008. Interpreting ambiguous stimuli: separating perceptual and judgmental biases. J. Exp. Soc. Psychol. 44, 1048–1056.

Wada, M., Ide, M., 2016. Rubber hand presentation modulates visuotactile interference effect especially in persons with high autistic traits. Exp. Brain Res. 234, 51–65.

Wallis, J.D., Kennerley, S.W., 2010. Heterogeneous reward signals in prefrontal cortex. Curr. Opin. Neurobiol. 20, 191–198.

Wallis, J.D., Miller, E.K., 2003. Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. Eur. J. Neurosci. 18, 2069–2081.

Wang, P.Y., Boboila, C., Chin, M., Higashi-Howard, A., Shamash, P., Wu, Z., Stein, N.P., Abbott, L.F., Axel, R., 2020. Transient and persistent representations of odor value in prefrontal cortex. Neuron 108, 1–16.

Watakabe, A., Hirokawa, J., 2018. Cortical networks of the mouse brain elaborate within the gray matter. Brain Struct. Funct. 223 (8), 3633–3652.

Watakabe, A., Ichinohe, N., Ohsawa, S., Hashikawa, T., Komatsu, Y., Rockland, K.S., Yamamori, T., 2007. Comparative analysis of layer-specific genes in Mammalian neocortex. Cereb. Cortex 17, 1918–1933.

Watanabe, K., Funahashi, S., 2014. Neural mechanisms of dual-task interference and cognitive capacity limitation in the prefrontal cortex. Nat. Neurosci. 17, 601–611.

White, M., Mackay, M., Whittaker, R.G., 2020. Taking optogenetics into the human brain: opportunities and challenges in clinical trial design. Open Access J. Clin. Trials 12, 33–41.

Wiers, R.W., Gladwin, T.E., Hofmann, W., Salemink, E., Ridderinkhof, K.R., 2013. Cognitive bias modification and cognitive control training in addiction and related psychopathology: mechanisms, clinical perspectives, and ways forward. Clin. Psychol. Sci. 1, 192–212.

Wikenheiser, A.M., Marrero-Garcia, Y., Schoenbaum, G., 2017. Suppression of ventral hippocampal output impairs integrated orbitofrontal encoding of task structure. Neuron 95, 1197-1207.e3.

Wilson, R.C., Takahashi, Y.K., Schoenbaum, G., Niv, Y., 2014. Orbitofrontal cortex as a cognitive map of task space. Neuron 81, 267–279.

Yamada, H., Imaizumi, Y., Matsumoto, M., 2021. Neural population dynamics underlying expected value computation. J. Neurosci. 41, 1684–1698.

Yamamori, T., 2011. Selective gene expression in regions of primate neocortex: implications for cortical specialization. Prog. Neurobiol. 94, 201–222.

Zador, A.M., 2019. A critique of pure learning and what artificial neural networks can learn from animal brains. Nat. Commun. 10.

Zhou, Y., Li, X., Zhang, M., Zhang, F., Zhu, C., Shen, M., 2012. Behavioural approach tendencies to heroin-related stimuli in abstinent heroin abusers. Psychopharmacology (Berl.) 221, 171–176.

Zhou, J., Jia, C., Montesinos-cartagena, M., Gardner, M.P.H., Wenhui, Z., Schoenbaum, G., 2020. Evolving schema representations in orbitofrontal ensembles during learning. Nature 590 (7847), 606–611.

Znamenskiy, P., Zador, A.M., 2013. Corticostriatal neurons in auditory cortex drive decisions during auditory discrimination. Nature 497, 482–485.