NEURAL MECHANISMS UNDERLYING VALUE COMPUTATION

LAW CHUN KIT

PhD

The Hong Kong Polytechnic University

2022

The Hong Kong Polytechnic University

Department of Rehabilitation Sciences

Neural Mechanisms Underlying Value Computation

Law Chun Kit

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

Doctor of Philosophy

August 2021

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

Law Chun Kit

Contents

Acknowledgements

It is unbelievable I am writing up my PhD thesis. I had never thought I could become a PhD one day. I still remember when I was an undergraduate, my cognitive psychology teacher Terence Lee said I should be a research person and I replied definitely "I NEVER do research!". Yet, eventually I listened to him and intrinsically like doing research.

There are different people that I have to thank for. I may not have the room to thank all of them one by one but I must express my deepest gratitude to my supervisor Bolton Chau. To be honest, I don't think I am capable of being a researcher and even his research student. But Bolton gave me the chance five years ago even I knew nothing about decision neuroscience at that time. He hired me as his RA, taught me lots of things, and even provided me guidance and help to pursue a PhD. His help to me is much greater than my contribution to his lab. The only thing I can do is providing better research (hopefully I am capable of doing so).

Another person that I want to thank for is my former labmate Michael Woo. We started our research degree at the same time and we went through the hard times that every research student must experience. Also, we have lots of good memories. We skipped lectures, skipped tutorials, played pokers at the office etc. Without him, I think my PhD study would be much more boring.

I would say it's a tough time over the four years. It's not easy to finish a PhD. But I realize life must be harder after I graduate. Hope I still like doing research in the future.

Finally, I have to thank God. Without Him, I won nothing, including this PhD.

Abstract

It is not hard to imagine making a choice between items of the same category, for example, two watches vary in price, accuracy, design, etc. A decision can be made by comparing these attributes. This kind of everyday decisions has been extensively investigated over the past decades and the underlying neural mechanisms were reviewed in **Chapter 1**. However, it is also ubiquitous to face decisions between ostensibly incomparable options (e.g. wages versus vacation, staying in the current company versus searching alternatives in the job market). A question arises, how could these heterogeneous options be compared? Since there are countless decisions, it is impossible for the brain to have individual mechanisms for every decision. I therefore propose the brain should respond to decisions by the classes of the options involved. I introduce an approach of classification according to the qualitative differences between the options. As a result, options are categorized into either items or environments. Items are concrete options that provide direct payoff whereas environments are options which lead to potential impacts in the future as opposed to direct payoff (e.g. searching in the job market provides an opportunity to get better jobs but not directly provide an offer). With this classification, I combined human behavioural testing, deep learning neural network, and brain imaging to address the central question of this thesis – whether there is a neural mechanism which can make comparison among all different kinds of option or there are different neural mechanisms to work with particular decisions.

To address the central question of this thesis, first it is necessary to investigate environment choice (i.e. decision between environments), which receives little attention and remains highly elusive. In **Chapter 2**, I examined human decision-making

specifically when decisions are made between environments via a behavioural study. In a binary decision-making task, environments were characterized by their complex structures – each environment was composed of 20 items and hence massive information was embedded. Yet, the results demonstrated that people were capable of integrating the information in the environments to guide their choices. Particularly, I found that people preferred environments with larger means or variances in their component item distributions. After that, the second study involved a two-stage decision-making task to directly contrast environment choice and item choice. Behavioural results reported in **Chapter 3** showed that context-dependent adaptation, an essential property of decision-making, was observed in both environment choice and item choice. It served as the basis for developing computational models in **Chapter 4** and examination of neural signals in **Chapter 5**.

To understand the neurocomputation during decision-making, in **Chapter 4,** I developed and tested different computational models to describe participants' choice behaviour in the studies reported in **Chapters 2** and **3**. Specifically, I employed convolutional neural network (CNN), general linear model, cumulative prospect theory, mean-variance-skewness model, power law model, and autoencoder to fit the behavioural data of those studies. Model comparison results showed that the CNN, a deep learning neural network, best describes participants' decision-making behaviour in both studies. The CNN possesses a strength over traditional computational models that it allows fewer a priori assumptions, by which implicit features of the valuation process can be captured and described even some they are not specified explicitly. Besides, it possesses multiple nodes and multiple layers for representation of option value, facilitating the examination

of multivariate neural signals in **Chapter 5**. A closer inspection using a series of representational similarity analysis (RSA), a multivariate analysis, ascertained that the CNN multi-nodal representations encode the complex information of the item distributions in the environments.

Chapter 5 reports a functional magnetic resonance imaging (fMRI) study that showed a double dissociation of the lateral frontopolar cortex (FPl) and ventromedial prefrontal cortex (vmPFC) in environment choice and item choice. Closer inspection revealed both the FPl and vmPFC signals exhibited essential properties of value comparison process (e.g. activity correlated with the difference in value between options; invariance to salience), suggesting the FPl and vmPFC did subserve the value comparison process. In addition, I performed a series of RSA to test the similarity of the FPl multi-voxel activation patterns during environment choice with the multi-nodal representations of CNN variants that employ single or multiple feature detectors for encoding environment value. Notably, the FPl was found similar to the CNN with multiple feature detectors only, and the FPl was the most similar to the CNN developed in **Chapter 4**. It implies the FPl and CNN shared similar multiple parallel encoding processes in valuation of environments and the FPl carried a multivariate coding for the complex environment information.

To conclude, in this thesis I proposed an approach of classification to categorize options into items or environments according to their qualitative differences. With the use of behavioural testing, computational models, and brain imaging, I demonstrated environment choice and item choice were indeed dissimilar and they involved distinct neural mechanisms. Dissociable roles of the FPl and vmPFC in environment choice and

item choice were revealed. It reflects a functional specialization in decision-making that the brain requires multiple neural substrates to work with particular types of decision as opposed to a single neural substrate to deal with all kinds of decision.

Chapter 1 Introduction

1.1 Background

Socrates asked "*O my friend, why do you who are a citizen of the great and mighty and wise city of Athens care so much about laying up the greatest amount of money and honor and reputation and so little about wisdom and truth and the greatest improvement of the soul, which you never regard or heed at all?"* (Apology, 29d-e).

From the ancient times, people have been asking how the valuation process works during decision-making. Even though making a living is the most predominant affair for the survival of human beings, people on occasion forgo the pursuit of basic needs and strive for higher-order affairs. Liberty, justice, equality, conscientiousness etc., without any one of these affairs people can still stay alive but it is not rare to see social activists fight for righteousness despite the risk to be jailed. Ostensibly, this is a matter of subjectivity that individual differences drive dissimilar valuation. Nonetheless, there is a more intrinsic issue embedded which is overlooked and worth much more endeavour to explore: why different affairs can be compared in spite of their diverse forms, disparate functions and/or heterogeneous natures? It is not hard to imagine making a choice between similar items, for example, two watches vary in price, accuracy, design, etc. A decision can be made by comparison in these attributes. However, decisions often involve items from different categories (e.g. buying a luxury watch or a high-end camera) and even heterogeneous, abstract options (e.g. vacation leave versus wages). It is unintuitive to articulate how a decision is made in these scenarios but practically people

manage to choose. It is intriguing to uncover how the valuation and comparison processes are carried out. Whether there is a single decision-making system universally applicable for all kinds of options or distinct systems are required for different scenarios remains elusive. This thesis aims to address this issue by investigating the neural mechanisms underlying value-based decision-making. Specifically, this thesis discussed the features and distinctiveness of different types of decisions, models for estimation of the value computation process, and the neural substrates underpinning valuation and value comparison were examined.

1.2 Operational definition

To systematically investigate the mechanisms of decision-making, the first crucial step is to categorize decisions. However, it is an equivalently difficult task because there are infinite options and thus limitless criteria for categorization. For instance, a consensus could be easily reached that choosing between an apple and an orange qualitatively differs from deciding between two career paths. Differentiation between these two kinds of decision could be accounted for by trivial decision versus important decision, or on the other hand decision involving concrete options versus decision involving abstract options. A well-structured classification is thus warranted. To this end, I proposed a classification via which options are categorized according to their qualitative differences, thereby decisions can be categorized with respective to their options involved. Consequently, in this thesis, options are categorized into two main classes: (1) item and (2) environment.

1.2.1 Item

To establish the basic understanding on decision-making, people started by investigating the simplest form of options, that is item. Most often, routine decisions in our daily life involve selection between items and it has been the mainstream of the existing literature. Items are referred to as options that are directly obtainable – the relation between selection of an item and acquisition of the item is straightforward. For example, one is picking between Japanese sushi and Korean kimbap for lunch from the fridge. A choice of sushi (kimbap) directly leads to its acquisition and hence the sushi (kimbap) is classified as an item. Apart from this example of concrete food items, items often exist in other less intuitive forms. It is worth noting that an option is classified as an item as long as selection of that option directly leads to acquisition of whom, regardless of whether the option is associated with deterministic or probabilistic outcomes. An illustrative example is a lottery of 50/50 chance to win a prize of $100 is also deemed as an item. The reason is if an individual buys a lottery ticket, s/he instantly obtains the lottery no matter it results in the prize or not.

1.2.2 Environment

By contrast, environments are characterized by their prospective nature, making them qualitatively different from items. Despite the ubiquity and importance of environments in real life, environments receive far less attention in the existing literature. Here, an environment is referred to as a choice set of multiple items that brings in the opportunities to attain potential items, as opposed to direct acquisition of items. In other words, selection of an environment only defines and constraints the items that can be encountered in the future. Further decisions are needed in order to obtain any item. For instance, a Japanese restaurant is an environment which serves with different Japanese

food items. For an individual who intends to have sushi for dinner, making a choice of Japanese restaurant over a British restaurant comes up with the possibility to attain sushi. The sushi can be obtained only if further ordering is made.

Although environments consist of multiple items, it is worth noting that environments should not be merely considered as aggregates of multiple items (e.g. assortments) although they both consist of multiple items. Crucially, environments are distinguishable from aggregates by their prospective nature. To illustrate, using the abovementioned example, a Japanese restaurant can be deemed as a collection of various food items. However, going in the restaurant only facilitates the attainment of sushi as opposed to act as a direct proxy of sushi (e.g. the sushi may have sold out so it is not equivalent to acquisition of sushi). By contrast, a snack assortment also provides a collection of various food items. However, purchase of a snack assortment directly leads to all the food items. In other words, the assortment can act as the proxy of any of its component item. On the other hand, there is a temporal distancing between selection of an environment and the final outcome (i.e. item acquisition). Nonetheless, it is a by-product of the prospective nature of environments. Fundamentally, temporal distancing is not a sufficient criterion to differentiate environments from items.

1.3 Decision-making frameworks

Valuation and value comparison are two indispensable processes for value-based decision and scrutinization of the functional specialization of these two processes are pivotal to the understanding of decision-making. In the visual perception literature, functional specialization is established by conclusive evidence that the fusiform face area

(FFA) and parahippocampal place area are regions selective for perception of faces and houses respectively (Kanwisher & Yovel, 2006; Pitcher et al., 2019; Tong et al., 1998). In a similar vein, the functional specialization during decision-making could operate according to the class of the options under the item-environment dichotomy –the valuation processes of items and environments are subserved by separate neural substrates. On the other hand, the FFA was also found activated when stimuli other than faces, that is objects which people were familiar with, were viewed (Gauthier et al., 1999, 2000). It suggests a generalization of encoding across divergent categories. It is also possible for the case of valuation – a neural substrate could have a generalizable coding to a variety of options. Apart from valuation, value comparison is another indispensable process in decision-making. Previous findings have shown that value signals could emerge even though no value-based decisions were required (Lebreton et al., 2009; Levy et al., 2011; Lopez-Persem et al., 2020). It implies the possiblity that the neural substrates subserving valuation and value comparison are dissociated. In order to tease out the functional specialization during decision-making, three candidate frameworks are proposed here by which the valuation and value comparison processes can be disentangled: (1) Option Homogeneity Framework, (2) Decision Homogeneity Framework, and (3) Neural Common Currency Framework.

Under the Option Homogeneity Framework, each class of option has an exclusive neural substrate for valuation and within-class value comparison (Fig. 1.1; left panel). As such, item choice and environment choice involve distinct neural substrates, and between-class decisions (i.e. item versus environment) will be subserved by another neural substrate. On the other hand, the Decision Homogeneity Framework is similar to

the Option Homogeneity Framework that different neural substrates are required for within-class and between-class decisions separately (Fig. 1.1; middle panel). However, the Decision Homogeneity Framework assumes all within-class decisions are subserved by one single neural substrate as opposed to distinct neural substrates. Finally, the Neural Common Currency Framework posits a single neural substrate is sufficient to deal with the valuation for options of all classes and decisions of any kinds (Fig. 1.1; right panel). The possibility of all these frameworks was extensively discussed in the following chapters.

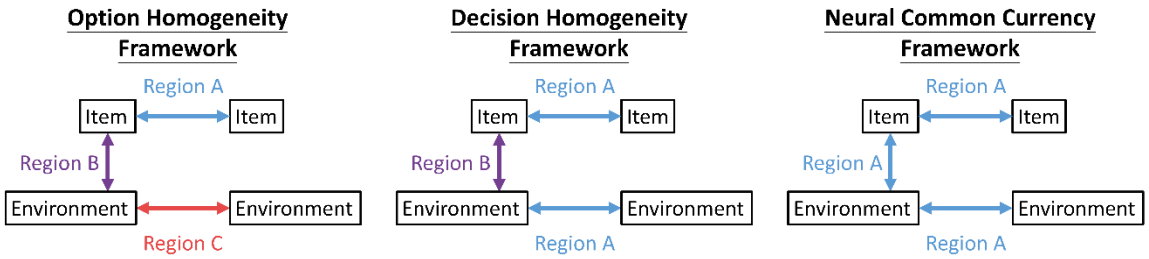


Figure 1.1. Frameworks of the mechanisms underlying different decisions. The Option Homogeneity Framework **(left panel)** suggests there are distinct brain regions responsible for the valuation and value comparison of each class of option and a separate region is rerquired for between-class decisions (i.e. item versus environment). In contrast, the Decision Homogeneity Framework **(middle panel)** asserts that only two regions are required; one region can deal with all kinds of within-class decisions (regardless of between items or between environments) and one region is responsible for between-class decisions. Finally, the Neural Common Currency Framework **(right panel)** assumes that a single brain region is sufficient to work for all kinds of decisiosn, independent of whether they are within-class or between-class.

1.4 Neuroanatomy

There are three candidate neural substrates to be involved in the abovementioned frameworks, namely the ventromedial prefrontal cortex (vmPFC), dorsal anterior prefrontal cortex (dACC), and frontopolar cortex (FPC). They have been shown closely tied to decision-making in the existing literature. Before scrutinization of their precise functional natures, the anatomy of these regions is reviewed here.

The human vmPFC is referred to the areas which are lying on the midline and located ventrally to the genu of the corpus callosum (Roberts & Clarke, 2019). It is homologous to primate medial orbitofrontal cortex (mOFC) or BA14 (Mackey & Petrides, 2010, 2014; Neubert et al., 2015). In primates, the OFC is primarily composed of three regions, namely BA11, BA 13, and BA14, but the OFC also includes parts of BA47/12 (Carmichael & Price, 1994; Petrides & Pandya, 1994). BA11 and BA47/12 are granular whereas BA13 and BA14 are agranular in nature (Roberts & Clarke, 2019). The OFC has strong connections with the adjacent dACC (Carmichael & Price, 1995) and the ventral striatum (Haber et al., 1995), another region which plays an important role in decision-making (e.g. reward value representation) (Bartra et al., 2013; Knutson et al., 2005).

The dACC is composed of BA24 and BA32 and they are parts of the agranular cortex (Petrides & Pandya, 1994). The dACC is located dorsally to the genu of the corpus callosum with a rostral border to the FPC and a caudal border to the posterior cingulate cortex (Heilbronner & Hayden, 2016). In addition to the connection with the

vmPFC, the dACC is strongly connected with the adjacent FPC and the ventral striatum (Barbas & Pandya, 1989; Vogt et al., 1987).

The FPC (also known as rostral prefrontal cortex and anterior prefrontal cortex) is the most anterior part of the granular prefrontal cortex (Semendeferi et al., 2001). It is composed of BA10 which includes the frontomarginal sulcus, the rostral superior frontal gyrus, and parts of the middle frontal gyrus (Brodmann, 1909). The FPC has a caudal border to BA46, a rostral border to BA11, and a mesial border to BA32. It shows strong connections with the adjacent OFC (Petrides & Pandya, 2007; Yeterian et al., 2012) and dACC (Carmichael & Price, 1996; Cavada et al., 2000).

1.5 Item choice

1.5.1 Neural substrate underlying valuation

Decision-making is about value comparison and there is a fundamental assumption that people always choose according to their preference. Such that the chosen option must be the preferable one or in other words the option possessing higher subjective value. The neural substrates underlying decision-making should show activity corresponding to this revealed preference. Therefore, searching for the active brain regions during making simple choices between concrete items could be the simplest approach to uncover the neural mechanisms underlying decision-making. An early study adopted this approach and showed that the vmPFC is the candidate region (Paulus & Frank, 2003). Participants underwent a binary item choice task in which they were asked to choose according to their preference to the items or physical description of the items. The vmPFC was found more active when decisions were based on participants'

preference, suggesting that the vmPFC activity was pertaining to decisions relying on values.

Despite the successful identification of the vmPFC, there is a critical limitation of revealed preference on the understanding of the role of the vmPFC in item choice. Revealed preference can only reflect the order of likeability among different items but a quantitative account is lacked. Essentially, a decision is made by comparison between the items. Quantifiable subjective values towards the items should hence exist such that the underlying neural computation can be executed. To this end, different approaches to probe the quantitative nature of item value were developed subsequently. Measuring the willingness-to-pay (WTP) is one of the nowadays widely used means (Plassmann et al., 2007). Preference towards an item can be numerically represented by the indication of how much money is willing to be paid in exchange for the item. Plassmann and colleagues (2007) required participants to bid different food items via WTP. Strikingly, the vmPFC activity was found correlated with the WTP of the food items (Plassmann et al., 2007). Moreover, this value signal was only observed when people were free to bid but absent when they were forced to bid with preassigned amount, revealing that the vmPFC activity reflecting the subjective value of the item. Apart from food items, several lines of studies concurrently demonstrated that the vmPFC activity also correlated with the preference ratings of faces, odour, and paintings (Harvey et al., 2010; O'Doherty et al., 2003; Rolls et al., 2003). Furthermore, the vmPFC could also signal the value of more abstract items. O'Doherty, Kringelbach, Rolls, Hornak, and Andrews (2001) adopted a reinforcement learning task in which abstract items (i.e. fractal patterns) were associated with different reward amount. Participants progressively learnt the association

through trial-and-error. During the course of learning, a self-generated value representation of each item arose and this self-generated value was found represented in the vmPFC.

With the considerable body of research revealing a value signal in the vmPFC responding to diverse items, a proposal of "neural common currency" was asserted. The common currency hypothesis advocates that a common scale exists through which valuation of all options are on the same lines (Montague & Berns, 2002). Hence comparison among different options, even in dissimilar categories, is feasible. To test the neural common currency signal, different experimental tasks which involved simultaneously viewing items from dissimilar categories were carried out and yielded congruent findings – a category-independent value representation was found in the vmPFC (Chib et al., 2009; Lebreton et al., 2009; McNamee et al., 2013). On top of these congruent findings from individual studies, two meta-analyses including up to 81 studies further provided affirming evidence about the presence of neural common currency signal in the vmPFC (Bartra et al., 2013; Clithero & Rangel, 2014).

1.5.2 Cognitive map coding

Apart from the long-held univariate coding, a growing body of studies points to a cognitive map coding in the vmPFC (e.g. Bongioanni et al., 2021; Constantinescu et al., 2016; Park et al., 2020; Schuck et al., 2016). Cognitive map coding, which allows multi-dimensional representation, indeed provide an advantageous account to delineate the value computation process because real-life item choices are most often multifaceted. For example, watch A is more elegant than watch B, but it is also costly. Integration of different attributes such as the design and price are necessary in order to

make a decision. Under the conventional univariate coding approach, value computation is carried out by integration of different attributes into a unidimensional representation (e.g. a weighted-sum of values in design and price). Explanation about how decisions are guided is limited especially when options are similar or even identical in integrated value. By contrast, cognitive map coding involves multiple dimensions for value representation with regard to different attributes. It provides a more precise account about how the integration process occurs along the attributes and thereby guides the decision behaviour.

Cognitive map coding is not a conventional approach for value-based decision-making but it is instead widely used in spatial navigation (Fyhn et al., 2007; Hafting et al., 2005; Leutgeb et al., 2005). Therefore, Constantinescu, O'Reilly, and Behrens (2016) attempted to examine whether conceptual information can also be represented by a cognitive map coding analogous to spatial information. They trained participants to associate particular items with different bird silhouettes of variable neck and leg lengths. The association was indeed organized in a 2D "bird space" that an item could be "navigated" in the "bird space" with parametric changes in neck and leg lengths (Fig. 1.2). Participants were instructed to learn the association by focusing on the neck and leg lengths, but critically they were not informed about the organization of the "bird space". During fMRI scanning, a video was displayed in which a bird morphed according to a particular ratio of neck length to leg length. Participants were required to imagine the outcome of morphing. In terms of "bird space", imagination of the morphing outcome implies a displacement from one item to another item. The vmPFC

activity was found modulated by the trajectory in the "bird space", providing evidence for the cognitive map coding hypothesis of conceptual information.

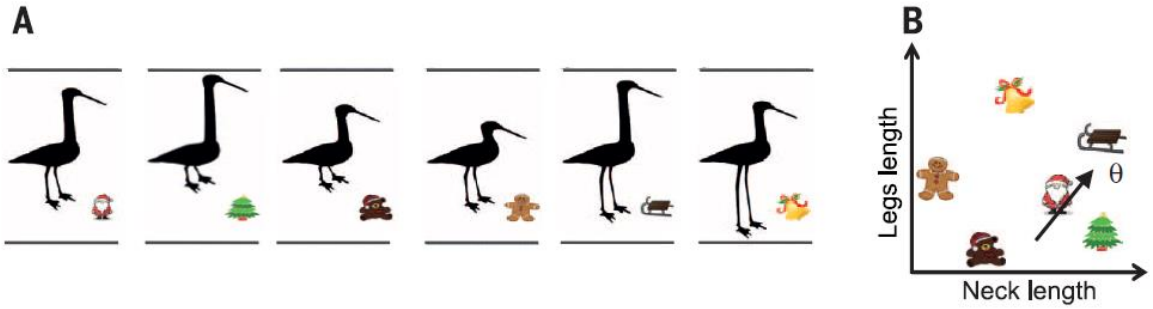


Figure 1.2. Stimuli used for studying the cognitive map coding in the vmPFC. (a) Participants learn to associate different items with bird silhouettes of variable neck and leg lengths. **(b)** The items were organized in a 2D map with respect to the neck length and legs length of their corresponding bird silhouettes. Adapted from Constantinescu et al., 2016.

Given the important role of the vmPFC in valuation of items, it is critical for the cognitive map coding in the vmPFC to be applicable to unexperienced items because in our daily lives it is inevitable to make decisions between novel options. Therefore, it is necessary to assure the cognitive map coding is not the result of parsimonious stimulus-outcome pairing. A recent study by Park and colleagues (2020) sheds light on the generalizability of cognitive map coding to novel items, by showing that two items which had never paired up could be compared with regard to their Euclidean distance in the cognitive map. Participants first learnt the social status of different people along two dimensions separately, namely popularity and competence. They were given various pairs of photos of the people and made decisions to indicate who possessed a higher social status in a given dimension. Feedback was provided such that the hierarchy could

be acquired across trials. Under fMRI scanning, participants were presented with novel pairs of photos and again chose the one possessed a higher social status in a given dimension. Notably, the vmPFC activity positively correlated with the Euclidean distance between the photos. It implies the locations of the items in the cognitive map were accessed for making decisions and it precludes the possibility that cognitive map coding is confounded by stimulus-outcome pairing. Similar results were also observed in non-human primates. Bongioanni and colleagues (2021) trained rhesus macaques to learn separately about the reward magnitude and reward probability of particular items. The items were presented in a form of a rectangular box with dots inside. Reward magnitude of an item was denoted by its colour (the more greenish [blueish], the smaller [larger] the reward) and the reward probability was denoted by the number of dots inside (the more the dots, the higher the reward probability). One set of training items varied in colour with dot number fixed while another set of training items varied in dot number with colour fixed. After training, monkeys were required to choose between unexperienced items with variable reward magnitude and reward probability. Surprisingly, the monkeys could generalize their understanding of colour and dot number to novel items and made decisions accordingly. It reflects the values of the items were not represented by a simple stimulus-outcome pairing manner. Moreover, if the items values are represented by a cognitive map coding, successive presentation of different items is analogous to displacement in the cognitive map. Hence, in a follow-up experiment, the monkeys were required to observe a single item on each trial under fMRI scanning to examine whether the inter-trial displacement in the cognitive map is reflected

in the vmPFC. The results showed that it was the case – the vmPFC activity was modulated by the trajectory in the cognitive map between successive trials.

Cognitive map coding was also found apart from contexts which require learning in advance in the vmPFC. In the study by Schuck and colleagues (2016), participants repeatedly viewed face-house rivalry images and judged the age of the face or house (young versus old). On the beginning trial, the category to be judged was cued (e.g. human face). The age of the cued category would remain the same for subsequent trials (e.g. a young human face). When age change occurred (e.g. a young man face changed to an old man face), it indicated switch of judging category. Starting from the next trial, participants had to focus on the alternative category (e.g. house) and made judgement. Therefore, it was essential to infer the task states about the category and age of both previous and current trial. Schuck and colleagues (2016) proposed there is a cognitive map where all task states are represented and transition among different task states are traced during the task. With the use of multi-voxel pattern analysis (MVPA), different task states were decoded from the vmPFC voxel activation patterns. It implies the presence of a cognitive map containing all task states in the vmPFC.

1.5.3 Neural substrate underlying value comparison

Existence of neural common currency signal in the vmPFC makes the vmPFC a plausible candidate region underlying value comparison during item choice. A considerable body of evidence did show that a signal of value difference between items was observed in the vmPFC. For instance, in a relatively parsimonious decision-making task, participants chose between two food items according to their preference on each trial (Lim et al., 2011). The vmPFC activity positively correlated with the

value difference between the items. In lines with the findings of neural common currency, apart from food items, value difference signal was also observed in the vmPFC during item choice involving a great variety of items, such as music or magazines (Lopez-Persem et al., 2016), lotteries (Chau et al., 2014, 2020; Hunt et al., 2012; Kolling et al., 2014), or even more abstract items whose value had to be learnt across trial (Boorman et al., 2009, 2013).

Notwithstanding, it is insufficient to affirm the role of the vmPFC in decision-making in spite of its encoding of value difference between items. To support the causal role of the vmPFC in value comparison during item choice, lesion studies could provide straightforward evidence. In fact, a series of lesion studies did demonstrate that patients with vmPFC lesion exhibited distorted revealed preference during item choice (Camille et al., 2011; Fellows & Farah, 2007). On the contrary, several lines of studies demonstrated that the value signal in the vmPFC was an automatic signal but not choice-dependent, opposing the findings from the lesion studies. For instance, the vmPFC was found to encode the item ratings during passive viewing of items from a variety of categories (Harvey et al., 2010; Levy et al., 2011). In another study where participants were presented with different faces and required to indicate the gender of the faces, the vmPFC activity was found positively correlated with the attractiveness ratings of the faces (O'Doherty et al., 2003). Nonetheless, it is worth noting that most of the evidence of automatic value signal in the vmPFC was derived from paradigms with single item presented. Automatic signal might only arise for valuation but not for value comparison. A study by Frömer and colleagues (2019) sheds light on this controversy. The authors conducted a study to disentangle task goal value from item

value by requiring the participants to make choices among the same set of items under two conditions: (1) choosing the most desirable items and (2) choosing the least desirable items. Therefore, the item value was constant but the task goal value was in the opposite manner across the two conditions. The results showed that the vmPFC encoded both item value and task goal value. Crucially, the vmPFC only signalled the value difference in task goal value but the value difference signal of item value was lacked. These findings provide a compatible account for the choice-independent value encoding of item value during passive viewing and the impaired value comparison revealed in lesion studies. Taken together, previous findings concurrently indicate that the vmPFC is a vital neural substrate underlying valuation and value comparison during item choice.

1.5.4 Mechanisms of value comparison

Having established the presence of value difference signal, the next question is how the value comparison process is executed. One class of theories is the diffusion models, and drift diffusion model (DDM) is an illustrative example (Ratcliff, 1978). DDM assumes two opposing decision boundaries corresponding to each option during a binary decision. When evidence which favours one of the options is provided, a particle drifts towards the corresponding decision boundary. Once the accumulated evidence reaches any decision boundary, a choice related to the boundary will be committed accordingly. DDM has been shown successful to describe behavioural choices (Krajbich et al., 2010; Lim et al., 2011; Lopez-Persem et al., 2016) but one limitation of DDM is specification of how the competition between the two choices occurs at the neural level is lacking. The biophysical cortical attractor network model

on the other hand sheds light on this issue. It suggests there are two pools of excitatory neurons which respectively receive input from the two options to be compared and encode their values through recurrent excitation (Wang, 2002). Meanwhile, there is a mutual inhibition via the interneurons connected between the two neuronal pools; such that heightened activity of a neuronal pool will suppress the activity of its counterpart. This mutual inhibition makes the value comparison process work in a winner-takes-all manner. As a result, the option corresponding to the winning pool will be chosen. Importantly, several lines of evidence have indicated that the biophysical cortical attractor network model could explain the value difference signal in the vmPFC empirically (Chau et al., 2014; Hunt et al., 2012). Single-unit recording findings however suggest valuation and value comparison take place within the same pool of vmPFC neurons and preclude the possibility of a two-pool theory (Strait et al., 2014; Yoo & Hayden, 2020). Particularly, these single-unit recording studies adopted a binary decision-making task with asynchronous presentation of items (i.e. in the first epoch, only one item was presented; in the second epoch, the second item appeared and both items were simultaneously shown). With this asynchronous presentation, valuation and value comparison could be partially separated (value comparison was only possible when the second item appeared). The authors found that the vmPFC neurons which exhibited value encoding in both epochs came from the same pool. Notably, the value encoding for the first item and second item was conducted in an antagonistic fashion – neurons exhibiting positive correlation with the value of the first item would show a negative correlation with the value of the second item, and vice versa. It suggests the same neurons first underwent option valuation but turned to

undergo value comparison when the second item appeared. Yoo and Hayden (2020) further demonstrated that values of the items were represented in subspaces in the vmPFC neurons. Crucially, distinct subspaces were found in the same neurons during two different item valuation epochs. The authors asserted that the tuning of coding from valuation to value comparison was carried out by subspace reorganization in the vmPFC neurons.

1.5.5 Properties of value signal

Thus far, large coverage has been given to the role of the vmPFC in valuation and value comparison processes during item choice. From behavioural studies, it has been revealed that human decision-making indeed possesses diverse properties (Kahneman & Tversky, 1979). It is intriguing whether the vmPFC as a key neural substrate of value processing also exhibits neural activities pertinent to these properties.

1.5.5.1 Differential responses under gains and losses. One noteworthy property during decision-making is people behave distinctively under gains and losses. In the preceding section, a large body of findings have been discussed which consistently indicate the vmPFC activity positively correlated with item value. However, most decision neuroscience studies only involve appetitive items or items leading to reward but seldom involve aversive items. It is unclear whether the vmPFC also responds to aversive items and carries a full spectrum of value signal from positive to negative. To this end, Plassmann and colleagues (2010) extended their previous work of a food item rating task (Plassmann et al., 2007) by expanding the rating scale from the positive end to the negative end. The results showed that the vmPFC differentially responded to appetitive and aversive items in the way that it positively correlated with the former but

negatively correlated with the latter, suggesting a unified spectrum to encode both positive and negative values in the vmPFC.

In addition to differentiation between appetitive and aversive coding of primary reward or punishment, similar coding was also found in the vmPFC for secondary reward or punishment. In the study by Tom and colleagues (2007), participants were offered items with a 50/50 chance leading to a monetary gain or loss and they could accept or reject the offer. The vmPFC exhibited negative correlation with the value of losses apart from the positive correlation with the value of gains. Furthermore, according to prospect theory people tend to be more sensitive to losses than gains of the same magnitude and particularly the sensitivity to losses is about twice to gains (Kahneman & Tversky, 1979). Tom et al. (2007) also found that the effect size of the vmPFC responses to losses was approximately two times to that to gains, providing a neural account of economic choices and importantly value processing of secondary items.

In spite of the findings that positive and negative values are discerned by the vmPFC, these studies are merely pertinent to the valuation process but inspection of the value comparison process is lacking. Value comparison only concerns which option is better and how much it is better than its counterpart, regardless of whether the options are indeed appetitive or aversive. Hence, it is crucial to scrutinize whether value comparison is carried out indifferently under gains and losses in the vmPFC to affirm the role of the vmPFC in computation of value difference. The findings by FitzGerald, Seymour, and Dolan (2009) did support this notion. In their study, participants were offered variable money-item pairs. In gain condition, they decided either to obtain the money or the item. On the contrary, in loss condition, they were provided with all the

money and items in advance and they had to decide which one to give up. The vmPFC was found signalling the value difference in both gain and loss conditions. The presence of this essential property of value difference signal in the vmPFC affirms its role in value comparison.

1.5.5.2 Independence of salience. Despite the large body of findings about the presence of value signal in the vmPFC, item value often covaries with salience especially those studies examine appetitive or rewarding items alone without aversive items. Salience is closely tied to attentional capture (Yantis & Jonides, 1984) and there are extensive works suggesting the impact of attention on perception during valuation and on choice bias during value comparison. For instance, scrutinizing the fixation patterns in item choice with DDM revealed that value of the attended item was amplified and the evidence accumulation towards the attended item was favoured (Krajbich et al., 2010; Krajbich & Rangel, 2011). Meanwhile, evidence accumulation of the unattended item was discounted (Thomas et al., 2019). Besides, taking both item value and salience into account rather than either factor in isolation has been shown to better depict item choice behaviour (Towal et al., 2013). In fact, attentional modulation is well-established in the perceptual decision-making literature. There are two long-held explanations: (1) encoding of visual information is enhanced by attention (Cohen & Maunsell, 2009; Mitchell et al., 2009; Ruff & Cohen, 2014); (2) read out of encoded visual information to guiding choices is tuned and improved by attention (Fries, 2015; Ruff & Cohen, 2016, 2017). Recently, a new hypothesis of attentional modulation asserts attention indeed does not improve the encoding of visual information nor the read out of encoded visual information, but reshapes the neural representation of the

encoded information to align with the current way of read out; such that behavioural responses are improved (Ruff & Cohen, 2019). This account is also plausible in the case of value-based decision-making – the representation of the encoded item value is reshaped to better suit the current way of read out such that the item value is utilized more efficiently to guiding choices. Illustrating the influence of attention on decision-making with all these lines of findings, it is thus critical to ensure whether the vmPFC activity during item choice is value-driven but not salience-driven.

To dissociate item value and salience, Litt and colleagues (2011) adapted a food rating task (Plassmann et al., 2007) by asking the participants to indicate their liking ratings to different food items (i.e. measure of value) and the extent they would like to consume the food at the end of experiment (i.e. measure of salience). The results showed that the vmPFC activity monotonically increased with value but not salience. On the other hand, Zhang and colleagues (2017) adopted a similar approach to address this issue. In their study, participants indicated their liking ratings to different stimuli. Crucially, stimuli included both reward (pleasant face images and monetary gain) and punishment (electric shock and monetary loss) in variable extent. Since the value and salience of the punishment were anti-correlated, such design could dissociate item value from salience. Akin to the findings by Litt and colleagues (2011), fMRI results revealed that the vmPFC exhibited a value signal but a lack of salience signal.

1.5.5.3 Relative rather than absolute coding of value. Thus far, all discussed findings point to a conclusion that the value signal in the vmPFC is absolute, for example, the vmPFC exhibits positive activity to gains but negative activity to losses. However, according to prospect theory, a mental reference point exists – how good or

bad an item being is determined by comparison with the reference point rather than its factual value (Kahneman & Tversky, 1979). It is possible that the neural value signal possesses a feature of relative coding similar to reference point. In fact, relative coding receives numerous empirical supports. For instance, in the study by Elliott, Agnew, and Deakin (2008), three items (in form of abstract patterns) were associated with large, intermediate, and small reward respectively. On each trial, a pair of items were displayed at first and then one of them disappeared after a delay. Participants were required to make value-independent responses to indicate which side the remaining item located at. Notably, there were two phases that the neurons for item valuation should respond: (1) when item pair was displayed, and (2) when only one item remained. If the value coding is executed in an absolute manner, the encoded value of the remaining item should be independent of the identity of the alternative. Interestingly, the vmPFC demonstrated higher activity when the intermediate item remained after the small-intermediate pair than after the large-intermediate pair. It implies the valuation to the intermediate item depended on the hierarchical position to the alternative item, supporting a relative coding hypothesis. In a similar vein, in the study by Kim and colleagues (2006), participants performed a binary decision-making task between items. Each item was associated with either a probabilistic reward or a probabilistic loss and one of the items possessed higher probability. On trials where both items were rewarding, unsurprisingly the vmPFC showed higher activity when the more probable reward was obtained. Intriguingly, acquisition of the less probable reward gave rise to lower vmPFC activity as if a loss is encountered. In contrast, on trials where both items were associated with losses, the vmPFC exhibited heightened activity to the avoidance of more probable loss but lowered

activity to avoidance of less probable loss. These results reflect that valuation of an item is regardless of its actual valence, but its relative value to the alternative item.

In theory, item value could go to infinity and beyond. Yet practically it is impossible for the valuation system to have a boundless representation of the item value and thus an efficient coding is necessary. Relative coding indeed forms the basis of efficient coding to cope with the dynamic environment. To investigate whether the valuation system possesses such adaptive nature, one means is by inspection of the value signal under different value ranges. If the value encoding is adaptive, the value-encoding brain region should exhibit similar maximum and minimum activities when it exposes to wide or narrow value ranges. An fMRI study revealed that the vmPFC activity did demonstrate range adaption (Cox & Kable, 2014). Participants performed a binary decision-making task between items block-by-block. Notably, the value range of the presented items changed across blocks. The vmPFC was found exhibited a larger regression slope when the value range was narrow whereas exhibited a smaller regression slope when the value range was wide. It reflects the vmPFC adaptively adjust its activity within the maximum and minimum borders. Taken together, all these lines of findings come to a conclusion that valuation of items is performed in a relative, rather than an absolute, manner.

1.5.5.4 Context dependence. Living in the changing world, it is vital to adjust the behaviour and adapt to the dynamic demand for the current environment. Yet, the "optimal" behavioural strategy may vary under different contexts. It is necessary for a decision maker to make choices according to the current context (e.g. re-evaluation of options time by time, focusing on relevant information). From an ecological perspective,

it is particularly important for the existence of an automatic context-dependent valuation process because it can help quickly discern whether approach or avoidance is needed. Several lines of findings indeed have revealed people's valuation as well as the corresponding value signal in the vmPFC were modulated by the context unintentionally. For example, in the study by Harvey and colleagues (2010), participants were asked to passively view different paintings during fMRI scanning but they were explicitly informed before experiment that their participation would be compensated by one of two companies. Paintings were presented with simultaneous display of either a sponsor or non-sponsor logo. Participants had no reason to upweight the paintings presented with sponsor logo because the only instruction they received was to view the paintings passively. However, surprisingly a post-scan rating task indicated that the ratings of the paintings that were presented with sponsor logo were higher than those presented with non-sponsor logo. Importantly, this was also reflected in the vmPFC activity. In a similar vein, Plassmann and colleagues (2008) presented some wines with different price labels to people and asked them to indicate their liking ratings towards the wines. Unbeknown to the participatns, two wines were presented twice with substantially different prices (wine 1: $5 & $45; wine 2: $10 & $90). Interestingly, the resutls revealed that participants perceived the same wine more appetitive when it was priced higher. It was not only observed behaviourally, but also exhibited at the neural level that the vmPFC showed heightened activity for the same wine when the labelled price was higher.

In additon to the modulation by concurrent context, another line of evidence demonstrated that the baseline activity of vmPFC was susceptible to task-independent

contextual information prior to stimulus onset. For instance, Abitbol and colleagues (2015) presented some music prior to a rating task of paintings to the participants on each trial. Participants were explicitly instructed to ignore the musical context preceding the presentation of the paintings. It was found that the vmPFC activity towards the ratings of paintings was modulated by the pleasantness of the musical context – a more pleasant musical context led to higher vmPFC activity to the subsequent painting. Apart from that, in the study by Losecaat Vermeer and colleagues (2014), participants could choose to accept or reject a 50/50 gamble that potentially resulted in a gain or loss. Prior to each gamble, they underwent a time-estimation task in which correct responses gave rise to a reward but incorrect responses led to a loss. As such, a pre-gamble context of gain or loss was provided by the time-estimation task. Behavioural results showed that people tended to avoid taking risks after receiving a gain and tended to be risk-seeking after encountering a loss. Interestingly, this "adaptation" of risk preference to the context was also reflected in the vmPFC. The vmPFC exhibited higher activity if people rejected to take the gamble after a gain, compared to if they accepted the gamble after a gain. Similarly, the vmPFC showed higher activity if people chose to gamble after a loss than if they refused to gamble after a loss.

1.6 Decision between items and environments (foraging)

In addition to the widely investigated item choice, how decisions are made between items and environments (also known as foraging in the existing literature) has

received growing concern in the last decade. On most occasions, decision between items and environments concerns the stay-switch problem of deciding to stay and engage currently available items or to explore an alternative environment for potentially better items. It is as ubiquitous as item choice in our daily lives, for instance, an individual can buy food from the only food stand in front or s/he can walk around the district to look for any restaurant providing better food. To directly examine the neural mechanisms of decision between items and environments, Kolling and colleagues (2012) designed a task in which participants were presented with two available items and an alternative environment composing of multiple items. They could make a choice between the two available items or forgo the items and search in the alternative environment to obtain another two items for selection. Behavioural results showed that participants were more likely to search when the value of the environment was high and contrarily less likely to search when the value of the available items was high. As discussed in preceding sections, the vmPFC has been implicated in decision-making among diverse items by exhibiting a neural common currency signal. It is plausible that the vmPFC also encodes the value of the environment and compares the environment with the available items to make a choice. Surprisingly, the dorsal anterior cingulate cortex (dACC), but not the vmPFC, was identified to show positively correlated activity with the environment value. In addition, the dACC activity decreased with the value of the available items. This implies the dACC carried a value difference signal (environment value minus item value). Importantly, participants whose dACC activity better predicted by environment value also showed higher tendency to search, suggesting the role of the dACC in driving searching. On the

contrary, when participants made choices between the two available items, the vmPFC demonstrated a value difference signal between the items. This dissociation between the roles of the dACC and vmPFC in value encoding of environments and items partially precludes the possibility of the Neural Common Currency Framework.

Furthermore, Kolling and colleagues (2018) reasoned that during evaluation of an environment, people also hold an expectation to obtain the desired composing items apart from estimation of how good overall the composing items are. To this end, they adapted the task by Kolling and colleagues (2012) to further study this prospective nature of environments. Participants were offered an item and an opportunity to search in another environment for alternative items. Crucially, the number of possible searches was manipulated and explicitly disclosed. Supposedly, in order to obtain the desired items, people are more likely to search when there is a long time horizon and vice versa. And there should be a neural value signal about the most rewarding item. It was indeed the case of the results. The value of the most rewarding item in the environment was represented in the dACC. It implies participants prospectively pursued for the most desired item by searching the environment. In addition, consistent to the previous findings by Kolling and colleagues (2012), a value signal of the environment (i.e. the average value of the composing items within the environment) was found in the dACC. These pieces of results revealed that the dACC was closely tied to the valuation of environments.

The role of the dACC in valuation of environments can also be implicated by studies focusing on environment-leaving. In terms of ecology, environment-leaving refers to the behaviour observed in wild animals that they forgo the current food

resources and search in the environment. In the nature, living in a region for a period of time, food resource will be depleted gradually. Animals have to decide whether to consume the diminishing food resource (i.e. exploiting the currently available items) or to search in the environment for alternative sources (i.e. exploring alternative environment for opportunity of potentially better items). Despite the similarities between environment-leaving and decision between items and environments, there is a subtle difference among these two processes. Instead of comparing a concrete item with an environment, environment-leaving concerns whether to leave an environment and engage in an alternative environment. Moreover, searching is accompanied by costs – there might be a period of time when the supply of food is lacking during searching. Additionally, the new sources of food could be a wealthier one but also it might yield fewer food than the current one. It is thus critical to estimate the trade-off between potential gains and costs of leaving an environment.

To simulate environment-leaving and investigate the underlying neural mechanisms, Wittmann and colleagues (2016) adopted a task in which participants were assigned an environment to engage in by default. Reward was delivered across time but the reward rate could either increase or decrease gradually. At certain time points, participants could make choices to stay in the current environment, or to leave the environment and engage an alternative environment which had known, constant reward rate instead. To decide whether to stay or leave, it is critical to estimate whether the reward rate of the current environment was trending to increase or decrease by comparing the most recent reward rate (LastRR) with the average past reward rate (AvgRR). It was found that participants were more likely to stay when LastRR was

high and less likely to stay when AvgRR was high. In other words, environment-leaving was guided by the value difference between LastRR and AvgRR and a value difference signal (LastRR minus AvgRR) was found in the dACC.

Kaiser and colleagues (2021) further tested the role of the dACC in environment-leaving by examination of the concentrations of glutamate and GABA, the neurotransmitters of excitation and inhibition, in the dACC. If the dACC is pertaining to environment-leaving, the balance between glutamatergic excitation and GABAergic inhibition (E/I balance) in the dACC should change with the value of environment-leaving. The authors adopted a two-stage decision-making task to mimic environment-leaving. Participants were first assigned to one of two simultaneously presented environments. They could choose to stay in the current environment or switch to the alternative environment. After a decision, the reward magnitudes of all available items within the environments were shown and two items were randomly drawn. These two items were also randomly assigned with different reward probabilities and participants chose an item to earn the reward. Notably, the reward in the current environment depleted over trials while the alternative environment was replenished, but switching was associated with a cost. To make a leaving decision, it was critical to consider the leaving advantage – the relative benefit of switching (i.e. value difference between alternative environment and current environment) to the cost associated with leaving. This leaving advantage on trials where the participants switched was correlated with the E/I balance in the dACC, indicating that the dACC was closely tied to guiding environment-leaving. Interestingly, the correlation between leaving advantage and the E/I balance in the vmPFC was also tested but yielded no significant results. On the

other hand, the second stage involved item choice and the value difference between the items was found correlated with the vmPFC E/I balance but not with the dACC E/I balance. This double dissociation echoed with the findings by Kolling and colleagues (2012) that the valuation processes of environments and items involve dissociable neural substrates.

Single-unit recordings also provide evidence of the role of the dACC in environment-leaving. Two rhesus macaques performed an environment-leaving task in which they could exploit an environment but the reward amount reduced every time it had been exploited, or they could leave the environment to replenish the environment with a cost of travel time (Hayden et al., 2011). The key consideration was thus to trace how long they had been staying in the environment. The dACC firing rates were found increased every time a decision of staying was made. Importantly, the authors identified a threshold of leaving – when the dACC firing rates reached the threshold, the subjects left the environment. These findings suggest that the dACC was pertaining to valuation of environments as well as guiding environment-leaving.

1.7 Items and environments are not confounded by component diversity or temporal distance

Thus far, the literature reviewed in the preceding sections suggests that the vmPFC and dACC underlie the valuation processes for items and environments respectively. It is worth taking a closer look at the distinction between these two classes of options to assure the dissociation. Environments are characterized by two features: (1)

environments are composed of a variety of items and (2) the relation between an environment choice and the final outcome is temporally distant (because selection of an environment gives rise to the opportunity to come across the component items rather than actual acquisition). It is easy to mix up environments with items associated with multiples components or delayed outcomes. However, the critical feature making environments distinct from items is their prospective nature. It is more than merely a diversity of components or the temporal distancing between decision and decision outcomes. In fact, several lines of studies identified the vmPFC in encoding of item value even the items were composed of multiples components or associated with delayed outcomes, reflecting association with a variety of components or delayed outcomes are not sufficient to discern items and environments.

In the study by Shenhav and Karmarkar (2019), participants were presented with a set of four items on each trial and they were required to rate the entire item set. Behavioural results revealed that the reported rating of the item sets was correlated with the average value of all their component items. Crucially, item sets lacked a prospective nature and the item set rating was found signalled by the vmPFC, which is incompatible with the findings that environment value in terms of the average value of all component items was signalled by the dACC (Kolling et al., 2012, 2018). In a similar vein, Symmonds and colleagues (2010) presented participants with an item that was associated with four possible amounts of reward (ranged from £0-£4) with equal probabilities on each trial. Participants could accept the item or choose the alternative with a fixed amount of £2 (not visually presented) and reward was directly delivered after decision. It was found that the vmPFC also encoded the expected value (EV) of the item, consistent

to the widely reported findings that the vmPFC encoded the value of simple items which only resulted in binary outcomes (Boorman et al., 2009; Chau et al., 2014; Jocham et al., 2012; Kolling et al., 2014). Taken together, component diversity is not a sufficient criterion for the differentiation between items and environments.

Intertemporal choice on the other hand allows inspecting the differentiation between items and environments in terms of temporal distancing. Intertemporal choice stems from economics and refers to the decisions in which the gains and costs are subjected to the time span (Loewenstein & Thaler, 1989). For instance, Kable and Glimcher (2007) presented participants with a delayed large reward which varied in amount of reward magnitude ($20.25-$110) and delay (6 hours to 180 days) across trials. Participants could accept the delayed larger reward or forgo it to select an immediate small reward (fixed at $20). Intertemporal choice requires integrating the costs and benefits of the delay which differ from the ubiquitous decisions between simple items that reward accompanies directly after decision. Surprisingly, the vmPFC activity was also found correlated with the value of the delayed reward. In a similar vein, Prévost and colleagues (2010) presented fuzzy erotic images to the participants and asked the participants to choose a default option by which they could view the images clearly for 1s after a short delay, or to wait for a long dealy but they could view the clear images for 3s. The vmPFC activity also increased with the value of the delayed reward even primary reward was used. A lesion study additionally provided evidence of the causal role of the vmPFC in value encoding of delayed reward (Sellitto et al., 2010). Patients with lesions in the vmPFC made choices between an immediate small reward and a delayed large reward. Their task performance was compared with that of control patients

with lesions outside the frontal lobe and healthy control. The vmPFC patients exhibited larger preference of the immediate reward than the two control groups. Closer inspection revealed the vmPFC patients possessed a larger value discounting to the delayed reward. In other words, their abilities to integrate the costs and benefit of a delayed reward were impaired. All these lines of findings from intertemporal choice consistently point to the role of the vmPFC in value encoding of delayed reward.

On the other hand, the multi-stage structure of environment choice entails environments a prospective nature. Any option that leads to further decision might be easily recognized as an environment. Nonetheless, it should be cautious that an environment is a choice set that provide the possibility to attain potential items. Consideration of the entire choice set instead of substituting the choice set as a proxy to a certain item is essential. Hence, option that is a distant proxy to a certain item via multiple stages of decision should not be deemed as an environment. A classical two-stage decision-making paradigm did affirm that choice sets which acted as the proxy to particular items did not give rise to any neural signal of environment value (Daw et al., 2011). The paradigm consists of a decision that requires planning (first stage) and a decision that requires learning (second stage) on each trial. There were two pairs of items in total in the second stage. Participants were presented with one of the pairs each time and made decision between the items. The items were associated with different reward probabilities, which were not explicitly depicted and had to be learnt across trials. By contrast, in the first stage participants chose between two items that were associated with the item pairs. Notably, each first-stage item led to one second-stage item pairs with high likelihood (i.e. 70%) and led to the remaining pair with low likelihood (i.e. 30%). Across

trials of learning, participants became more familiar with the reward probabilities of all the second-stage items and they could choose a first-stage item that had high likelihood to reach the most rewarding second-stage item thereby maximizing their earning. Crucially, the values of the second-stage items were represented in the vmPFC during first-stage decision. It implies the first-stage items had been deemed as proxies to particular second-stage items. On top of intertemporal choice, this finding about items relating with multiple stages demonstrated that it is not sufficient to discern environments and items by temporal distancing between selection and final outcome alone.

In this chapter, I have reviewed multiple lines of studies which indicated the vmPFC also encoded the item value even item was composed of multiple components or associated with distant outcomes. It in other words suggests that the critical difference between environments and items do not stem from the diversity of components nor the temporal distancing between decision and decision outcomes. Besides, findings of item choice and environments and environment-leaving reported an absence of environment value signal in the vmPFC. Taken together, it can be assured that the critical feature making environments distinct from items should be their prospective nature. In my two studies that were reported in subsequent chapters, this prospective nature of environments was examined and environment choice and item choice were discerned.

1.8 Approaches to model decision-making behaviour

A considerable body of studies has implicated the role of the vmPFC and dACC in valuation of items and environments respectively. These previous studies indeed employed many different approaches to derive value estimates for seeking the neural

correlates of option value. Yet, are these approaches sufficient to reflect the underlying valuation process? Historically, people's economic choices were explained by normative models, which suggest how decisions should be made. One typical example is modelling people's choices by expected value, a widely used approach in the existing literature (e.g. Chau et al., 2014; Hunt et al., 2012; Kolling et al., 2012). It assumes that people always select the option yielding the highest payoff. However, the same option may be perceived differently across people, for example a lottery of winning $1000 appears much more valuable to a poor person than a billionaire. Therefore, expected utility theory was proposed which states that the value of an option should be in a form of utility, a subjective value transformed by a non-linear function, and people should make decisions according to the expected utility instead of expected value (Bernoulli, 1954). The introduction of expected utility did provide improved modelling of human choices. Nonetheless, empirically many economic choices could not be accounted for by normative models. For instance, to make a choice between a 50/50 gamble of winning $100 and a sure-win option of $49, a risk-averse individual might prefer the sure-win option despite its lower expected payoff. Descriptive models were thus developed in order to vividly describe how people subjectively evaluate the options to give rise to their choices. Therefore, estimating people's subjective value using descriptive models greatly helps unveil the neural mechanisms underlying decision-making and different models which are related to the valuation of items or environments were discussed below.

Cumulative prospect theory (CPT) is a canonical model to describe decision-making behaviour under risk (Tversky & Kahneman, 1992). It accounts for the value

estimation process during economic choices in a variety of scenarios. One of the most influential account by CPT is the formulation of loss aversion. Conventionally, the value estimation processes to both gains and losses are assumed to perform in the same manner. CPT instead suggests the value estimations of gains and losses are computed by a concave value function and a convex value function respectively (Fig. 1.3a). Moreover, the function for losses is steeper, indicating a loss aversion – loss in a given amount is perceived twice as strongly as a gain in objectively equivalent amount. Besides, CPT provides account for the value estimation in probability. CPT states that small probabilities are overweighted and moderate to large probabilities are underweighted (Fig. 1.3b), which can explain a variety of behaviour patterns (e.g. purchase of an insurance plan to a rare event, more willing to take risk when experiencing a loss).

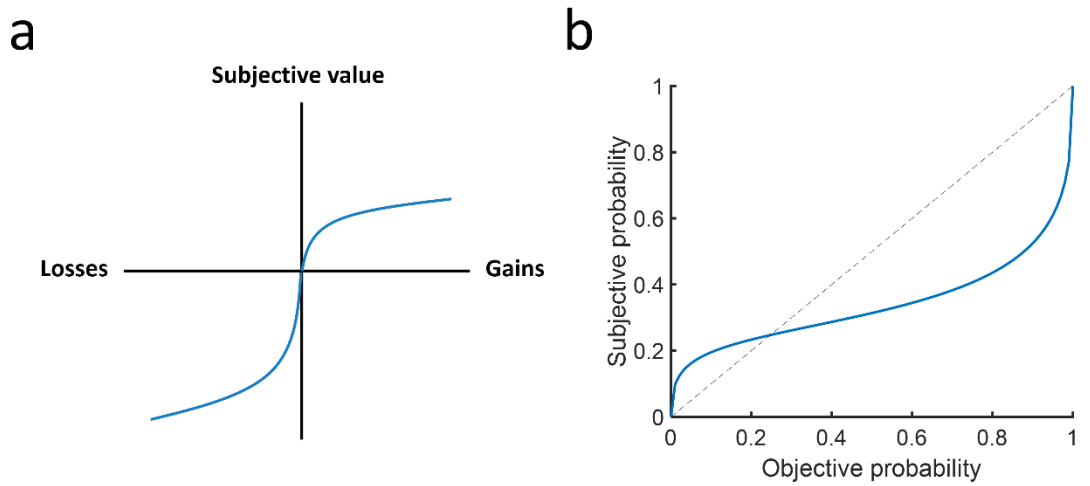


Figure 1.3. Transformation by cumulative prospect theory. (a) People possess larger sensitivity to gains than losses. The subjective value to a loss is approximately twice of that to a gain in objectively equivalent amount **(b)** People possess unequal weightings on probabilities. Small probabilities are overweighted whereas moderate and large probabilities are underweighted.

CPT most often is applied for individual quantity. For choices involving evaluation of a distribution of numbers, the value estimation becomes complex because integration of massive information is needed. CPT is not developed for describing evaluation of distribution and hence alterative models are warranted. One intuitive way is to estimate the mean value for each distribution. Besides, distributions can be characterized by different statistical moments (e.g. variance, skewness). It is plausible that evaluation of distributions is influenced by the statistical moments. The mean-variance-skewness (MVS) model (Symmonds et al., 2011, 2013; Wright, Symmonds, Morris, et al., 2013) provides a direct account of this issue. It posits evaluation of a distribution primarily relies on the mean value but the preferences of variance and skewness additionally contribute to the value estimate in a linear sum manner. Nevertheless, it has been discussed that value estimation of individual quantity could be better explained by non-linear functions (e.g. expected utility theory). It might also hold true for evaluation of distributions, especially people have insufficient cognitive resources to perfectly estimate the values of all quantities within a distribution. It is plausible that people perceive the quantities within a distribution with unequal weights. To this end, the power law model proposes each quantity within a distribution is transduced by an exponent. As a result, when estimating the mean value of a distribution, the small and large quantities are underweighted.

In recent years, a growing body of neuroscience research have been applying deep learning to develop models of cognitive processes that are hard to be portrayed by conventional computational models (e.g. object classification) (Yang & Wang, 2020). In a similar vein, there could be implicit cognitive processes during value estimation

that are overlooked by the traditional computational models, especially for distributions which involve massive information. Therefore, deep learning provides potential solutions to modelling the value estimation process. Autoencoder (Kiarashinejad et al., 2020) and convolutional neural network (CNN; Krizhevsky et al., 2012; Lindsay, 2020) are two widely used deep learning neural networks. Autoencoder is distinguished for its capability in compression of high-dimensional data into concise representations (i.e. hidden nodes) with minimal information loss. Specifically, high-dimensional data are fed as the input of the autoencoder. In the subsequent hidden layers, features of the input are extracted, meanwhile redundant information is removed. Notably, the resulting hidden nodes are highly representative to the original input but with lower dimensionality. On the other hand, the CNN is successful in extracting and integrating complicated visual features from images to solve classification problems. Images are fed as the input of the CNN. In the subsequent convolutional layer, the input is convolved by various feature detectors – features in the input are extracted and represented by different feature maps. In other words, the feature maps characterize the original images in particular aspects. Afterwards, the feature maps are aggregated in the fully-connected layer and then activated by a softmax function to generate choice predictions. Since autoencoder and CNN both involve feature extraction and integration of massive information which are akin to the evaluation of distributions, they are plausible to provide good estimate of the subjective value of distribution, thereby shed light on the value estimation process.

1.9 Research question

This thesis aims to address the question whether there is a neural mechanism which can make comparison among all different kinds of option or there are different neural mechanisms to work with particular decisions. A considerable body of research has implicated the vmPFC to decisions between a wide variety of items. It provides evidence for the Neural Common Currency Framework which assumes that all kinds of decision can be dealt with by a single neural substrate. Yet, the increasing works on the decision between items and environments in the last decade reveal that the dACC, but not the vmPFC, is pertaining to the guide item choice and environments. It has refuted the Neural Common Currency Framework. On the other hand, both the Option Homogeneity Framework and Decision Homogeneity Framework assume multiple neural mechanisms for within-class decisions (i.e. item versus item, environment versus environment) and between-class decisions (i.e. items versus environments). However, the Option Homogeneity Framework posits within-class decisions involve exclusive neural substrates for each class of option whereas the Decision Homogeneity Framework postulates a single neural substrate can work with all within-class decisions. Since the existing literature extensively investigated item choice and little endeavour was made on environment choice, the underlying neural mechanism of environment choice remains highly unclear. To discern the functional specialization during decision-making, in this thesis, the neural mechanism underlying environment choice was examined and contrasted to the neural mechanism underlying item choice.

Chapter 2 Evaluation of environments with complex information

Chapter highlights

- This chapter aims to test the evaluation process during environment choice
- People were capable of integrating complex information in environments to guide decision-making
- People preferred environments with larger means and variances of the item distributions in the environments

2.1 Introduction

Lots of endeavour in the last two decades has been made on the underlying neural mechanisms of decision between simple items. There are also growing concerns of foraging, which is about the stay-switch problem – to exploit current items or to explore an alternative environment for potentially better items. A large body of findings about item choice and foraging has been discussed in **Chapter 1**. Yet, environment choice receives far less attention in the existing literature. How the decision-making system works, specifically, whether a neural substrate universally copes with all different kinds of decision or multiple neural substrates are required to handle particular types of decision, remains elusive. To address this question, it is important to fill in the missing piece of environment choice. The first crucial step is to examine how evaluation of

environments is carried out. Environments possess two essential features which constitute to their distinctiveness. First, an environment comprises of multiple items; valuation of an environment requires consideration of the distribution of its component items – whether overall the items are good and whether there is a great variety of items to choose from. Second, selection of an environment gives rise to the opportunity to come across the component items rather than actual payoff. This study was conducted to test whether these essential features of environments are exhibited when people undergo decision-making between environments. In particular, previous studies showed that people's economic choices were biased by statistical moments (e.g. mean, variance, skewness) when options were presented in forms of distribution (Wright et al., 2012; Wright, Symmonds, & Dolan, 2013; Wright, Symmonds, Morris, et al., 2013). It is hypothesized that during environment choice, statistical moments of the item distributions within the environments are also considered and influence people's environment choices.

To address the evaluation process of environments, this study involved a parsimonious behavioural decision-making task which required participants to choose between two environments repeatedly. Each environment was composed of 20 items and selection of an environment gave rise to the opportunity to obtain one of its component items. The results showed that participants were capable of integrating the multiplex information embedded in the environments to guide decisions. Importantly, elementary findings on the evaluation process of environments were unveiled – environments were evaluated with regard to the statistical moments (i.e. mean, variance) of their component

item distributions. Environments with larger means and variances in their item distributions were preferred.

2.2 Methods

2.2.1 Participants

Twenty healthy young adults (11 females), aged 18-27 years with normal or corrected-to-normal vision and no current or history of neurological or psychiatric conditions, were recruited via advertisement in the university and participants' referral. Written informed consent for each participant was obtained before experiment. The experiment was approved by the Human Subjects Ethics Committee of the Hong Kong Polytechnic University.

2.2.2 Experimental task

This study involved a binary decision-making task between environments (Fig. 2.1a). In each block, participants chose repeatedly between two environments for 10 trials. On each trial, two environments were displayed – each of which was presented in a form of a series of 20 bars and a number. Each bar represented a lottery item and the height of each bar was related to the probability of winning the lottery. The number indicated the mean reward magnitude of all lotteries in the same environment (reward magnitude of an item ranged from -$100 to $100 in HKD). After selection of an environment, the chosen environment was surrounded by a blue frame (0.5 second). One item from the chosen environment was randomly drawn and highlighted in yellow after a delay of 0.7 second. Reward was delivered according to the reward magnitude and

reward probability of the drawn lottery item. Decision outcomes were disclosed at once after all 10 trials of the same block were completed. Participants were told that points earned in the task would be converted into monetary reward after the experiment. There was a total of 400 trials.

In **Chapter 1**, it has been discussed in detail about the neural value difference signal should be invariant to gains or losses. Despite the lack of measurement of neural data, this experiment also considered this essential property of the value comparison process. Specifically, this experiment included environments that led to losses (reward magnitude and probability were coloured in red) in addition to environments that led to gains (coloured in green). 50% of the trials involved two gain environments, 30% of the trials involved two loss environments, and 20% of the trials involved one gain and one loss environment.

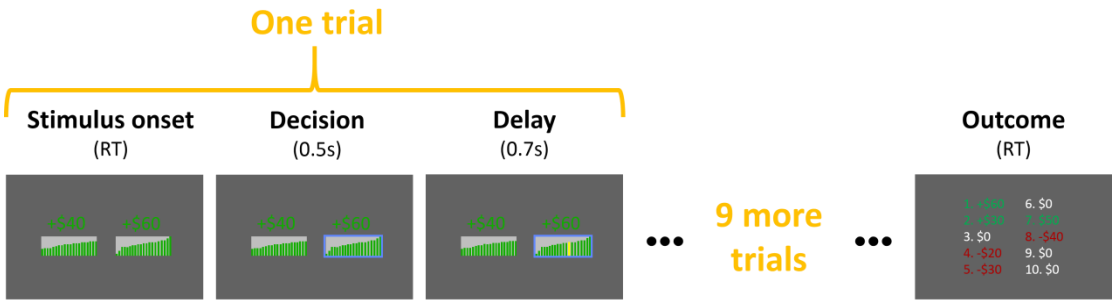


Figure 2.1. Task schematic of the experiment. Within a block, two environments were presented on each trial (*Stimulus onset*). After selection of an environment, a blue frame appeared and surrounded the chosen environment (*Decision*). A lottery item was randomly drawn from the chosen environment and highlighted in yellow after a delay (*Delay*). Decision outcomes of all trials were displayed at the end of the block (*Outcome*).

2.2.3 Statistical analysis

A logistic regression (GLM1) was applied to predict whether participants' choices (to the right=1 and left=0) were biased by the environments' reward magnitudes and probabilities respectively:

$$logit\frac{P(R)}{1-P(R)} = \beta_0 + \beta_1 Rew_R + \beta_2 Prob_R + \beta_3 Rew_L + \beta_4 Prob_L$$

where *P(R)* denotes the probability of choosing the option on the right. Rew_R and $Prob_R$ denotes the reward magnitude and reward probability of the rightward environment respectively; Rew_L and $Prob_L$ denotes the reward magnitude and reward probability of the leftward environment respectively. This logistic regression was separately applied to gain trials and loss trials.

Another logistic regression (GLM2) was performed to test whether people integrated both reward magnitude and reward probability of each environment to guide their decisions:

$$logit\frac{P(R)}{1-P(R)} = \beta_0 + \beta_1 EV_{(R-L)}$$

where *P(R)* denotes the probability of choosing the rightward environment. $EV_{(R-L)}$ denotes the difference in expected value (reward magnitude multiplied by reward probability) between the rightward and leftward environments. Apart from participants' choices, how their reaction time (RT) was affected by EV difference was also inspected (GLM3):

$$RT = \beta_0 + \beta_1 EV_{(R-L)}$$

Given each environment was composed of a distribution of items, to test whether participants' choices were influenced by the statistical moments of the item distributions, the following logistic regression (GLM4) was applied:

$$logit\frac{P(R)}{1-P(R)} = \beta_0 + \beta_1 EV_{(R-L)} + \beta_2 Mean_{(R-L)} + \beta_3 Variance_{(R-L)} + \beta_4 Skewness_{(R-L)}$$

where $Mean_{(R-L)}$, $Variance_{(R-L)}$, $Skewness_{(R-L)}$ denote the differences in mean, variance, and skewness of environment probability respectively. A one-sample t-test was performed to examine whether each beta weight was significantly different from zero across participants in all regressions.

2.3 Results

First, it is important to assure participants held correct understanding to the environments given their complex structures. This study included both environments that led to gains and environments that led to losses. Supposedly, the preference of large probability should differ between gain and loss trials – options with large probabilities should be preferred on gain trials and options with small probabilities should be preferred on loss trials. A logistic regression (GLM1) which included the respective reward magnitudes and reward probabilities of the two environments as regressors was performed to predict participants' choices on gain and loss trials separately. As expected, it was found that participants preferred environments with large reward magnitude on both gain and loss trials ($|\beta|$s>3.062, $|t|$s>8.053, Ps<2.228×10^{-7}; Fig. 2.2). Notably, the participants exhibited differential preferences of reward probability when the

environments leading to gains and when the environments leading to losses. On gain trials, participants were more likely to choose the rightward environment when its reward probability was large (β=0.841, t_{18}=6.476, P=4.317×10^{-6}; Fig. 2.2a) and less likely to choose the rightward environment when the reward probability of the leftward environment was large (β=-0.403, t_{18}=-3.679, P=0.002; Fig. 2.2a). In contrast, on loss trials, participants were less likely to select the rightward environment when its reward probability was large (β = -0.451, t_{15} = -3.884, P = 0.002; Fig. 2.2b) and more likely to select the rightward environment when the reward probability of the leftward environment was large (β = 0.548, t_{15} = 4.048, P = 0.001; Fig. 2.2b). All these pieces of results suggest participants comprehended the values of the environments despite their complex structures and behaved conformingly to the task instruction. Interestingly, participants exhibited opposite patterns of side bias on gain and loss trials – they preferred the leftward environments on gain trials (β = -0.906, t_{18} = -7.550, P = 5.537×10^{-7}; Fig. 2.2a) and rightward environments on loss trials (β = 1.328, t_{15} = 5.018, P = 1.528×10^{-4}; Fig. 2.2b). By design, the values of both reward magnitude and reward probability were counterbalanced across sides. Moreover, the side bias refers to the preference to a particular side when all other attributes (i.e. reward magnitude and reward probability) have been accounted. This unexplained side bias did not confound the crucial findings that participants were capable of making choices according to the environments' values.

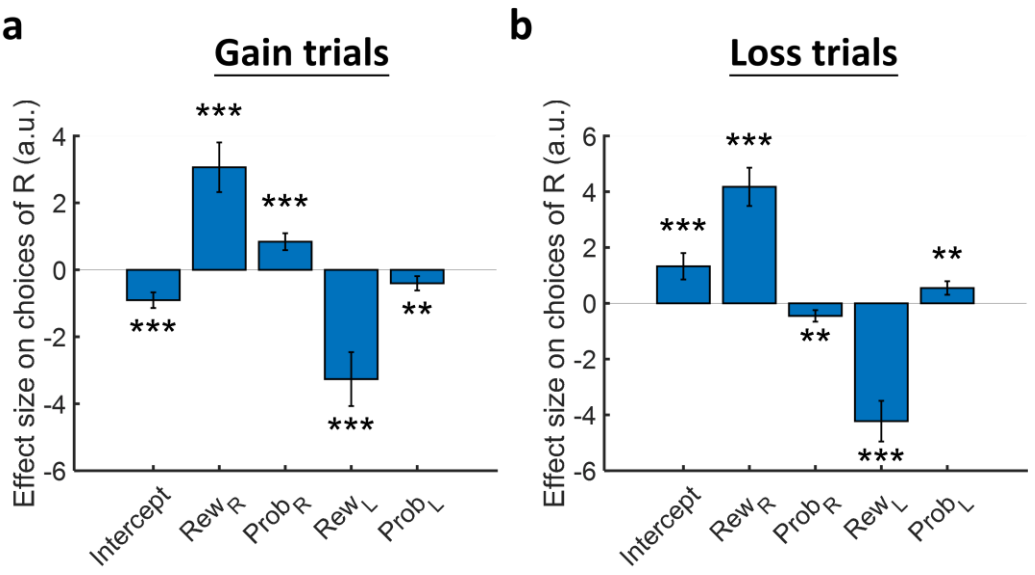


Figure 2.2. Differential preferences of reward probability under gains and losses. (a) Environments with larger reward magnitude and reward probability were preferred on gain trials. **(b)** On loss trials, environments with larger reward magnitude were still favoured but environments with larger reward probability were avoided. Rew_R=reward magnitude of the rightward environment; $Prob_R$=reward probability of the rightward environment; Rew_L=reward magnitude of the leftward environment; $Prob_L$=reward probability of the leftward environment. *** denotes $P<0.001$ and ** denotes $P<0.01$. Error bars represent means±s.e.m.

Next, it was tested in detail how environments are evaluated to guide decision-making. One conventional approach used in the literature of item choice is to regress item choices against the difference in EV between the items (Chau et al., 2014; Hunt et al., 2012). In this study, participants' environment choices were similarly regressed against the difference in EV between environments (GLM2). The results showed that participants were more likely to choose the rightward environment when its EV was relatively larger than that of the leftward environment ($EV_{(R-L)}$; β=7.356, t_{19}=6.883,

P=1.451×10^{-6}; Fig. 2.3a,b). Decisions were also made faster when the EV difference between the environments was large (β=-0.033, t_{19}=-3.557, P=0.002; Fig. 2.3c).

Having established the basis of a typical positive effect of EV difference on choices, a closer inspection of the effects by other less obvious parameters was taken. Particularly, during environment choice, it is critical to integrate information of the component items in each environment to form representations of statistical moments (e.g. the mean, variance, skewness of the item distribution). Hence, differences in the statistical moments between environments (i.e. Mean$_{(R-L)}$, Variance$_{(R-L)}$, Skewness$_{(R-L)}$) were also included as regressors in addition to the EV difference to examine whether the participants considered any of these pieces of information during decisions between environments. Consistent to the results of GLM2, environments with larger EVs were preferred (EV$_{(R-L)}$: β=7.641, t_{19}=6.989, P=1.172×10^{-6}; Fig. 2.3d). On the other hand, there was an absence of effect of mean probability (Mean$_{(R-L)}$: β=0.174, t_{19}=1.711, P=0.103). Nonetheless, it has been shown participants' choices were indeed biased by mean probability because as shown in Figure 2.2, the mean probability imposed positive and negative effects on gain trial and loss trials respectively. The effect of mean probability was cancelled out when all trials were pooled together and hence an absence of effect was observed. Interestingly, a positive effect of variance on environment choices was also observed (Variance$_{(R-L)}$: β=0.139, t_{19}=2.393, P=0.027; Fig. 2.3d). In other words, environments with more widespread distributions of item probabilities were favoured. Besides, skewness did not show any significant effect on environment choice (Skewness $_{(R-L)}$: β=0.073, t_{19}=1.215, P=0.239; Fig. 2.3d). Nonetheless, skewness is a less intuitive statistical moment than the mean and variance. One possibility for the absence

of effect could be a more sensitive test is necessary in order to capture the effect of skewness.

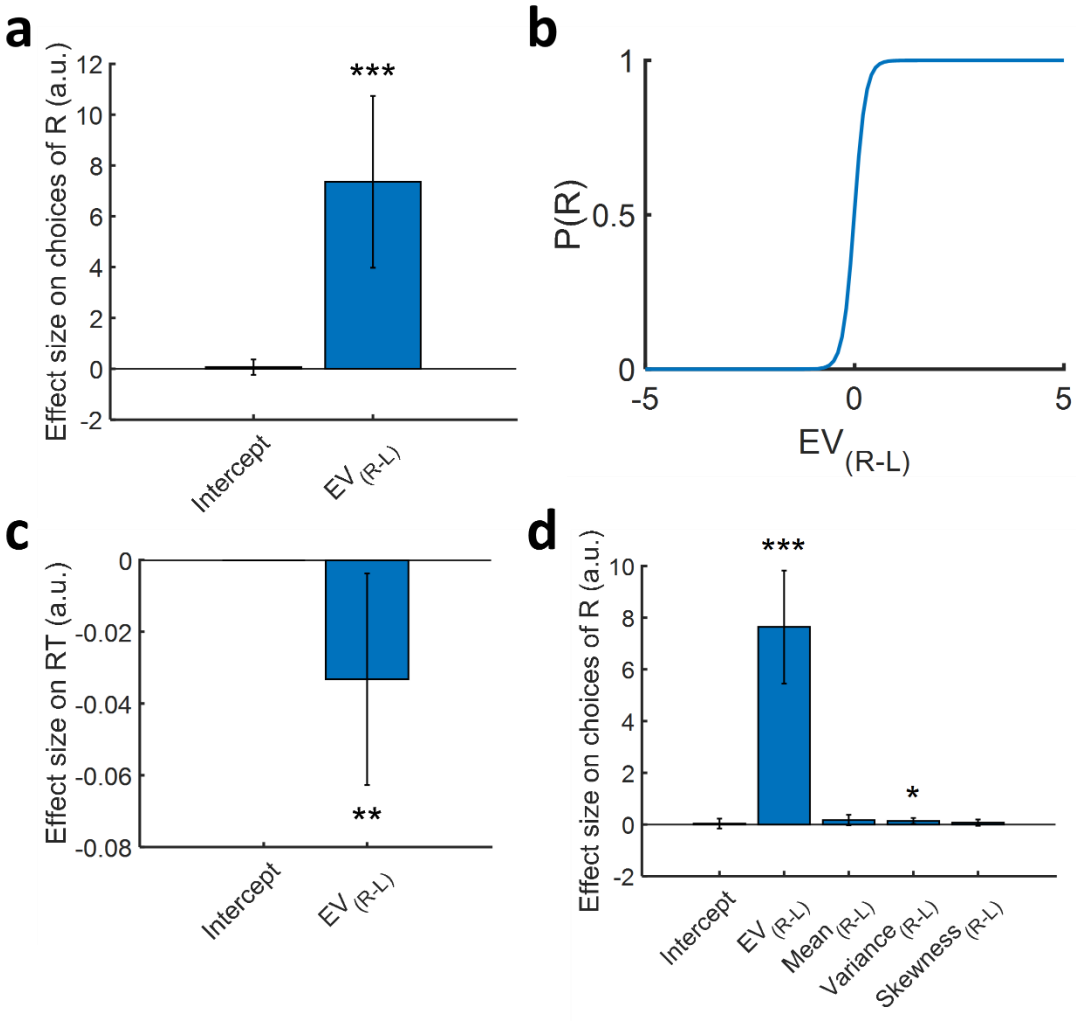


Figure 2.3. Effects of EV and statistical moments on environment choice. (a) A logistic regression (GLM2) indicated that there was a positive $EV_{(R-L)}$ effect on participants' environment choices. **(b)** Psychometric curve illustrating that participants were more likely to choose the environment with larger EV. **(c)** A linear regression (GLM3) indicated that there was a negative $EV_{(R-L)}$ effect on the speed of making environment choices. **(d)** Another logistic regression (GLM4) examining the effect of different statistical moments on environment choices revealed that participants also preferred environments with larger variance in addition to larger EV. $EV_{(R-L)}$=difference in expected value; $Mean_{(R-L)}$, $Variance_{(R-L)}$,

$Skewness_{(R-L)}$=difference in mean, variance, and skewness of reward probability between environments respectively; RT=reaction time. *** denotes $P<0.001$, ** denotes $P<0.01$, and * denotes $P<0.05$. Error bars represent means±s.e.m.

2.4 Discussion

This study provides fundamental understanding on several behavioural aspects of environment choice, which are important to the subsequent investigation of neural mechanisms in the fMRI study reported in **Chapter 5**. First, this study revealed that people were able to integrate different attributes of environments to guide decisions despite the complex structure of the environments (Fig. 2.3a,b,d). Specifically, each environment was composed of a distribution of items and positive effects of EV and variance in the item distribution on environment choices were observed (Fig. 2.3d). Preference of large EV is intuitive because in this paradigm the reward magnitude and mean reward probability were explicitly portrayed. It is worth noting that, variance, the less obvious second statistical moment, also significantly biased participants' environment choices. A preference of large variance was observed during environment choice but it is contradictory to the findings from economic studies. Previous findings revealed that economic choices with larger variance are perceived as riskier and thus a preference of small variance was resulted in (Kahneman & Tversky, 1979; Symmonds et al., 2011). This discrepancy could stem from the particular importance of larger variance to environment choice. Most often in real world decisions between environments (e.g. choosing between restaurants; each of which is an environment offering a collection of food), people have to consider whether there is a diversity of items to choose from (e.g. how many types of food are offered in each restaurant) in addition to whether overall the

items are good. The variance preference revealed in this study was in accordance with the previous findings that people preferred environments in which diverse lotteries were available over environments which only provided identical lotteries for selection (Norton & Liljeholm, 2020). Therefore, an opposite effect of variance was shown in evaluation of environments when compared to that observed in typical economic choices.

Moreover, it was found that participants preferred environments with higher reward probability when the environments leading to gains (i.e. larger environment EV), but avoided environments with higher reward probability when the environments leading to losses (i.e. smaller environment EV). Although this finding is unsurprising at the behavioural aspect, it is of great importance for scrutinizing the neural value comparison process between environments in the fMRI study. One critical signature of a neural signal of value comparison is the invariance to gains and losses. For a value comparison signal, the only relevant information is how much one option is better than the alternative. The neural substrate underlying the value comparison process should reflect the value difference regardless of whether the options are rewarding or aversive. Existing literature of item choice revealed this is indeed the case – although the vmPFC exhibits a positive value coding of rewarding items as well as negative value coding of aversive items (Plassmann et al., 2010), it signals the value difference between the available items during decision-making indifferently under gains and losses (FitzGerald et al., 2009). The inclusion of environments leading to losses in addition to environments leading to gains allows direct testing whether this essential property of value comparison signal also holds true during environment choice. Properties of the neural signals were discussed in **Chapter 5**.

Chapter 3 Behavioural hallmarks of environment choice and item choice

Chapter highlights

- This chapters aims to test the behavioural hallmarks in environment choice and item choice
- Human participants were tested on a two-stage decision-making that mimic environment choice and item choice
- Context-dependent adaption was observed in environment choice, as well as item choice
- Each environment was considered as an aggregate of multiple items, as opposed to a proxy to a particular component item
- People preferred environments with larger means and variances of the item distributions in the environments

3.1 Introduction

In **Chapter 2**, I have demonstrated people were able to compare the values of environments for making decisions given their complexity. It provides the basis for examination of the underlying neural mechanisms. Notably, as per the findings of neural value comparison process during item choice that were reviewed in **Chapter 1**, the neural signal of the value difference between options possesses two critical properties: (1) invariance to salience and (2) context dependence. These properties are the essence of

value-based decision-making and should also be observed in environment choice. To this end, a study that involved fMRI was designed with careful manipulation with regards to these properties. Invariance to salience can be tested by inclusion of both environments that lead to gains and environments that lead to losses. The first study reported in **Chapter 2** has shown that people could distinguish these this property behaviourally, despite the lack of neural data. The fMRI study therefore adopted a similar design. On the other hand, to probe the nature of context dependence that has not been investigated in the first study, a "bonus", an additional reward apart from the item reward, was included in this study. The acquisition of bonus varied under conditions, by which context dependence can be tested – whether people exhibit behaviour flexibly adapted to the conditions to obtain the bonus.

Furthermore, the primary aim of this thesis is to elucidate whether a single neural substrate is sufficient to deal with all different kinds of decision or multiple neural substrates are necessary. To this end, the fMRI study adopted a more sophisticated design – it involved a two-stage decision-making task with the first stage mimicking environment choice and the second stage mimicking typical item choice. Specifically, decision outcome of the first stage affected the items available for selection in the second stage as if a real-life scenario of choosing an environment (e.g. a restaurant) determines what items can be encountered for subsequent decision (e.g. food in the restaurant that can be ordered). This experimental design can discern environment choice from item choice and allow direct testing on whether a unitary neural substrate or multiple neural substrates are underlying to deal with different kinds of decisions.

This study focused on the behavioural results of the fMRI study. It revealed people could envisage the potential items that were later available after selection of an environment. Besides, preference in statistical moment (i.e. variance) was observed which implies each environment was considered holistically with respect to its item distributions, as opposed to being evaluated as a proxy to a particular item. More importantly, people demonstrated adaptation to the context related to the bonus, suggesting the property of context dependence. All these properties as well as other behaviourally unexplained properties were discussed in **Chapter 5** with neural data.

3.2 Methods

3.2.1 Participants

Twenty-six healthy young adults (15 females), aged 19-39 years with normal or corrected-to-normal vision and no current or history of neurological or psychiatric conditions, were recruited via advertisement in the university and participants' referral. Written informed consent for each participant was obtained before experiment. This study was approved by the Human Subjects Ethics Committee of The Hong Kong Polytechnic University. One participant withdrew because of contraindications to MRI scans. Data of another participant was excluded because the fMRI session was broken into two separate runs and it was not possible to be analysed in the same way as data of other participants.

3.2.2 Experimental task

Participants underwent a two-stage decision-making task. Each block of trials included one Stage 1 trial, followed by zero to three Stage 2 trials (Fig. 3.1a). Stage 2 trials were similar to a typical binary decision-making task, which required participants to choose between two items associated with different reward magnitudes (between -£10 and +£10) and reward probabilities (proportional to the height of the coloured bars) (Fig. 3.1b). On Stage 1 trials, participants chose between two environments, instead of two items, each of which was an aggregate of 20 items, presented in a form of a series of 20 bars and a number. The height of each bar was related to the reward probability of an item in the environment (i.e. a total of 20 items) and the number indicated the mean reward magnitude of all items in the same environment (Fig. 3.1b). Notably, the decision made during Stage 1 determined the items that were made available on the subsequent Stage 2 trial(s) in the same block, as opposed to decisions in Stage 2 that resulted in earning rewards directly.

To be specific, when a block began, participants first chose between two environments. After selection of an environment, the chosen environment was surrounded by a red frame (0.5 second) and then presented at the centre; two items from which were pseudo-randomly drawn (highlighted in yellow) after a delay of 2-7 seconds, indicating that these two items would be offered on the subsequent Stage 2 trial. For each chosen environment, participants were pseudo-randomly given zero to three Stage 2 trials; each of which involved two items drawn from the same environment. On each Stage 2 trial, participants chose between two items for reward earning. The chosen item was then surrounded by a red frame (0.5 second). Reward was delivered according to the

reward magnitude and reward probability of the chosen item. Decision outcomes of all Stage 2 trials related to the same environment were simultaneously displayed after all Stage 2 trials of the same block were completed. Participants had to maximize their earning by choosing a more rewarding item in Stage 2. In turn, they also had to seek for more advantageous items in Stage 2 by choosing a more rewarding environment in Stage 1. There was a total of 200 Stage 1 trials and 200 Stage 2 trials in this task.

This study intended to scrutinize the neural mechanisms underlying environment choice. A neural substrate which underpins environment choice should possess the following properties: (1) it should exhibit a signal reflecting the difference in values between the environments; (2) the signal should be independent of salience; (3) the signal should show a context-dependent modulation. To address the first two properties, akin to the study in **Chapter 2**, in this study environments/items leading to gains (Fig. 3.1c; green bars) and losses (Fig. 3.1c; red bars) were both included. This manipulation can orthogonalize option value and option salience. Specifically, there were 53.5% (55%) of the trials involved two gain environments (items), 31.5% (30%) of the trials involved two loss environments (items) and 15% of the trials involved one gain and one loss environment/item.

Besides, to examine the context-dependent adaptation, this study included an opportunity to acquire additional rewards, the bonus, via different ways in two different Bonus Conditions (Fig. 3.1c). On each Stage 1 trial, a bonus between -£4 and +£6 was displayed at the lower centre and stayed the same on the subsequent Stage 2 trials of the same block. In Linked Condition (indicated by two arrows pointing from the bonus to the two environments/items), the bonus was obtained when the chosen Stage 2 item led to

reward. Hence, in the presence of bonuses that were particularly rewarding, participants were incentivized to choose environments composed of items with larger probabilities, such that eventually they would be offered these items to win the bonus. And they also ought to select items with higher reward probability in Stage 2. In contrast, in Unlinked Condition (indicated by absence of arrows), the bonus was delivered unconditionally such that participants' choices ought not to be affected by the bonus.

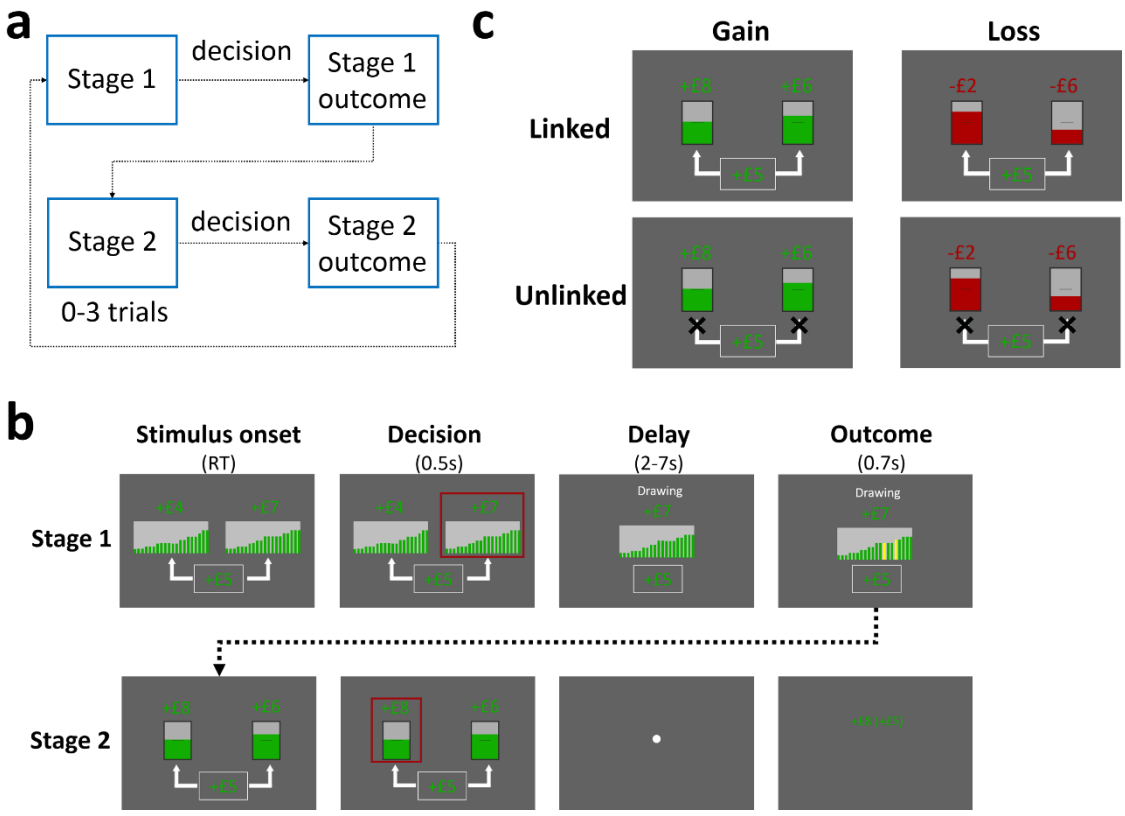


Figure 3.1. A two-stage decision-making task. (a) Each block began by a Stage 1 trial that involved environment choice and followed by zero to three Stage 2 trials that involved item choice. **(b)** On each Stage 1 trial, participants chose between two environments, each of which was composed of 20 items (*Stimulus onset*). For each environment, the heights of the bars were related to the reward probabilities of the component items and the number above the bars was related to the average reward magnitude of the items. After an environment was chosen, it was surrounded by a red frame (*Decision*) and then was

presented at the centre (*Delay*). After a delay, two items were drawn pseudo-randomly and highlighted in yellow (*Outcome*). These two items were presented on the subsequent Stage 2 trial (*Stimulus onset*). A red frame surrounded the chosen item after decision (*Decision*). A fixation dot was presented (*Delay*). After that, the amount of probabilistic reward earned from the chosen item was shown (*Outcome*). **(c)** Each block involved a bonus (number at the bottom) that was delivered in one of the two Bonus Conditions. In Linked Bonus Condition (upper row; indicated by two arrows), the bonus was delivered only when the probabilistic reward of the chosen item in Stage 2 was earned. In Unlinked Bonus Condition (lower row; indicated by absence of arrows), the bonus was delivered unconditionally on each Stage 2 trial. Each block was also assigned to one of the three Valence Conditions (Gain, Loss or Mixed), in which the two options were either related to gains (left column), losses (right column), or a combination of gains and losses (not shown).

3.2.3 Statistical analysis

For both Stage 1 and Stage 2 decisions, a logistic regression (GLM5) was applied to predict whether participants' choices (to the right=1 and left=0) were biased by the options' expected value and probability, the magnitude of the bonus, and Bonus Condition:

$$logit\frac{P(R)}{1-P(R)} = \beta_0 + \beta_1 EV_{(R-L)} + \beta_2 Prob_{(R-L)} + \beta_3 Bon + \beta_4 Cond +$$

$$\beta_5(EV_{(R-L)} \times Bon) + \beta_6\left(Prob_{(R-L)} \times Bon\right) + \beta_7(Bon \times Cond) +$$

$$\beta_8\left(EV_{(R-L)} \times Bon \times Cond\right) + \beta_9\left(Prob_{(R-L)} \times Bon \times Cond\right)$$

where *P(R)* denotes the probability of choosing the option on the right. $Prob_{(R\text{-}L)}$ denotes the difference in reward probabilities (i.e. the average of item probabilities within environment in Stage 1 or item probability in Stage 2) between the options on the right

and left; $EV_{(R-L)}$ denotes the difference in expected value between the options on the right and left; *Bon* and *Cond* represent the bonus value and Bonus Condition (Linked Condition=1, Unlinked Condition=0) respectively.

Stage 1 environments were composed of distributions of items. To test whether participants' choices were affected by the statistical moments of the item distributions, another logistic regression (GLM6) was performed:

$$logit\frac{P(R)}{1-P(R)} = \beta_0 + \beta_1 EV_{(R-L)} + \beta_2 Mean_{(R-L)} + \beta_3 Variance_{(R-L)} + \beta_4 Skewness_{(R-L)}$$

where $Mean_{(R-L)}$, $Variance_{(R-L)}$, $Skewness_{(R-L)}$ denote the differences in mean, variance, and skewness of environment probability respectively. A one-sample t-test was performed to examine whether each beta weight was significantly different from zero across participants in both regressions.

3.3 Results

3.3.1 Environment choice and item choice flexibly modulated by context

This study adopted a two-stage decision-making task to investigate the neural mechanisms underpinning environment choice, and to discern environment choice from item choice. First, it is important to assure some fundamental properties of decision-making were successfully probed. Therefore, both types of decision were tested behaviourally by a logistic regression (GLM5) to examine (1) whether participants chose options with larger values and (2) flexibly adapted their choices to earn large bonuses.

Under the current experimental design, environment choice in Stage 1 did not give rise to direct reward outcome. Nonetheless, the selected environment determined the consequential items in Stage 2 which were directly related to reward earning. A prospective thinking was thus required to opt for a more facilitative environment to maximize reward earning in Stage 2 when undertaking Stage 1. Not surprisingly, the results of GLM5 revealed when the EV of the rightward environment was relatively larger than that of the leftward environment (i.e. large $EV_{(R-L)}$), participants were also more likely to choose the rightward environment (β=6.264, t_{23}=6.196, P=2.543×10^{-6}; Fig. 3.2a). On the other hand, in Linked Condition, the acquisition of bonus was dependent on whether the reward associated with the chosen item was won or not. Therefore, environments composing of items with large reward probabilities should be preferred in Stage 1 in order to maximize their opportunities of obtaining the large bonus on the subsequent Stage 2 trials. Contrarily, in Unlinked Condition, the bonus was bounded to be delivered and hence choices should not be influenced by the bonus. This was indeed the case of the results – there was a three-way interaction effect between probability difference ($Prob_{(R-L)}$), bonus value (*Bon*), and Bonus Condition (*Cond*; Linked vs Unlinked) (β=0.248, t_{23}=3.069, P=0.005; Fig. 3.2a,b). This reflects that on Linked trials with large bonuses, participants preferred environments with large reward probability such that on the upcoming Stage 2 trials they were more likely to encounter items with large reward probabilities and then earn the bonus.

A similar logistic regression was applied to analyze the item choice in Stage 2. Akin to the Stage 1 results, items with lager EVs were preferred (β=4.420, t_{23}=11.582, P=4.468×10^{-11}; Fig. 3.2a) and there was also a significant $Prob_{(R-L)}$×*Bon*×*Cond*

interaction effect (β=0.760, t_{23}=4.731, P=9.113×10^{-5}; Fig. 3.2a,b). Taken together, in both Stages 1 and 2 participants did make decision according to option value and adapt their choices to the context by exhibiting a stronger preference to choose options with greater probability when there was a large bonus in Linked Condition. Importantly, these results also revealed that participants realized the offers on the subsequent Stage 2 trials were determined by the environments they selected in Stage 1. In other words, the prospective nature of environments was successfully probed.

3.3.2 Environment choices biased by statistical moments

Similar to the study reported in **Chapter 2**, it is critical to scrutinize whether different statistical moments (e.g. mean, variance, skewness) were represented and integrated to guide decisions between complex environments. To this end, another logistic regression (GLM6) was carried out and it showed an absence of effect of mean probability on environment choices (β=0.089, t_{23}=1.121, P=0.274; Fig. 3.2c). This was because the component items could lead to either gains or losses. The preferences for mean probability were opposite under gains and losses such that the effects of mean probability in these two conditions were cancelled out. On the other hand, a positive effect of variance was found (β=0.126, t_{23}=2.248, P=0.035; Fig. 3.2c) but no effect of skewness on choices was observed (β=0.074, t_{23}=1.290, P=0.210; Fig. 3.2c). It implies that, akin to the findings of the study reported in **Chapter 2**, environments with more widespread item distribution were favoured.

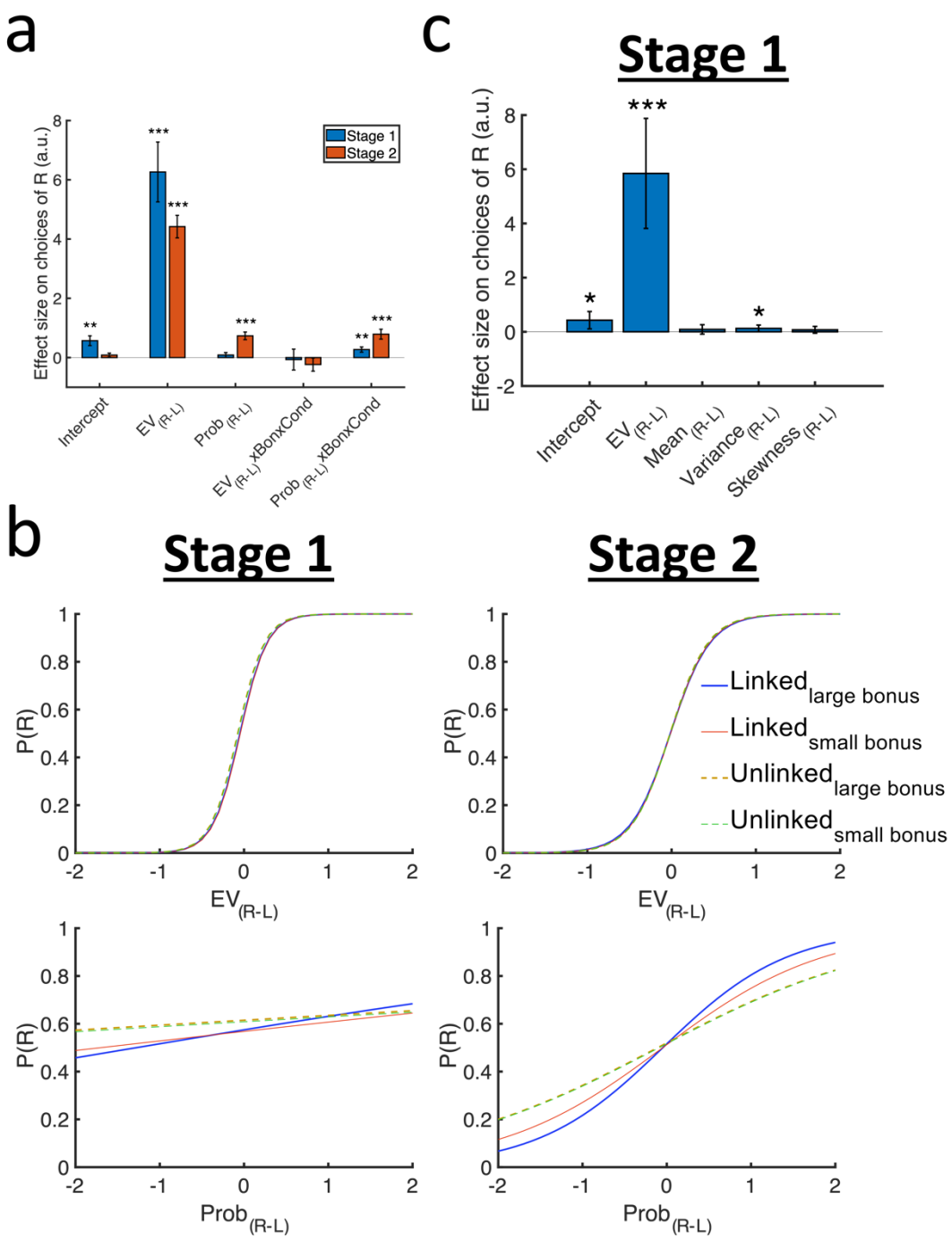


Figure 3.2. Logistic regression results. (a) A logistic regression (GLM5) indicated that there were positive $EV_{(R-L)}$ and $Prob_{(R-L)}$×Bon×Cond effects on participants' choices in Stage 1 environment choice and Stage 2 item choice. **(b)** Psychometric curves illustrating that, in Stage 1, participants were more likely to choose environments with larger EVs (top-left panel). In Linked Condition, when the bonus value

became larger, there was a stronger preference for choosing environments with larger reward probabilities (bottom-left panel; blue vs red). In contrast, in Unlinked Condition, the preference for larger reward probabilities was no longer modulated by the bonus value (bottom left panel; orange vs green). Similar choice patterns were observed in Stage 2 (right column). **(c)** In Stage 1, participants showed a preference for environments with large variances, in addition to a preference for those with larger EVs. $EV_{(R-L)}$=difference in expected value; $Prob_{(R-L)}$=difference in reward probability; Bon=bonus value; Cond=Bonus Condition; $Mean_{(R-L)}$, $Variance_{(R-L)}$, $Skewness_{(R-L)}$=difference in mean, variance and skewness of reward probability between environments respectively. *** denotes $P<0.001$, ** denotes $P<0.01$, and * denotes $P<0.05$. Error bars represent means±s.e.m.

3.4 Discussion

The primary goal of this thesis is to investigate whether a single neural substrate deals with all kinds of decision or distinct neural substrates are required for different types of decision. Therefore, in this study, environment choice as well as item choice were included for direct comparison. Environment choice and item choice are similar in the way that they both involve making choices between options of the same class. Given this resemblance, they should show a number of similar properties of decision-making. The behavioural results did provide evidence that these two types of decision do share some fundamental properties. First, in both Stage 1 environment choice and Stage 2 item choice, participants chose according to the EV difference between the options (environments in Stage1 and items in Stage 2; Fig. 3.2a). It reflects participants integrated both attributes of reward magnitude and reward probability to guide decision-making, which is a well-established property in economic choice (von Neumann &

Morgenstern, 1944). Second, a property of context dependence was observed in both stages. In the context where the acquisition of bonus was related to reward probability (i.e. Linked Condition), participants preferred options with larger probability when the bonus value was large (Fig. 3.2b, bottom row, blue versus red). On the contrary, in the context where the bonus was unconditionally delivered (i.e. Unlinked Condition), participants choices were not affected by the reward probability nor the bonus value (Fig. 3.2b, bottom row, yellow versus green).

Showing the similarities among environment choice and item choice, the next step is to scrutinize their differences. This chapter only focuses on the behavioural results of the fMRI exp. Without neural data, it is insufficient to discern these two types of decision. Nevertheless, the behavioural results revealed that these two different types of decision were being simulated and measured, providing the preliminary basis to further discern them at the neural level. As discussed in **Chapter 1**, there are two essential features of environment: (1) an environment is an aggregate of multiple items and (2) it brings in the opportunity to come across its component items rather than direct acquisition of the final payoff associated with the items. The behavioural results affirmed that these two features of environment were successfully probed. First, consistent finding to the study reported in **Chapter 2** was observed – environments with larger EV and larger variance were preferred (Fig. 3.2c). It implies the distribution of items within an environment was considered holistically and in other words suggests that each environment was viewed as a collection of multiple items, as opposed to a proxy to a specific item. While the second feature of environments concerns the interdependent relation between the choice of an environment, the prospective items that will be encountered, and the final payoff. In this

study, Stage 1 choices did not give rise to reward (i.e. final payoff) but gave rise to the opportunity to encounter potentially better items in the subsequent Stage 2. By contrast, Stage 2 decisions directly led to reward earning and the decision outcome might affect the acquisition of bonus (in Linked Condition a bonus was delivered if a reward was obtained; in Unlinked Condition the bonus was bound to be delivered). Therefore, a context-dependent adaptation was needed for Stage 2 decisions and this was the case of the results. Interestingly, although bonus acquisition was not directly dependent on Stage 1 decisions in either condition, an interaction effect between the mean probability of environment, bonus value, and Bonus Condition was also observed during Stage 1. It implies participants did carry a prospective thinking about the consequence of environment choices – they flexibly chose environments composing of items with high reward probability to facilitate the acquisition of bonus in the upcoming Stage 2 in Linked Condition but not in Unlinked Condition. This provides the basis for discerning the neural signal related to environment choice from that related to item choice.

Thus far, results of this study demonstrated that behaviourally both environment choice and item choice were successfully probed, and these two types of decision did share some fundamental properties of decision-making. To disentangle whether these types of decision involve separate neural substrates or both of them are executed by a single neural substrate, the neural data were inspected and reported in **Chapter 5**. Specifically, three frameworks which hold different assumptions on within-class (item versus item or environment versus environment) and between-class (item versus environment) decisions were tested. The Neural Common Currency Framework posits all kinds of decision, regardless of within-class or between-class decisions, are

underpinned by a single neural substrate. On the other hand, the Option Homogeneity Framework postulates there are distinct neural substrates underlying the within-class decisions for each class of option and a separate neural substrate subserves the between-class decision. In a similar vein, the Decision Homogeneity Framework also assumes separate neural substrates for between-class and within-class decisions. It however suggests that all within-class decisions are subserved by the same neural substrate, irrespective of what the options involved are. A previous study which involved both decision between items and environments has partially precluded the Neural Common Currency Framework by showing that separate neural substrates were involved in item choice and item choice and decision between items and environments (Kolling et al., 2012). Hence, for the neural data analysis in **Chapter 5**, the Option Homogeneity Framework and Decision Homogeneity Framework were focused.

Chapter 4 Estimating value computation process during decision-making

Chapter highlights

- This chapter aims to develop computational models to describe the value computation process
- A convolutional neural network (CNN) best predicts people's decision-making behaviour compared to general linear model, cumulative prospect theory, mean-variance-skewness model, power model, autoencoder
- The CNN encodes the means, variances, and skewness of the item distributions in the environments even though these parameters are not explicitly specified to the CNN
- The CNN involves multiple feature detectors to extract and integrate features of the complex information in the environments

4.1 Introduction

Thus far, all behavioural analyses in **Chapters 2** and **3** assume participants perfectly estimated the option value according to some typically used parameters such as expected value (EV), variance, and skewness. However, actual human decision-making may not necessarily follow these computations, especially for the case of complex environments. Computational models, on the contrary, can vividly describe how people subjectively evaluate the options. Critically, these accounts of the subjectivity facilitate the

examination of the neural computation underlying decision-making. Multiple lines of studies have demonstrated that describing people's choices using computational models as opposed to EV better explained the neural value signals (Hsu et al., 2009; Symmonds et al., 2011). Therefore, before analysis of the neural data, I developed and compared different computational models to scrutinize how the values of the environments or items were estimated in the studies reported in **Chapters 2** and **3**.

Models that are commonly used for modelling economic value or value in distribution were adapted in this study, such as general linear model, cumulative prospect theory, mean-variance-skewness model, and power law model. Apart from conventional computational models, I also adapted the cutting-edge deep learning neural networks from the field of artificial intelligence. There are several merits of deep learning neural networks over conventional computational models. First, conventional models require explicit specification of every feature-of-interest within the data. While deep learning neural networks allow fewer a priori assumptions. Only the essential information of environments is needed and the deep learning neural networks will decide what features to extract. Implicit features or patterns within the data can also be extracted even not specified. Second, deep learning neural networks possess a multi-nodal architecture via which environment values can be represented by multiple nodes. It provides a means to delineate the multivariate coding during valuation. Examination of multivariate neural signals is permitted and it facilitates the disentanglement of functional specialization among different brain regions during decision-making in **Chapter 5**.

For both studies reported in **Chapters 2** and **3**, the convolutional neural network (CNN) outperforms other alternative models in predicting environment choices.

Furthermore, it was found that the CNN carries information about statistical moments – the between-environment differences in mean, variance, and skewness of the item distributions. It was consistent to the findings observed from the behavioural analyses in **Chapters 2** and **3**.

4.2 Methods

4.2.1 Experimental task

The behavioural study involved a parsimonious binary decision-making task between environments. Details about the task are described in **Chapter 2**. In brief, participants chose repeatedly between two environments for 10 trials in each block. Each environment was composed of a number (representing the average reward magnitude of 20 items) and 20 bars (representing the individual reward probabilities of 20 items). After selection of an environment, one item from the chosen environment was randomly drawn. Reward was then delivered according to the reward magnitude and reward probability of the drawn item.

On the other hand, the fMRI experiment involved a two-stage decision-making task. Details about the task are described in **Chapter 3**. In brief, in each block there was a Stage 1 trial accompanied by 0-3 Stage 2 trials. Stage 1 was similar to the behavioural study. Each environment was composed of a number (representing the average reward magnitude of 20 items) and 20 bars (representing the individual reward probabilities of 20 items). Notably, after selection of an environment, two items from the chosen environment were pseudo-randomly drawn and offered for selection on the subsequent Stage 2 trial. Participants chose between the two items which were associated with

different reward magnitudes and reward probabilities. Reward was delivered accordingly after the choice of an item. Furthermore, there was an additional reward, the bonus, whose acquisition was determined by the Bonus Condition. In Linked Condition, the bonus was delivered if the chosen item on the same Stage 2 trial led to reward. In Unlinked Condition, the bonus was delivered unconditionally.

4.2.2 General linear model (GLM)

Computational modelling aims to delineate the valuation process. Despite the differences in the underlying computation among models, all models commonly generate a decision value (DV) for each environment for choice prediction.

GLM is typical approach to fit economic choices. Environment choice data of the behavioural study were therefore fitted by a GLM with the three essential attributes of the environments, namely reward magnitude (*Rew*), reward probability (*Prob*), and EV (i.e. *Rew* multiplied by *Prob*). DVs of the environments are estimated by:

$$DV_R = \beta_0 + \beta_1 Rew_{(R-L)} + \beta_2 Prob_{(R-L)} + \beta_3 (Rew_R \times Prob_R - Rew_L \times Prob_L)$$

Eq. 1

where DV_R is the DV of the rightward environment; $Rew_{(R-L)}$ and $Prob_{(R-L)}$ denote the differences in reward magnitude and mean reward probability between the rightward and leftward environments respectively.

The environments in the fMRI study were highly similar to those in the behavioural study, except that information about bonus value and Bonus Condition were additionally included. Besides, the fMRI study involved item choice as well. To fit the behavioural

data of both environment choice and item choice in the fMRI study, Eq. 1 was modified as follow:

$$DV_R = \beta_0 + \beta_1 Rew_{(R-L)} + \beta_2 Prob_{(R-L)} + \beta_3 Bon + \beta_4 Cond +$$
$$\beta_5(Rew_R \times Prob_R - Rew_L \times Prob_L) + \beta_6\left(Prob_{(R-L)} \times Bon \times Cond\right)$$

Eq. 2

where DV_R is the DV of the rightward option (environment or item for the corresponding choice data); $Rew_{(R-L)}$ denotes the difference in reward magnitude between the rightward and leftward options; $Prob_{(R-L)}$ denotes the difference in reward probabilities between the rightward and leftward options; *Bon* and *Cond* represent the bonus value and Bonus Condition (Linked Condition=1, Unlinked Condition=0) respectively.

4.2.3 Mean-variance-skewness (MVS) model

Given an environment can be characterized by the mean, variance, and skewness of the distribution of its component items, the subjective probability of a given environment can be estimated with regard to these statistical moments by the MVS model (Symmonds et al., 2011, 2013; Wright, Symmonds, Morris, et al., 2013):

$$\widehat{Prob^k} = \frac{1}{20}\sum_{q=1}^{20} Prob_q^k + \rho Var(Prob^k) + \sigma Skew(Prob^k)$$

Eq. 3

where $\widehat{Prob^k}$ is the subjective probability of the k^{th} environment (either the leftward or rightward environment) while $Prob^k{}_q$ is the q^{th} individual item probability in the k^{th}

environment. Preferences for variance and skewness are reflected by two free parameters ρ and σ respectively. *DV* of the k^{th} environment is formulated as follow:

$$DV_k = \beta_1 Rew^k + \beta_2 \widehat{Prob}^k + \beta_3 (Rew^k \times \widehat{Prob}^k)$$

Eq. 4

where Rew^k is the reward magnitude of the k^{th} environment.

Similarly, Eq. 4 was adapted to estimate environments' subjective reward probabilities in the fMRI study and the computation of DVs is as follow:

$$DV_k = \beta_1 Rew^k + \beta_2 \widehat{Prob}^k + \beta_3 (Rew^k \times \widehat{Prob}^k) + \beta_4 Bon + \beta_5 Cond + \beta_6 \left(\widehat{Prob}^k \times Bon \times Cond\right)$$

Eq. 5

4.2.4 Power law model

Previous findings showed that the value estimation of an array of stimuli can be accounted for by a power law model (Li et al., 2017). The power law model was therefore applied to estimate the subjective probabilities of the environments:

$$\widehat{Prob}^k = \frac{1}{20} \sum_{q=1}^{20} (Prob_q^k)^g$$

Eq. 6

where $\widehat{Prob}^k$ is the subjective probability of the k^{th} environment while $Prob^k_q$ is the q^{th} individual item probability in the k^{th} environment. Each individual item probability in an

environment is transduced by a power *g*. The computation of the *DV*s is identical to Eq. 4.

For the fMRI study, the estimation of environments' subjective reward probabilities is identical to Eq. 6 and the computation of DVs is identical to Eq. 5.

4.2.5 Cumulative prospect theory (CPT)

The cumulative prospect theory (Tversky & Kahneman, 1992) is a classical model of subjective value estimation towards item value. It was also adopted to estimate the subjective probability of each item within a given environment. Different from the abovementioned models, CPT provides the capability for estimating the subjective value of reward magnitude in addition to the reward probability. Moreover, it has discrete transformation functions for value estimation under gains and losses. Subjective reward magnitudes of the environments were estimated by the following two utility functions for gains and losses respectively:

$$\widehat{Rew^k} = \begin{cases} (Rew^k)^{\alpha} & if\ Rew^k \geq 0 \\ -\lambda(-Rew^k)^{\beta} & if\ Rew^k < 0 \end{cases}$$

Eq. 7

where Rew^k is the reward magnitude of the k^{th} environment and $\widehat{Rew^k}$ is the corresponding subjective reward magnitude; α and β denote the sensitivity to value change in gains or losses; λ denotes the degree of loss aversion. There are also two weighting functions for estimation of subjective probability under gains and losses

respectively:

$$\widehat{Prob^k} = \begin{cases} \frac{1}{20}\sum_{q=1}^{20} \frac{(Prob_q^k)^{\gamma}}{((Prob_q^k)^{\gamma}+(1-Prob_q^k)^{\gamma})^{\frac{1}{\gamma}}} & if\ Rew^k \geq 0 \\ \frac{1}{20}\sum_{q=1}^{20} \frac{(Prob_q^k)^{\Delta}}{((Prob_q^k)^{\Delta}+(1-Prob_q^k)^{\Delta})^{\frac{1}{\Delta}}} & if\ Rew^k < 0 \end{cases}$$

Eq. 8

where γ and Δ denotes the extent of upweighting of small probability and downweighting of large probability, under gains and losses respectively. *DV* of the k^{th} environment is computed with the subjective reward magnitude and subjective reward probability:

$$DV_k = \beta_1\widehat{Rew^k} + \beta_2\widehat{Prob^k} + \beta_3(\widehat{Rew^k} \times \widehat{Prob^k})$$

Eq. 9

The CPT used for estimation of subjective reward magnitude and subject reward probability in the fMRI study is identical to Eq. 7-9, except that the computation of DVs involved the bonus value and Bonus Condition:

$$DV_k = \beta_1\widehat{Rew^k} + \beta_2\widehat{Prob^k} + \beta_3(\widehat{Rew^k} \times \widehat{Prob^k}) + \beta_4 Bon + \beta_5 Cond + \beta_6\left(\widehat{Prob^k} \times Bon \times Cond\right)$$

Eq. 10

where $\widehat{Rew^k}$ and $\widehat{Prob^k}$ denote the subjective reward magnitude and subjective reward probability of the k^{th} option (environment or item for the corresponding choice data) respectively; *Bon* is the bonus value and *Cond* is the Bonus Condition.

4.2.6 Convolutional neural network (CNN)

In addition to the abovementioned models, deep learning models such as CNN were also applied to fit environment choices. One characteristic of the CNN is its capability to integrate bundles of information into digested representation of features. It is hence applicable to describe the feature extraction and integration of the complex information embedded in the environments. There are four essential parts in a CNN: (1) input layer; (2) hidden layers; (3) choice-predicting fully-connected layer; (4) output layer (Fig. 4.1a). Environment value (i.e. reward magnitude and reward probability) is fed in as the input layer of the CNN. Next, reward probability in the input layer, which is stored in a form of matrix, undergoes convolution with four feature detectors and results in four sets of feature maps (Fig. 4.1b). To be specific, each feature detector which involves a set of weights, superimposes on the matrix of reward probability and slides along the entire matrix stride by stride. In each stride, the superimposed portion in the matrix is multiplied by the feature detector weights and the weighted-sum becomes the output of that stride. Feature map derived from this convolution process can be formulated as follow:

$$Feature\ map_{ij}^{A} = b_{A}^{(1)} + \sum_{m=1}^{M}\sum_{n=1}^{N} W_{mn}^{(1)} \odot X_{(i+m-1)(j+n-1)}$$

Eq. 11

where $b_A^{(1)}$ is the bias term for the A^{th} feature map. X is the matrix containing information about reward probability with a size of $I \times J$; $W^{(1)}$ is the weight matrix of the feature detector with a size of $M \times N$ and $W_{mn}^{(1)}$ refers to the element in the m^{th} row, n^{th} column;

The i^{th} row, j^{th} column of a feature map is the convolution outcome of a stride ($i \in \{1,2,\ldots,[I-m+1]\}$; $j \in \{1,2,\ldots,[J-n+1]\}$).

Features maps are subsequently multiplied by the reward magnitude to take the interaction in between into account, similar to the EV (i.e. reward magnitude multiplied by reward probability). Afterwards, feature maps, reward magnitude, and feature maps × reward magnitude are concatenated (denoted by *concat*). In an early fully-connected layer (*fc1*), *concat* is vectorized and then converted into 512 nodes and all these nodes are activated by a rectified linear unit (*ReLU*) function:

$$Y_N = W_N^{(2)} \cdot concat + b_N^{(2)}$$

Eq. 12

$$ReLU(Y) = \begin{cases} 0 \; if \; Y < 0 \\ Y \; if \; Y \geq 0 \end{cases}$$

Eq. 13

where $b_N^{(2)}$ is the bias term and $W_N^{(2)}$ is the weight matrix for the N^{th} node respectively.

Afterwards, in the choice-predicting fully-connected layer (*final fc*), all nodes are vectorized and then weigh-summed to form the DVs. The DV for the k^{th} environment is computed by:

$$DV_k = W_k^{(3)} \cdot \mathrm{Y} + b_k^{(3)}$$

Eq. 14

where $W_k^{(3)}$ are the weights in *final fc* and $b_k^{(3)}$ is the bias term for the k^{th} environment. Finally, in the output layer, the DV of each environment is transformed into a choice probability via a softmax function:

$$P(Environment_k) = \frac{e^{DV_k}}{\sum_k^2 e^{DV_k}}$$

Eq. 15

On the other hand, a CNN was developed to fit the environment choice data of the fMRI study. It was adapted from the CNN for the behavioural study and it is highly similar except several differences (Fig. 4.1c). First, since the fMRI study involved an additional component of bonus, information about bonus value and Bonus Condition are fed in to the input layer. Feature maps are multiplied by bonus value and Bonus Condition to take the interaction between reward probability and the bonus into account. This interaction is also concatenated into *concat* in addition to feature maps, reward magnitude and feature maps × reward magnitude. Next, the information about the environments represented in *concat* is integrated through *fc1* and *ReLU*. It is concatenated with the information about bonus value and Bonus Condition into *concat2* to account for the existence of context (i.e. bonus). The remaining divisions are identical to the CNN for the behavioural study. This CNN for environment choice data of the fMRI study was then adapted to fit the item choice data. However, due to the limitation in data structure (each item has only one number for reward probability as opposed to an environment which has 20 numbers), convolution is not feasible for item choice and the CNN is reduced into an artificial neural network (ANN) by removing the convolution part (Fig. 4.1d).

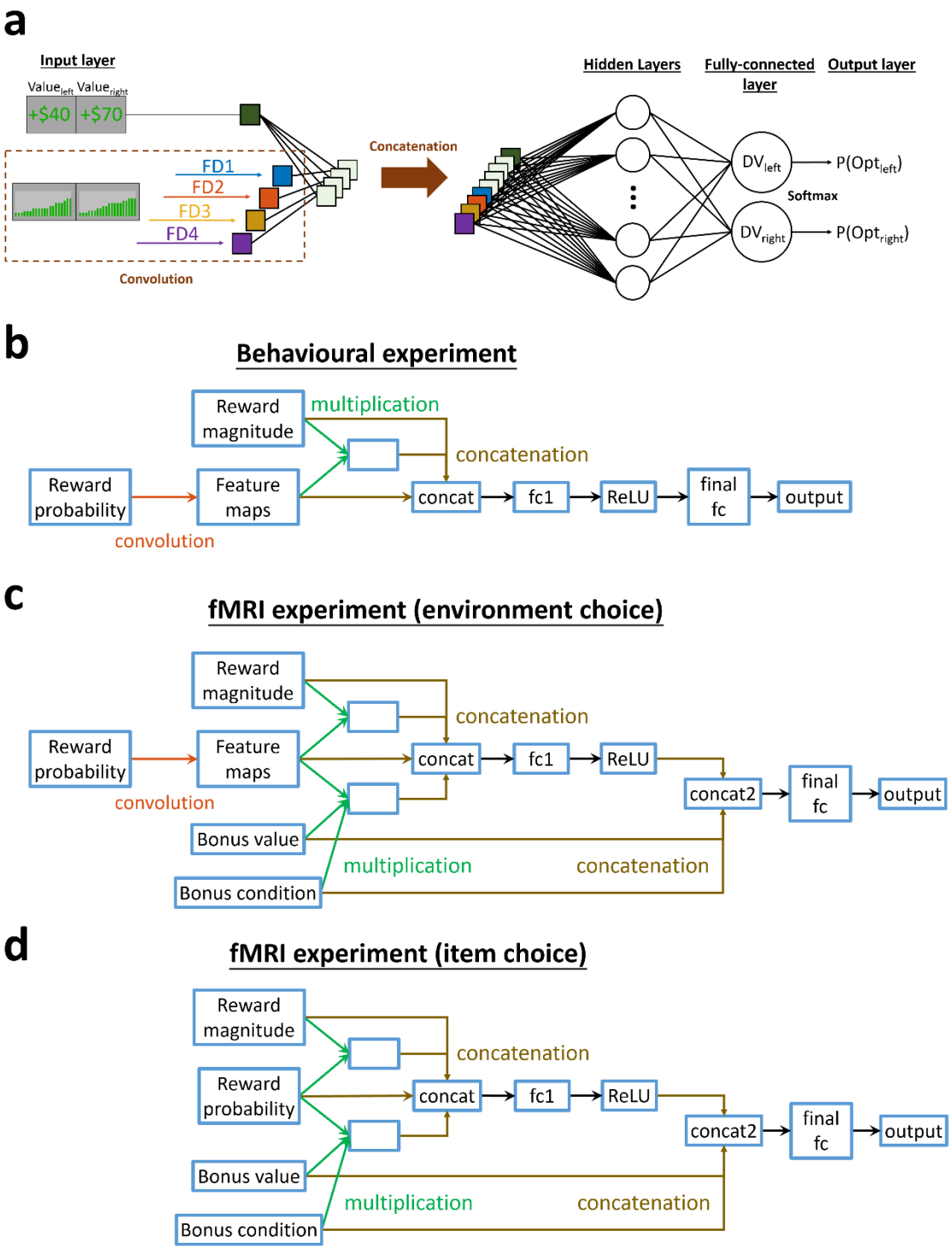


Figure 4.1. Deep learning neural network architectures. (a) Simplified schematic of the convolutional neural network (CNN) for the behavioural study. The CNN receives information of each environment as the input. During convolution, four feature detectors extract the features embedded in the item distributions

of the environments, giving rise to four sets of feature maps. Feature maps are multiplied by the reward magnitude and then feature maps, reward magnitude, and their interaction are concatenated. The concatenated features are represented by 512 nodes and integrated via a couple of hidden layers. Finally, in the fully-connected layer, all nodes are weighted and added up to form decision values (DVs) for the leftward or rightward environments. DVs are subsequently transformed by a softmax function into choice probabilities of the environments in the output layer. **(b-d)** Full schematics of the **(b)** CNN for behavioural study, **(c)** CNN for environment choice data of fMRI study, and **(d)** artificial neural network (ANN) for item choice data of fMRI study.

In addition, to investigate whether the CNN integrates information about statistical moments of the environments, four "simplified" CNN models, namely Mean Model, Mean+Var Model, Mean+Skew Model and Mean+Var+Skew Model, were developed. In the original CNN, the values of all component items in each environment are individually specified in the input. In contrast, in the Mean Model, the input involves the mean of the item distribution, instead of individual item values. In the Mean+Var Model, the input involves the mean and variance of the item distribution. In the Mean+Skew Model, the input involves the mean and skewness of the item distribution. In the Mean+Var+Skew Model, the input involves the mean, variance, and skewness of the item distribution.

4.2.7 Autoencoder

Apart from CNN, another deep learning neural network – autoencoder, was used for fitting the environment choice of the behavioural study. There are three essential layers in the autoencoder: (1) input layer; (2) hidden layer; (3) output layer. Environment value is fed in as the input layer of the autoencoder (Fig. 4.2). In the hidden layer, features

embedded in the environments are extracted as such the input layer is transformed into 10 hidden nodes. Afterwards, every hidden node is activated by a sigmoid function:

$$Z_h = Sigmoid(W_h^{(4)} \cdot X_{Autoenc} + b^{(4)})$$

Eq. 16

$$Sigmoid(z) = \frac{1}{1 + e^{-z}}$$

Eq. 17

where $X_{Autoenc}$ is the input; $W_h^{(4)}$ is the h^{th} weight matrix transforming $X_{Autoenc}$ into the h^{th} hidden node Z_h; $b^{(4)}$ is a bias term. All hidden nodes are then weigh-summed to form the *DV*. The *DV* for the k^{th} option is computed by:

$$DV_k = W_k^{(5)} \cdot Z + b_k^{(5)}$$

Eq. 18

where $W_k^{(5)}$ is the weight matrix for the k^{th} option; $b_k^{(5)}$ is the bias term for the k^{th} option. In the output layer, the DV of each environment is transformed into a choice probability via Eq. 15.

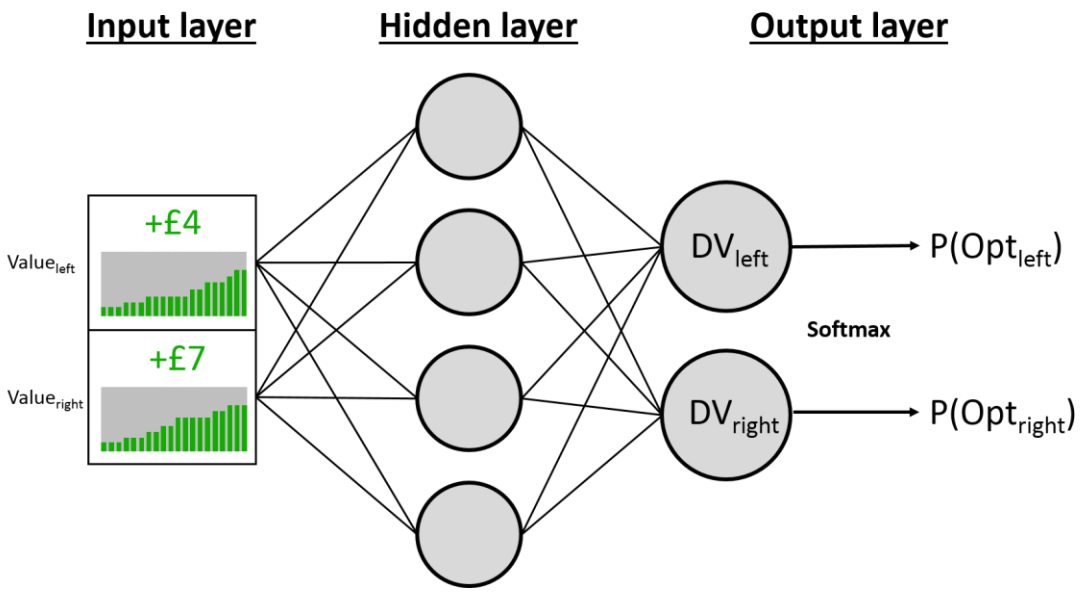


Figure 4.2. Schematic of the autoencoder. In the input layer, all attributes of the environments were fed as the input. In the subsequent hidden layer, feature extraction is performed. As a result, the input is transformed into 10 hidden nodes (not all of them are shown here). In the output layer, the hidden nodes are weighted and added up to form a decision value (DV) for each environment, which are then transformed by a softmax function to generate the choice probability.

The autoencoder employed in the behavioural study was adapted for fitting the data of the fMRI study. The primary changes are: (1) *Bon*, *Cond*, and *Prob*×*Bon*×*Cond* in addition to *Rew*, *Prob*, and EV were fed in as the input; (2) a different number of hidden nodes (i.e. 512) was used. Similarly, the autoencoder was applied to the item choice data with an architecture of 512 hidden nodes.

4.2.8 Parameter optimization and model comparison

All CNNs and autoencoders were fitted to participants' choices using Matlab Deep Learning Toolbox (The MathWorks, 2018) whereas the other models were fitted using

MatlabVBA-toolbox (Daunizeau et al., 2014). For each model, a 20-fold cross-validation procedure was performed to prevent overfitting. Specifically, data from all participants were collapsed and was randomly split into subsets. All but one subsets were used for training a model and the remaining subset was selected to test the trained model. This procedure repeated 20 times by changing the testing subset each time. The performance of the models was compared by their negative log-likelihood (NLL).

4.2.9 Representational similarity analysis (RSA)

RSA (Kriegeskorte et al., 2008) was performed to inspect the relatedness between the CNN and the simplified CNN, by comparing their representational dissimilarity matrices (RDMs). First, RDMs of different datasets for each subject were computed. RSA was then conducted in a subject-by-subject manner. Specifically, trial-by-trial nodal activations of a given layer (e.g. a feature map) in each CNN model were extracted. Next, a self-correlation was performed using Pearson correlation. The resulting correlation matrix was then subtracted from 1 to become an RDM. RSA was performed by computing Spearman correlations between a pair of RDMs. Noting that only the lower triangle (or equivalently the upper triangle) of the RDMs was used for RSA because the upper and lower triangles within an RDM are symmetrical.

4.3 Results

Since each environment is composed of multiple items, to scrutinize how the complex information within the environments was integrated to guide decision-making, different computational models were applied to fit participants' choices and their goodness-of-fit was compared. It is worth noting that the performance of the CNN and

autoencoder is susceptible to the architectures employed. Therefore, before model comparison, different architectures of the CNN and autoencoder were tested. The optimized models were then selected for model comparison.

4.3.1 Parameterization of CNN/ANN

Environment choice requires encoding and integrating the complex information of the environments. CNN is a candidate model that could provide remarkable description about these processes because CNN is efficient in feature extraction from complex visual information to decode image identity (Krizhevsky et al., 2012; Lindsay, 2020). Therefore, the CNN was adopted to decode environment choices from the multiplex environments' values. Yet, the architectures of the CNN can greatly affect its performance. To this end, the sizes of the feature detector (from small to large: 2×1, 5×1, 10×1, 20×1), the number of feature detector (from 1 to 5 detectors), and the number of nodes in *fc1* (32, 64, 128, 256, 512) were systematically varied to identify the optimized CNN architecture. For the behavioural study, no significance difference among the CNN variants were observed ($F(99,1900)=0.099$, $P=1$; Fig. 4.3a). The CNN with four feature detectors in a size of 2×1 and 512 nodes in *fc1* yielded the lowest NLL (39.935±4.318) and hence it was selected for subsequent analyses.

Different CNN variants for environment choice of the fMRI study were also inspected and the model comparison results showed a marginally significant difference among the variants ($F(99,1900)=1.199$, $P=0.094$; Fig. 4.3b). The variant with four feature detectors in a size of 2×1 and 512 nodes in *fc1* yielded the lowest NLL (33.235± 5.5628) and this variant was selected for subsequent analyses. Interestingly, this CNN

variant possesses identical parameters (i.e. four feature detectors in a size of 1×2 and 512 nodes in *fc1*) to the CNN variant that best describes the environment choice data of the behavioural study. This provides further evidence that the delineation of environment choice behaviour by the CNN is reliable irrespective of data set, in addition to the cross-validation procedure that prevents overfitting (Section *Parameter optimization and model comparison* of **Chapter 4)**. Similarly, the optimized ANN architecture for item choice of the fMRI study was tested by varying the number of nodes in *fc1* (32, 64, 128, 256, 512). It was found that there was no significant difference among the ANN variants ($F(4,95)=0.071$, $P=0.991$; Fig. 4.3c). Nonetheless, the variant with 512 nodes in *fc1* yielded the lowest NLL (46.195±2.918) and thus it was selected for subsequent analyses.

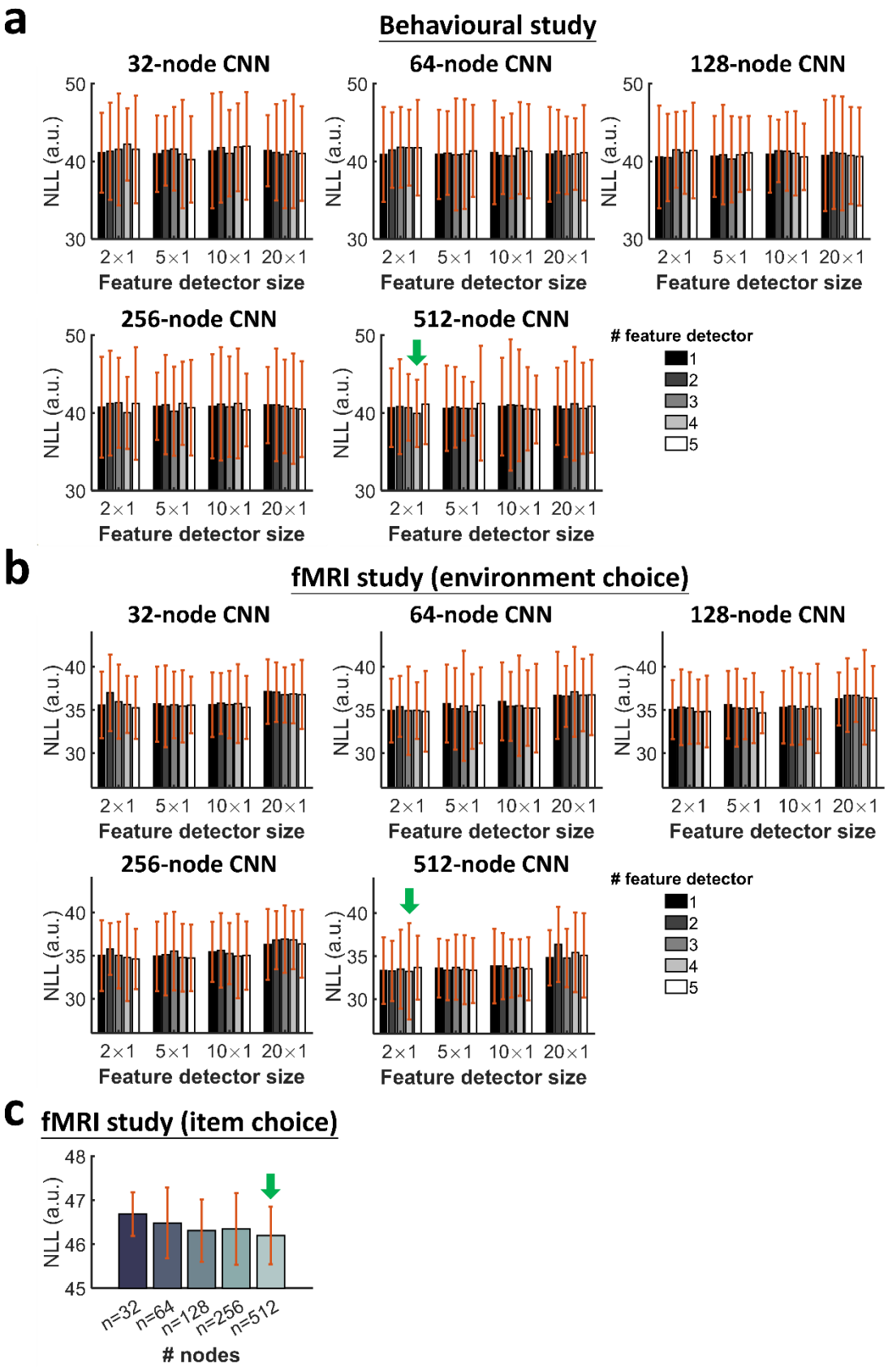
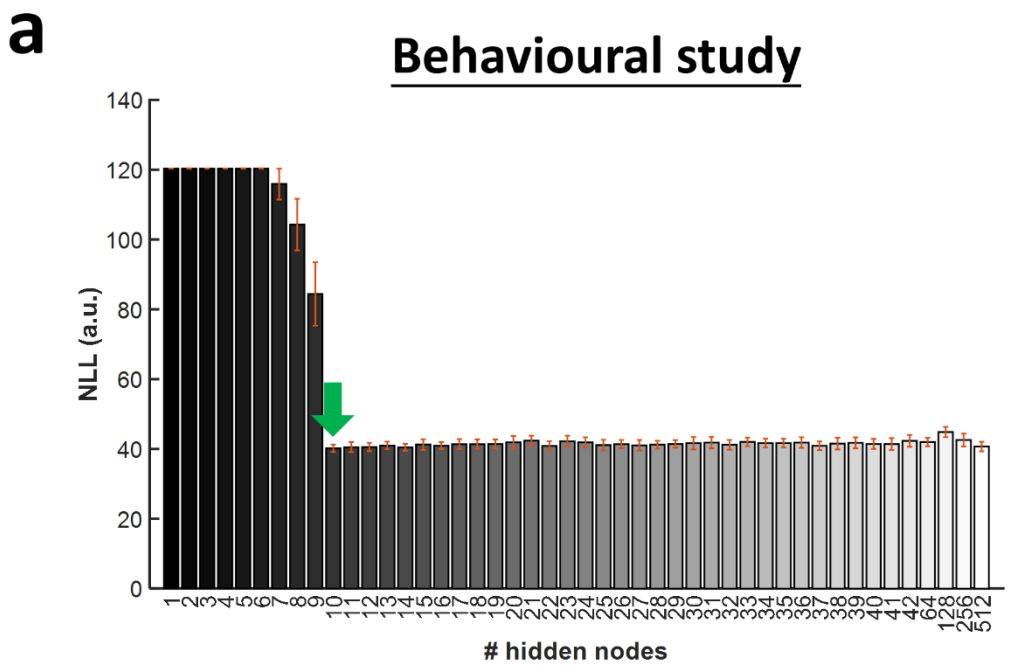

a
Behavioural study
32-node CNN
64-node CNN
128-node CNN
256-node CNN
512-node CNN
NLL (a.u.)
Feature detector size
2×1
5×1
10×1
20×1
feature detector
1
2
3
4
5
b
fMRI study (environment choice)
c
fMRI study (item choice)
n=32
n=64
n=128
n=256
n=512
# nodes

Figure 4.3. Performance of different CNN/ANN architectures. (a) The CNN involving four feature detectors in a size of 1×2 and 512 nodes in *fc1* (green arrow) yielded the lowest negative log-likelihood (NLL; 39.935) for the environment choice of the behavioural study. **(b)** Identical CNN architecture (green arrow) also yielded the lowest NLL (33.235) for environment choice of the fMRI study. **(c)** ANN with 512 nodes in *fc1* yielded the lowest NLL (46.195) for item choice of the fMRI study. Error bars represent means±s.e.m.

4.3.2 Parameterization of autoencoder

The autoencoder is characterized by its capability in dimensionality reduction. The autoencoder reduces the high-dimensional input into low-dimensional hidden nodes representing the features of the original input with minimal information loss (Kiarashinejad et al., 2020). Different autoencoder architectures were tested by varying the number of hidden nodes. Since there were 42 numbers representing the value of reward magnitude and reward probability (one for reward magnitude and 20 for reward probability for each environment) in the input layer, the number of hidden nodes being tested ranged from one to 42 and further increment (i.e. 64, 128, 256, and 512) was arbitrarily set. Interestingly, relatively high NLLs were yielded for autoencoders with less than 10 nodes and the NLL significantly dropped when 10 nodes were used (ts$_{38}$<-4.823, Ps<2.301×10^{-5}; Fig. 4.4a). From 10 nodes onwards, there was no significant difference in NLL (F(36,703)=0.314, P=1; Fig. 4.4a). Nevertheless, the autoencoder with 10 nodes yielded the lowest NLL (40.177±4.687) among all autoencoder variants and thus it was selected for comparison with other computational models.

For environment choice of the fMRI study, the variant with 512 hidden nodes yielded the lowest NLL (37.175±3.589; Fig. 4.4b). This 512-node variant outperformed about half the tested variants, especially when fewer number of hidden nodes were used (i.e. variants with number of nodes of 1-20, 24, 27; ts_{38}<-2.133, Ps<0.040; Fig. 4.4b). Item choice data was fitted with the same procedures. The variant with 512 hidden nodes also yielded the lowest NLL (46.638±4.95; Fig. 4.4c). In addition, it significantly outperformed all other variants (ts_{38}<-2.500, Ps<0.017; Fig. 4.4c), except when 43-45 hidden nodes were used (ts_{38}>-0.755, Ps>0.454; Fig. 4.4c).



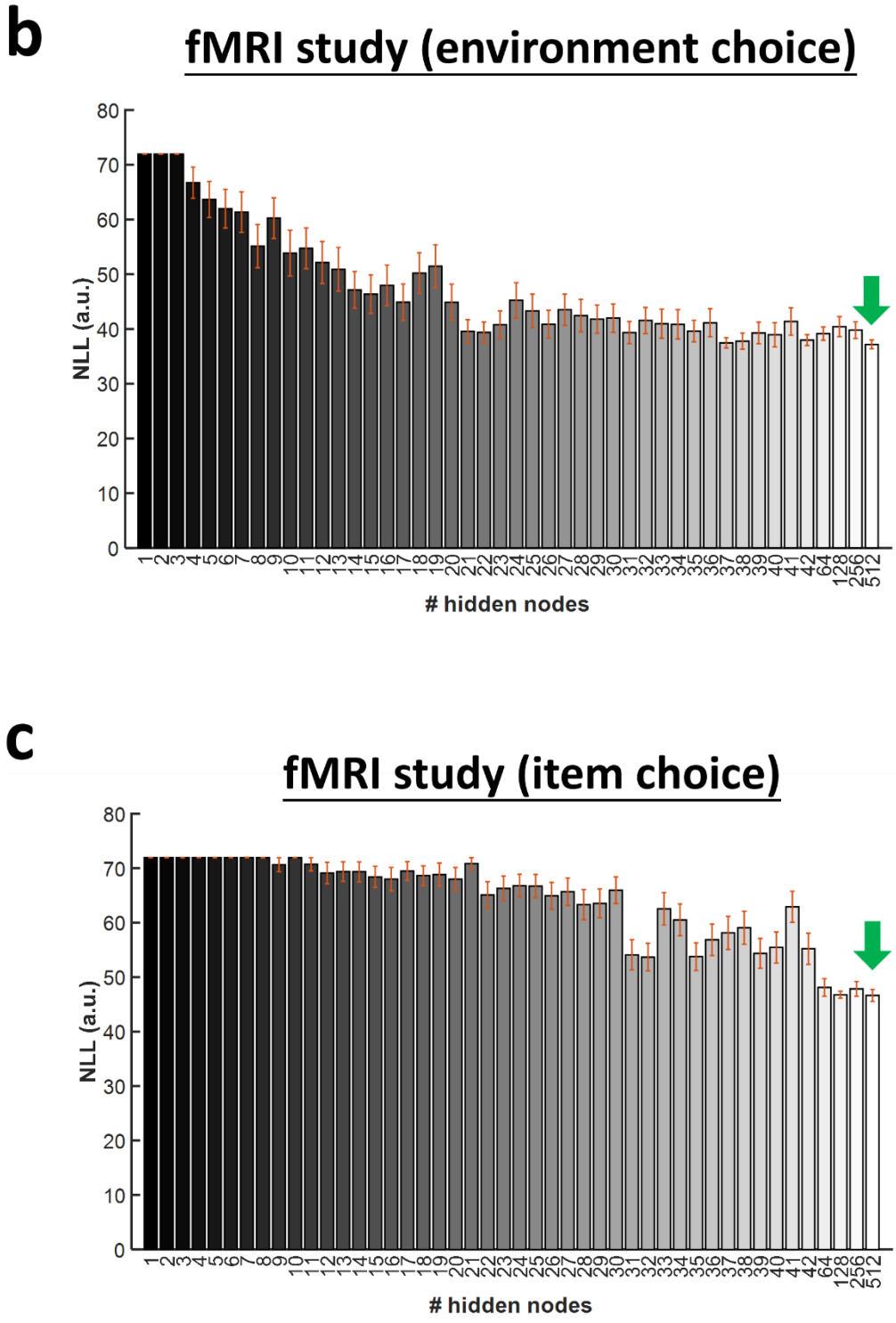


Figure 4.4. Performance of different autoencoder architectures. (a) The autoencoder involving 10 hidden nodes (green arrow) yielded the lowest negative log-likelihood (NLL; 40.177) for the behavioural

study. **(b)** For the environment choice of the fMRI study, the autoencoder involving 512 hidden nodes (green arrow) yielded the lowest NLL (37.175). **(c)** The autoencoder involving 512 hidden nodes (green arrow) also yielded the lowest NLL (46.638) for item choice of the fMRI study. Error bars represent means±s.e.m.

4.3.3 Model comparison revealed CNN/ANN best predicts participants' choices

After identifying the optimized architectures, these optimized CNNs, ANN, and autoencoders were selected and compared with all other models. Model comparison results of the behavioural study revealed that the CNN outperformed all other models (NLL=39.935; ts_{38}<-2.222, *P*s<0.033; Fig. 4.5a) except the autoencoder (t_{38}=-0.170, *P*=0.866). Surprisingly, when the environment choice data of the fMRI study was inspected, the CNN also significantly outperformed all other alternative models (NLL=33.235; ts_{38}<-2.662, *P*s<0.012; Fig. 4.5b). Finally, the same set of models was applied to predict participants' item choices (except MVS model and power law model that are not applicable). Model comparison revealed the ANN outperformed other alternative models in predicting item choices (NLL=46.195, ts_{38}<-3.226, *P*s<0.003; Fig. 4.5c) except the autoencoder (ts_{38}=-0.345, *P*=0.732; Fig. 4.5c).

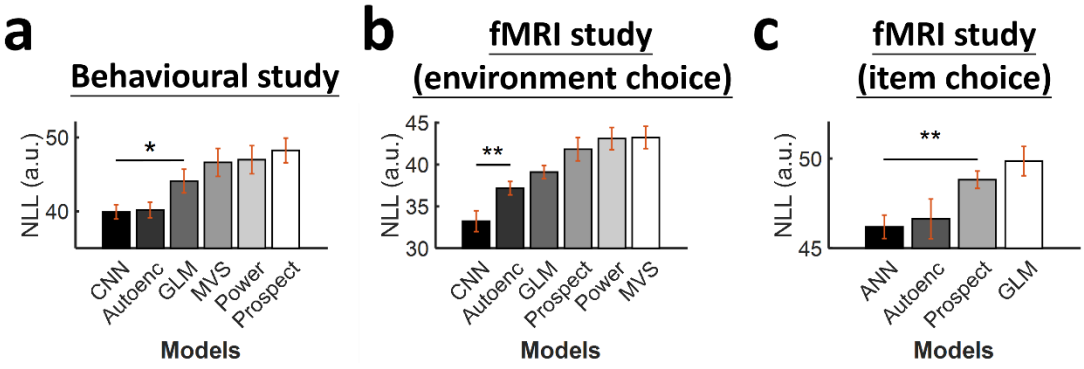


Figure 4.5. Model comparison results. The CNN/ANN outperforms other alternative models in predicting participants' decision-making behaviour in all data sets: **(a)** behavioural study, **(b)** environment choices of the fMRI study, **(c)** item choices of the fMRI study. CNN=convolutional neural network; ANN=artificial neural network; Autoenc=autoencoder; GLM=general linear model; MVS=mean-variance-skewness model; Power=power law model; Prospect=cumulative prospect theory. ** denotes $P<0.01$ and * denotes $P<0.05$. Error bars represent means±s.e.m.

4.3.4 CNN encodes statistical moments within the environments

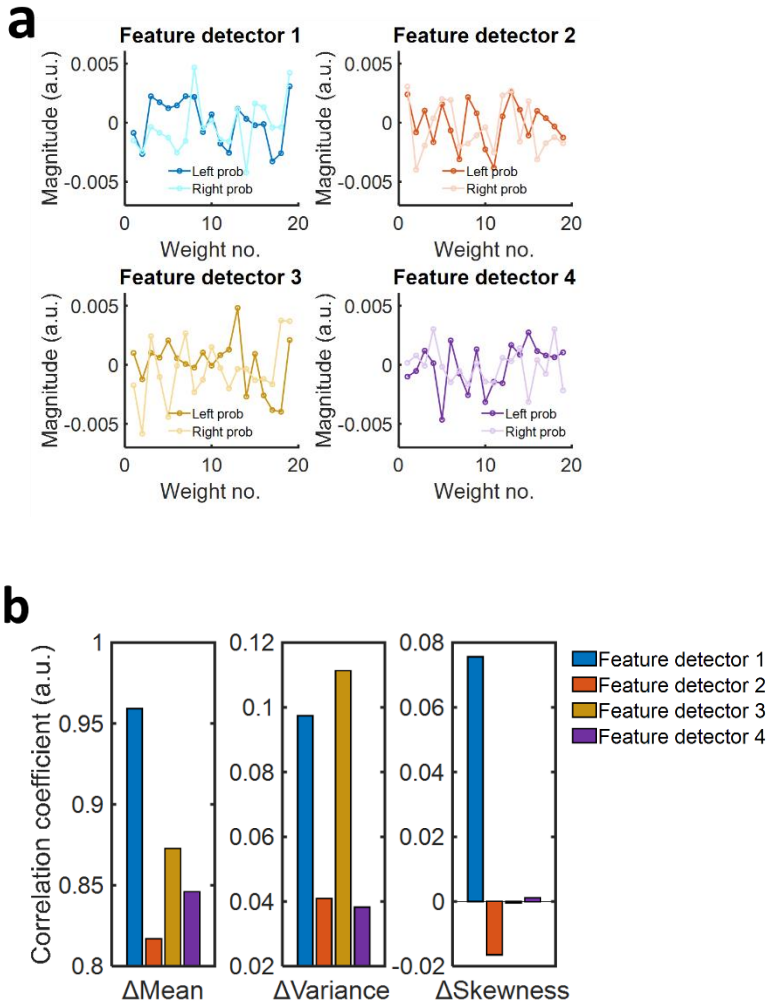
Thus far, the CNN was found outperforming all other models in predicting participant's choices across data sets. It is worth noting that the behavioural analysis results in **Chapters 2** and **3** also revealed participants' environment choices were guided by statistical moments of the environments. Yet, whether the CNN involves any information about statistical moments remains elusive. In order to scrutinize the neural mechanisms during environment choice in the next chapter, hereafter all analyses only focused on the CNN for the fMRI study.

I have showed that the CNN consisting of four feature detectors best predicts participants' environment choices. During convolution, each feature map extracts the features about the item distribution in each environment by assigning a weight for every item probability (Fig. 4.6a). Given the extracted features are ultimately integrated into a DV for each environment to guide decision, the DVs should contain information about the statistical moments. Hence, I inspected the DVs derived from each feature detector individually to examine the corresponding contribution of each feature detector in extraction of the statistical moments. The relationships between the between-

environment difference in DV and the differences in statistical moments of the environments were tested. Significant correlations with the between-environment differences in mean (rs>0.816, Ps<0.001; Fig. 4.6b) and variance (rs>0.038, Ps<0.009) were found for all feature detectors but significant correlation with difference in skewness was only observed for Feature detector 1 (r=0.076, P=1.658×10^{-7}). Besides, when focusing on the correlations with differences in mean and variance, interestingly these feature detectors exhibit disparate correlation patterns. Feature detector 1 possesses more prominent correlation with the difference in mean and relatively stronger correlation with the difference in variance than other feature detectors. While Feature detector 3 shows the strongest correlation with the difference in variance among all feature detectors. Feature detectors 2 and 4 exhibit very similar correlations on difference in variance but Feature detectors 4 shows a relatively stronger correlation with difference in mean. These differential correlation patterns enable the CNN to encode the statistical moments of the item distribution in the environments during environment choice.

Apart from that, a diagnostic test was performed to examine whether the CNN encodes any information about statistical moments. First, four "simplified" CNN models were trained. Instead of receiving 20 component items of each environment as input, these models receive either the item mean (Mean Model), the item mean and variance (Mean+Var Model), item mean and skewness (Mean+Skew Model), or item mean, variance and skewness (Mean+Var+Skew Model). Supposedly the original CNN resembles the fourth model the most and the results of RSA revealed it was the case. There were significant correlations between the original CNN and these four "simplified"

CNN models (*rhos*>0.978; *Z*s>4.299, signed-rank *P*s<1.708×10^{-5}, Fig. 4.6c) and importantly, the correlation was strongest with the Mean+Var+Skew Model (*Z*s>4.413, signed-rank *P*s<1.017×10^{-5}; Fig. 4.6c). These suggest that the CNN is most similar to a model that receives input of the mean, variance and skewness of the environments.



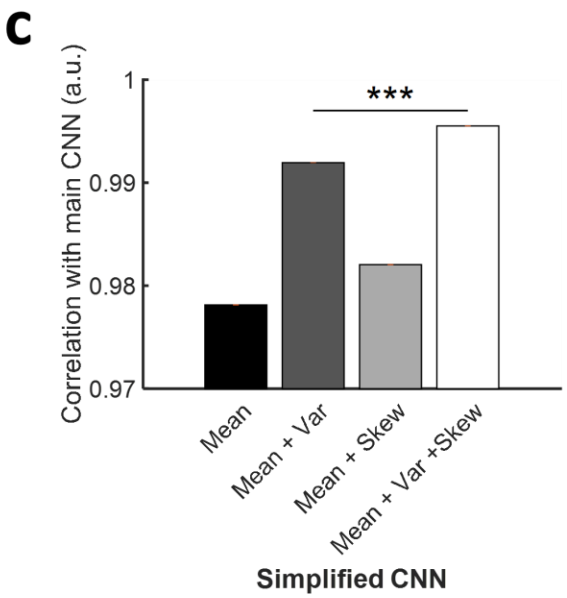


Figure 4.6. The CNN contains information about statistical moments in environments. (a) All feature detectors possess a weight for every item probability in each environment. **(b)** Since the feature detectors have different weights on the item probabilities, the DVs (i.e. the ultimate output) of the feature detectors show different degrees of correlation to the differences in mean, variance, and skewness of the environments. **(c)** A representational similarity analysis (RSA) was performed to compare the similarities of the original CNN with different "simplified" CNNs, which directly receive the actual mean, variance, and/or skewness of the environments as input. The original CNN shows greatest similarity to the simplified model that receive a combination of mean, variance, and skewness. Error bars represent means±s.e.m.

4.4 Discussion

It is a critical step to scrutinize how the complex information embedded in the environments are integrated before examination of the neural computation underlying decision between environments. Each environment was consisted of a distribution of 20 items and the behavioural analysis results in **Chapters 2** and **3** have shown that statistical

moments such as the variance could bias environment choices in human behaviour. Therefore, different models were applied to investigate how these pieces of multiplex values in the environments are estimated. By careful testing of different architectures, it was found that the CNN with four feature detectors in a size of 2×1 and 512 nodes in *fc1* outperforms other models, in both the behavioural study and fMRI study. This could be due to the high degree of resemblance of environment choice to computer vision that the CNN is widely used for. In artificial intelligence, the CNN has been found successful in computer vision. It is not only because it can analyze complex visual information to decode image identity, but also it resembles the hierarchical neural process of the human ventral visual pathway (Krizhevsky et al., 2012; Lindsay, 2020). On one hand, visual object recognition involves integrating complex information from the retina. Similarly, environment choice requires integration of complex choice information to guide decisions. This similarity might explain why the CNN is also successful in decoding environment choices from the multiplex environments.

Another important finding of the CNN is that it carries information about statistical moments within the environments. Although the behavioural analysis results in **Chapters 2 and 3** have demonstrated that participants' choices were biased by statistical moments, the behavioural analysis as well as other alternative models in this chapter (e.g. MVS model, power law model) assume the participants hold perfect estimation of the statistical moments. Contrarily, CNN does not require explicit specification of the statistical moments and the CNN representations are not perfectly correlated with the statistical moments indeed. Yet, this should better resemble the actual behaviour that people are unable to carry out a perfect estimation of the statistical moments due to their

limited cognitive resources. This might explain why the CNN outperforms the other models, in addition to the similarity among environment choice and computer vision. Moreover, the GLMs reported in **Chapters 2** and **3** only revealed effects of mean and variance of the item distributions on choices. Now I showed that the outperforming CNN encodes the skewness in addition to the mean and variance of the item distributions although to a lesser extent. It reflects environment choice is also biased by the less intuitive skewness and this effect can only be delineated with a more sensitive model such as the CNN.

This chapter shows that participants' environment choice behaviour was best described by the CNN among different computational models. Importantly, the CNN provides value estimates of the environments as the basis for examination of the neural signals during environment choice. The neural substrates underlying environment choice could be identified with the use of the DVs derived from the CNN. Furthermore, the CNN possesses multiple nodes and multiple layers for representation and integration of multiplex environment values. The precise mechanisms underlying environment choice could be inspected with the use of CNN. All these details of neural data analysis were discussed in **Chapter 5**.

Chapter 5 Contrasting roles for lateral frontopolar cortex and ventromedial prefrontal cortex in environment choice and item choice

Chapter highlights

- This chapter aims to test the neural substrates underlying environment choice and item choice
- A functional magnetic resonance imaging study showed a double dissociation of lateral frontopolar cortex (FPl) and ventromedial prefrontal cortex (vmPFC) in environment choice and item choice
- The FPl signals exhibited two essential properties of neural value comparison: (1) context-dependent modulation and (2) invariance to salience
- FPl was similar to the convolutional neural network in encoding environment value via parallel processing
- FPl employed a multivariate coding for representation and integration of environment value

5.1 Introduction

The central question of this dissertation is, given the diversity of decisions, whether a single neural substrate deals with all different kinds of decision or distinct neural substrates are necessary for specific kinds of decision. However, the exiting

literature primarily focuses on item choice whereas environment choice receives far less attention and remains obscure. To address the central question, it is crucial to investigate the neural mechanisms underlying environment choice. In the preceding chapters, I have demonstrated the most important feature of environment choice, the prospective thinking, was successfully probed behaviourally. Besides, I fitted people's environment choices with different computational models and found the convolutional neural network (CNN) best described their environment choice behaviour. These behavioural findings provide the basis for further examination of the neural mechanisms underlying environment choice. With the use of CNN, the neural computation involved during environment choice can be estimated. Critically, this fMRI study also included decisions between items. By examining and contrasting the neural mechanisms during environment choice and item choice, whether these two types of decision are subserved by the same neural substrate or separate neural substrates can be directly tested.

The CNN is a deep learning neural network which allows estimation of the implicit cognitive processes during environment choice that are hard to be portrayed by conventional computational models. The CNN by means of multiple feature detectors extracts and integrates the multiplex information embedded in the environments into decision values (DVs) thereby to guide environment choices. Given the CNN has been shown to best predict participants' environment choices, the neural substrate underlying environment choice supposedly involves similar processes to the CNN by showing two characteristics. First, the underlying neural substrate should exhibit activity that correlates with the CNN DVs. Second, the CNN possesses multiple nodal activations and the underlying neural substrate should exhibit a similar multivariate coding, which

can be revealed by a representational similarity analysis (RSA). RSA is a multivariate analysis that can test the similarity between different data sets such as behavioural data and neural data, by which the closeness between the trial-by-trial brain activation patterns and the trial-by-trial CNN representations can be inspected. The neural substrate underlying environment choice supposedly shows correlated activation patterns with the CNN representations in RSA.

In this chapter, the functional specialization in environment choice and item choice was examined. Whole-brain analysis results revealed a double dissociation in the neural substrates underpinning environment choice and item choice – the lateral frontopolar cortex (FPl) was involved in the former and the ventromedial prefrontal cortex (vmPFC) was involved in the latter. Region-of-interest analysis results further affirmed this double dissociation. Notably, the FPl signal during environment choice exhibited two essential properties of value comparison process, namely (1) context-dependent modulation and (2) invariance to salience. Moreover, with the use of RSA, the FPl was found underwent a parallel value encoding process that is similar to the CNN to extract information from environments. The extracted information was subsequently represented and integrated by a multivariate coding in the FPl.

5.2 Methods

5.2.1 Experimental task and computational modelling

Details please refer to Section *Experimental task* of **Chapter 3**. In brief, this fMRI study involved a two-stage decision-making task. In each block, there was a Stage 1 trial followed by 0-3 Stage 2 trials. Stage 1 was related to environment choice. On each

Stage 1 trial, two environments were presented; each of which was composed of a number (representing the average reward magnitude of 20 items) and 20 bars (representing the individual reward probabilities of 20 items). After selection of an environment, two items from the chosen environment were pseudo-randomly drawn and became the items available for selection on the subsequent Stage 2 trial. On each Stage 2 trial, participants made a decision between the items. Reward was then delivered according to the reward magnitude and reward probability associated with the selected item. Moreover, there were two Bonus Conditions to determine the acquisition of an additional bonus. In Linked Condition, the bonus was delivered if the chosen item on the same Stage 2 trial led to reward. In Unlinked Condition, the bonus was delivered unconditionally.

Details about the CNN can refer to the Section *Convolutional neural network (CNN)* in **Chapter 4**. Briefly, there are four essential parts in the CNN: (1) input layer; (2) hidden layers; (3) choice-predicting fully-connected layer; (4) output layer. Essential environment value (i.e. reward magnitude and reward probability) is fed in as the input layer of the CNN. Afterwards, reward probability in the input layer is convolved by four feature detectors and transformed into four sets of feature maps. Feature maps are then multiplied by the reward magnitude and the bonus information (i.e. bonus value and Bonus Condition) to account for the interactions between reward probability and these attributes. Feature maps, reward magnitude, feature maps × reward magnitude, and feature maps × bonus value × Bonus Condition are concatenated and converted 512 nodes. Finally, these 512 nodes are weigh-summed to form the decision values (DVs) in

the choice-predicting fully-connected layer. DVs are subsequently transformed by a softmax function to generate choice probabilities in the output layer.

5.2.2 Neuroimaging data acquisition and preprocessing

A Siemens 3T MRI scanner was used to collect the neuroimaging data. FMRI data were collected with a multi-band echo planar imaging sequence: $2.5 \times 2.5 \times 2.5$ mm^3 voxel- resolution, TR=1.6s, TE1=15ms, TE2=36.19ms, TE3=57.38ms, flip angle=70°, slice angle=-30° (T>C). Field maps were collected to correct for signal distortions using a dual echo 2D gradient echo sequence: $2.5 \times 2.5 \times 2.5$ mm^3 voxel resolution, TR=590ms, TE1=4.92ms, TE2=7.38ms. T1-weighted structural images were acquired using an MPRAGE sequence: $1 \times 1 \times 1$ mm^3 voxel resolution, $174 \times 192 \times 192$ grid, TR=1900ms, TE=3.97ms, TI=904ms.

FMRIB's Software Library (FSL) (Smith et al., 2004) was used to analyze fMRI data. FMRI data were preprocessed by the following procedures: motion correction by using FMRIB's Linear Image Registration Tool (Jenkinson et al., 2002), brain extraction by Brain Extraction Tool (Smith, 2002), field map correction for distorted signal (Jenkinson, 2003), Gaussian spatial smoothing with fullwidth at half maximum of 5mm, and high-pass temporal filtering (3 dB cutoff of 100s). FMRI data were registered to the high-resolution structural images of individual participants, which were then normalized into the standard Montreal Neurological Institute (MNI) space (Jenkinson & Smith, 2001).

5.2.3 Whole-brain analysis

A univariate GLM approach was used for all whole-brain analyses. At individual level, a GLM was applied to fit the fMRI data of every participant. At group level, FMRIB's local analysis of mixed effects (Beckmann et al., 2003; Woolrich et al., 2004) was applied with outlier deweighting (Woolrich, 2008). Significant clusters were identified using a cluster-based thresholding of $Z>3.1$ and significance threshold of $P<0.05$ (Worsley et al., 1992), unless otherwise stated.

The whole-brain GLM included two regressors-of-interest, the $\Delta DV_{environment}$ (i.e. difference in DV between the chosen and unchosen environments in Stage 1) and ΔDV_{item} (i.e. difference in DV between the chosen and unchosen items in Stage 2), to examine the signals during environment choice and item choice respectively. These two regressors were time-locked to the stimulus (environments in Stage 1 and items in Stage 2) onset and convolved with a haemodynamic response function. A total of six nuisance regressors were additionally included: (1) a parametric regressor of the bonus value, time-locked to block bonus onset; (2) a parametric regressor about the value of the chosen environment for modelling the neural signal pertaining to the chosen environment prior to the onset of each Stage 2 trial, time-locked to the onset of Stage 1 delay phase; (3) two box car regressors related to left-hand and right-hand responses, time-locked to participants' responses; (4) two regressors related to the average BOLD signal in the cerebrospinal fluid (CSF) and white matter (WM).

5.2.4 Region-of-interest analysis

To inspect the time courses of the key brain regions identified from the whole-brain GLM, region-of-interest (ROI) analysis was conducted. Whole-brain analysis showed that activities of the lateral frontopolar cortex (FPl) and the ventromedial prefrontal cortex (vmPFC) were correlated with the ΔDV during environment choice and item choice respectively. Therefore, a mask was generated by centering a sphere of 3mm radius at the coordinates pertaining to information-seeking to extract the FPl activity (Zajkowski et al., 2017), whereas another mask was generated by centering a sphere of 3mm radius at the coordinates pertaining to item choice to extract the vmPFC activity (Juechems et al., 2017). Besides, the dorsal posterior cingulate cortex (dPCC) was identified in both environment choice and item choice. Activity of the dPCC was extracted using an atlas which is developed based on findings of reward-guided decision-making (Neubert et al., 2015). A ten times upsampling was performed by cubic spline interpolation and time-locked to stimulus onset (either Stage 1 or Stage 2) for the extracted ROI time courses. For each participant, a GLM was applied to regress the ROI activities at every time point to obtain time courses of beta weight for each term in the GLM. The beta weight time courses were group averaged and plotted. To test whether each peak in the group time courses was significantly different from zero, the size of each peak of each participant was extracted and tested using a one-sample t-test. To avoid bias in estimating the peak size, a leave-one-subject-out procedure was employed (Chau et al., 2015; Trudel et al., 2020). In particular, a time window for extracting a peak was defined by the full-width half-maximum of that peak in the group time course. Next, for a given participant, the time point of peak extraction was defined by the position of the peak within this time window when the group time course had the beta weights from that

participant removed. This leave-one-out procedure was repeated for all participants to extract the peak for statistical tests.

The first ROI analysis (GLM7) involved using the ΔDV to regress the ROI activities:

$$ROI\ time\ courses = \beta_0 + \beta_1 \Delta \mathrm{DV}$$

where ΔDV denotes $\Delta DV_{environment}$ and ΔDV_{item} for Stage 1 environment choice and Stage 2 item choice respectively.

The second ROI analysis (GLM8) tested whether the ROI signal time courses exhibited differential responses in different Bonus Conditions. First, the DV term in GLM7 was split into two terms: (1) a DV term that concerns all choice information except the bonus (i.e. $DV_{bonus\ excluded}$) and (2) a DV term that only concerns the bonus-related information (i.e. $DV_{bonus\ only}$). Afterwards, $\Delta DV_{bonus\ excluded}$, $\Delta DV_{Linked\ bonus}$, and $\Delta DV_{Unlinked\ bonus}$ were computed and included as regressors. $\Delta DV_{Linked\ bonus}$ was the bonus value in the Linked Condition. In contrast, $\Delta DV_{Unlinked\ bonus}$ was the bonus value in the Unlinked Condition but it was computed as if the bonus occurred in the Linked Condition. GLM8 is formulated as follow:

$$ROI\ time\ courses = \beta_0 + \beta_1 \Delta \mathrm{DV}_{bonus\ excluded} + \beta_2 \Delta \mathrm{DV}_{Linked\ bonus} + \beta_3 \Delta \mathrm{DV}_{Unlinked\ bonus}$$

Finally, GLM9 tested the salience invariance of the ΔDV signal by including regressors of ΔDV, DV sum and salience (i.e. sum of the two options' absolute values). Noting that DV sum and salience were computed using expected value but not CNN-derived DV. It is because CNN involves normalization and non-linear transformation

such that information about the sign of the option values would be lost. GLM9 is formulated as follow:

$$ROI\ time\ courses = \beta_0 + \beta_1 \Delta \mathrm{DV}_{bonus\ excluded} + \beta_2 DV\ sum + \beta_3 Salience$$

5.2.5 Psychophysiological interaction (PPI) analysis

Since coactivation among different regions (i.e. FPl, vmPFC, dPCC) was observed in the whole-brain analysis, a PPI analysis was performed to examine the functional coupling between these regions during environment choice and item choice. PPI is an analysis of functional connectivity by which how the task-dependent activities among different brain regions interact is tested (Friston et al., 1997). Basically, a PPI analysis included a psychological regressor (i.e. the task variable convolved with a haemodynamic response function), a physiological regressor (i.e. the time course of an ROI), an interaction term between the psychological regressor and physiological regressor. Because ROI analysis revealed the dPCC exhibited correlated activity with the ΔDV during both environment choice and item choice, the dPCC was set as the seed region and its functional coupling with FPl and vmPFC during different choice were tested by GLM10:

$$dPCC\ time\ courses = \beta_0 + \beta_1 Stage + \beta_2 PHYS_{FPl} + \beta_3 (Stage \times PHYS_{FPl}) + \beta_4 PHYS_{vmPFC} + \beta_5 (Stage \times PHYS_{vmPFC})$$

where *Stage* is the psychological regressor (a dummy variable; Stage 1 environment choice=1, Stage 2 item choice=0), $PHYS_{FPl}$ and $PHYS_{vmPFC}$ are the physiological regressors of FPl and vmPFC time courses respectively.

5.2.6 Representational similarity analysis (RSA)

The procedures of RSA and the computation of the representational dissimilarity matrices (RDMs) were largely similar to the description in Section *Representational similarity analysis* of **Chapter 4**. First, for the CNN representations, trial-by-trial nodal activations of a given layer (e.g. a feature map) in the CNN were extracted. A self-correlation was then performed using Pearson correlation and the resulting correlation matrix was subtracted by 1 to become an RDM. Similarly, for the neural data, trial-by-trial voxel activations in different ROIs were extracted. The ROIs included the FPl, vmPFC, dPCC, primary visual cortex (V1), and cerebrospinal fluid (CSF). Voxel activations in the ROIs were extracted using built-in atlases of the FSL (Smith et al., 2004), except for the FPl, vmPFC, and dPCC that the corresponding ROI masks mentioned in the previous section *ROI analysis* were used. Since one participant did not finish the experimental task within an fMRI run, there were fewer number of trials for neural data than behavioural data. For RSA of that participant, the CNN RDM was computed by excluding the trials that lacked fMRI data. Finally, RSA was performed by computing Spearman correlations between a pair of RDMs.

5.3 Results

5.3.1 FPl, but not vmPFC, involved in environment choice

Chapter 4 shows that the CNN outperforms other computational models in describing behavioural environment choices. Particularly, the CNN is characterized in integrating multiplex information embedded in the environments (e.g. statistical moments) into single DVs to guide decisions. The CNN-derived DVs were thus applied

to identify the neural substrates underlying environment choice. Multiple lines of previous work have demonstrated that during item choice, the difference in option DVs was represented in the vmPFC (Boorman et al., 2009, 2013; Chau et al., 2014; Hunt et al., 2012; Levy & Glimcher, 2011). It is intriguing whether environment choice and item choice involved the same vmPFC region or dissociable neural substrates. To this end, a whole-brain GLM was performed to examine the fMRI data, which included the DV difference in Stage 1 environment choice (i.e. $\Delta DV_{environment}$) and the DV difference in Stage 2 item choice (i.e. ΔDV_{item}) as regressors. Surprisingly, a double dissociation of the DV difference signals was revealed. First, the widely reported finding that the vmPFC activity correlated with ΔDV_{item} was replicated (MNI=[2, 40, -10], cluster-based thresholding $Z>3.1$, $P=7.77\times10^{-21}$; Fig. 5.1b). In contrast, there was an absence of $\Delta DV_{environment}$ signal in the vmPFC when the Stage 1 environment choice trials were inspected. On the contrary, there was a positive signal of the Stage 1 $\Delta DV_{environment}$ (MNI=[38, 52, 22], cluster-based thresholding $Z>3.1$, $P=1.19\times10^{-7}$; Fig. 5.1a), but absence of Stage 2 ΔDV_{item}, in the FPl. Other significant regions are reported in Table 1. An additional whole-brain contrast of ($\Delta DV_{environment}$ - ΔDV_{item}) further identified the FPl (MNI=[28, 44, 36], note that this applied a more lenient threshold of cluster-based thresholding $Z>2.3$, $P=0.021$ because it did not reach significance at threshold of $Z>3.1$; Fig. 5.1c). Besides, it is a conventional approach to estimate DVs using expected value (i.e. the product between reward probability and magnitude), instead of using DVs derived from a CNN. Similar results were obtained even when I took this conventional approach. A $\Delta DV_{environment}$ signal in the FPl (MNI=[38, 52, 24], cluster-based thresholding $Z>3.1$, $P=2.65\times10^{-5}$, Fig. 5.1d) and an absence of $\Delta DV_{environment}$ signal in the

vmPFC were observed during environment choice. Also, a ΔDV_{item} signal in the vmPFC (MNI=[2, 40, -12], cluster-based thresholding $Z>3.1$, $P=1.5\times10^{-12}$, Fig. 5.1e) and an absence of ΔDV_{item} in the FPl were observed during item choice. It revealed that the results by the CNN-derived DVs were unlikely due to statistical artefact.

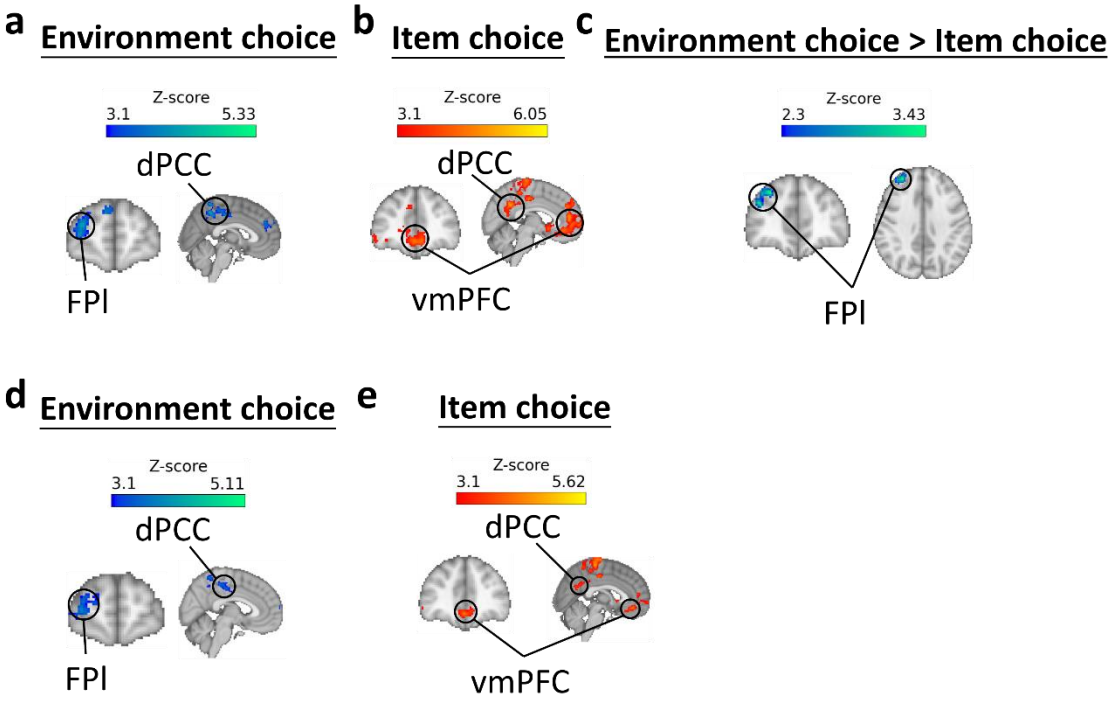


Figure 5.1. Dissociable roles of FPl and vmPFC in environment choice and item choice. (a) A whole-brain analysis showed that the lateral frontopolar cortex (FPl) signal was related to the difference in DV between environments ($\Delta DV_{environment}$) in Stage 1 environment choice and **(b)** the ventromedial prefrontal cortex (vmPFC) signal was related to the difference in DV between items (ΔDV_{item}) in Stage 2 item choice. **(c)** An additional whole-brain contrast indicated that the $\Delta DV_{environment}$ signal in environment choice was significantly stronger than the ΔDV_{item} signal in item choice in the FPl (threshold of cluster-based thresholding $Z>2.3$, $P<0.05$). **(d-e)** A conventional whole-brain analysis approach using difference in EV (ΔEV) instead of ΔDV also yielded similar results to **(a-b)**.

Table 1. Whole-brain analysis results						
	X	Y	Z	Max Z-score	P-value	# voxels
$\Delta DV_{environment}$						
Lateral frontopolar cortex	38	52	22	4.31	1.19×10^{-7}	486
Dorsal posterior cingulate cortex	16	-32	42	4.65	2.44×10^{-13}	1074
Frontal eye field	4	44	30	4.12	0.001	180
Inferior frontal gyrus	56	28	0	4.29	0.035	100
Lateral orbitofrontal cortex	42	26	-18	4.46	0.004	151
Postcentral gyrus	24	-40	66	4.11	0.001	188
Inferior parietal lobule	48	-62	32	5.28	7.09×10^{-29}	3281
Superior temporal gyrus	-62	-30	20	4.63	8.53×10^{-09}	598
Right cerebellum	18	-76	-32	3.9	0.014	122
Left cerebellum	-22	-86	-30	-5	1.22×10^{-13}	780
ΔDV_{item}						
Ventromedial prefrontal cortex	2	40	-10	4.47	7.77×10^{-21}	2055
Dorsal posterior cingulate cortex	8	-52	36	5.26	1.73×10^{-24}	2599
Temporoparietal junction	52	-20	-10	5.99	<0.0001	6932
Supramarginal gyrus	-54	-38	24	4.95	1.22×10^{-18}	1749
Medial frontopolar cortex	-16	62	12	3.61	2.47×10^{-4}	233
Temporal fusiform cortex	-34	-34	-24	3.98	0.014	123
Angular gyrus	-50	-68	28	3.69	0.015	121
Right cerebellum	28	-74	-36	3.77	0.038	99
Left cerebellum	-24	-74	-36	5.11	3.84×10^{-12}	947
$\Delta DV_{environment}$ - ΔDV_{item}^						
Lateral frontopolar cortex	28	44	36	3.4	0.021	348

$\Delta DV_{environment}$=decision value difference between environments; ΔDV_{item}=decision value difference between items.

^Cluster-based thresholding $Z>2.3$

5.3.2 Time courses of FPl and vmPFC affirmed their dissociable roles

Despite the findings of dissociable FPl and vmPFC signals in environment choice and item choice, it is critical to ascertain the absence of ΔDV_{item} signal in the FPl during item choice and absence of $\Delta DV_{environment}$ signal in the vmPFC during environment choice were not because of the conservative statistical corrections in the whole-brain analysis. An ROI analysis (GLM7) was thus performed by placing a mask over the FPl and vmPFC and tested their signal time courses using one-sample t-tests. To affirm any null effects identified, a Bayesian t-test was also applied to each ROI analysis because of the limitation of conventional frequentist inference in confirming null effects. GLM7 yielded broadly consistent results to the whole-brain analysis. First, there was a $\Delta DV_{environment}$ signal in the FPl during environment choice which ramped up and peaked at 3.507s after stimulus onset (β=0.046, t_{23}=3.032, P=0.006, BF_{10}=7.526; Fig. 5.2a, top panel), and no significant ΔDV_{item} signal was found in the FPl during item choice (β=0.023, t_{23}=1.409, P=0.172, BF_{10}=0.514; Fig. 5.2b, top panel). On the contrary, a ΔDV_{item} signal during item choice was observed in the vmPFC which peaked at 7.813s (β=0.067, t_{23}=3.806, P=9.105×10^{-4}, BF_{10}=38.074; Fig. 5.2b, bottom panel) but no significant $\Delta DV_{environment}$ signal was found during environment choice in the vmPFC (β=0.018, t_{23}=1.011, P=0.323, BF_{10}=0.339; Fig. 5.2a, bottom panel). To further ascertain the double dissociation, sizes of the ΔDV signals observed in the FPl and vmPFC during environment choice and item choice were compared by a two-way ANOVA (2 regions × 2 stages). The ANOVA yielded a significant interaction effect (F(1,23)=16.026, P=0.001), implying the ΔDV signal in the FPl was stronger during environment choice than during item choice and the vmPFC exhibited the opposite pattern.

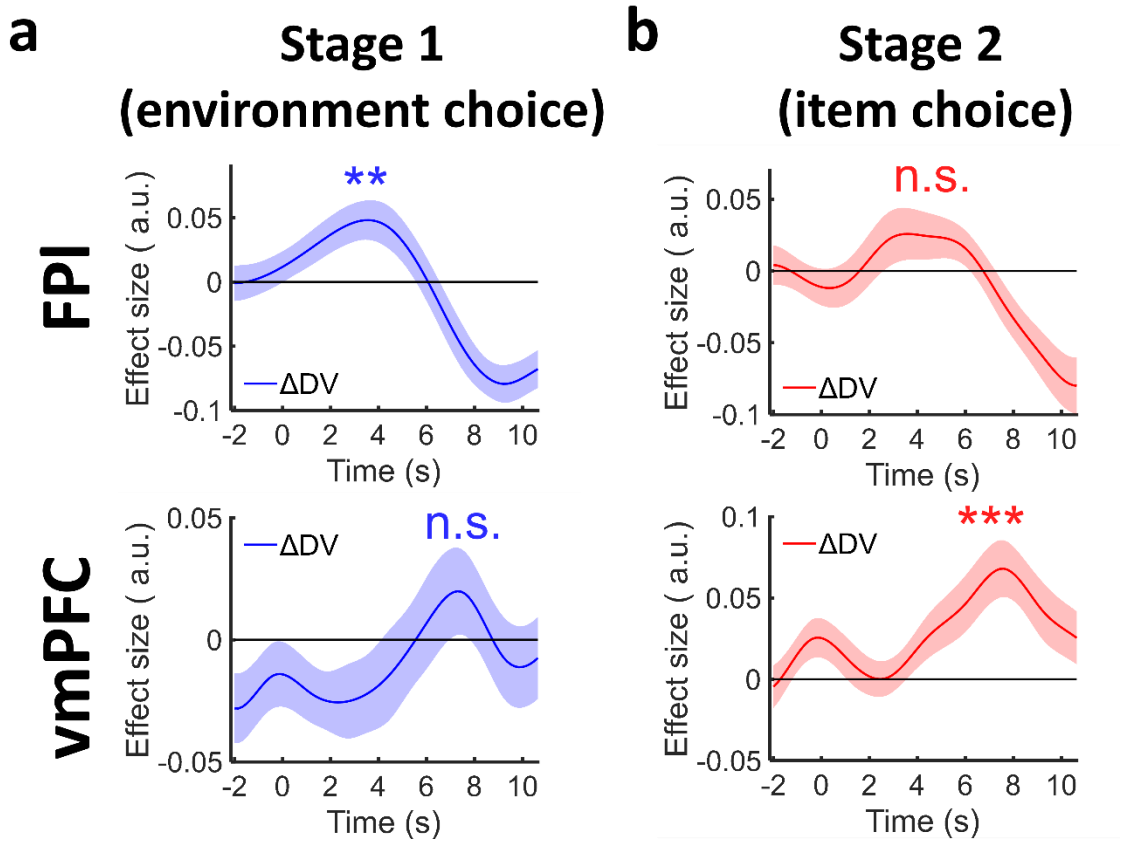


Figure 5.2. ROI analysis affirmed the dissociable roles of FPl and vmPFC. (a) Significant Stage 1 $\Delta DV_{environment}$ signal was only observed in the FPl (top panel), but not in the vmPFC (bottom panel). **(b)** Contrarily, significant Stage 2 ΔDV_{item} signal was only found in the vmPFC (bottom panel), but not in the FPl (top panel). *** denotes $P<0.001$ and ** denotes $P<0.001$. Shading represent means±s.e.m.

Apart from that, the dissociable roles of the FPl and vmPFC in environment choice and item choice were supported by a closer inspection of their mean activities. Both the FPl and vmPFC are parts of the default mode network. Supposedly, an overall deactivation (i.e. negative signal in the mean activity) is observed in the composing regions of the default mode network when they engaged in a task (Buckner et al., 2008; Raichle & Raichle, 2001). If the vmPFC exclusively engaged in Stage 2 item choice but

not in Stage 1 environment choice, it should demonstrate an overall deactivation in Stage 2 only. The results revealed this was the case – the vmPFC exhibited a negative mean signal in Stage 2 but not in Stage 1 (β=-0.218, t_{23}=-7.423, P=1.512×10^{-7}, BF_{10}=1.026×10^{5}; Fig. 5.3b, magenta line). In contrast, if the FPl was only involved in environment choice exclusively, an opposite deactivation pattern in mean signal to the vmPFC should be shown in the FPl. The results in the FPl, however, were less conclusive. In Stage 1, despite the presence of an overall deactivation at a time point close to that of the $\Delta DV_{environment}$ signal (at 4.840s; β=-0.182, t_{23}=-5.874, P=5.491×10^{-6}, BF_{10}=3.752; Fig. 5.3a, cyan line), there was a positive effect in the mean signal before Stage 1 onset (β=0.172, t_{23}=6.770, P=6.633×10^{-7}, BF_{10}=2.613×10^{4}; Fig. 5.3a, cyan line). Whereas in Stage 2 an opposite pattern was observed – the FPl showed a deactivation after Stage 2 onset (β=-0.125, t_{23}=-6.183, P=2.623×10^{-6}, BF_{10}=7.380×10^{3}; Fig. 5.3a, magenta line) but its activity ramped up approximately 5s later (β=0.232, t_{23}=6.526, P=1.169×10^{-6}, BF_{10}=1.551×10^{4}; Fig. 5.3a, magenta line). It is less clear why there was a combination of both overall activation and deactivation. Nonetheless, it is unambiguous that the FPl showed opposite patterns in the mean signal during environment choice and item choice. Taken together, the ROI analysis results provided further evidence for the functional dissociation between the FPl and vmPFC in environment choice and item choice.

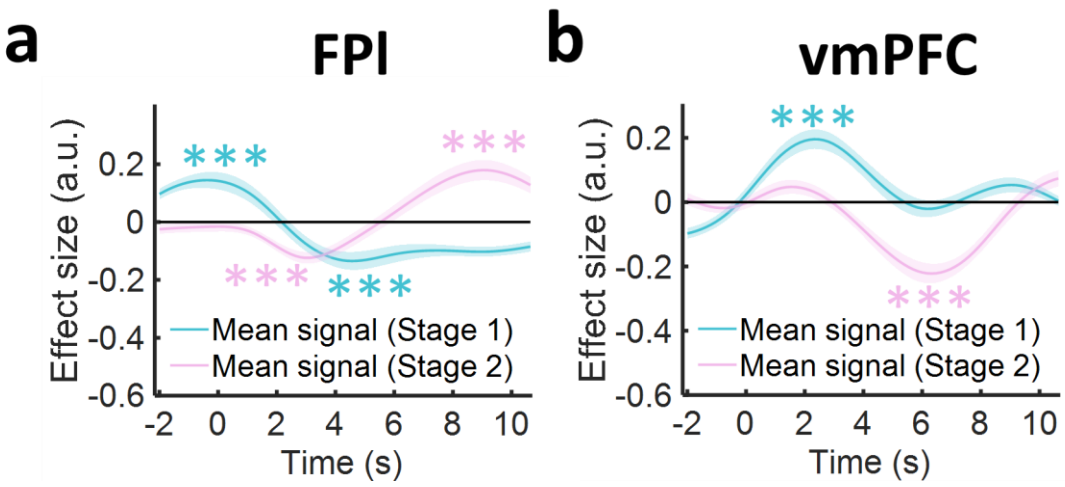


Figure 5.3. Mean signals of FPl and vmPFC shed light of their dissociable roles. (a) The FPl exhibited a mixture of ramping up and down in the mean signal in both Stage 1 environment choice and Stage 2 item choice. **(b)** The vmPFC demonstrated an overall activation of mean signal in Stage 1 environment choice (cyan) and an overall deactivation in Stage 2 item choice (magenta). *** denotes $P<0.001$. Shading represent means±s.e.m.

$\Delta DV_{environment}$ and ΔDV_{item} signals were also observed in the dPCC during environment choice and item choice respectively (Environment choice: MNI=[16, 32, -42], cluster-based thresholding $Z>3.1$, $P=2.44\times10^{-13}$; Fig 5.1a,d; Item choice: MNI=[8, -52, 36], cluster-based thresholding $Z>3.1$, $P=1.73\times10^{-24}$; Fig. 5.1b,e) on top of the signals identified in FPl and vmPFC. It was consistent to the roles of dPCC in value encoding of environments (Kolling et al., 2018) and value encoding of items (Lopez-Persem et al., 2016) in the existing literature. ROI analysis (GLM7) also reproduced these two signals in the dPCC ($\Delta DV_{environment}$: $\beta=0.039$, $t_{23}=2.199$, $P=0.038$, $BF_{10}=1.609$; Fig. 5.4a, left panel; ΔDV_{item}: $\beta=0.051$, $t_{23}=3.363$, $P=0.003$, $BF_{10}=14.818$; Fig. 5.4a, right panel). Besides, dPCC is also part of default mode network, similar to the FPl and vmPFC

(Buckner et al., 2008; Raichle & Raichle, 2001). It should exhibit overall deactivations during both environment choice and item choice. However, the dPCC only showed an overall deactivation during item choice (β=-0.108, t_{23}=-4.919, P=5.708×10^{-5}, BF_{10}=448.050; Fig. 5.4b, magenta line) but an absence of deactivation during environment choice (β=0.019, t_{23}=0.840, P=0.410, BF_{10}=0.295; Fig. 5.4b, cyan line).

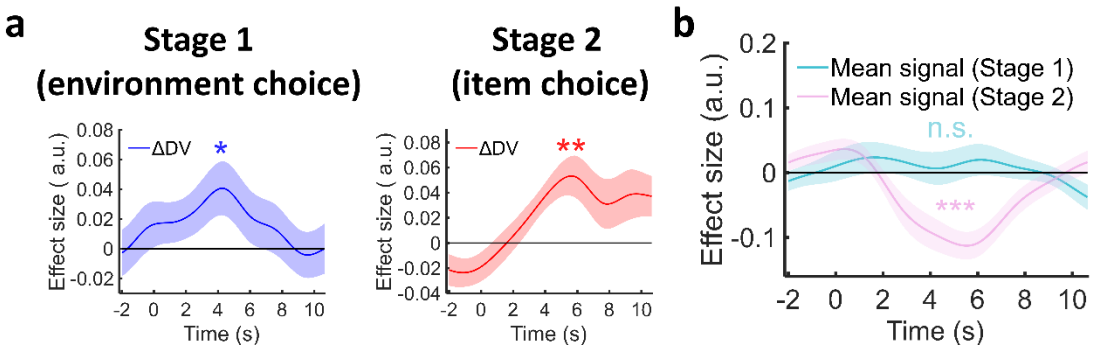


Figure 5.4. Value signals of dPCC during environment choice and item choice. (a) The dPCC showed correlated activity with ΔDV during Stage 1 environment choice (left panel) and during Stage 2 item choice (right panel) respectively. **(b)** An overall deactivation of dPCC was observed during Stage 2 item choice (magenta) but it was absent during Stage 1 environment choice (cyan). *** denotes P<0.001, ** denotes P<0.01, and * denotes P<0.05. Shading represent means±s.e.m.

Since all the FPl, vmPFC, and dPCC exhibited correlated activities with the ΔDV, a PPI analysis (GLM10) was performed to further test the interplay between these regions when undergoing different decisions (Fig. 5.5a). PPI analysis results revealed that there was a positive effect of *Stage* (a dummy variable; Stage 1 environment choice=1, Stage 2 item choice=0) on the functional connectivity between FPl and dPCC (β=0.048, t_{23}=2.529, P=0.019, BF_{10}=2.869; Fig. 5.5b), suggesting that the coactivation between FPl and dPCC was more related during environment choice than during item choice. On the

contrary, there was a negative effect of *Stage* on the functional connectivity between vmPFC and dPCC (β=-0.081, t_{23}=-2.388, P=0.026, BF_{10}=2.227; Fig. 5.5c). It implies the coactivation between vmPFC and dPCC was more related during item choice than during environment choice. Taken together, the changes in functional connectivity of the dPCC with FPl and vmPFC affirmed its engagement during environment choice and item choice.

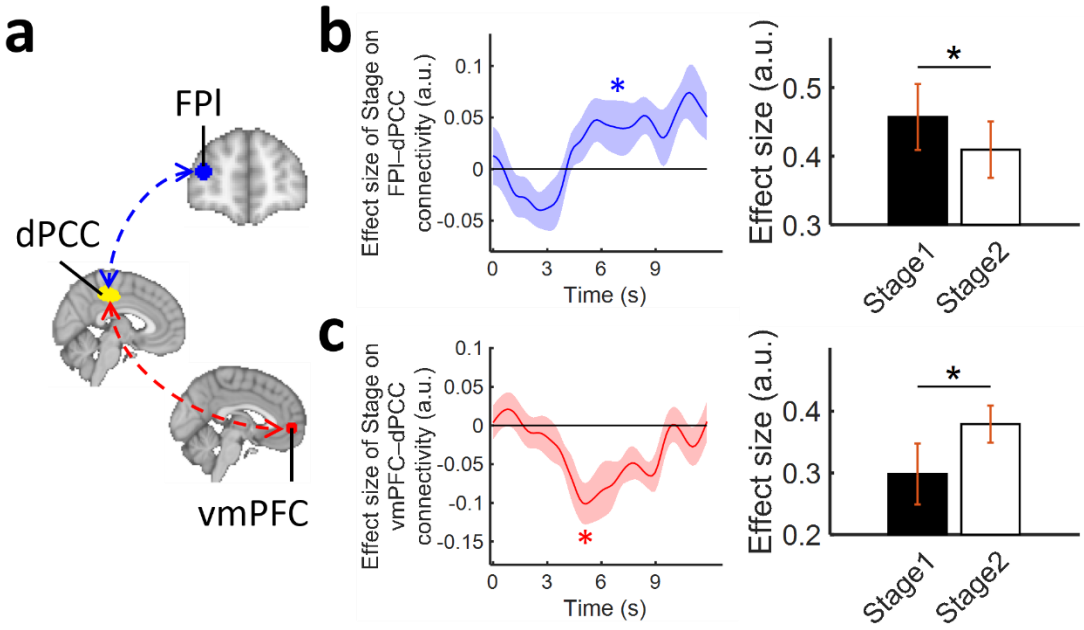


Figure 5.5. Psychophysiological interaction (PPI) analysis of dPCC with FPl and vmPFC. (a) The ROIs involved in the PPI analysis. **(b)** The FPl and dPCC were more functionally connected during environment choice than during item choice. The left panel demonstrates the time courses of *Stage* (a dummy variable; Stage 1 environment choice=1, Stage 2 item choice=0) while the right panel illustrates the corresponding simple effects. **(c)** Contrarily, the vmPFC and dPCC were more functionally connected during item choice than during environment choice. * denotes P<0.05. Shading represent means±s.e.m.

5.3.3 FPl signals exhibited essential properties of value comparison

Results of the whole-brain and ROI analyses consistently indicate that the FPl exclusively engaged in environment choice, while the vmPFC exclusively engaged in item choice. Next, it is crucial to affirm that these signals observed in the FPl and vmPFC possess two essential properties of the neural value comparison signal: (1) a context-dependent modulation to flexibly respond to current environment and (2) invariance to salience. To this end, closer inspection of the signals in the FPl and vmPFC was taken based on the GLM7 reported above.

5.3.3.1 Context-dependent modulation. In this experiment, there were two conditions (i.e. Linked and Unlinked Conditions) determining the ways to obtain the bonus. Particularly, in Linked Condition, the bonus was obtained if the chosen item on the same Stage 2 trial resulted in a reward. Yet, Stage 1 environment choice outcome would have an indirect influence on the bonus acquisition because it affected what items were available in Stage 2 item choice. In contrast, in Unlinked Condition, the bonus was delivered unconditionally. Behavioural results reported in **Chapter 3** have demonstrated that participants did make choices in response to the Bonus Conditions in order to obtain the bonus, in both Stage 1 and Stage 2. If the FPl and vmPFC signals possess the property of context dependence, they should show responses adapted to the Bonus Condition – the signals should change with the bonus value in Linked Condition only, but not in the Unlinked Condition.

To address this issue, GLM8 was carried out in which the DV ($DV_{environment}$ in Stage 1 and DV_{item} in Stage 2) was split into two terms: (1) a DV term that involves all choice information but the bonus (i.e. $DV_{bonus\ excluded}$) and (2) a DV term that only involves the

combination of bonus value and Bonus Condition (i.e. $DV_{bonus\ only}$). Akin to the findings of GLM7, during environment choice the FPl signal was related to the $\Delta DV_{bonus\ excluded}$ (β=0.047, t_{23}=2.967, P=0.007, BF_{10}=6.621; Fig. 5.6, left panel). Notably, the FPl also encoded the bonus value in Linked Condition (β=0.037, t_{23}=3.030, P=0.006, BF_{10}=7.506; Fig. 5.6, middle panel). In contrast, in the Unlinked Condition the bonus acquisition was independent of the choices that the participants made and the FPl was found not encode the bonus value in Unlinked Condition (β=-0.013, t_{23}=-0.975, P=0.340, BF_{10}=0.329; Fig. 5.6, right panel). These findings revealed that during environment choice, the FPl selectively integrated the choice-relevant information. On the other hand, in Stage 2 item choice, no $\Delta DV_{bonus\ excluded}$ signal (β=0.027, t_{23}=1.551, P=0.135, BF_{10}=0.613; Fig. 5.6, left panel) nor any bonus related signals ($\Delta DV_{Linked\ bonus}$: β=0.023, t_{23}=1.497, P=0.148, BF_{10}=0.572; Fig. 5.6, middle panel; $\Delta DV_{Unlinked\ bonus}$: β=-0.011, t_{23}=-0.993, P=0.331, BF_{10}=0.334; Fig. 5.6, right panel) were observed in the FPl. It reflects the FPl was unrelated to the value comparison process during item choice.

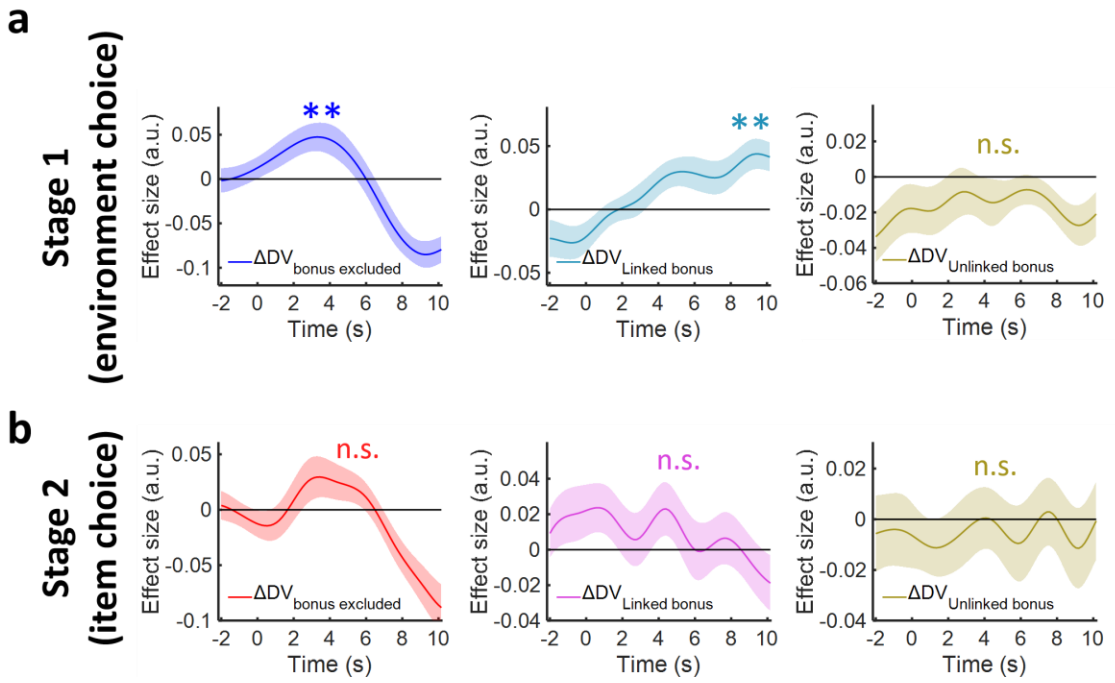


Figure 5.6. FPl encoded choice-relevant information during environment choice. The ΔDV term was split into three components – the ΔDV that excluded the bonus ($\Delta DV_{bonus\ excluded}$), the rest of the ΔDV that was contributed by the bonus in Linked Condition ($\Delta DV_{Linked\ bonus}$), and that in Unlinked Condition ($\Delta DV_{Unlinked\ bonus}$). **(a)** The FPl signal exhibited a bonus adaption in Stage 1 environment choice – the FPl signal modulated by the choice-relevant $\Delta DV_{bonus\ excluded}$ (left panel) and $\Delta DV_{Linked\ bonus}$ (middle panel), but not by the choice-irrelevant $\Delta DV_{Unlinked\ bonus}$ (right panel). **(b)** In Stage 2 item choice, no $\Delta DV_{bonus\ excluded}$ (left panel), $\Delta DV_{Linked\ bonus}$ (middle panel) nor the $\Delta DV_{Unlinked\ bonus}$ signals were found in the FPl. ** denotes $P<0.01$. Shading represent means±s.e.m.

The same analysis was applied to scrutinize the neural signals in the vmPFC. Consistent to the results of GLM7, during Stage 2 item choice, a significant $\Delta DV_{bonus\ excluded}$ signal ($\beta=0.070$, $t_{23}=3.970$, $P=6.061\times10^{-4}$, $BF_{10}=54.458$; Fig. 5.7b, left panel) was observed in the vmPFC. Surprisingly, there were no significant $\Delta DV_{bonus\ only}$ signals in Linked Condition ($P>0.05$; Fig. 5.7b, middle panel) nor in Unlinked Condition ($P>0.05$;

Fig. 5.7b, right panel). The vmPFC signal during Stage 1 environment choice was also examined. Similar to the results of GLM7, no $\Delta DV_{bonus\ excluded}$ signal was observed in the vmPFC (β=0.023, t_{23}=1.320, P=0.200, BF_{10}=0.463; Fig. 5.7a, left panel). Besides, the vmPFC did not signal any bonus information in environment choice ($\Delta DV_{Linked\ bonus}$: β=-0.020, t_{23}=-1.492, P=0.149, BF_{10}=0.569; Fig. 5.7a, middle panel; $\Delta DV_{Unlinked\ bonus}$: β=0.015, t_{23}=1.378, P=0.182, BF_{10}=0.495; Fig. 5.7a, right panel). It reflects the vmPFC was unrelated to the value comparison process between environments.

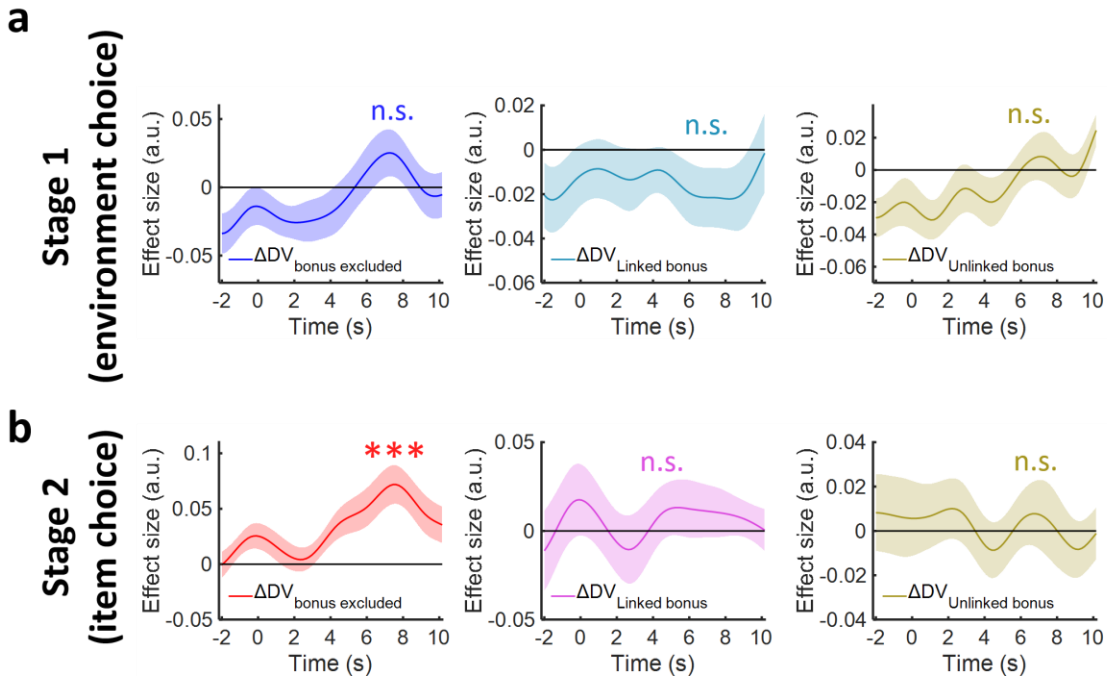


Figure 5.7. vmPFC signals under different Bonus Conditions. (a) In Stage 1, there was an absence of $\Delta DV_{bonus\ excluded}$ signal nor any $\Delta DV_{bonus\ only}$ signal, suggesting the vmPFC was unrelated to environment choice. **(b)** In Stage 2 item choice, the vmPFC showed a significant $\Delta DV_{bonus\ excluded}$ signal and but no $\Delta DV_{Linked\ bonus\ nor}$ $\Delta DV_{Unlinked\ bonus}$ signals were found. *** denotes P<0.001. Shading represent means±s.e.m.

5.3.3.2 Invariance to salience. To affirm a neural signal reflecting the value comparison process, it is particularly important to show that it is not confounded by salience. Hence, in GLM9, a regressor of salience (i.e. the sum of the options' absolute values) was included apart from ΔDV. A nuisance regressor about the sum of the options' DVs was also included to take the remaining variance by the options' DVs into account. The results revealed that the FPl still encoded the $\Delta DV_{environment}$ in spite of the inclusion of salience as a regressor (β=0.043, t_{23}=2.829, P=0.010, BF_{10}=5.051; Fig. 5.8a, left panel). Critically, the FPl activity was found independent of salience (P>0.05; Fig. 5.8a, right panel). And consistently there was an absence of ΔDV_{item} signal in the FPl during Stage 2 item choice (β=0.021, t_{23}=1.194, P=0.245, BF_{10}=0.405; Fig. 5.8b, left panel).

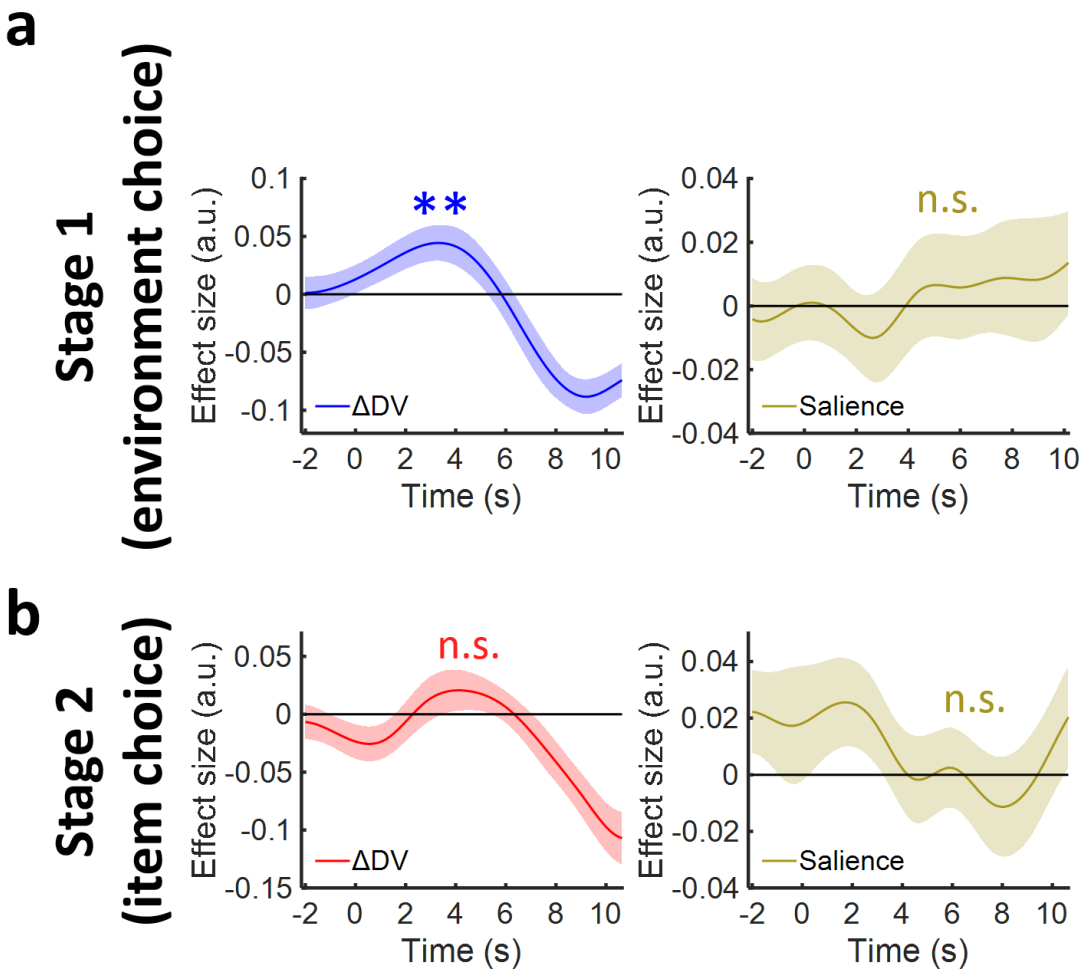


Figure 5.8. $\Delta DV_{environment}$ signal in FPl was invariant to salience. (a) ΔDV signal of the FPl in Stage 1 environment choice was orthogonal to the option salience. **(b)** No ΔDV signal was found in the FPl during Stage 2 item choice. ** denotes $P<0.01$. Shading represent means±s.e.m.

In a similar vein, when the same analysis was applied to test the vmPFC activity, a significant ΔDV_{item} signal during Stage 2 item choice was found ($\beta=0.086$, $t_{23}=4.125$, $P=4.124\times10^{-4}$, $BF_{10}=76.536$; Fig. 5.9b, left panel) and there was an absence of salience signal ($\beta=0.021$, $t_{23}=1.456$, $P=0.159$, $BF_{10}=0.544$; Fig. 5.9b, right panel). Also, there was

an absence of $\Delta DV_{environment}$ signal in the vmPFC during Stage 1 environment choice (β=0.024, t_{23}=1.261, P=0.220, BF_{10}=0.434; Fig. 5.9a, left panel).

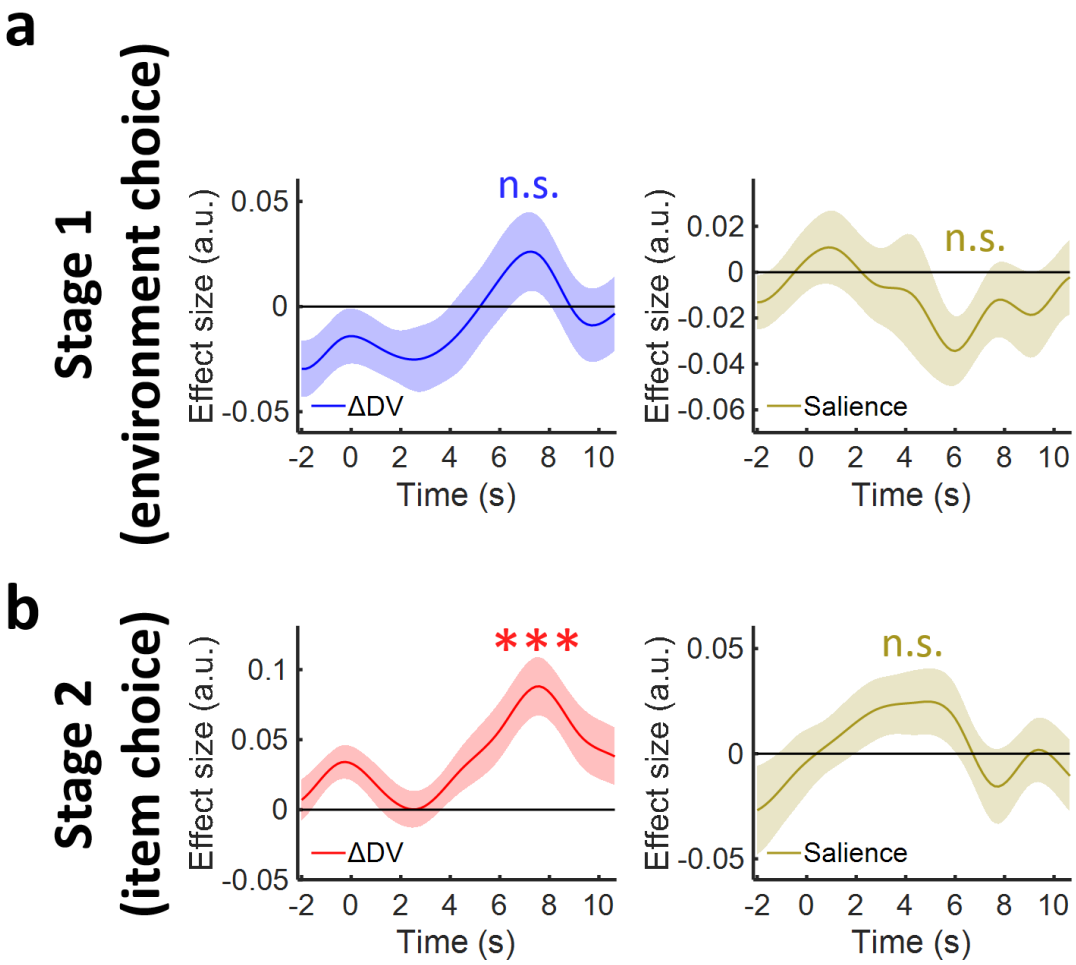


Figure 5.9. ΔDV_{item} signal in vmPFC was invariant to salience. (a) There was a absence of ΔDV signal in the vmPFC during Stage 1 environment choice. **(b)** Importantly, the vmPFC showed a significant DV signal during Stage 2 item choice and this signal was orthogonal to the option salience. *** denotes P<0.001. Shading represent means±s.e.m.

5.3.4 FPl shared similar computation to CNN

With the use of the CNN-derived DV in whole-brain analysis as well as ROI analysis, it was found that the FPl carried a univariate signal reflecting the value comparison process between environments. Nonetheless, results from the whole-brain and ROI analyses merely revealed the FPl and CNN shared a similar computational output. It remains obscure whether the FPl undergoes similar computational processes as if the CNN to encode the multiplex information embedded in the environments to guide decision-making. **Chapter 4** shows that the CNN extracts the information about statistical moments of the item distribution in the environments and integrates into concise DVs. Specifically, four feature detectors are utilized in the CNN to extract different pieces of information from the environments. Furthermore, during the course of evolution from the input environment value to the final DVs, CNN employs multiple nodes and multiple layers for value representation and value integration. To inspect whether the FPl shared similar computational mechanisms to the CNN, the relationship of the multi-voxel neural activation patterns of the FPl with the multi-nodal representations of the CNN was examined.

5.3.4.1 Multivariate coding for value representation and value integration of complex environment value. First, an RSA was performed to test whether the FPl employed a multivariate coding for representation of the multiplex environment value, like the CNN. RSA results showed that the multi-voxel activation patterns of FPl was correlated with the CNN (*rho*=0.007, signed-rank *P*=0.016; Fig. 5.10a). The same RSA was applied to other control regions (i.e. vmPFC, dPCC, V1, and CSF) and no significant correlation was yielded (*rho*s<0.004, signed-rank *P*s>0.345; Fig. 5.10a). It provides

preliminary evidence of similar mechanisms in value encoding between the FPl and CNN. Next, a series of RSAs were performed with regard to separate CNN layers to further scrutinize the computations in the FPl. It began with testing the first layer – the input layer. The input layer contains rich visual information of the environments and hence it is not surprising that a marginally significant correlation was observed with V1 (*rho*=0.003, signed-rank *P*=0.056; Fig. 5.10b, top left panel). In contrast, the FPl and other control regions were not correlated with the input layer (*rho*s<0.003, signed-rank *P*s>0.290; Fig. 5.10b, top left panel). Second, the hidden layers where value integration undergoes were inspected. Intriguingly the FPl was correlated with all hidden layers (*rho*s>0.007, signed-rank *P*s<0.035; Fig. 5.10b, top middle panel) while no other regions were similar to any of the hidden layers (*rho*s<0.004, signed-rank *P*s>0.317; Fig. 5.10b, top middle panel). Finally, the choice-predicting fully-connected layer where DVs are generated was examined. Surprisingly, this layer was similar to the dPCC (*rho*=0.002, signed-rank *P*=0.022; Fig. 5.10b, top right panel) but not the FPl (*rho*=0.001, signed-rank *P*=0.568; Fig. 5.10b, top right panel) nor other control regions (*rho*s<0.002, signed-rank *P*s>0.492; Fig. 5.10b, top right panel). Taken together, all these lines of results imply the FPl was more concerned in the value integration process whereas the dPCC was more concerned in representing the ultimate decision values for guiding decision.

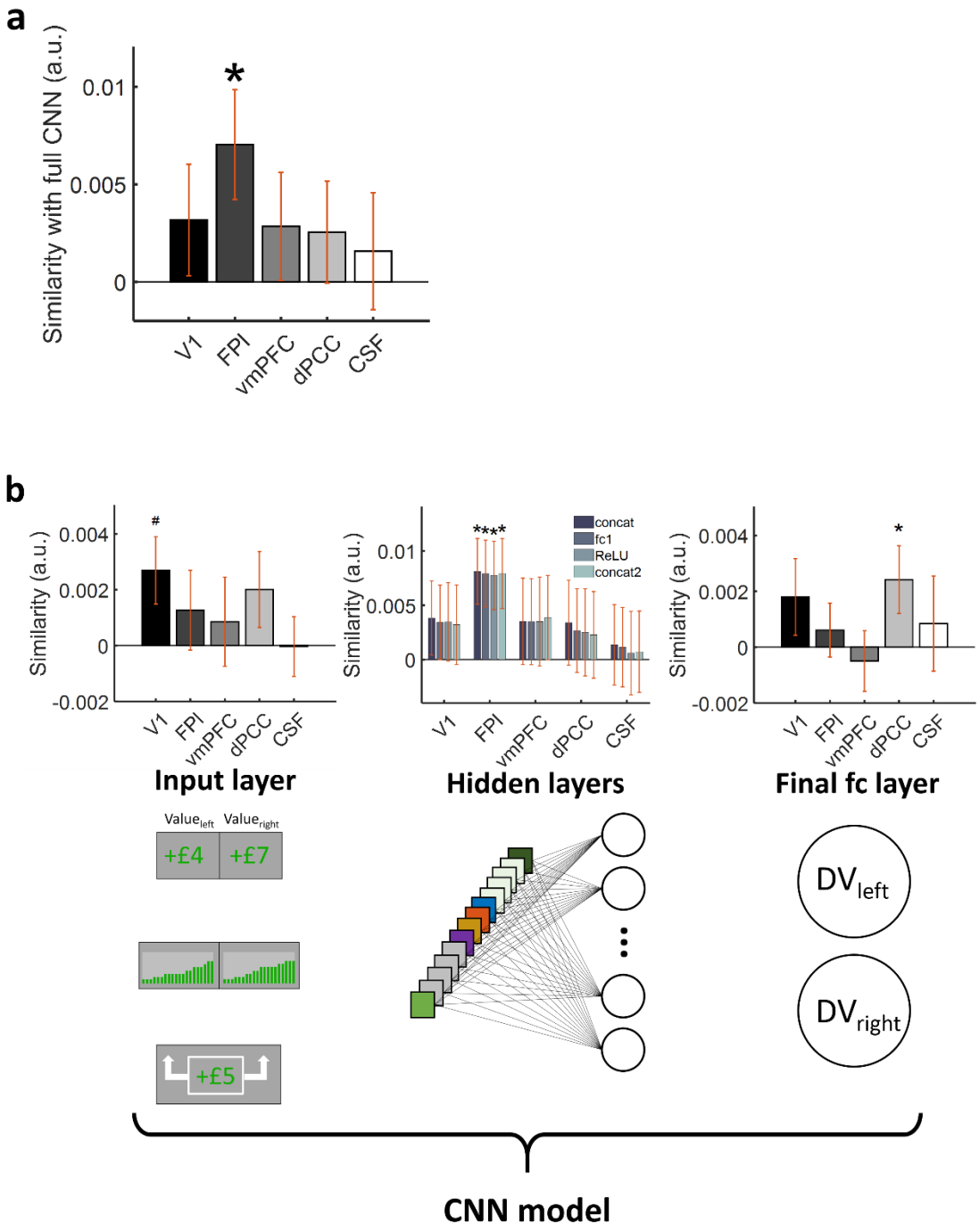


Figure 5.10. Multivariate coding of environment value in the FPl. (a) An RSA showed that the full CNN (consisting of multiple layers and multiple nodes) was similar to the multi-voxel activation patterns of the FPl, but not other control regions. **(b)** A series of RSAs were performed with respect to different layers of the CNN (Details please refer to Section *Convolutional neural network (CNN)* of **Chapter 4**). V1=

primary visual cortex; FPl=lateral frontopolar cortex; vmPFC=ventromedial prefrontal cortex; dPCC=dorsal posterior cingulate cortex; CSF=cerebrospinal fluid.

5.3.4.2 Parallel feature detection processes during encoding of complex environment value.

It is worth noting that the CNN involves four parallel feature detectors to extract information from the environments. Given the multiplex information embedded in the environments, it is plausible that this parallel manner is necessary for efficient value encoding. To further affirm this is indeed the case, the FPl activity was additionally tested with CNN variants that involve single feature detector as well as variants involve different numbers of feature detectors. RSA was performed by focusing on the feature maps because they are the first representation after feature detection. It was found that the FPl was similar to all CNN variants (*rho*s>0.005, signed-rank *P*s<0.026; Fig. 5.11), except the variant with single feature detector only (*rho*=-0.0002, signed-rank *P*=0.977; Fig. 5.11). It suggests the FPl employed multiple parallel processes to encode the multiplex information of the environments. Besides, the FPl was the most similar to the 4-feature-detecoter CNN among all the variants (*rho*=0.008) although a Kruskal-Wallis test did not reveal significant differences among the variants ($H(9)=4.254$, $P=0.894$).

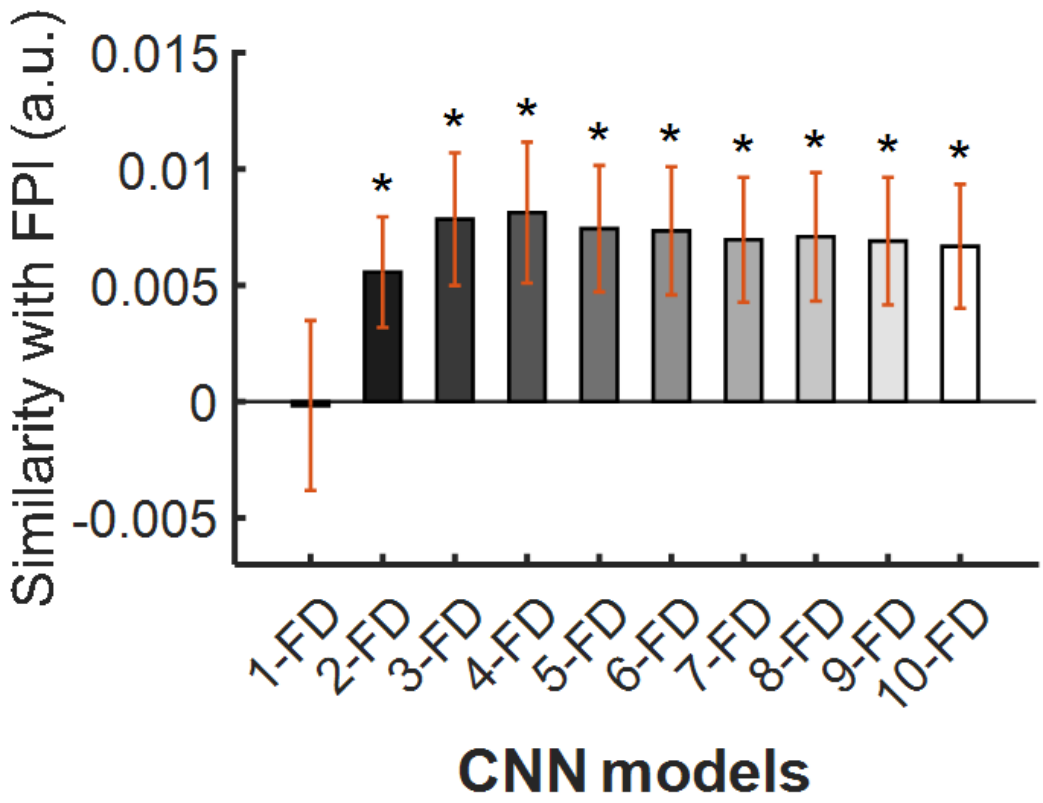


Figure 5.11. RSA between FPl and CNN variants that involve different number of feature detectors. FPl voxel activation patterns were correlated with the feature maps of all CNN variants, except the variant with single feature detector (i.e. 1-FD CNN). * denotes $P<0.05$.

5.4 Discussion

The vmPFC is widely accepted as the key neural substrate underlying decision-making (Boorman et al., 2009, 2013; Chau et al., 2014; Hunt et al., 2012; Lopez-Persem et al., 2016). However, it was established on the basis of a large body of previous findings about item choice. This study, by adopting a two-stage decision-making task isolating environment choice and item choice, revealed that these two kinds of decision

involved distinct underlying neural mechanisms: a double dissociation between the FPl and vmPFC.

To affirm the roles of the FPl and vmPFC in decision-making, different signatures of the value comparison process were scrutinized. First, for a neural substrate underlying decision-making, it should signal the difference in value between the options being compared. Second, value comparison signal should purely reflect the difference in value and not confounded by salience. Third, it has to adapt to the changing environment and flexibly respond in a productive way according to the context. All these signatures of value comparison signal have been widely observed in the vmPFC (e.g. Bartra et al., 2013; Chau et al., 2014; Hunt et al., 2012; Litt et al., 2011; Plassmann et al., 2008; Zhang et al., 2017) during item choice and consistent findings except the context-dependent modulation were reproduced in this study. In fact, the context-dependent modulation of the vmPFC is less conclusive because it is primarily observed when contextual changes occur externally to the ongoing task (Abitbol et al., 2015; Harvey et al., 2010; Losecaat Vermeer et al., 2014; Plassmann et al., 2008). By contrast, in my study contextual information (i.e. bonus information) had to be integrated deliberately within the ongoing task in order to make responses. Further investigation is needed to discern the context dependent modulation in the vmPFC under difference conditions. Nonetheless, crucially, the vmPFC was not identified, but the FPl, during decision between complex environments. Closer inspection of the FPl signal revealed that it exhibited all the abovementioned signatures of the value comparison process. First, the FPl activity correlated with the DV difference between the environments. Moreover, the FPl signal was not confounded by salience. It also demonstrated a selective encoding of choice-

relevant information (i.e. reward probability and bonus value) depending on the Bonus Conditions. The presence of these properties in the FPl signal implicates its role underlying decision-making, importantly it is exclusive in environment choice.

Environment choice requires integrating information from multiple items within each environment and envisaging the potential outcomes that the items can lead to. Previous findings have implicated the role of FPl in parallel processing of massive information such as maintaining decision strategies and multiple goals simultaneously (Mansouri et al., 2017; Schuck et al., 2015), and the encoding of multiple goals could be carried out in a multivariate manner (Haynes et al., 2007). In lines with the previous findings, here I showed that the FPl shared similar computational mechanisms to the CNN that multiple parallel feature detectors were employed to encode the complex information embedded in the environments. Furthermore, the multi-voxel activation patterns of the FPl were correlated with the CNN. In particular, the FPl activation patterns were correlated with all hidden layers of the CNN, which are the layers where value integration occurs, suggesting the FPl possessed a multivariate coding for value representation and value integration. Crucially, this finding of multivariate coding also facilitates the functional specialization between FPl and dPCC. By the conventional approach of focusing on the univariate DV difference signals, it is unable to tease out the functional roles of FPl and dPCC in environment choice due to their coactivation. With the use of RSA, a multivariate analysis, I demonstrated that the FPl was more tied to value integration whereas the dPCC was more tied to representation of the final decision value. Taken the results of the univariate and multivariate analyses together, FPl is an important neural substrate underlying environment choice – it is involved in encoding

and integrating complex environment value as well as the value comparison process to guide decisions.

Chapter 6 General discussion

We face numerous types of decisions in our daily lives, ranging from trivial ones to influential ones. Traditional economics theories suggest that decisions of all kinds are made according to the options' utility, which is a ubiquitous scale that reflects the subjective preference for any given option. In decision neuroscience, early studies have attempted to search in the brain and identify a specific region that encodes such ubiquitous utility. However, recent evidence suggests that the brain does not rely on a single brain region for value encoding. Instead, multiple regions subserve value encoding in their specialized way. To systematically investigate whether a unitary neural substrate is sufficient to work with decisions of all kinds or multiple neural substrates are required for particular decisions, in this thesis I introduced an approach of classification. Options can be classified into two classes: (1) item and (2) environment. Under this classification, three candidate frameworks are proposed to test the functional specialization in decision-making, namely Neural Common Currency Framework, Option Homogeneity Framework, and Decision Homogeneity Framework. The Neural Common Currency Framework presumes a neural substrate is compatible with decisions of all kinds. On the other hand, both the Option Homogeneity Framework and Decision Homogeneity Framework assumes within-class decisions (i.e. item versus item, environment versus environment) and between-class decisions (i.e. items versus environments) involve separate neural substrates. However, the Decision Homogeneity Framework assumes a unitary neural substrate copes with all within-class decisions whereas the Option Homogeneity Framework posits within-class decisions involve exclusive neural substrates for each class of option.

The Neural Common Currency Framework is favoured by the generalizability of the vmPFC on value encoding of a wide variety of items. A considerable body of findings in the existing literature has demonstrated that the vmPFC exhibited a "neural common currency" signal – items from disparate categories whose values were commonly represented in the vmPFC (Chib et al., 2009; Lebreton et al., 2009; McNamee et al., 2013). Besides, the vmPFC signalled the difference in value when decision was made between dissimilar items (Levy & Glimcher, 2011; Lopez-Persem et al., 2016). All these lines of findings make the vmPFC plausible to encode the value of environments and making decisions involving environments. Nevertheless, the Neural Common Currency Framework has been refuted by the findings from studies of decision between items and environments or environment-leaving that emerged in the last decade. Environment value was found represented by the dACC but not the vmPFC (Hayden et al., 2011; Kaiser et al., 2021; Kolling et al., 2012, 2018; Wittmann et al., 2016). Making choices between items and environments was also shown to be driven by the dACC (Kolling et al., 2012). These lines of findings have precluded the possibility of the Neural Common Currency Framework and the next step is to discern the Option Homogeneity Framework and Decision Homogeneity Framework.

Despite the absence of the vmPFC signal during decision between items and environments, it is worth noting that decision between items and environments involves comparison between options of different classes. In contrast, both item choice and environment choice involve comparison between options of the same class. According to the Decision Homogeneity Framework, it is feasible that the vmPFC is not exclusive for item choice but within-class decision. Otherwise, as the Option Homogeneity

Framework predicts, the vmPFC will merely show involvement in item choice. To discern the Option Homogeneity Framework and Decision Homogeneity Framework, in **Chapters 2-5,** I conducted different studies to unveil the underlying mechanisms of environment choice and made direct contrast with item choice. The first study was a pilot behavioural study of the fMRI study. It involved a binary decision-making task which required participants to choose between environments. The results showed that participants were capable of integrating complex information in the environments to guide decisions. In particular, participants made environment choices with consideration of the statistical moments (i.e. mean, variance) of the item distributions within the environments, which sheds light on the evaluation of environments. In **Chapter 5**, I reported an fMRI study to test the neural mechanisms underlying environment choice. Participants performed a two-stage decision-making task which included both environment choice and item choice while underwent fMRI scanning. In spite of the similarities shared by environment choice and item choice, surprisingly a double dissociation was revealed. I showed that the vmPFC signal emerged during item choice, consistent to the well-documented findings in the existing literature. Crucially, there was an absence of the vmPFC signal during environment choice. On the contrary, a FPl signal was identified during environment choice but it was absence when people choosing between items. This double dissociation provides direct evidence for the Option Homogeneity Framework and refutes the Decision Homogeneity Framework.

One distinctive feature of environment choice is it necessitates a prospective thinking to envisage the interdependent relation between an environment choice, the consequential items encountered, and the final payoff. Information is maintained and decays across

time in neurons dynamically (Runyan et al., 2017). The neuronal timescales (i.e. the time that information can be sustained) of the brain regions are thus closely tied to their functionalities. A neural substrate underlying environment choice must possess a longer intrinsic timescale to maintain loads of information for such computation. By contrast, a choice of an item directly leads to the final payoff and a neural substrate with shorter intrinsic timescale is sufficient for item choice. Single-unit recordings from rhesus macaques did show that the FPl had longer timescale than the vmPFC (Fascianelli et al., 2019; Maisson et al., 2021). It on the other hand provides evidence of the functional dissociation between the FPl and vmPFC in environment choice and item choice.

Given the existing literature largely focuses on item choice and decision between items and environments, I complemented the missing piece of understanding on environment choice. I showed that environment choice was subserved by the FPl, a brain region that has been implicated to prospective processing. For instance, the FPl was found to carry a representation of internal task state during tracing the current position in a succession of events (Desrochers et al., 2015). It provides evidence of the capability of the FPl in envisioning the consequence of an environment choice. The role of the FPl in prospective processing can also be revealed from studies on explore-exploit dilemma. Environment choice and exploration indeed are similar in the way that both of them are executed to affect future events to facilitate acquisition of the final payoff. Several lines of findings have shown that the FPl signal is pertaining to exploration of options which can lead to potentially larger reward in the future (Beharelle et al., 2015; Daw et al., 2006; Zajkowski et al., 2017). Non-invasive brain stimulation over FPl also revealed that excitation or disruption of the FPl resulted in more frequent and less frequent exploration

respectively (Beharelle et al., 2015; Zajkowski et al., 2017). All these lines of findings provide supporting evidence of the capability of the FPl in environment choice.

Despite the findings of the FPl in my study, intriguingly the dACC was identified in previous studies on environment-related decisions. Those studies generally focused on stay-switch choices in which participants decided to stay and exploit current items or to search for better items in an alternative environment. The dACC activity was found pertaining to the value of the items' average or the most rewarding item in the alternative environment (Fouragnan et al., 2019; Hayden et al., 2011; Kolling et al., 2012, 2018; Sarafyazd & Jazayeri, 2019). In fact, the dACC is a region having direct connection with the FPl (Carmichael & Price, 1996; Saleem et al., 2014). The dACC and FPl are implicated in a variety of functional similarities, such as integration of task-relevant information, hierarchical reasoning, task state monitoring, and envisioning the choice prospect (Heilbronner & Hayden, 2016; Mansouri et al., 2017). Nevertheless, direct contrast in the functional distinction between these two closely connected regions is seldom made. My studies focused on choices between simultaneously presented environments and identified the FPl, but not the dACC. It suggests that, despite the highly overlapping in the functions of the FPl and dACC, the FPl is more connected to parallel processing demands in particular types of decision, such as environment choice.

I also demonstrated that the use of deep learning neural networks, such as the CNN, can predict value-based decision-making behaviour with better performance than conventional computational models. One limitation of the conventional models is that they greatly rely on a priori assumptions such that it might not be able to capture the actual underlying processes. For example, the expected value theory presumes people

compute the expected value (i.e. reward magnitude multiplied by reward probability) for each option and make choices by comparing the expected values (Tobler & Weber, 2013). Any factors which are not explicitly specified (e.g. attentional capture derived from the option salience) cannot be described. On the contrary, deep learning neural networks offers a model-free approach to model behaviour with fewer a priori assumptions. For example, I demonstrated information about statistical moments (i.e. mean, variance, skewness) within the environments are extracted by the CNN even though the CNN does not receive information about these parameters directly. Besides, the CNN involves multiple parallel feature detectors to extract the multiplex information from the environments. Although the CNN was trained purely by the behavioural data without any knowledge about the brain, interestingly, its resulting multi-nodal representations resemble the multi-voxel activation patterns of the FPl. It implies the FPl similarly employed an approach of parallel processing to encode the value of the environments. This property is difficult to reveal using the conventional computational models, such as expected value theory (Tobler & Weber, 2013). Furthermore, univariate analysis (e.g. whole-brain analysis) is a typical approach to investigate the neural mechanisms underlying decision-making. It is not easy to discern the roles of different brain regions when they show coactivation. The multi-nodal representation structure of the CNN allows examination of multivariate neural signals, by which inspection of the functional specialization is facilitated. In addition to my univariate analysis findings about the functional dissociation between FPl and vmPFC, I found that both the FPl and dPCC exhibited correlated activity with the environment value. With the use of multivariate analysis (i.e. RSA), the roles of the FPl and dPCC in environment choice

were discerned – the FPl is more tied to value integration during environment choice while the dPCC is more tied to representation of ultimate decision value. Taken together, these findings suggest the use of deep learning neural networks provides more biologically plausible account of the underlying mechanisms because it can capture the implicit cognitive processes involved during decision-making.

6.1 Conclusions and Suggestions for Future Research

This thesis reveals that item choice and environment choice involve dissociable neural mechanisms. The long-held role of the vmPFC in valuation and value comparison are shown exclusive to items only but not as universal as the neural common currency hypothesis (Montague & Berns, 2002) predicted. In contrast, the valuation and value comparison of complex environments require the FPl, a brain region which is closely tied to prospective processing. This functional distinction might stem from the hierarchy in intrinsic timescales of the vmPFC and FPl.

On the other hand, environment choice is of great importance because in daily lives, influential decisions often involve choosing between environments (e.g. choose between two cities which to migrate to by comparing the available jobs they potentially provide). Despite its importance, environment choice receives little attention in the existing literature. This thesis has scrutinized the neural mechanisms underlying environment choice and shed light on the role of the FPl. To further provide causal evidence, future studies about lesion or inhibition by brain stimulation over the FPl are warranted. If the FPl is indeed underlying environment choice, FPl lesions should show impaired decision-making behaviour. Alternatively, transcranial magnetic stimulation (TMS), a non-

invasive brain stimulation, provides a means to the alter the neural activity (temporal enhancement or disruption) by changing the cortical excitability (Pascual-Leone et al., 1999). Causal role of the FPl can be established by TMS over the FPl during environment choice.

References

Abitbol, R., Lebreton, M., Hollard, G., Richmond, B. J., Bouret, S., & Pessiglione, M. (2015). Neural mechanisms underlying contextual dependency of subjective values: Converging evidence from monkeys and humans. *Journal of Neuroscience*, *35*(5), 2308–2320. https://doi.org/10.1523/JNEUROSCI.1878-14.2015

Barbas, H., & Pandya, D. N. (1989). Architecture and intrinsic connections of the prefrontal cortex in the rhesus monkey. *Journal of Comparative Neurology*, *286*(3), 353–375. https://doi.org/10.1002/cne.902860306

Bartra, O., McGuire, J. T., & Kable, J. W. (2013). The valuation system: A coordinate-based meta-analysis of BOLD fMRI experiments examining neural correlates of subjective value. *NeuroImage*, *76*, 412–427. https://doi.org/10.1016/j.neuroimage.2013.02.063

Beckmann, C., Jenkinson, M., & Smith, S. (2003). General multilevel linear modeling for group analysis in FMRI. *NeuroImage*, *20*(2), 1052–1063. https://doi.org/10.1016/S1053-8119(03)00435-X

Beharelle, A. R., Polanía, R., Hare, T. A., & Ruff, C. C. (2015). Transcranial stimulation over frontopolar cortex elucidates the choice attributes and neural mechanisms used to resolve exploration–exploitation trade-offs. *Journal of Neuroscience*, *35*(43), 14544–14556. https://doi.org/10.1523/JNEUROSCI.2322-15.2015

Bernoulli, D. (1954). Exposition of a New Theory on the Measurement of Risk. *Econometrica*, *22*(1), 23–36.

Bongioanni, A., Folloni, D., Verhagen, L., Sallet, J., Klein-Flügge, M. C., & Rushworth, M. F. S. (2021). Activation and disruption of a neural mechanism for novel choice in monkeys. *Nature*, *591*(October 2019). https://doi.org/10.1038/s41586-020-03115-5

Boorman, E. D., Behrens, T. E. J., Woolrich, M. W., & Rushworth, M. F. S. (2009). How Green Is the Grass on the Other Side? Frontopolar Cortex and the Evidence in Favor of Alternative Courses of Action. *Neuron*, *62*(5), 733–743. https://doi.org/10.1016/j.neuron.2009.05.014

Boorman, E. D., Rushworth, M. F., & Behrens, T. E. (2013). Ventromedial prefrontal and anterior cingulate cortex adopt choice and default reference frames during sequential multi-alternative choice. *The Journal of Neuroscience*, *33*(6), 2242–2253. https://doi.org/https://doi.org/10.1523/JNEUROSCI.3022-12.2013

Brodmann, K. (1909). *Vergleichende Lokalisationslehre der Großhirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues.* Verlag von Johann Ambrosius Barth.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network: Anatomy, function, and relevance to disease. *Annals of the New York Academy of Sciences*, *1124*, 1–38. https://doi.org/10.1196/annals.1440.011

Camille, N., Griffiths, C. A., Vo, K., Fellows, L. K., & Kable, J. W. (2011). Ventromedial frontal lobe damage disrupts value maximization in humans. *Journal of Neuroscience*, *31*(20), 7527–7532. https://doi.org/10.1523/JNEUROSCI.6527-10.2011

Carmichael, S. T., & Price, J. L. (1994). Architectonic subdivision of the orbital and medial prefrontal cortex in the macaque monkey. *Journal of Comparative Neurology*, *346*(3), 366–402. https://doi.org/10.1002/cne.903460305

Carmichael, S. T., & Price, J. L. (1995). Limbic connections of the orbital and medial prefrontal cortex in macaque monkeys. *Journal of Comparative Neurology*, *363*(4), 615–641. https://doi.org/10.1002/cne.903630408

Carmichael, S. T., & Price, J. L. (1996). Connectional networks within the orbital and medial prefrontal cortex of macaque monkeys. *Journal of Comparative Neurology*, *371*(2), 179–207. https://doi.org/10.1002/(sici)1096-9861(19960722)371:2<179::aid-cne1>3.0.co;2-%23

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. *Cerebral Cortex*, *10*(3), 220–242. https://doi.org/10.1093/cercor/10.3.220

Chau, B. K. H., Kolling, N., Hunt, L. T., Walton, M. E., & Rushworth, M. F. S. (2014). A neural mechanism underlying failure of optimal choice with multiple alternatives. *Nature Neuroscience*, *17*(3), 463–470. https://doi.org/10.1038/nn.3649

Chau, B. K. H., Law, C.-K., Lopez-Persem, A., Klein-Flügge, M. C., & Rushworth, M. F. (2020). Consistent patterns of distractor effects during decision making. *ELife*, *9*, 1–36. https://doi.org/10.7554/eLife.53850

Chau, B. K. H., Sallet, J., Papageorgiou, G. K., Noonan, M. A. P., Bell, A. H., Walton, M. E., & Rushworth, M. F. S. (2015). Contrasting Roles for Orbitofrontal Cortex and Amygdala in Credit Assignment and Learning in Macaques. *Neuron*, *87*(5), 1106–1118. https://doi.org/10.1016/j.neuron.2015.08.018

Chib, V. S., Rangel, A., Shimojo, S., & O'Doherty, J. P. (2009). Evidence for a common representation of decision values for dissimilar goods in human ventromedial prefrontal cortex. *Journal of Neuroscience*, *29*(39), 12315–12320. https://doi.org/10.1523/JNEUROSCI.2575-09.2009

Clithero, J. A., & Rangel, A. (2014). Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience*, *9*(9), 1289–1302. https://doi.org/10.1093/scan/nst106

Cohen, M. R., & Maunsell, J. H. R. (2009). Attention improves performance primarily by reducing interneuronal correlations. *Nature Neuroscience*, *12*(12), 1594–1600. https://doi.org/10.1038/nn.2439

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. J. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*(6292), 1464–1468. https://doi.org/10.1126/science.aaf0941

Cox, K. M., & Kable, J. W. (2014). Bold subjective value signals exhibit robust range adaptation. *Journal of Neuroscience*, *34*(49), 16533–16543. https://doi.org/10.1523/JNEUROSCI.3927-14.2014

Daunizeau, J., Adam, V., & Rigoux, L. (2014). VBA: A Probabilistic Treatment of Nonlinear Models for Neurobiological and Behavioural Data. *PLoS Computational Biology*, *10*(1). https://doi.org/10.1371/journal.pcbi.1003441

Daw, N. D., Gershman, S. J., Seymour, B., Dayan, P., & Dolan, R. J. (2011). Model-based influences on humans' choices and striatal prediction errors. *Neuron*, *69*(6), 1204–1215. https://doi.org/10.1016/j.neuron.2011.02.027

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879. https://doi.org/10.1038/nature04766

Desrochers, T. M., Chatham, C. H., & Badre, D. (2015). The Necessity of Rostrolateral Prefrontal Cortex for Higher-Level Sequential Behavior. *Neuron*, *87*(6), 1357–1368. https://doi.org/10.1016/j.neuron.2015.08.026

Elliott, R., Agnew, Z., & Deakin, J. F. W. (2008). Medial orbitofrontal cortex codes relative rather than absolute value of financial rewards in humans. *European Journal of Neuroscience*, *27*(9), 2213–2218. https://doi.org/10.1111/j.1460-9568.2008.06202.x

Fascianelli, V., Tsujimoto, S., Marcos, E., & Genovesio, A. (2019). Autocorrelation Structure in the Macaque Dorsolateral, but not Orbital or Polar, Prefrontal Cortex Predicts Response-Coding Strength in a Visually Cued Strategy Task. *Cerebral Cortex*, *29*(1), 230–241. https://doi.org/10.1093/cercor/bhx321

Fellows, L. K., & Farah, M. J. (2007). The role of ventromedial prefrontal cortex in decision making: Judgment under uncertainty or judgment per se? *Cerebral Cortex*, *17*(11), 2669–2674. https://doi.org/10.1093/cercor/bhl176

FitzGerald, T. H. B., Seymour, B., & Dolan, R. J. (2009). The role of human orbitofrontal cortex in value comparison for incommensurable objects. *Journal of Neuroscience*, *29*(26), 8388–8395. https://doi.org/10.1523/JNEUROSCI.0717-09.2009

Fouragnan, E. F., Chau, B. K. H., Folloni, D., Kolling, N., Verhagen, L., Klein-Flügge, M., Tankelevitch, L., Papageorgiou, G. K., Aubry, J. F., Sallet, J., & Rushworth, M.

F. S. (2019). The macaque anterior cingulate cortex translates counterfactual choice value into actual behavioral change. *Nature Neuroscience*, *22*(5), 797–808. https://doi.org/10.1038/s41593-019-0375-6

Fries, P. (2015). Rhythms for Cognition: Communication through Coherence. *Neuron*, *88*(1), 220–235. https://doi.org/10.1016/j.neuron.2015.09.034

Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). *Psychophysiological and Modulatory Interactions in Neuroimaging*.

Frömer, R., Dean Wolf, C. K., & Shenhav, A. (2019). Goal congruency dominates reward value in accounting for behavioral and neural correlates of value-based decision-making. *Nature Communications*, *10*(1), 1–11. https://doi.org/10.1038/s41467-019-12931-x

Fyhn, M., Hafting, T., Treves, A., Moser, M. B., & Moser, E. I. (2007). Hippocampal remapping and grid realignment in entorhinal cortex. *Nature*, *446*(7132), 190–194. https://doi.org/10.1038/nature05601

Gauthier, I., Skudlarski, P., Gore, J. C., & Anderson, A. W. (2000). Expertise for cars and birds recruits brain areas involved in face recognition. *Nature Neuroscience*, *3*(2), 191–197. https://doi.org/10.1038/72140

Gauthier, I., Tarr, M. J., Anderson, A. W., Skudlarski, P., & Gore, J. C. (1999). Activation of the middle fusiform 'face area' increases with expertise in recognizing novel objects. Nature Neuroscience 2, 6 (. *June*, *2*(6), 568–573.

Haber, S. N., Kunishio, K., Mizobuchi, M., & Lynd-Balta, E. (1995). The orbital and medial prefrontal circuit through the primate basal ganglia. *Journal of Neuroscience*, *15*(7 I), 4851–4867. https://doi.org/10.1523/jneurosci.15-07-04851.1995

Hafting, T., Fyhn, M., Molden, S., Moser, M. B., & Moser, E. I. (2005). Microstructure of a spatial map in the entorhinal cortex. *Nature*, *436*(7052), 801–806. https://doi.org/10.1038/nature03721

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System. *Science*, *324*.

Harvey, A. H., Kirk, U., Denfield, G. H., & Montague, P. R. (2010). Monetary favors and their influence on neural responses and revealed preference. *Journal of Neuroscience*, *30*(28), 9597–9602. https://doi.org/10.1523/JNEUROSCI.1086-10.2010

Hayden, B. Y., Pearson, J. M., & Platt, M. L. (2011). Neuronal basis of sequential foraging decisions in a patchy environment. *Nature Neuroscience*, *14*(7), 933–939. https://doi.org/10.1038/nn.2856

Haynes, J. D., Sakai, K., Rees, G., Gilbert, S., Frith, C., & Passingham, R. E. (2007). Reading Hidden Intentions in the Human Brain. *Current Biology*, *17*(4), 323–328. https://doi.org/10.1016/j.cub.2006.11.072

Heilbronner, S. R., & Hayden, B. Y. (2016). Dorsal Anterior Cingulate Cortex: A Bottom-Up View. *Annual Review of Neuroscience*, *39*(1), 149–170. https://doi.org/10.1146/annurev-neuro-070815-013952

Hsu, M., Krajbich, I., Zhao, C., & Camerer, C. F. (2009). Neural response to reward anticipation under risk is nonlinear in probabilities. *Journal of Neuroscience*, *29*(7), 2231–2237. https://doi.org/10.1523/JNEUROSCI.5296-08.2009

Hunt, L. T., Kolling, N., Soltani, A., Woolrich, M. W., Rushworth, M. F. S., & Behrens, T. E. J. (2012). Mechanisms underlying cortical activity during value-guided choice. *Nature Neuroscience*, *15*(3), 470–476. https://doi.org/10.1038/nn.3017

Jenkinson, M. (2003). Fast, automated, N-dimensional phase-unwrapping algorithm. *Magnetic Resonance in Medicine*, *49*(1), 193–197. https://doi.org/10.1002/mrm.10354

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage*, *17*(2), 825–841. https://doi.org/10.1016/S1053-8119(02)91132-8

Jenkinson, M., & Smith, S. (2001). A global optimisation method for robust affine registration of brain images. *Medical Image Analysis*, *5*(2), 143–156. https://doi.org/10.1016/S1361-8415(01)00036-6

Jocham, G., Hunt, L. T., Near, J., & Behrens, T. E. J. (2012). A mechanism for value-guided choice based on the excitation-inhibition balance in prefrontal cortex. *Nature Neuroscience*, *15*(7), 960–961. https://doi.org/10.1038/nn.3140

Juechems, K., Balaguer, J., Ruz, M., & Summerfield, C. (2017). Ventromedial Prefrontal Cortex Encodes a Latent Estimate of Cumulative Reward. *Neuron*, *93*(3), 705-714.e4. https://doi.org/10.1016/j.neuron.2016.12.038

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience*, *10*(12), 1625–1633. https://doi.org/10.1038/nn2007.TheThe neural correlates of subjective value during intertemporal choice

Kahneman, D., & Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk. *Econometrica*, *47*(2), 263–291.

Kaiser, L. F., Gruendler, T. O. J., Speck, O., Luettgau, L., & Jocham, G. (2021). Dissociable roles of cortical excitation-inhibition balance during patch-leaving versus value-guided decisions. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-020-20875-w

Kanwisher, N., & Yovel, G. (2006). The fusiform face area: A cortical region specialized for the perception of faces. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *361*(1476), 2109–2128. https://doi.org/10.1098/rstb.2006.1934

Kiarashinejad, Y., Abdollahramezani, S., & Adibi, A. (2020). Deep learning approach based on dimensionality reduction for designing electromagnetic nanostructures. *Npj Computational Materials*, *6*(1), 1–12. https://doi.org/10.1038/s41524-020-0276-y

Kim, H., Shimojo, S., & O'Doherty, J. P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biology*, *4*(8), 1453–1461. https://doi.org/10.1371/journal.pbio.0040233

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., & Glover, G. (2005). Distributed neural representation of expected value. *Journal of Neuroscience*, *25*(19), 4806–4812. https://doi.org/10.1523/JNEUROSCI.0642-05.2005

Kolling, N., Behrens, T. E. J., Mars, R. B., & Rushworth, M. F. S. (2012). Neural Mechanisms of Foraging. *Science*, *336*(6077), 95–98. https://doi.org/10.1126/science.1216930.Neural

Kolling, N., Scholl, J., Chekroud, A., Trier, H. A., & Rushworth, M. F. S. (2018). Prospection, Perseverance, and Insight in Sequential Behavior. *Neuron*, *99*(5), 1069-1082.e7. https://doi.org/10.1016/j.neuron.2018.08.018

Kolling, N., Wittmann, M., & Rushworth, M. F. S. (2014). Multiple neural mechanisms of decision making and their competition under changing risk pressure. *Neuron*, *81*(5), 1190–1202. https://doi.org/10.1016/j.neuron.2014.01.033

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*(10), 1292–1298. https://doi.org/10.1038/nn.2635

Krajbich, I., & Rangel, A. (2011). Multialternative drift-diffusion model predicts the relationship between visual fixations and choice in value-based decisions. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(33), 13852–13857. https://doi.org/10.1073/pnas.1101328108

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*(NOV), 1–28. https://doi.org/10.3389/neuro.06.004.2008

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems*, *25*, 1106–1114. https://doi.org/10.1145/3383972.3383975

Lebreton, M., Jorge, S., Michel, V., Thirion, B., & Pessiglione, M. (2009). An Automatic Valuation System in the Human Brain: Evidence from Functional Neuroimaging. *Neuron*, *64*(3), 431–439. https://doi.org/10.1016/j.neuron.2009.09.040

Leutgeb, S., Leutgeb, J. K., Barnes, C. A., Moser, E. I., McNaughton, B. L., & Moser, M. B. (2005). Neuroscience: Independent codes for spatial and episodic memory in hippocampal neuronal ensembles. *Science*, *309*(5734), 619–623. https://doi.org/10.1126/science.1114037

Levy, & Glimcher, P. W. (2011). Comparing Apples and Oranges : Using Reward-Specific and Reward-General Subjective Value Representation in the Brain. *Journal of Neuroscience*, *31*(41), 14693–14707. https://doi.org/10.1523/JNEUROSCI.2218-11.2011

Levy, I., Lazzaro, S. C., Rutledge, R. B., & Glimcher, P. W. (2011). Choice from Non-Choice: Predicting Consumer Preferences from Blood Oxygenation Level-Dependent Signals Obtained during Passive Viewing. *Journal of Neuroscience*, *31*(1), 118–125. https://doi.org/10.1523/JNEUROSCI.3214-10.2011

Lim, S.-L., O'Doherty, J. P., & Rangel, A. (2011). The decision value computations in the vmPFC and striatum use a relative value code that is guided by visual attention. *The Journal of Neuroscience*, *31*(37), 13214–13223. https://doi.org/10.1523/JNEUROSCI.1246-11.2011

Lindsay, G. W. (2020). Convolutional Neural Networks as a Model of the Visual System: Past, Present, and Future. *Journal of Cognitive Neuroscience*, 1–15. https://doi.org/10.1162/jocn_a_01544

Litt, A., Plassmann, H., Shiv, B., & Rangel, A. (2011). Dissociating valuation and saliency signals during decision-making. *Cerebral Cortex*, *21*(1), 95–102. https://doi.org/10.1093/cercor/bhq065

Li, V., Herce Castañón, S., Solomon, J., Vandormael, H., & Summerfield, C. (2017). Robust averaging protects decisions from noise in neural computations. *PLOS Computational Biology*, *13*(8), e1005723. https://doi.org/https://doi.org/10.1371/journal.pcbi.1005723

Loewenstein, G., & Thaler, R. H. (1989). Anomalies: Intertemporal Choice. *Journal of Economic Perspectives*, *3*(4), 181–193. https://doi.org/10.1257/jep.3.4.181

Lopez-Persem, A., Bastin, J., Petton, M., Abitbol, R., Lehongre, K., Adam, C., Navarro, V., Rheims, S., Kahane, P., Domenech, P., & Pessiglione, M. (2020). Four core properties of the human brain valuation system demonstrated in intracranial signals. *Nature Neuroscience*, *23*(5), 664–675. https://doi.org/10.1038/s41593-020-0615-9

Lopez-Persem, A., Domenech, P., & Pessiglione, M. (2016). How prior preferences determine decision-making frames and biases in the human brain. *ELife*, *5*, e20317. https://doi.org/10.7554/eLife.20317

Losecaat Vermeer, A. B., Boksem, M. A. S., & Sanfey, A. G. (2014). Neural mechanisms underlying context-dependent shifts in risk preferences. *NeuroImage*, *103*, 355–363. https://doi.org/10.1016/j.neuroimage.2014.09.054

Mackey, S., & Petrides, M. (2010). Quantitative demonstration of comparable architectonic areas within the ventromedial and lateral orbital frontal cortex in the human and the macaque monkey brains. *European Journal of Neuroscience*, *32*(11), 1940–1950. https://doi.org/10.1111/j.1460-9568.2010.07465.x

Mackey, S., & Petrides, M. (2014). Architecture and morphology of the human ventromedial prefrontal cortex. *European Journal of Neuroscience*, *40*(5), 2777–2796. https://doi.org/10.1111/ejn.12654

Maisson, D. J.-N., Cash-Padgett, T. V., Wang, M. Z., Hayden, B. Y., Heilbronner, S. R., & Zimmermann, J. (2021). Choice-relevant information transformation along a ventrodorsal axis in the medial prefrontal cortex. *Nature Communications*, *12*(1), 1–14. https://doi.org/10.1038/s41467-021-25219-w

Mansouri, F. A., Koechlin, E., Rosa, M. G. P., & Buckley, M. J. (2017). Managing competing goals - A key role for the frontopolar cortex. *Nature Reviews Neuroscience*, *18*(11), 645–657. https://doi.org/10.1038/nrn.2017.111

McNamee, D., Rangel, A., & O'Doherty, J. P. (2013). Category-dependent and category-independent goal-value codes in human ventromedial prefrontal cortex. *Nature Neuroscience*, *16*(4), 479–485. https://doi.org/10.1038/nn.3337

Mitchell, J. F., Sundberg, K. A., & Reynolds, J. H. (2009). Spatial Attention Decorrelates Intrinsic Activity Fluctuations in Macaque Area V4. *Neuron*, *63*(6), 879–888. https://doi.org/10.1016/j.neuron.2009.09.013

Montague, P. R., & Berns, G. S. (2002). Neural economics and the biological substrates of valuation. *Neuron*, *36*(2), 265–284. https://doi.org/10.1016/S0896-6273(02)00974-1

Neubert, F.-X., Mars, R. B., Sallet, J., & Rushworth, M. F. S. (2015). Connectivity reveals relationship of brain areas for reward-guided learning and decision making in human and monkey frontal cortex. *Proceedings of the National Academy of Sciences*, *112*(20), E2695–E2704. https://doi.org/10.1073/pnas.1410767112

Norton, K. G., & Liljeholm, M. (2020). The rostrolateral prefrontal cortex mediates a preference for high-agency environment. *Journal of Neuroscience*, *40*(22), 4401–4409. https://doi.org/10.1523/JNEUROSCI.2463-19.2020

O'Doherty, J., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience*, *4*(1), 95–102. https://doi.org/10.1038/82959

O'Doherty, J., Winston, J., Critchley, H., Perrett, D., Burt, D. M., & Dolan, R. J. (2003). Beauty in a smile: The role of medial orbitofrontal cortex in facial attractiveness. *Neuropsychologia*, *41*(2), 147–155. https://doi.org/10.1016/S0028-3932(02)00145-8

Park, S. A., Miller, D. S., Nili, H., Ranganath, C., & Boorman, E. D. (2020). Map Making: Constructing, Combining, and Inferring on Abstract Cognitive Maps. *Neuron*, *107*(6), 1226-1238.e8. https://doi.org/10.1016/j.neuron.2020.06.030

Pascual-Leone, A., Bartres-Faz, D., & Keenan, J. (1999). Transcranial magnetic stimulation: studying the brain-behaviour. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *354*(1387), 1229–1238.

Paulus, M. P., & Frank, L. R. (2003). Ventromedial prefrontal cortex activation is critical for preference judgments. *NeuroReport*, *14*(10), 1311–1315. https://doi.org/10.1097/01.wnr.0000078543.07662.02

Petrides, M., & Pandya, D. N. (1994). Comparative architectonic analysis of the human and macaque frontal cortex. In F. Boller & J. Grafman (Eds.), *Handbook of Neuropsychology* (pp. 17–57). Elsevier.

Petrides, M., & Pandya, D. N. (2007). Efferent association pathways from the rostral prefrontal cortex in the macaque monkey. *Journal of Neuroscience*, *27*(43), 11573–11586. https://doi.org/10.1523/JNEUROSCI.2419-07.2007

Pitcher, D., Ianni, G., & Ungerleider, L. G. (2019). A functional dissociation of face-, body- and scene-selective brain areas based on their response to moving and static stimuli. *Scientific Reports*, *9*(1), 1–9. https://doi.org/10.1038/s41598-019-44663-9

Plassmann, H., O'Doherty, J. P., & Rangel, A. (2007). Orbitofrontal Cortex Encodes Willingness to Pay in Everyday Economic Transactions. *Journal of Neuroscience*, *27*(37), 9984–9988. https://doi.org/10.1523/JNEUROSCI.2131-07.2007

Plassmann, H., O'Doherty, J. P., & Rangel, A. (2010). Appetitive and aversive goal values are encoded in the medial orbitofrontal cortex at the time of decision making. *Journal of Neuroscience*, *30*(32), 10799–10808. https://doi.org/10.1523/JNEUROSCI.0788-10.2010

Plassmann, H., O'Doherty, J. P., Shiv, B., & Rangel, A. (2008). Marketing actions can modulate neural representations of experienced pleasantness. *Proceedings of the National Academy of Sciences*, *105*(3), 1050–1054. https://doi.org/10.1073/pnas.0706929105

Prévost, C., Pessiglione, M., Météreau, E., Cléry-Melin, M. L., & Dreher, J. C. (2010). Separate valuation subsystems for delay and effort decision costs. *Journal of Neuroscience*, *30*(42), 14080–14090. https://doi.org/10.1523/JNEUROSCI.2752-10.2010

Raichle, M. E., & Raichle, M. E. (2001). Searching for a baseline: Functional imaging and the resting human brain. *Nature Reviews Neuroscience*, *2*(10), 685–694. https://doi.org/10.1038/35094500

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*(2), 59–108. https://doi.org/10.1037/0033-295X.85.2.59

Roberts, A. C., & Clarke, H. F. (2019). Why we need nonhuman primates to study the role of ventromedial prefrontal cortex in the regulation of threat- And reward-elicited responses. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(52), 26297–26304. https://doi.org/10.1073/pnas.1902288116

Rolls, E. T., Kringelbach, M. L., & De Araujo, I. E. T. (2003). Different representations of pleasant and unpleasant odours in the human brain. *European Journal of Neuroscience*, *18*(3), 695–703. https://doi.org/10.1046/j.1460-9568.2003.02779.x

Ruff, D. A., & Cohen, M. R. (2014). Attention can either increase or decrease spike count correlations in visual cortex. *Nature Neuroscience*, *17*(11), 1591–1597. https://doi.org/10.1038/nn.3835

Ruff, D. A., & Cohen, M. R. (2016). Attention increases spike count correlations between visual cortical areas. *Journal of Neuroscience*, *36*(28), 7523–7534. https://doi.org/10.1523/JNEUROSCI.0610-16.2016

Ruff, D. A., & Cohen, M. R. (2017). A normalization model suggests that attention changes the weighting of inputs between visual areas. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(20), E4085–E4094. https://doi.org/10.1073/pnas.1619857114

Ruff, D. A., & Cohen, M. R. (2019). Simultaneous multi-area recordings suggest that attention improves performance by reshaping stimulus representations. *Nature Neuroscience*. https://doi.org/10.1038/s41593-019-0477-1

Runyan, C. A., Piasini, E., Panzeri, S., & Harvey, C. D. (2017). Distinct timescales of population coding across cortex. *Nature*, *548*(7665), 92–96. https://doi.org/10.1038/nature23020

Saleem, K. S., Miller, B., & Price, J. L. (2014). Subdivisions and connectional networks of the lateral prefrontal cortex in the macaque monkey. *Journal of Comparative Neurology*, *522*(7), 1641–1690. https://doi.org/10.1002/cne.23498

Sarafyazd, M., & Jazayeri, M. (2019). Hierarchical reasoning by neural circuits in the frontal cortex. *Science*, *364*(6441). https://doi.org/10.1126/science.aav8911

Schuck, N. W., Cai, M. B., Wilson, R. C., & Niv, Y. (2016). Human Orbitofrontal Cortex Represents a Cognitive Map of State Space. *Neuron*, *91*(6), 1402–1412. https://doi.org/10.1016/j.neuron.2016.08.019

Schuck, N. W., Gaschler, R., Wenke, D., Heinzle, J., Frensch, P. A., Haynes, J. D., & Reverberi, C. (2015). Medial prefrontal cortex predicts internally driven strategy shifts. *Neuron*, *86*(1), 331–340. https://doi.org/10.1016/j.neuron.2015.03.015

Sellitto, M., Ciaramelli, E., & Di Pellegrino, G. (2010). Myopic discounting of future rewards after medial orbitofrontal damage in humans. *Journal of Neuroscience*, *30*(49), 16429–16436. https://doi.org/10.1523/JNEUROSCI.2516-10.2010

Semendeferi, K., Armstrong, E., Schleicher, A., Zilles, K., & Van Hoesen, G. W. (2001). Prefrontal cortex in humans and apes: A comparative study of area 10. *American Journal of Physical Anthropology*, *114*(3), 224–241. https://doi.org/10.1002/1096-8644(200103)114:3<224::AID-AJPA1022>3.0.CO;2-I

Shenhav, A., & Karmarkar, U. R. (2019). Dissociable components of the reward circuit are involved in appraisal versus choice. *Scientific Reports*, *9*(1), 1–12. https://doi.org/10.1038/s41598-019-38927-7

Smith, S. M. (2002). Fast robust automated brain extraction. *Human Brain Mapping*, *17*(3), 143–155. https://doi.org/10.1002/hbm.10062

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, *23*(SUPPL. 1), 208–219. https://doi.org/10.1016/j.neuroimage.2004.07.051

Strait, C. E., Blanchard, T. C., & Hayden, B. Y. (2014). Reward value comparison via mutual inhibition in ventromedial prefrontal cortex. *Neuron*, *82*(6), 1357–1366. https://doi.org/10.1016/j.neuron.2014.04.032

Symmonds, M., Bossaerts, P., & Dolan, R. J. (2010). A behavioral and neural evaluation of prospective decision-making under risk. *The Journal of Neuroscience*, *30*(43), 14380–14389. https://doi.org/10.1523/JNEUROSCI.1459-10.2010

Symmonds, M., Wright, N. D., Bach, D. R., & Dolan, R. J. (2011). Deconstructing risk: Separable encoding of variance and skewness in the brain. *NeuroImage*, *58*(4), 1139–1149. https://doi.org/10.1016/j.neuroimage.2011.06.087

Symmonds, M., Wright, N. D., Fagan, E., & Dolan, R. J. (2013). Assaying the Effect of Levodopa on the Evaluation of Risk in Healthy Humans. *PLoS ONE*, *8*(7). https://doi.org/10.1371/journal.pone.0068177

Thomas, A. W., Molter, F., Krajbich, I., Heekeren, H. R., & Mohr, P. N. C. (2019). Gaze bias differences capture individual choice behaviour. *Nature Human Behaviour*, *3*(6), 625–635. https://doi.org/10.1038/s41562-019-0584-8

Tobler, P. N., & Weber, E. U. (2013). Ch9 Valuation for Risky and Uncertain Choices. In *Neuroeconomics: Decision Making and the Brain: Second Edition* (pp. 149–172). Elsevier Inc. https://doi.org/10.1016/B978-0-12-416008-8.00009-7

Tom, S. M., Fox, C. R., Trepel, C., & Poldrack, R. A. (2007). The neural basis of loss aversion in decision-making under risk. *Science, 315*(5811), 515–518. https://doi.org/10.1126/science.1134239

Tong, F., Nakayama, K., Vaughan, J. T., & Kanwisher, N. (1998). Binocular rivalry and visual awareness in human extrastriate cortex. *Neuron, 21*(0896-6273 (Print) LA-eng PT-Journal Article PT-Research Support, Non-U.S. Gov't PT-Research Support, U.S. Gov't, P.H.S SB-IM SB-S), 753–759.

Towal, R. B., Mormann, M., & Koch, C. (2013). Simultaneous modeling of visual saliency and value computation improves predictions of economic choice. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(40). https://doi.org/10.1073/pnas.1304429110

Trudel, N., Scholl, J., Klein-Flügge, M. C., Fouragnan, E., Tankelevitch, L., Wittmann, M. K., & Rushworth, M. F. S. (2020). Polarity of uncertainty representation during exploration and exploitation in ventromedial prefrontal cortex. *Nature Human Behaviour*. https://doi.org/10.1038/s41562-020-0929-3

Tversky, A., & Kahneman, D. (1992). Advances in Prospect Theory: Cumulative Representation of Uncertainty. *Journal of Risk and Uncertainty*, *5*, 297–323.

Vogt, B. A., Pandya, D. N., & Rosene, D. L. (1987). Cingulate cortex of the rhesus monkey: I. Cytoarchitecture and thalamic afferents. *Journal of Comparative Neurology*, *262*(2), 256–270. https://doi.org/10.1002/cne.902620207

von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton UP.

Wang, X. J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. *Neuron*, *36*(5), 955–968. https://doi.org/10.1016/S0896-6273(02)01092-9

Wittmann, M. K., Kolling, N., Akaishi, R., Chau, B. K. H., Brown, J. W., Nelissen, N., & Rushworth, M. F. S. (2016). Predictive decision making driven by multiple time-linked reward representations in the anterior cingulate cortex. *Nature Communications*, *7*, 12327. https://doi.org/10.1038/ncomms12327

Woolrich, M. (2008). Robust group analysis using outlier inference. *NeuroImage*, *41*(2), 286–301. https://doi.org/10.1016/j.neuroimage.2008.02.042

Woolrich, M., Behrens, T. E. J., Beckmann, C. F., Jenkinson, M., & Smith, S. M. (2004). Multilevel linear modelling for FMRI group analysis using Bayesian inference. *NeuroImage*, *21*(4), 1732–1747. https://doi.org/10.1016/j.neuroimage.2003.12.023

Worsley, K. J., Evans, A. C., Marrett, S., & Neelin, P. (1992). A three-dimensional statistical analysis for CBF activation studies in human brain. *Journal of Cerebral Blood Flow and Metabolism*, *12*(6), 900–918. https://doi.org/10.1038/jcbfm.1992.127

Wright, N. D., Symmonds, M., & Dolan, R. J. (2013). Distinct encoding of risk and value in economic choice between multiple risky options. *NeuroImage*, *81*, 431–440. https://doi.org/10.1016/j.neuroimage.2013.05.023

Wright, N. D., Symmonds, M., Hodgson, K., Fitzgerald, T. H. B., Crawford, B., & Dolan, R. J. (2012). Approach-Avoidance Processes Contribute to Dissociable Impacts of Risk and Loss on Choice. *Journal of Neuroscience*, *32*(20), 7009–7020. https://doi.org/10.1523/JNEUROSCI.0049-12.2012

Wright, N. D., Symmonds, M., Morris, L. S., & Dolan, R. J. (2013). Dissociable influences of skewness and valence on economic choice and neural activity. *PLoS ONE*, *8*(12). https://doi.org/10.1371/journal.pone.0083454

Yang, G. R., & Wang, X. J. (2020). Artificial Neural Networks for Neuroscientists: A Primer. *Neuron*, *107*(6), 1048–1070. https://doi.org/10.1016/j.neuron.2020.09.005

Yantis, S., & Jonides, J. (1984). Abrupt visual onsets and selective attention: Evidence from visual search. *Journal of Experimental Psychology: Human Perception and Performance*, *10*(5), 601–621. https://doi.org/10.1037/0096-1523.10.5.601

Yeterian, E. H., Pandya, D. N., Tomaiuolo, F., & Petrides, M. (2012). The cortical connectivity of the prefrontal cortex in the monkey brain. *Cortex*, *48*(1), 58–81. https://doi.org/10.1016/j.cortex.2011.03.004

Yoo, S. B. M., & Hayden, B. Y. (2020). The Transition from Evaluation to Selection Involves Neural Subspace Reorganization in Core Reward Regions. *Neuron*, *105*(4), 712-724.e4. https://doi.org/10.1016/j.neuron.2019.11.013

Zajkowski, W., Kossut, M., & Wilson, R. C. (2017). A causal role for right frontopolar cortex in directed, but not random, exploration. *ELife*, 79–84. https://doi.org/10.1101/064741

Zhang, Z., Fanning, J., Ehrlich, D. B., Chen, W., Lee, D., & Levy, I. (2017). Distributed neural representation of saliency controlled value and category during anticipation of rewards and punishments. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-02080-4