Neural Mechanisms For Learning In Dynamic Environments

Chang-Hao Kao

University of Pennsylvania

Recommended Citation
https://repository.upenn.edu/edissertations/4229

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/4229
For more information, please contact repository@pobox.upenn.edu.
Neural Mechanisms For Learning In Dynamic Environments

Abstract
I investigated the neural mechanisms for learning in dynamic environments. In dynamic environments, where the underlying state of the environment can change, inferring the current state of the environment is important to guide adaptive behavior. People should maintain their beliefs about the environmental state when it is stable and they should quickly update their beliefs when the environmental state changes. Belief updating can be guided by prediction errors, which are the difference between the expected observation and the actual observation. How people use prediction errors to update their beliefs is determined by their learning rate. Learning rates can be influenced by belief surprise and belief uncertainty. Belief surprise reflects how unlikely an observation is, given the person's belief about the current state. Belief uncertainty reflects how imprecise a person's belief is about the current state. In three studies, I investigated how the brain detects state changes and guides subsequent behavioral adaptation. In the first study, I examine the roles of physiological arousal during learning in two kinds of dynamic environments. In both environments, prediction errors enhanced learning rates and induced pupil dilation. Among different measures of physiological arousal (pupil dilation, skin conductance, heart rate and respiration rate), only pupil dilation consistently predicted trial-by-trial learning rates in both environments. Furthermore, pupil dilation mediated the relationship between prediction errors and learning rates and predicted variance in learning rates that could not be accounted for by prediction errors. In the second study, I investigated how whole-brain functional networks reconfigure for the adjustment of learning. Learning rates were influenced by belief surprise and belief uncertainty, and these two variables also modulated the integration between fronto-parietal and other brain networks. This modulation of functional networks was also associated with behavioral adaptation across individuals. In the third study, I further distinguished the functional roles of frontal and parietal regions during learning. Using multi-voxel pattern classification, I showed that posterior parietal cortex encoded prediction errors in a task-dependent manner while frontal cortex predicted the subsequent behavioral shifts in response to errors. From these studies, I demonstrated how different neural systems contribute to learning in dynamic environments.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Psychology

First Advisor
Joseph W. Kable

Keywords
adaptive learning, functional network, physiological arousal, prediction error, surprise, uncertainty

Subject Categories
Neuroscience and Neurobiology | Psychology

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/4229
NEURAL MECHANISMS FOR LEARNING IN DYNAMIC ENVIRONMENTS

Chang-Hao Kao

A DISSERTATION

in

Psychology

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Supervisor of Dissertation

_______________________

Joseph W. Kable, Baird Term Professor, Psychology

Graduate Group Chairperson

_______________________

Sara Jaffee, Professor, Psychology

Dissertation Committee

Sudeep Bhatia, Assistant Professor, Psychology

Michael L. Platt, James S. Riepe University Professor, Psychology

Joshua I. Gold, Professor, Neuroscience
ACKNOWLEDGMENT

It has been a long trip to pursue my PhD. During this period, I have received a lot of help from many people. I would not be able to complete my PhD without you.

I would like to thank my family in Taiwan to support my decision to pursue the PhD in the US. They provide me lots of mental support when I am far from Taiwan. I would also thank my girlfriend, Yun-Yen Yang. She has been supportive for my whole PhD period while I am distressed in research or in life.

My advisor, Joe, is very nice and supportive. I am glad to work with him for my PhD, and I learn a lot of experience from him. Whenever I discuss with him about my difficulties, he can always give me helpful feedback and guide me to a clear direction. I would especially thank my close collaborator, Josh. He has given me many critical and helpful feedback on all of my dissertation studies and made me improve them a lot. I would also thank my other committee members, Sudeep and Michael. They have provided me helpful feedback for my dissertation studies during my PhD period.

Being in the Kable lab is fun. Everyone in the lab is friendly and I have received a lot of help from them. I would especially thank Arthur, who is around during my PhD. Talking research and life with him always inspires my thinking and makes me relaxed.

Last, I would like to thank every person who I have collaborated with and who I asked for help.
ABSTRACT

NEURAL MECHANISMS FOR LEARNING IN DYNAMIC ENVIRONMENTS

Chang-Hao Kao

Joseph W. Kable

I investigated the neural mechanisms for learning in dynamic environments. In dynamic environments, where the underlying state of the environment can change, inferring the current state of the environment is important to guide adaptive behavior. People should maintain their beliefs about the environmental state when it is stable and they should quickly update their beliefs when the environmental state changes. Belief updating can be guided by prediction errors, which are the difference between the expected observation and the actual observation. How people use prediction errors to update their beliefs is determined by their learning rate. Learning rates can be influenced by belief surprise and belief uncertainty. Belief surprise reflects how unlikely an observation is, given the person’s belief about the current state. Belief uncertainty reflects how imprecise a person’s belief is about the current state. In three studies, I investigated how the brain detects state changes and guides subsequent behavioral adaptation. In the first study, I examine the roles of physiological arousal during learning in two kinds of dynamic environments. In both environments, prediction errors enhanced learning rates and induced pupil dilation. Among different measures of physiological arousal (pupil dilation, skin conductance, heart rate and respiration rate), only pupil dilation consistently predicted trial-by-trial learning rates in both environments. Furthermore, pupil dilation mediated the relationship between prediction errors and learning rates and predicted
variance in learning rates that could not be accounted for by prediction errors. In the second study, I investigated how whole-brain functional networks reconfigure for the adjustment of learning. Learning rates were influenced by belief surprise and belief uncertainty, and these two variables also modulated the integration between fronto-parietal and other brain networks. This modulation of functional networks was also associated with behavioral adaptation across individuals. In the third study, I further distinguished the functional roles of frontal and parietal regions during learning. Using multi-voxel pattern classification, I showed that posterior parietal cortex encoded prediction errors in a task-dependent manner while frontal cortex predicted the subsequent behavioral shifts in response to errors. From these studies, I demonstrated how different neural systems contribute to learning in dynamic environments.
TABLE OF CONTENTS

ACKNOWLEDGMENT ...................................................................................... II
ABSTRACT ........................................................................................................ III
LIST OF TABLES .............................................................................................. VII
LIST OF ILLUSTRATIONS ............................................................................ VIII
CHAPTER 1 - INTRODUCTION ..................................................................... 1
CHAPTER 2 – PUPIL DIAMETER, BUT NOT OTHER MEASURES OF AROUSAL, TRACKS THE ADJUSTMENT OF LEARNING ACROSS DIFFERENT DYNAMIC ENVIRONMENTS .......................................................... 13

Abstract ........................................................................................................ 13
Introduction .................................................................................................... 14
Methods ........................................................................................................ 17
Results .......................................................................................................... 31
Discussion ..................................................................................................... 47

CHAPTER 3 – FUNCTIONAL BRAIN NETWORK RECONFIGURATION DURING LEARNING IN A DYNAMIC ENVIRONMENT .............................................. 60

Abstract ........................................................................................................ 60
Introduction .................................................................................................... 61
Methods ........................................................................................................ 64
Results .......................................................................................................... 80
Discussion ..................................................................................................... 101
CHAPTER 4 – NEURAL ENCODING OF TASK-DEPENDENT ERRORS
DURING ADAPTIVE LEARNING..................................................... 130

Abstract.............................................................................................. 130
Introduction.......................................................................................... 131
Methods............................................................................................... 133
Results................................................................................................. 150
Discussion ............................................................................................ 167

CHAPTER 5 – GENERAL DISCUSSION ............................................. 184

Summary of studies ............................................................................. 184
Task-dependent influence on learning in dynamic environments .......... 184
Integrated neural systems .................................................................... 186
Future directions .................................................................................. 188
Conclusion ............................................................................................ 191

BIBLIOGRAPHY ................................................................................. 192
LIST OF TABLES

Table 4-1 BIC of behavior models ................................................................. 155
Table 4-2 Parameters of behavior models ....................................................... 156
Table 4-3 Summary of fMRI results: error magnitude and past error ............. 163
Table 4-4 Summary of fMRI results: behavior change .................................. 166
# LIST OF ILLUSTRATIONS

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Task design</td>
<td>20</td>
</tr>
<tr>
<td>2-2</td>
<td>Behavior results</td>
<td>34</td>
</tr>
<tr>
<td>2-3</td>
<td>Demonstration for different measures of physiological arousal</td>
<td>37</td>
</tr>
<tr>
<td>2-4</td>
<td>Regression of task factors on phasic change</td>
<td>40</td>
</tr>
<tr>
<td>2-5</td>
<td>Regression of task factors on tonic change</td>
<td>41</td>
</tr>
<tr>
<td>2-6</td>
<td>Functional roles of pupil tonic change</td>
<td>45</td>
</tr>
<tr>
<td>3-1</td>
<td>Overview of the task and theoretical model of belief updating (McGuire et al., 2014)</td>
<td>66</td>
</tr>
<tr>
<td>3-2</td>
<td>Schematic overview of the method</td>
<td>85</td>
</tr>
<tr>
<td>3-3</td>
<td>Patterns of connectivity in subgraphs</td>
<td>87</td>
</tr>
<tr>
<td>3-4</td>
<td>Temporal expression of subgraph 4 was related to task factors and individual differences</td>
<td>90</td>
</tr>
<tr>
<td>3-5</td>
<td>The contribution of between-system and within-system edges to effects of task factors and individual differences on subgraph 4 expression</td>
<td>95</td>
</tr>
<tr>
<td>3-6</td>
<td>Relationship between edge strength of subgraph 4 and univariate task activations</td>
<td>99</td>
</tr>
<tr>
<td>4-1</td>
<td>Overview of task and experimental design</td>
<td>136</td>
</tr>
<tr>
<td>4-2</td>
<td>Behavioral results</td>
<td>153</td>
</tr>
<tr>
<td>4-3</td>
<td>Representations of error and error magnitude</td>
<td>161</td>
</tr>
<tr>
<td>4-4</td>
<td>Representations of errors on the previous trial conditional on different types of current trials (columns)</td>
<td>162</td>
</tr>
<tr>
<td>4-5</td>
<td>Representations of subsequent behavioral choices (switch versus stay) after ambiguous small errors in the noisy condition</td>
<td>165</td>
</tr>
</tbody>
</table>
Making an adaptive decision requires a precise belief about the current state in the environment. In real life, this state is usually not observable and we need to infer it based on noisy evidence. For example, whether to go to a restaurant depends on our belief about the quality (i.e., state) of this restaurant. As it is difficult to directly observe this quality, we can form and adjust our belief based on our experience (i.e., noisy observations) in this restaurant. That is, we decide whether to go to this restaurant based on our belief, and update our belief after we receive a new observation. As we receive more and more observations, our belief about this restaurant would be more precise.

**Learning in dynamic environments**

Belief updating can be guided by a prediction error, which reflects the difference between the expected observation and the actual observation, and a learning rate, which determines how much we use this prediction error to update the belief (Sutton & Barto, 1998). By this way, our belief would be updated toward the actual observation.

In dynamic environments, the state may change sometimes. For example, the quality of a restaurant can change when they change chefs. Thus, when the state changes, people should detect this change and dynamically adjust their belief (Nassar, Wilson, Heasly, & Gold, 2010). That is, when a state is stable, people gradually learn the state and eventually their belief should not be influenced by noisy observations (i.e., learning rates equals to 0). When the state has changed, they should quickly update their belief (i.e., high learning rates) toward the new state.
Recent studies have provided much evidence on how beliefs are updated in environments where a relatively stable state occasionally undergoes sudden changes (Behrens, Woolrich, Walton, & Rushworth, 2007a; Krishnamurthy, Nassar, Sarode, & Gold, 2017; McGuire, Nassar, Gold, & Kable, 2014; Nassar et al., 2012; Nassar et al., 2010). In such change-point environments, people should quickly adjust belief when they detect the change of the state; otherwise, they should keep their belief. Ideal learning in this dynamic environment can be captured by the framework of Bayesian decision theory (Behrens et al., 2007a). However, this framework is computationally expensive and people need to track the entire past history of observations. A reduced Bayesian model is more computational trackable and can achieve similar predictive performance for behavior as the complete Bayesian model (Nassar et al., 2010). In the reduced Bayesian model, learning is influenced by two sources: belief surprise and belief uncertainty. Belief surprise reflects how unlikely an observation is to have been generated from the state in our belief, while belief uncertainty reflects the imprecision of our belief about the state. Learning rate is high when either surprise or uncertainty is high.

**Uncertainty**

Belief uncertainty reflects the variance of the belief distribution. Belief uncertainty can be reduced as people receive more and more precise observations about the state from the environment (Kobayashi & Hsu, 2017; Nassar et al., 2010; O'Reilly, 2013). For example, in a novel environment (e.g., a new city), people do not know the probability of rain. They have to collect observations from the environment and form their belief about the probability of rain. If people form a belief that the probability of rain is 0.5 based on the observations of 5 days, they would be more uncertain about their belief. On the other
hand, if they form this belief based on the observations of 50 days, they would be less uncertain about their belief. Thus, belief uncertainty is influenced by past history of observations or errors.

Belief uncertainty guides the integration between the belief and the new observation. Belief uncertainty reflects the reliability of the belief and determine how much people should update toward the new observation (Angelaki, Gu, & DeAngelis, 2009; Berniker, Voss, & Kording, 2010; Ernst & Banks, 2002; Knill & Pouget, 2004; Körding & Wolpert, 2004; O’Reilly, Jbabdi, & Behrens, 2012; Ting, Yu, Maloney, & Wu, 2015; Vilares, Howard, Fernandes, Gottfried, & Kording, 2012). That is, if the belief uncertainty is high, people should update more toward the new observation (i.e., high learning rates); otherwise, they should tend to keep their current belief (i.e., low learning rates).

In different dynamic environments, belief uncertainty is influenced by different sources of uncertainty. One of the sources is the noise from the observation. In real life, the observation generated from the state is usually noisy. This noise introduces uncertainty to inferring the state, but this uncertainty would gradually decrease as more and more observations are received (Kobayashi & Hsu, 2017; Nassar et al., 2010; O’Reilly, 2013). In a change-point environment, another source of uncertainty comes from the change of the state. Thus, belief uncertainty would be high after a change-point, and gradually decrease as people receive more and more observations (Nassar et al., 2012; Nassar et al., 2010).

Surprise
Belief surprise can be quantified as how unlikely an observation is to have been generated from the state in one’s belief. If a state has changed, the new observation would usually lead to a large prediction error (i.e., large mismatch between the belief and the observation). In a change-point environment, where the state is mostly stable but undergoes occasional sudden changes, such large prediction errors elicit high belief surprise and imply that this new observation is more likely from a new state than from the current belief (Nassar et al., 2010). Belief surprise is also related to unexpected uncertainty (Payzan-LeNestour & Bossaerts, 2011). That is, this uncertainty is deviated from the expected uncertainty due to the noise of observation. Such unexpectedness might indicate a change of the state. As belief surprise is high, people should quickly update their belief toward the new observation (i.e., high learning rates).

Belief surprise can reflect either unexpectedness or novelty (Barto, Mirolli, & Baldassarre, 2013; Reisenzein, Horstmann, & Schützwohl, 2017). Unexpectedness that can be explained by noise does not drive changes in belief or behavior (d’Acremont & Bossaerts, 2016; Nassar, Bruckner, & Frank, 2019; O’Reilly et al., 2013). Only the surprise provided by the signal of novelty such as a change of the state drives belief updating (Krishnamurthy et al., 2017; McGuire et al., 2014; Nassar et al., 2012; Nassar et al., 2010; O’Reilly et al., 2013).

Task factors

In dynamic environments, task factors determine how the state changes and how the observation is generated from the state. Evaluating these task factors is important to infer the state. Past studies have usually used several task factors such as environmental noise and hazard rate. The environmental noise determines the variance
of observations generated from the current state. In a change-point environment, the same prediction error is related to higher belief surprise when environment noise is low than when environment noise is high (McGuire et al., 2014; Nassar et al., 2012; Nassar et al., 2010). Hazard rate indicates the frequency that the state would change in a change-point environment. A small prediction error is more likely to indicate the change of the state in an environment with a high hazard rate that an environment with a low hazard rate (Y. S. Li, Nassar, Kable, & Gold, 2019). Thus, evaluating the task factors guides people how to use prediction errors for belief updating.

**Incidental stimuli**

Learning can also be influenced by other non-normative factors. These factors are not related to the state or the observation but they can modulate learning rates. For example, many studies have showed that incidental reward can enhance learning rates (S. Lee, Gold, & Kable, 2020; McGuire et al., 2014). Compared to trials with neutral values, trials with monetary values enhanced learning rates. In another study, the change of auditory stimuli during the presentation of outcomes modulated learning rates (Nassar et al., 2012).

**Neural systems during adaptive learning**

**Physiological arousal**

Pupil dilation responds to learning-related components and learning rates in dynamic environments. In a change-point environment, pupil dilation is associated with prediction errors, belief surprise and belief uncertainty (Browning, Behrens, Jocham, O'Reilly, & Bishop, 2015; Filipowicz, Glaze, Kable, & Gold, 2020; Krishnamurthy et al.,
Pupil dilation also predicts learning rates or hazard rates (Browning et al., 2015; Krishnamurthy et al., 2017; Nassar et al., 2010; O'Reilly et al., 2013). Furthermore, pupil dilation tracks the variance of learning rates that cannot be accounted for by prediction errors or task factors (Krishnamurthy et al., 2017).

Pupil diameter can reflect general arousal regulated by the autonomic nervous system (ANS), which includes the sympathetic nervous system and parasympathetic nervous system (Cannon, 1929). The sympathetic nervous system promotes the "fight-or-flight" response, leading to reaction to environmental events such as threat. On the other hand, the parasympathetic nervous system promotes the "rest and digest" response, leading to regular bodily function. Similar to general arousal, pupil dilation can be induced by salient stimuli such as emotional stimuli (Bradley, Miccoli, Escrig, & Lang, 2008) or an auditory oddball (Hong, Walz, & Sajda, 2014; Murphy, O'Connell, O'Sullivan, Robertson, & Balsters, 2014; Murphy, Robertson, Balsters, & O'Connell, 2011).

In addition, pupil diameter is potentially regulated by the locus coeruleus-norepinephrine system (LC-NE) (Berridge, 2008; Sara & Bouret, 2012). Pupil diameter was found to covary with the activity of LC-NE system in nonhuman studies (Joshi, Li, Kalwani, & Gold, 2016; Varazzani, San-Galli, Gilardeau, & Bouret, 2015) and human functional magnetic resonance imaging (fMRI) studies (de Gee et al., 2017; Murphy et al., 2014). Furthermore, electrical microstimulation on LC neurons leads to the dilation of pupil diameter (Joshi et al., 2016).

Although no direct anatomical connections have been shown between the LC-NE system and ANS, the activity of both systems are correlated. These parallel responses suggest similar inputs project to ANS and LC-NE system separately and activate them separately (Nieuwenhuis, De Geus, & Aston-Jones, 2011).
Local brain regions

Several frontal and parietal regions have been shown to activate during learning in dynamic environments. Both belief surprise and belief uncertainty are represented in posterior parietal cortex (PPC), anterior cingulate cortex (ACC), dorsomedial frontal cortex (DMFC), and dorsolateral prefrontal cortex (DLPFC) (McGuire et al., 2014). PPC responds to errors or surprise (Fischer & Ullsperger, 2013; Gläscher, Daw, Dayan, & O'Doherty, 2010; McGuire et al., 2014; Nassar, Bruckner, et al., 2019; Nassar, McGuire, Ritz, & Kable, 2019; O'Reilly et al., 2013; Payzan-LeNestour, Dunne, Bossaerts, & O'Doherty, 2013) and past history of outcomes and stimuli (Akrami, Kopec, Diamond, & Brody, 2018; Brody & Hanks, 2016; FitzGerald, Moran, Friston, & Dolan, 2015; Furl & Averbeck, 2011; Hanks et al., 2015; Hayden, Nair, McCoy, & Platt, 2008; Hwang, Dahlen, Mukundan, & Komiyama, 2017). OFC and DMFC are related to the mental representation of the state in the environment (Chan, Niv, & Norman, 2016; Hunt et al., 2018; Karlsson, Tervo, & Karpova, 2012; Nassar, McGuire, et al., 2019; Schuck, Cai, Wilson, & Niv, 2016; Wilson, Takahashi, Schoenbaum, & Niv, 2014). The activity in frontopolar cortex and DMFC increases during exploratory choices (Blanchard & Gershman, 2018; Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006; Kolling, Behrens, Mars, & Rushworth, 2012; Kolling et al., 2016; Muller, Mars, Behrens, & O'Reilly, 2019). These past studies emphasize the importance of frontal and parietal regions during learning in dynamic environments.

Functional brain network

In addition to activation in local brain regions, recent studies have identified the importance of communication between brain regions during cognitive functions (Bassett
& Sporns, 2017; Bullmore & Sporns, 2009; Medaglia, Lynall, & Bassett, 2015). The integration between different networks is critical for complex cognitive functions such as memory, decision-making and learning (Bertolero, Yeo, & D'Esposito, 2015; Cohen & D'Esposito, 2016; Shine et al., 2016; Shine & Poldrack, 2017). Many studies have showed the contribution of the fronto-parietal network during learning. From the early phase of learning to the late phase of learning, the integration between the fronto-parietal network and other networks gradually decreases (Bassett, Yang, Wymbs, & Grafton, 2015; Büchel, Coull, & Friston, 1999; Sun, Miller, Rao, & D'esposito, 2006). The fronto-parietal network is also thought to be a control system that adjusts behaviors in response to changes in task requirements (Cole, Repovš, & Anticevic, 2014; Power & Petersen, 2013). This suggests that, at the beginning of learning, cognitive control is required to adapt to the new task requirement or task rule, and this requirement gradually decreases as learning progresses. Effective learning also requires the flexibility to adjust functional connectivity (Bassett et al., 2011; Gerraty et al., 2018).

Open questions

Recent studies have provided some evidence about neural mechanisms for learning in dynamic environments. However, there are some open questions.

First, the roles of physiological arousal across different dynamic environments is not well-established. Many recent studies have shown the involvement of pupil dilation during learning in a change-point environment (Browning et al., 2015; Filipowicz et al., 2020; Krishnamurthy et al., 2017; Nassar et al., 2012; O'Reilly et al., 2013), but there is little evidence about the involvement of physiological arousal in other types of dynamic environments such as drifting environments. In a drifting environment, the state keeps
shifting, with the current state generated from the previous state (Daw et al., 2006; S. Lee et al., 2020). People should update belief based on belief uncertainty, which is influenced by the drift of state and the noise of the observation. However, little is known about the involvement of physiological arousal in this environment. Moreover, past studies have paid most attention to pupil dilation and few studies investigated other measures of physiological arousal (e.g., skin conductance, heart rate and respiration rate) during learning. It is also not clear whether physiological arousal is a response to task factors (e.g., prediction errors) or whether it plays a direct role in influencing learning.

Second, it is unknown how functional networks change during learning in dynamic environments. Recent studies have shown the association between the integration of the fronto-parietal network and learning in stable environments (Bassett et al., 2015; Büchel et al., 1999; Sun et al., 2006). This relationship may suggest that the integration of the fronto-parietal network can be flexibly adjusted as learning is needed to be re-initiated in response to the change of the state. Furthermore, pupil dilation has been associated with the integration of the fronto-parietal network (Shine et al., 2016). Thus, I will examine the plausible association between dynamic functional networks and adaptive learning.

Third, little is known about the distinct functional roles of frontal and parietal regions during learning in a dynamic environment. Many studies have shown the involvement of several frontal regions (e.g., ACC, DMFC, DLPFC, OFC) and posterior parietal cortex during adaptive learning (Behrens et al., 2007a; McGuire et al., 2014; Nassar, McGuire, et al., 2019; O’Reilly et al., 2013; Payzan-LeNestour et al., 2013). However, it is not well established what their distinct functional roles are.
In my dissertation, I aim to understand the neural mechanisms for learning in dynamic environments. Three questions will be addressed: (1) how is physiological arousal associated with learning across different dynamic environments? (2) how do functional networks change in response to the adjustment of learning? (3) what are the distinct functional roles for frontal and parietal regions during learning in a dynamic environment?

Research overview

In Chapter 2, I investigated the involvement of physiological arousal during learning across different dynamic environments. I focused on learning in the change-point and drifting environments since people should adapt their behavior differently in these environments. To compare different computational processes between the two environments, I examined how people’s learning rates were guided by observation noise, current prediction errors and previous prediction errors in the two environments. During learning, I collected four measures of physiological arousal: pupil diameter, skin conductance, heart rate and respiration rate. Moreover, I separated each measure into two types of response: phasic change and tonic change. Phasic change reflects the fast change after the presentation of the outcome while tonic change reflects the slow change between the presentation of the outcome and the next prediction. Across the two environments, current prediction errors consistently drove learning rates and also induced pupil tonic change. This pupil tonic change was not only a response to prediction errors during learning. Pupil tonic change also directly influenced learning rates. First, it mediated the association between the current prediction errors and learning rates in the change-point environment. Second, it tracked the variance of
learning rates that cannot be accounted for by task factors (e.g., prediction errors). I also investigated whether incidental manipulations can influence learning rates and physiological arousal. Both incidental reward and incidental auditory oddballs induced pupil phasic change. However, only incidental reward influenced learning rates as shown in previous studies (S. Lee et al., 2020; McGuire et al., 2014).

In Chapter 3, I examined dynamic whole-brain functional networks during learning in a dynamic environment. The association between pupil dilation and adaptive learning and the association between pupil dilation and dynamic functional networks suggests an association between adaptive learning and dynamic functional networks. I re-analyzed fMRI data from a previous study (McGuire et al., 2014) where participants performed a predictive-inference task in a change-point environment. I calculated the functional connectivity over time between every two ROIs from a pre-defined atlas (Power et al., 2011). Then, I implemented non-negative matrix factorization to decompose this whole-brain matrix of functional connectivity over time into subgraphs, which reflect patterns of whole-brain functional connectivity, and their time-dependent magnitude. One subgraph, which characterized strong integration between the fronto-parietal network and other networks, responded to trial-by-trial belief surprise and belief uncertainty, which guide the adjustment of learning. In addition, the modulation of this subgraph by belief surprise and belief uncertainty was associated with the extent that participants used these two components to update beliefs.

In Chapter 4, I further dissociated the functional roles of frontal and parietal regions during learning. Two types of change-point environment were created. In the low noise and high hazard rate environment (referred as the unstable condition), participants should change their subsequent behavior in response to any errors, which indicated a
change of the state. In the high noise and low hazard rate environment (referred as
noisy condition), small prediction errors (error magnitude of 1 or 2) can indicate either a
cchange of the state or environmental noise. In the noisy condition, the results showed
that people integrated past error history to guide their behavioral adaptation for these
small errors. In the fMRI data, I implemented multi-voxel pattern classification to
evaluate different neural encoding between these two conditions. Posterior parietal
cortex encoded errors in a task-dependent manner, with stronger decoding performance
in the noisy condition for error magnitudes and for past errors conditional on current
small errors. On the other hand, frontal regions including ACC, DMFC, DLPFC and OFC
encoded behavioral change in response to small errors in the noisy condition.

From these three studies, I extended past findings and provided new insights
about the neural mechanism for learning in dynamic environments. I demonstrated (1)
that pupil dilation can directly influence the adjustment of learning across dynamic
environments, (2) that the integration between the fronto-parietal network and other
networks changes in response to adaptive learning, and (3) that parietal and frontal
regions respond to errors in a task-dependent manner and predict subsequent
behavioral adaptation, respectively.
CHAPTER 2 – Pupil diameter, but not other measures of arousal, tracks the adjustment of learning across different dynamic environments

Chang-Hao Kao, Yixin Chen, Frewine Ogbaselase, Joshua I. Gold, Joseph W. Kable

Abstract

Learning is guided by prediction errors, which reflects the deviation of the new observation from the belief. In dynamic environments, learning should be adaptively adjusted in response to the change of the state. People have shown to adapt different strategies in the change-point environment (where the stable state undergoes sudden changes sometimes) and in the drifting environment (where the current state was shifted from the previous state). However, there was little attention on the involvement of physiological arousal in the drifting environment. In this study, we collected different measures of physiological arousal (e.g., pupil diameter, skin conductance, heart rate and respiration rate), and examined the involvement of physiological arousal across the change-point and the drifting environments. Among different measures, only pupil diameter consistently predicted trial-by-trial learning rates and responded to the current prediction errors across the two environments. Moreover, pupil diameter mediated the relationship between the current prediction errors and learning rates in the change-point environment, and it tracked the variance of learning rates that cannot be accounted for by task factors across both environments. In this study, we demonstrated the importance of pupil diameter across different dynamic environments. Such learning-related pupil diameter may reflect the activity in the locus coeruleus-norepinephrine system rather than general arousal regulated by autonomic nervous system.
Introduction

Learning in dynamic environments

Adaptive decisions rely on forming a precise belief about the state of the current environment. For example, deciding whether to go out with an umbrella relies on a precise weather forecast. In real life, the state of the environment is usually not directly observed but rather inferred based on noisy observations (e.g., humidity or cloud) generated by the state (e.g., weather). Belief about the state can be updated by prediction errors, which reflect the deviation between observations and predictions based on one’s beliefs, with a learning rate determining how much people update their beliefs by the prediction errors (Sutton & Barto, 1998). In environments where the state can undergo sudden change-points, learning rates should be dynamically adjusted (Behrens, Woolrich, Walton, & Rushworth, 2007; Nassar et al., 2012; Nassar, Wilson, Heasly, & Gold, 2010). When the state has changed, belief should be updated quickly (i.e., high learning rate). When the state remains stable, belief should be kept (i.e., low learning rate).

Physiological arousal and learning

Many recent studies have shown the involvement of physiological arousal during learning. In change-point environments, pupil diameter has been associated with learning-related components such as prediction errors, surprise and uncertainty. (Browning, Behrens, Jocham, O’Reilly, & Bishop, 2015; Filipowicz, Glaze, Kable, & Gold, 2020; Krishnamurthy, Nassar, Sarode, & Gold, 2017; Nassar et al., 2010). Physiological
arousal can also track learning rates per se. Many studies have shown that dynamic changes in learning rates were associated with dynamic changes in pupil diameter (Browning et al., 2015; Krishnamurthy et al., 2017; Nassar et al., 2010; O’Reilly et al., 2013) and skin conductance (Li, Schiller, Schoenbaum, Phelps, & Daw, 2011). In addition, pupil diameter also tracks variance in learning rates that cannot be accounted for by other learning-related variables (Krishnamurthy et al., 2017). These association may not simply reflect the response of physiological arousal to learning-related variables. These relationships may imply that physiological arousal can serve as a bridge between prediction errors and learning rates.

Pupil diameter can reflect general arousal or the activity in the locus coeruleus-norepinephrine (LC-NE) system. During cognitive functioning, various physiological arousal has shown to reflect our bodily response (Satpute, Kragel, Barrett, Wager, & Bianciardi, 2018). Physiological arousal can be measured as pupil diameter, skin conductance, heart rate and respiration rate, and reflect general arousal regulated by autonomous nervous system (ANS) (Cannon, 1929). Moreover, pupil diameter can be also regulated by the LC-NE system (Berridge, 2008; Sara & Bouret, 2012). Pupil diameter was shown to covary with the activity of LC-NE system in nonhuman studies (Joshi, Li, Kalwani, & Gold, 2016; Varazzani, San-Galli, Gilardeau, & Bouret, 2015) and human functional magnetic resonance imaging (fMRI) studies (de Gee et al., 2017; Murphy, O’Connell, O’Sullivan, Robertson, & Balsters, 2014). Although no direct anatomical connections have been shown between the LC-NE system and ANS, the activity of both systems are correlated. This parallel responses suggested similar inputs project to ANS and LC-NE separately and activate them separately (Nieuwenhuis, De Geus, & Aston-Jones, 2011).
It is not well-established about the common mechanism of physiological arousal across different dynamic environments. Recent studies have shown that people adaptively adjust their learning in two types of dynamic environments: change-point and drifting. In a change-point environment, stable states would undergo sudden change sometimes (Behrens et al., 2007; Nassar et al., 2010; O’Reilly et al., 2013). People’s belief updating should be guided by the likelihood of state change (i.e., belief surprise) and the uncertainty of their belief about the state (i.e., belief uncertainty) (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Nassar et al., 2012). As either surprise or uncertainty is high, people should update their belief quicker. In a drifting environment, the state continuously drifts with the current state shifting from the previous state (Daw, O’Doherty, Dayan, Seymour, & Dolan, 2006; S. Lee, Gold, & Kable, 2020). In such environment, belief updating is solely driven by people’s belief uncertainty, which come from the drift of the state and the noise of observation. As mentioned above, there were many evidences about the involvement of physiological arousal in the change-point environment but little is known about this involvement in the drifting environment.

Incidental manipulations also showed to modulate the learning rates but few studies investigated their influence on physiological arousal. For example, compared with trials with neutral value, people enhanced learning rates for trials with monetary value (S. Lee et al., 2020; McGuire, Nassar, Gold, & Kable, 2014). The sudden change of the auditory stimuli during the presentation of outcome can also influence learning rates and pupil diameter (Nassar et al., 2012).

*The current study*
Although recent studies showed the involvement of physiological arousal during learning in dynamic environments, there are some unanswered questions. In this study, we aimed to systematically investigate the involvement of physiological arousal during learning in dynamic environments. First, we investigated both change-point and drifting environments. We aimed to examine the common mechanism of physiological arousal across different dynamic environments. Second, little is known whether pupil diameter during learning reflects the involvement of LC-NE system or ANS. To address this issue, in addition to pupil diameter, we also collected other measures of physiological arousal (e.g., skin conductance, heart rate and respiration rate) which reflects general arousal regulated by ANS. Third, little is known about the functional roles of physiological arousal during learning. We examined whether physiological arousal is just a by-product induced by learning-related factors or it can directly influence learning. Specifically, we examined whether physiological arousal mediates the relationship between prediction errors and learning rates, and examined whether physiological arousal tracked learning rates that cannot be accounted for by task factors. Fourth, it is not clear about the influence of incidental stimuli on learning and physiological arousal. To have systematical understanding about the effects of incidental stimuli, we introduced incidental auditory oddball and incidental reward during the presentation of outcome and examined their influence on physiological arousal and belief updating.

**Methods**

*Participants*
Participants performed a predictive-inference task in one of two environments: change-point and drifting. In the change-point environment, there were 29 participants: 19 females and 10 males, mean age = 23.03 years (SD = 3.31; range 18-31). Three additional participants were excluded from the analyses: two for insufficient trials (below 30%) of good pupil measure, and one for trial-by-trial learning rates consistently 1 (with 0 for one trial only), indicating a potential misunderstanding of the task. In the drifting environment, there were 27 participants: 15 females and 12 males, mean age = 22.30 years (SD = 3.31; range 18-29). Five additional participants were excluded from the analyses: one for the loss of timing triggers during data collection, and four for insufficient trials of good pupil measure. Human subject protocols were approved by the Internal Review Board in University of Pennsylvania. All participants provided informed consent before the experiment.

*Predictive-inference task*

Participants performed a predictive-inference task during the collection of physiological measures (Fig. 1A). In this task, participants predicted a number between 0 and 300. Then, a number was sampled from a Gaussian distribution with a mean number corresponding to the current state and a standard deviation corresponding to the environment noise. There were six 40-trial runs, and the noise level was either low (SD = 10) or high (SD = 25) in each run. Additionally, there were two incidental manipulations: value and sound. For value, each trial was randomly assigned to have either monetary value or neutral value. Only in trials with monetary value did participants receive an additional incentive bonus based on the distance between their prediction and the outcome number. The color (yellow or gray) of a square around the outcome number
determined whether the value was monetary or neutral, with colors counterbalanced across participants. For the sound, participants heard either a standard sound ($p=0.9$) or an oddball sound ($p=0.1$) along with the presentation of the outcome number. Assignment of the two types of sound to standard or oddball was counterbalanced across participants.
Figure 2-1 Task design

(A) Predictive-inference task. On each trial, participants were required to make a prediction for the following outcome number between 0 and 300. After an inter-stimulus-interval (ISI), an outcome was sampled according to the current state and environmental noise. Different color (counterbalanced across participants) of squares around the outcome determined whether the current trial has monetary or neutral value. In a trial with monetary value, participants received extra money depended on how close their prediction to the outcome. Additionally, when the outcome presented, either a standard or an oddball sound played. For the physiology analysis, we focused on the period of 9 seconds after the onset of outcome and used 100 milliseconds before the onset of outcome as the baseline. (B) Dynamic environments. Participants performed the predictive-inference task in one of two dynamic environments: change-point or drifting environments. In the change-point environment, the state was mostly stable but underwent a sudden change occasionally. In the drifting environment, the state continuously shifted, with the current state being sampled from the previous state. In both environments, the state was hidden but an outcome sampled from a Gaussian distribution (mean = state, and SD = environmental noise) was shown to participants. In different runs, low and high environmental noise was used.
To make the best prediction, participants should infer the current state, which changed dynamically in this task. Participants were assigned to one of two dynamic environments: change-point and drifting (Fig. 1B). In the change-point environment, the state was mostly stable but underwent a sudden change occasionally. The state was stable for three trials after a change-point but changed with a probability of 0.125 on each trial thereafter. As a change-point occurred, the current state was sampled from a uniform distribution between 0 and 300. In the drifting environment, on the other hand, the state continuously shifted as a Gaussian random walk. That is, the current state was determined by a Gaussian distribution with a mean of the previous state and a standard deviation of a drift rate (D = 11).

To minimize the change of pupil diameter due to the change of luminance, the luminance of visual stimuli was the same as the luminance of the background. The background was a checkerboard with black and white, and the fixation and number were gray. Moreover, the luminance of the yellow and gray squares around the number were the same.

**Behavior analysis**

We investigated the influence of task factors on learning rates. Learning rates were calculated based on the delta learning rule (Sutton & Barto, 1998):

\[
\delta_t = X_t - B_t \quad (1)
\]
\[
B_{t+1} = B_t + \alpha_t \times \delta_t \quad (2)
\]
where $\delta_t$ is the prediction error, which is the difference between the current outcome number ($X_t$) and the current prediction ($B_t$), $B_{t+1}$ is the next prediction and $\alpha_t$ is the learning rate. Based on Eq. 2, trial-by-trial learning rates can be calculated as belief updating proportional to the prediction error ($\frac{B_{t+1}-B_t}{\delta_t}$). For the following behavior and physiological analyses, we removed trials with unstable estimates of learning rates: the first four trials were removed because learning rates were still converging in the drifting environments; trials with prediction errors smaller than 2 were removed due to low resolution of learning rates (i.e., learning rate can only be 0 or 1).

In both environments, we focused on how task factors influence learning rates. First, we investigated the average learning rates between the two noise conditions. Second, we examined how the current prediction errors influenced learning rates. We binned trials based on different range of prediction errors (i.e., 2-5, 6-10, 11-15, 16-20, 21-25, 26-30, 31-35, 36-40, 41-45, 46-50, 51-55, 56-60, 61-100, 101-140, 141-180, 181-220, 221-260, 261-300), and calculated the average learning rate for each binned trial and each noise condition. Third, we examined whether learning rates gradually decrease after an unexpected large prediction error, which was larger than twice the environment noise. We calculated the average learning rates of trial $t-2$ to $t+4$ corresponding to the trial of an unexpected large prediction error for each noise condition. Fourth, we examined the influence of monetary value on learning rates by comparing the average learning rate in the trials with monetary value with that in the trials with neutral values. Last, we investigated whether oddball sound influences learning rates by comparing the average learning rate in the trials with oddball sound and that in the trials with standard sound.
Moreover, we investigate the influence of task factors on learning rates in one regression model:

\[ \alpha_t = \beta_0 + \beta_1 I_{Low} + \beta_2 PE_t + \beta_3 PE_{t-1} + \beta_4 Value + \beta_5 Oddball + \beta_6 I_{Edge} \]  \hspace{1cm} (3)

where \( I_{Low} \) indicates whether the noise condition is low, \( PE_t \) reflects the current prediction error, \( PE_{t-1} \) reflects the previous prediction error, \( Value \) indicates whether the current value is monetary, \( Oddball \) indicates whether the current sound is an oddball, and \( I_{Edge} \) indicates whether the outcome number is close to the edge of the scale (0 or 300). The indicator of \( I_{Edge} \) captured participants’ tendency to avoid adjusting the belief toward the edge. We made \( I_{Edge} \) as 1 if the distance between the outcome number and the edge is smaller than the current environment noise. We fitted this multiple regression using linear mixed effects model (R packages of lme4 and lmerTest) for the change-point and drifting environments separately. In the model, we also estimated random intercepts and slopes for participants.

Behavior model

Our analyses focused on the influence of task factors on learning rates but we also implemented normative models to show the ideal observer’s performance in the two environments.

Dynamic learning rates in a change-point environment
In a change-point environment, learning rates were influenced by two normative factors (belief surprise and belief uncertainty) (McGuire et al., 2014; Nassar et al., 2012; Nassar et al., 2010):

\[ \alpha_t = \Omega_t + (1 - \Omega_t) \times \tau_t \]  \hspace{1cm} (4)

where \( \Omega_t \) is the change-point probability (CPP; belief surprise) and \( \tau_t \) is the relative uncertainty (RU; belief uncertainty).

CPP indicates the likelihood that a change-point has occurred:

\[ \Omega_t = \frac{U(X_t | 0, 300)^H}{U(X_t | 0, 300)^H + N(X_t | B_t, \sigma^2_t)(1-H)} \]  \hspace{1cm} (5)

where \( U(X_t | 0, 300) \) reflects the probability of \( X_t \) from a uniform distribution between 0 and 300, \( N(X_t | B_t, \sigma^2_t) \) reflects the probability of \( X_t \) from a Gaussian distribution with a mean of \( B_t \) and a variance of \( \sigma^2_t \), \( \sigma^2 \) reflects the variance of the predictive distribution of the outcome number, and \( H \) is the average rate of a change-point (0.1) in this task.

Thus, learning rates increase as an unexpected large prediction error occurs.

RU indicates the uncertainty about the current state relative to the environment noise:

\[ \tau_{t+1} = \frac{\Omega_t \sigma^2_t + (1 - \Omega_t) \tau_t \sigma^2_t + \Omega_t(1 - \Omega_t)(\delta_t(1 - \tau_t))^2}{\Omega_t \sigma^2_t + (1 - \Omega_t) \tau_t \sigma^2_t + \Omega_t(1 - \Omega_t)(\delta_t(1 - \tau_t))^2 + \sigma^2_H} \]  \hspace{1cm} (6)
where \( \sigma_N^2 \) is the variance (i.e., environment noise) of the distribution for generating the outcome number. In both the numerator and denominator, there are three terms representing different sources of uncertainty. The first term is the variance of the state distribution conditional on a change-point whereas the second term is the variance of the state distribution conditional on no change-point. The third term reflects the variance resulted from the difference of mean between these two conditional distributions. Thus, RU increases after a high CPP and then gradually decreases. That is, learning rates gradually decrease after an unexpected large prediction error.

**Dynamic learning rates in a drifting environment**

In a drifting environment, the state continuously shifted. The Kalman filter has been shown to describe the ideal observer in such drifting environments (Daw et al., 2006; S. Lee et al., 2020).

\[
\alpha_t = \frac{\sigma_{s,t}^2}{\sigma_{s,t}^2 + \sigma_N^2}
\]  

(7)

where \( \sigma_{s,t}^2 \) is the variance of the state distribution on the current trial. The learning rate is high when participants are more uncertain about the current state or the observed outcome is less uncertain. The variance of the state is updated as the following formula:

\[
\sigma_{s,t+1}^2 = (1 - \alpha_t)\sigma_{s,t}^2 + D
\]  

(8)
where $D$ is the drifting rate. That is, the variance of the state is discounted as a new observed outcome is received but is enhanced due to the drifting of the state.

Based on Eq. 7 and 8, learning rates are not influenced by prediction errors but are influenced by $\sigma^2_N$ and $D$ only. Given fixed $\sigma^2_N$ and $D$, namely, there is an asymptotic learning rate (S. Lee et al., 2020). The ideal learning rate is either high ($\alpha = 0.65$ for $\sigma_N = 10$ and $D = 11$) or low ($\alpha = 0.35$ for $\sigma_N = 25$ and $D = 11$) for the runs with different environment noise. Through the simulation, trial-by-trial learning rates in the Kalman filter converged at trial 5 (with the difference of learning rates between the adjacent trials smaller than 0.01).

**Physiological data acquisition**

During each run, we continuously collected four physiological measures: pupil diameter, skin conductance, electrocardiogram (ECG), and respiration. A Tobii T60XL eye tracker was used to collect pupil diameter at 60Hz while a Biopac MP150 system was used to collect skin conductance, ECG and respiration at 1000Hz. For skin conductance, two Ag/AgCl sensors were attached to the distal phalanges of the first and middle figures of the left hand. For ECG, three Ag/AgCl sensors were attached to the left arm, right arm and left leg. For respiration, a belt was tightened under participants’ chest.

**Physiological data analysis**

**Data preprocessing**

We preprocessed physiological measures for each run separately, and focused on pupil diameters, skin conductance (SC), heart rates (HR) and respiration rates (RSP)
in our analyses. For pupil diameters, we removed time points of eye blinks or missing
data, interpolated the removed time points, and removed the influence of gaze positions
on pupil diameters by regressing out the distance between the gaze positions and the
screen center. For heart rates, we calculated the intervals between two adjacent peaks
of ECG, transformed each interval into beats per minute (BPM), and aligned the
transformed BPM to the midpoint of the interval, and implemented interpolation to
acquire continuous BPM at each time point. For respiration rates, we calculated the
intervals between two adjacent cycles of respiration, and then implemented the same
procedures as heart rates to acquire continuous BPM at each time point.

We aimed to investigate fast and slow change of physiological arousal, so we
implemented two types of first-order Butterworth filters to extract phasic and tonic
signals, respectively. We used a band-pass filter with a cutoff of 0.02-4Hz to extract
phasic signals and used a low-pass filter with a cutoff of 0.02Hz to extract tonic signals
(Knapen et al., 2016). We then z-scored these two signals within each run, and down-
sampled all the measures to 10Hz. To avoid the influence of interpolation on pupil phasic
change, we only kept trials with good pupil measure for the following analyses related to
pupil phasic change. A trial with good pupil was defined as the proportion of interpolated
time points (due to the remove of eye blinks or missing data) between 0 and 3 seconds
being smaller than 50%. Participants with fewer than 30% of trials with good pupil
measures were removed from all the analyses.

*Phasic and tonic change*

We investigated the relationship between task factors and physiological arousal
and the relationship between learning rates and physiological arousal. Thus, we focused
on the physiological arousal during the period of 9 seconds after an outcome, and we used 100ms before an outcome as the baseline (Fig. 1A).

For each measure, we used phasic change and tonic change to capture fast and slow changes of arousal from the baseline, respectively. Phasic change and tonic change were calculated based on phasic signals and tonic signals, respectively. Pupil phasic change was calculated as the difference between the maximum peak within the period of 0-3 seconds and the average of phasic baseline. SC phasic change was calculated as the difference between the maximum peak within the period of 1-8 seconds and the average of phasic baseline. HR phasic change was calculated as the difference between the minimum peak within the period of 0-3 seconds and the average of phasic baseline. RSP phasic change was calculated as the difference between the minimum peak within the period of 5-8 seconds and the average of phasic baseline. On the other hand, tonic change of all the measures was calculated as the average within the period of 8-9 seconds and the average of tonic baseline.

To evaluate the relationship between different physiological measures, we calculated the correlation coefficients between these four measures for phasic change and tonic change separately. Correlation coefficients were calculated within each participant, and tested at the group level via two-tailed t-tests.

**Relationship between physiological arousal and task factors**

We examine how task factors influenced physiological arousal. Similar to the behavior analysis, we included five factors in a regression model: $I_{Low}$, $PE_t$, $PE_{t-1}$, $Value$ and, *Oddball*. In addition to these factors, several control factors were also included: tonic baseline on the current trial, tonic baseline on the previous trial, phasic change on
the previous trial, tonic change on the previous trial. We used these factors to predict phasic change and tonic change of each measure separately. These regression models were fitted through the linear mixed effects model separately. Random intercepts and random slopes were also estimated for participants.

We also implemented similar analyses on phasic time course to validate our results of phasic change. Phasic time course was baseline-corrected by phasic baseline. In the regression model, we included the following factors: \( I_{low} \), \( PE_t \), \( PE_{t-1} \), \( Value \), Oddball, and tonic baseline. The regression model was fitted through the linear mixed effects model at each time point between 0 and 9 seconds. Random intercepts and random slopes were also estimated for participants. For the correction of multiple comparison across time points, we implemented cluster-based permutation tests. We first identified clusters, which were formed by adjacent significant time point with \( p<0.05 \). For each cluster, we calculated the cluster mass as the summation of \( t \) value. We repeated one-sample permutation on participants’ regression coefficients 5,000 times at all the time points to form a null distribution of cluster mass. We then acquired \( p \) values of identified clusters from this null distribution.

**Relationship between physiological arousal and learning rates**

We evaluated the relationship between physiological arousal and learning rates. In a regression model, phasic change and tonic change of the four measures were included. To take into account the negative relationship between tonic baseline and phasic change (Gilzenrat, Nieuwenhuis, Jepma, & Cohen, 2010; Joshi et al., 2016), tonic baseline of the four measures were also included as control variables. This regression
model was fitted through linear mixed effects model for each dynamic environment separately. Random intercepts and random slopes were also estimated for participants.

**Mediation analysis**

We investigated whether physiological arousal served as a mediator between task factors and learning rates. Linear mixed effects mediation analysis was implemented through a MATLAB mediation toolbox (Wager, Davidson, Hughes, Lindquist, & Ochsner, 2008; Wager et al., 2009).

**Improvement for the prediction of learning rates**

In this analysis, we investigated whether physiological arousal per se improve the prediction of learning in addition to task factors. Specifically, we compared the performance of three models that were used to predict learning rates. These models were fitted using linear mixed effects. The first model is a base model which includes constants only. The second model is the same behavior model we used in the behavior analysis. The third model is the behavior model plus one type of physiological arousal. We repeated this analysis by including different type of physiological arousal. We used Log-likelihood ratio test to compare the model performance between these three models.

**Relationship between physiological arousal and model factors in the change-point environment**

To link the current results with the model-based effects on pupil diameter in the change-point environment in our previous studies (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Nassar et al., 2012), we also examine how physiological arousal encodes
CPP and RU. Trial-by-trial CPP and RU were calculated from Eq. 5 and Eq. 6, respectively. We included four factors in a regression model: CPP, RU, Value and, Oddball. In addition to these factors, several control factors were also included: tonic baseline on the current trial, tonic baseline on the previous trial, phasic change on the previous trial, tonic change on the previous trial. We used these factors to predict phasic change and tonic change of each measure separately. In addition, we investigated the influence of these four factors on tonic baseline, and included tonic baseline on the previous trial, phasic change on the previous trial, tonic change on the previous trial as control variables. These regression models were fitted through linear mixed effects model separately. Random intercepts and random slopes were also estimated for participants.

We also implemented similar analyses on phasic time course, which was baseline-corrected by phasic baseline. In the regression model, we included the following factors: CPP, RU, Value, Oddball, and tonic baseline. The regression model was fitted through linear mixed effects model at each time point between 0 and 9 seconds. Random intercepts and random slopes were also estimated for participants. Similar to the above analysis on the time course, statistical tests were implemented through cluster-based permutation tests.

**Results**

*Influence of task factors on learning rates*

To compare learning between the two dynamic environments, we investigated the influence of task factors shared across both environments on learning rates. We first
described the prediction of an ideal observer. In the change-point environment (Fig. 2A), environmental noise had little influence on ideal learning rates. Ideal learning rates are modulated by the current prediction errors since the size of the prediction error can reflect the likelihood that the state has changed. After an unexpected large prediction error (which may indicate a change-point), learning rates would gradually decrease as more precise estimates of the new state are received (which reflects the decrease of belief uncertainty). In the drifting environment (Fig. 2B), ideal learning rates were affected by the environment noise only, with higher learning rates in a low-noise condition than in a high-noise condition. Additionally, ideal learning rates were not affected by the incidental manipulations (i.e., monetary value and oddball sound) since these manipulations were not related to the current outcome or state. The influence of task factors on ideal learning rates can be also seen in the regression analysis, which include all the task factors (Fig. 2C). After controlling other factors, ideal learning would be higher in the low-noise condition than in the high-noise condition in the change-point environment.

Participants’ belief updating was partially consistent with the ideal observer. In the change-point environment (Fig. 2A), participants’ average learning rates were significantly higher than the ideal learning rates in both low-noise condition (mean±SEM=0.126±0.045, t=2.80, p=0.009) and high-noise condition (mean±SEM=0.123±0.046, t=2.68, p=0.012). In the drifting environment (Fig. 2B), participants’ average learning rates was similar to the ideal learning rate in the low-noise condition (p=0.10) but were higher than the ideal learning rate in the high-noise condition (mean±SEM=0.182±0.032, t=5.78, p<0.001). In the regression analysis in the change-
point environment (Fig. 2C), consistent with the ideal observer, participants’ learning rates were modulated by the noise conditions (mean±SEM=0.055±0.015, t=3.57, p=0.002), current prediction errors (mean±SEM=0.040±0.005, t=8.69, p<0.001) and previous prediction errors (mean±SEM=0.001±0.000, t=3.64, p=0.002). In the regression analysis in the drifting environment (Fig. 2C), similar to an ideal observer, participants’ learning rates were modulated by the noise conditions (mean±SEM=0.089±0.029, t=3.07, p=0.005) but this modulation was weaker than that on the ideal learning rates. However, inconsistent with an ideal observer, participants’ learning rates were also enhanced by the current prediction errors (mean±SEM=0.048±0.011, t=4.22, p<0.001).

Unlike an ideal observer, participants’ learning rates were influenced by the incidental reward (Fig. 2C). Incidental reward enhanced learning rates in both the change-point (mean±SEM=0.033±0.009, t=3.81, p=0.001) and the drifting environment (mean±SEM=0.036±0.011, t=3.16, p=0.004). In contrast, incidental auditory oddball showed no effect on the learning rates in either change-point (p=0.16) or drifting environments (p=0.92).
Figure 2-2 Behavior results

(A) Summary of learning rates in the change-point environment. In the first panel, the average learning rates were shown. Each data represented participants’ learning rates. The gray line connected one participant’s learning rates between the two noise conditions. The diamond reflected the group average. The dashed line indicated an ideal observer. Participants’ learning
rates were higher than the ideal learning rates in both noise conditions. In the second panel, participants' learning rates increased as prediction errors (PEs) increased. Error bars indicated one SEM. In the third panel, learning rates increased as an unexpected large PE (which was larger than twice of the environmental noise) occurred and learning rates gradually decreased thereafter. In the fourth panel, learning rates were higher in the trials with monetary value than in the trials with neutral value. In the fifth panel, learning rates were not different between trials with oddball or standard sound. (B) Summary of learning rates in the drifting environment. In the first panel, the average learning rate was close to the ideal learning rate in the low-noise condition but was higher than the ideal learning rate in the high-noise condition. In the second panel, participants' learning rates increased as PEs increased. In the third panel, learning rates increased as a large PE occurred. In the fourth panel, learning rates were higher in the trials with monetary value than in the trials with neutral value. In the fifth panel, learning rates were not different between trials with oddball or standard sound. (C) Regression of task factors on learning rates. For the purpose of display, the coefficients of PE_t and PE_{t-1} were multiplied by 10. Error bars indicated one SEM. The diamond indicates the regression coefficients of an ideal observer. **p<0.01, ***p<0.001.
Relationships between physiological measures

We examined phasic change and tonic change from each measure of physiological arousal (Fig. 3A). The phasic change reflects the fast change after the outcome while the tonic change reflects the slow change between the outcome and the next prediction. We evaluated the relationships between different physiological measures for phasic and tonic change. For the phasic change (Fig. 3B; Supplementary Fig. 1A), pupil diameter was significantly positively correlated with skin conductance in both change-point (mean±SEM=0.063±0.021, t=3.08, p=0.005) and drifting environments (mean±SEM=0.077±0.024, t=3.28, p=0.003). For the tonic change (Fig. 3C; Supplementary Fig. 1B), pupil diameter was significantly positively correlated with heart rates (change-point: mean±SEM=0.131±0.025, t=5.31, p<0.001; drifting: 0.192±0.036, t=5.41, p<0.001), skin conductance (change-point: mean±SEM=0.034±0.016, t=2.21, p=0.035; drifting: 0.042±0.016, t=2.55, p=0.017) and respiration rates (change-point: mean±SEM=0.047±0.021, t=2.23, p=0.034; drifting: 0.094±0.034, t=2.80, p=0.009) in both environments.
Figure 2-3 Demonstration for different measures of physiological arousal

(A) Time course of physiological measures. For each measure, the time course was z-scored, baseline-corrected, and averaged across all participants. The gray shaded under the time course indicated one SEM. The yellow area indicated the period of baseline, which was calculated as the average of 100 milliseconds before the onset of outcome. The red area indicated the period of phasic change. For the pupil (0-3s) and skin conductance (1-8s), the phasic change was calculated as the difference between the maximum peak in this period and the baseline. For the heart rate (0-3 s) and respiration rate (5-8s), the phasic change was calculated as the difference between the minimum peak in this period and the baseline. The blue area indicated the period of tonic change. In all the physiological measures (8-9s), the tonic change was calculated as the difference between the average in this period and the baseline. (B) Correlations of phasic change between physiological measures. In both environments, pupil was positively correlated with skin conductance. (C) Correlations of tonic change between physiological measures. In both
environments, pupil was positively correlated with heart rates, skin conductance and respiration rates. *p<0.05, **p<0.01, ***p<0.001.
Physiological arousal encoded task factors

We examined how different types of physiological arousal encode different task factors. Similar to the behavior model, we focused on the noise condition, the current prediction error, the previous prediction error, monetary value and oddball sound. Among different measures of physiological arousal, only pupil diameter consistently showed the same response to the same task factors across the two environments. For phasic change (Fig. 4), pupil diameter responded to monetary value (change-point: mean±SEM=0.124±0.041, t=3.02, p=0.006; drifting: 0.108±0.043, t=2.50, p=0.018) and oddball sound (change-point: mean±SEM=0.138±0.049, t=2.82, p=0.006; drifting: 0.233±0.078, t=2.99, p=0.006) consistently in both environments. Time course analyses for phasic change are shown in Supplementary Fig. 2. For tonic change (Fig. 5), pupil diameter responded to the current prediction error consistently in both environments (change-point: mean±SEM=0.012±0.003, t=4.42, p<0.001; drifting: 0.029±0.008, t=3.51, p=0.002). Additionally, pupil tonic change was reduced in response to the previous prediction error in the change-point environments (mean±SEM=-0.007±0.003, t=-2.53, p=0.019).
Figure 2-4 Regression of task factors on phasic change

(A) Regression of task factors on pupil phasic change. For the purpose of display, the coefficients of PE_t and PE_{t-1} were multiplied by 10. Error bars indicated one SEM. (B) Regression of task factors on heart rate phasic change. (C) Regression of task factors on skin conductance phasic change. (D) Regression of task factors on respiration rate phasic change. *p<0.05, **p<0.01.
Figure 2-5 Regression of task factors on tonic change

(A) Regression of task factors on pupil tonic change. For the purpose of display, the coefficients of PE_t and PE_{t-1} were multiplied by 10. Error bars indicated one SEM. (B) Regression of task factors on heart rate tonic change. (C) Regression of task factors on skin conductance tonic change-point. (D) Regression of task factors on respiration rate tonic change.
change. (D) Regression of task factors on respiration rate tonic change. *\(p<0.05\), **\(p<0.01\), ***\(p<0.001\).
Physiological arousal predicted learning rates

We investigated which type of physiological arousal predicts trial-by-trial learning rates. Among different types of physiological arousal, only pupil tonic change consistently predicted learning rates in both environments (Fig. 6A; change-point: mean±SEM=0.046±0.012, t=3.76, p<0.001; drifting: 0.039±0.012, t=3.15, p=0.004).

Pupil tonic change mediated the relationship between the current prediction errors and learning rates

From the previous analyses, we identified relationships between the current prediction errors, pupil tonic change and learning rates in both environments. We further examined whether pupil tonic change mediated the relationship between the current prediction errors and learning rates. In a mediation analysis (Fig. 6B), pupil tonic change mediated this relationship in the change-point environment (p=0.002) but not in drifting environment (p=0.49).

Pupil tonic change improved the prediction of learning rates

We further investigated whether pupil tonic change per se provides additional information for the prediction of the dynamic learning rates. We compared three models for the prediction of learning rates: base (intercepts only), behavior, and behavior plus pupil tonic change (Fig. 6C). Through the log-likelihood ratio test, both behavior (change-point: improvement of log-likelihood=671.39, p<0.001; drifting: 289.99, p<0.001) and behavior plus pupil tonic change model (change-point: improvement of log-likelihood=683.92, p<0.001; drifting: 305.54, p<0.001) performed better than the base model in both environments. Furthermore, adding the pupil tonic change to the behavior
model (behavior + pupil tonic change) improved the model performance for predicting learning rates (change-point: improvement of log-likelihood=12.53, \( p=0.003 \); drifting: 15.56, \( p<0.001 \)).
**Figure 2-6 Functional roles of pupil tonic change**

(A) Regression of different physiological responses on learning rates. In both environments, pupil tonic change was positively correlated with learning rates. (B) Path analysis between current prediction errors, pupil tonic change and learning rates. In the change-point environment, pupil tonic change mediated the relationship between current prediction errors and learning rates. (C) Model improvement by pupil tonic change. Adding pupil tonic change to the behavioral model
Fig. 2C) improved the model predictions for belief updating in both environments. *$p<0.05$, **$p<0.01$, ***$p<0.001$. 
Physiological arousal encoded CPP and RU in the change-point environment

To link the current results with the model-based effects on pupil diameter in the change-point environment in our previous studies (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Nassar et al., 2012), we also examine how physiological arousal encode CPP and RU. Specifically, we investigated the tonic baseline before the outcome, the phasic change and the tonic change. RU was encoded by pupil baseline (Supplementary Fig. 3A) and pupil phasic change (Supplementary Fig. 4A). This effect of RU was also seen during the presentation of outcome in the pupil phasic time course (Supplementary Fig. 4B). On the other hand, CPP was encoded by pupil tonic change (Supplementary Fig. 5A). The results on other measures of physiological arousal can be seen in Supplementary Fig. 3, Supplementary Fig. 4 and Supplementary Fig. 5.

Discussion

Summary of results

Among different measures of physiological arousal, pupil diameter was consistently reflected task factors and predicted learning rates across the two different kinds of dynamic environments. Pupil tonic change not only served as response to the current prediction errors but also directly tracked learning rates. First, pupil tonic change served as a mediator between the current prediction errors and learning rates in the change-point environment. Second, pupil tonic change captured the variance of learning rates that cannot be account for by task factors in both environments. In both environments, pupil phasic change was induced by both incidental reward and incidental auditory oddball but only incidental reward enhanced learning rates.
Learning in dynamic environments

In the change-point environment, where the stable state undergoes sudden changes, people’s learning should be driven by belief surprise and belief uncertainty (Krishnamurthy et al., 2017; McGuire et al., 2014; Nassar et al., 2012; Nassar et al., 2010). Belief surprise and belief uncertainty was related to several task factors in this study. Participants’ learning rates increased as the current prediction errors increased. This relationship suggested the influence of unexpected uncertainty on belief updating (Payzan-LeNestour & Bossaerts, 2011). That is, an unexpected large prediction errors reflected unexpected uncertainty and indicated the change of state (Nassar et al., 2010). Participants’ learning rates were also influenced by belief uncertainty, which was related to the noise condition and the previous prediction errors. Participants had higher learning rates in the low-noise condition since the lower noise would make belief uncertainty higher (Nassar et al., 2010). In addition, participants’ learning rates increased as the previous prediction errors increased since these prediction errors reflected the imprecise estimates about the state in their belief (Nassar et al., 2012; Nassar et al., 2010).

In the drifting environment, where the current state is shifted from the previous state, people should adjust learning according to belief uncertainty alone (Daw et al., 2006; S. Lee et al., 2020). The uncertainty of participants’ belief come from two sources: the uncertainty due to the drift of state and the uncertainty due to the generation of outcome. Given fixed drifting rate and fixed noise level in one condition, an ideal observe should use a consistent learning rate and should not be influenced by prediction errors (Daw et al., 2006; S. Lee et al., 2020). As predicted, participants’ learning rates were higher in the low-noise condition than in the high-noise condition. However, inconsistent
with an ideal observer and our previous study (S. Lee et al., 2020), participants’ learning rates were modulated by the current prediction errors. Though the influence of the current prediction errors was lower than that in the change-point environment. One possibility would be the different task presentation between this study and our previous study. That is, the drifting environment in this study was the same as that in our previous study but we used different visual presentation from our previous study. Our previous study presented the task as a helicopter video game, where participants estimated the prediction errors as the visual distance on the screen. However, this study used specific numbers so participants need to mathematically calculate the prediction errors as the difference between numbers. Thus, it may be easier to implement a consistent learning rate strategy with the visual cue in our previous study. In addition, participants may be influenced by the fluctuation of unexpected uncertainty in this study. That is, the uncertainty that comes from the drift of state and the noise of outcome should be stable on average but the trial-by-trial uncertainty would be fluctuated due to the fluctuation of current prediction errors and introduced unexpected uncertainty. Mathematically calculating the difference between numbers may enlarge the influence of unexpected uncertainty on learning rates.

Incidental reward enhanced learning rates but there was no influence from incidental auditory oddball. Even though the incidental reward provided no information about the current state, monetary value enhanced participants’ learning rates compared with neutral value (S. Lee et al., 2020; McGuire et al., 2014). Moreover, this incidental reward was not shown between large monetary value and small monetary value (Kao, Lee, Gold, & Kable, 2020). These results suggested that monetary value regardless of its size can modulate learning rates. Unlike our previous study (Nassar et al., 2012), we
did not identify the effect of incidental auditory oddball. This difference may be due to the
difference of manipulation for auditory stimuli. In our previous study (Nassar et al.,
2012), the incidental auditory stimulus would occasionally change to another new one,
randomly selected from a library of sounds. In this study, we used two auditory stimuli
with one as an oddball sound and the other as a standard sound. The frequency of
change was similar (~0.1) for this study and our previous study. Future studies should
further which type of manipulation for incidental auditory stimuli can effectively
modulated learning rates.

*The roles of physiological arousal during learning*

Only pupil diameter was associated with task factors and learning rates across
the two dynamic environments. This association may imply that pupil diameter is a more
sensitive measure of arousal. However, this might not be the main cause since different
measures of physiological arousal shared similar components of arousal during the task.
That is, pupil phasic change was correlated with skin conductance phasic change and
pupil tonic change was correlated with tonic change of the other three measures. This
may suggest that all the measures reflected general arousal regulated by ANS but
learning-related pupil diameter reflected activity of LC-NE system.

Pupil diameter responded to task factors in the two dynamic environments. Pupil
phasic change was induced by incidental stimuli. Extending from our previous studies
(S. Lee et al., 2020; McGuire et al., 2014), we showed that the monetary value induced
higher pupil phasic change than the neutral value. Similar to our previous study (Nassar
et al., 2012), pupil phasic change responded to the change of the auditory stimulus.
On the other hand, pupil tonic change responded to the current prediction errors, which modulated learning rates in both environments. These prediction errors potentially indicated surprise or unexpected uncertainty (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Nassar et al., 2012; O’Reilly et al., 2013; Preuschoff, ’t Hart, & Einhauser, 2011). In the model-based analysis, pupil tonic change also responded CPP in the change-point environment. Although an ideal observer should not respond to these prediction errors in the drifting environment, the response of pupil tonic change to these prediction errors may reflect the fluctuation of unexpected uncertainty. These results were also associated with the role of LC-NE system for unexpected uncertainty (Yu & Dayan, 2005).

Unlike previous studies, we showed the encoding of surprise on a later time point after the outcome. Past studies showed that surprise induced fast evoked change (~2 seconds after an outcome) (Browning et al., 2015; Filipowicz et al., 2020; Krishnamurthy et al., 2017; Nassar et al., 2012; O’Reilly et al., 2013; Preuschoff et al., 2011), while we showed this effect in a later time (~8 seconds after an outcome). The major difference between our study and previous studies is the time interval between the outcome and the subsequent prediction. That is, past studies usually have a short interval, while we have a long interval. Thus, across our studies and past studies, surprise-induced pupil change occurred right before when participants need to use surprise to update their belief.

Pupil tonic change mediated the relationship between the current prediction errors and learning rates, and tracked the variance of learning rates that cannot be accounted for by task factors. These results suggested that pupil tonic change worked more than just response to the current prediction errors. Moreover, it may play a causal
role on learning. First, pupil diameter served as a bridge between the current prediction errors and learning rates. This evidence may reflect the plausible function of neural gain regulated by arousal and LC-NE system (Aston-Jones & Cohen, 2005; Mather, Clewett, Sakaki, & Harley, 2015). Neural gain means that the activation of a neuron that receives excitatory input would increase but the activation of a neuron that receives inhibitory input decreases. Recent studies showed arousal (such as pupil diameter) enhanced attention or performance for salient stimuli and inhibited attention or performance for non-salient stimuli (Eldar, Cohen, & Niv, 2013; T.-H. Lee, Itti, & Mather, 2012; T.-H. Lee, Sakaki, Cheng, Velasco, & Mather, 2014). During learning, the prediction error reflected the deviation of the new observation from the belief and may serve as a salient signal. Arousal may enhance the processing of this salient information, leading to the increase of learning rates. Second, pupil tonic change tracked the variance of learning rates that cannot be accounted for by task factors. This result suggested that the manipulation of arousal level or activity in LC-NE system may drive the change of belief. Recent studies showed that directly stimulating LC or increase the level of NE would lead to more exploration behavior or the change of belief (Jepma et al., 2018; Jepma et al., 2016; Kane et al., 2017). This fluctuation of arousal may drive the network reset in the brain (Bouret & Sara, 2005; Servan-Schreiber, Printz, & Cohen, 1990). Such network reset was associated with the change of belief in the dynamic environment (Kao, Kambhati, et al., 2020; Nassar, McGuire, Ritz, & Kable, 2019).

Replication of the effects of pupil diameter on previous studies

We replicated the effects of pupil diameter on previous studies. Pupil diameter responded to prediction errors and, combined with the evidence in this study and our
past study (Filipowicz et al., 2020), we showed this encoding might be important right before the subsequent prediction. In the model-based analysis, this effect of prediction errors was associated with surprise or unexpected uncertainty (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Nassar et al., 2012; O’Reilly et al., 2013; Preuschoff et al., 2011). Similar to the previous studies, we also showed that belief uncertainty was encoded on pupil baseline and evoked change after the outcome (Filipowicz et al., 2020; Krishnamurthy et al., 2017; Muller, Mars, Behrens, & O’Reilly, 2019; Nassar et al., 2012; Urai, Braun, & Donner, 2017). As belief uncertainty is high, people should rely more on the new observation for belief updating (i.e., high learning rates). This response on pupil diameter may suggest the evaluation of belief uncertainty for information integration. We also showed that pupil diameter played an important role to predict learning rates (Krishnamurthy et al., 2017; Nassar et al., 2012).
Supplementary figure 1 (A) Distribution for correlations of phasic change between physiological measures. The red line indicated the mean of correlation coefficient. (B) Distribution for correlations of tonic change between physiological measures.
Supplementary figure 2 (A) Regression of task factors on pupil phasic time course. For the purpose of display, the coefficients of $PE_t$ and $PE_{t-1}$ were multiplied by 10. Shaded areas indicated one SEM. Colorful horizontal lines indicated the significant time points after the correction by the cluster-based permutation test. (B) Regression of task factors on heart rate phasic time course. (C) Regression of task factors on skin conductance phasic time course. (D) Regression of task factors on respiration rate phasic time course.
Supplementary figure 3 (A) Regression of model factors on pupil baseline. Error bars indicated one SEM. (B) Regression of model factors on heart rate baseline. (C) Regression of model factors on skin conductance baseline. (D) Regression of model factors on respiration rate baseline. ***p<0.001.
Supplementary figure 4 (A) Regression of model factors on pupil phasic change. Error bars indicated one SEM. (B) Regression of model factors on pupil phasic time course. Shaded areas indicated one SEM. Colorful horizontal lines indicated the significant time points after the correction by the cluster-based permutation test. (C) Regression of model factors on heart rate phasic change. (D) Regression of model factors on heart rate phasic time course. (E) Regression of model factors on skin conductance phasic change. (F) Regression of model factors on skin conductance phasic time course. (G) Regression of model factors on respiration rate phasic change. (H) Regression of model factors on respiration rate phasic time course.
conductance phasic time course. (G) Regression of model factors on respiration rate phasic change. (H) Regression of model factors on respiration rate phasic time course. *$p<0.05$, **$p<0.01$. 
**Supplementary figure 5** (A) Regression of model factors on pupil tonic change. Error bars indicated one SEM. (B) Regression of model factors on heart rate tonic change. (C) Regression of model factors on skin conductance tonic change. (D) Regression of model factors on respiration rate tonic change. *p<0.05, **p<0.01.
CHAPTER 3 – Functional brain network reconfiguration during learning in a dynamic environment


Abstract

When learning about dynamic and uncertain environments, people should update their beliefs most strongly when new evidence is most informative, such as when the environment undergoes a surprising change or existing beliefs are highly uncertain. Here we show that modulations of surprise and uncertainty are encoded in a particular, temporally dynamic pattern of whole-brain functional connectivity, and this encoding is enhanced in individuals that adapt their learning dynamics more appropriately in response to these factors. The key feature of this whole-brain pattern of functional connectivity is stronger connectivity, or functional integration, between the fronto-parietal and other functional systems. Our results provide new insights regarding the association between dynamic adjustments in learning and dynamic, large-scale changes in functional connectivity across the brain.
Introduction

Human decisions are guided by beliefs about current features of the environment. These beliefs often must be inferred from indirect and uncertain evidence. For example, deciding to go to a restaurant typically relies on a belief about its current quality, which can be inferred from past experiences at that restaurant. This inference process is particularly challenging in dynamic environments whose features can change unexpectedly (e.g., a new chef was just hired). In these environments, people tend to follow normative principles and update their beliefs dynamically and adaptively, such that beliefs are updated more strongly when existing beliefs are weak or irrelevant, and/or the new evidence is strong or surprising (McGuire et al., 2014; Nassar et al., 2012; Nassar et al., 2010). Recent studies have identified potential neural substrates of this adaptive belief-updating process, including univariate and multivariate activity patterns for uncertainty and surprise in several brain regions including dorsomedial frontal cortex, anterior insula, lateral prefrontal cortex, and lateral parietal cortex (Behrens et al., 2007b; McGuire et al., 2014; Meder et al., 2017; Nassar, McGuire, et al., 2019; O’Reilly et al., 2013). The goal of the present study was to gain deeper insights into how these representations might interact dynamically to support adaptive belief updating.

We focused on how changes in belief updating relate to changes in functional connectivity between brain regions with task-relevant activity modulations. Functional connectivity reflects statistical dependencies between regional activity time series (Friston, 2011) and can form functional-connectivity networks that provide new perspectives on brain function (Bassett & Sporns, 2017; Bullmore & Sporns, 2009; Medaglia et al., 2015). Many recent studies of learning have focused on brain network
reconfigurations occurring between naïve and well-learned phases in various domains such as motor, perceptual, category, spatial, or value learning (Antzoulatos & Miller, 2014; Baeg et al., 2007; Bassett et al., 2011; Bassett et al., 2015; Büchel et al., 1999; Fatima, Kovacevic, Misic, & McIntosh, 2016; Gerraty et al., 2018; Lewis, Baldassarre, Committeri, Romani, & Corbetta, 2009; Mattar, Thompson-Schill, & Bassett, 2018; Mattar, Wymbs, et al., 2018; Sun et al., 2006). In these cases, functional connectivity associated with the fronto-parietal system decreased gradually as learning progressed and this change in connectivity was associated with individual learning or performance (Bassett et al., 2015; Büchel et al., 1999; Sun et al., 2006). In dynamic environments, however, people progressively learn the current state and then re-initialize learning once the state changes. Thus, we expected frequent reconfigurations in functional connectivity, as learning shifts between slower and faster updating in response to changes in uncertainty and surprise. Additionally, although brain regions that encode uncertainty and surprise participate in multiple networks, including the fronto-parietal system, dorsal attention system, salience system, and memory system (Behrens et al., 2007b; McGuire et al., 2014; Meder et al., 2017; Nassar, McGuire, et al., 2019; O’Reilly et al., 2013), based on previous network analyses of learning in stable environments we hypothesized that the fronto-parietal system would serve a particularly important role in network reconfiguration during learning in dynamic environments.

In the current study, we aimed to identify such frequent reconfigurations in functional connectivity during adaptive belief updating. A key to our approach was the use of an unsupervised machine-learning technique known as non-negative matrix factorization (NMF) (D. D. Lee & Seung, 1999). NMF decomposes the whole-brain network into subgraphs, which describe patterns of functional connectivity across the
entire brain, and the time-dependent magnitude with which these subgraphs are expressed. Briefly, a subgraph is a weighted pattern of functional interactions that statistically recurs as the brain network evolves over time. We chose NMF because it provides two key advantages over other approaches to matrix factorization, such as principal components analysis (PCA) or independent components analysis (ICA) (Chai et al., 2017; Khambhati, Mattar, Wymbs, Grafton, & Bassett, 2018). First, NMF yields a parts-based representation of the network, in which the individual components are strictly additive – a constraint that is not present in PCA and ICA. This important feature enables interpretation of the resulting subgraph and time-dependent expression coefficients on the basis of their positive distance from zero. Second, NMF does not enforce an orthogonality or independence constraint and, therefore, allows subgraphs to overlap in their structure. This property may more effectively model distinct subgraphs that may be jointly related via weak connections and better account for the flexibility of neural systems, such that one connection between regions can be involved in multiple systems or cognitive functions. Recently, NMF has been used to identify network dynamics during rest and task states (Khambhati, Mattar, et al., 2018; Khambhati, Medaglia, Karuza, Thompson-Schill, & Bassett, 2018) and to determine how these dynamics vary across development (Chai et al., 2017). Here we extend the use of this technique to examine changes in network dynamics linked to task variables and individual differences.

Our results show that that uncertainty and surprise, task variables that drive the adjustment of learning, are related to the temporal expression of specific patterns of functional connectivity (i.e., specific subgraphs). These specific patterns of functional connectivity prominently involve the fronto-parietal network. We also show that the
dynamic modulation of these patterns of functional connectivity (i.e., subgraph expression) are associated with individual differences in learning.

Methods

Participants

The dataset has been described in our previous reports (McGuire et al., 2014). Thirty-two individuals participated in the fMRI experiment: 17 females and 15 males, mean age = 22.4 years (SD = 3.0; range 18-30). Human subject protocols were approved by the Internal Review Board in University of Pennsylvania. All participants provided informed consent before the experiment.

Task

Each participant completed four 120-trial runs during functional magnetic resonance imaging. In each run, participants performed a predictive-inference task (Fig. 1a). On each trial, participants made a prediction about where the next bag would be dropped from an occluded helicopter by positioning a bucket along the horizontal axis (0-300) of the screen. The location of the bag was sampled from a Gaussian distribution with a mean (the location of the helicopter) and a standard deviation (noise). The standard deviation was high (SD = 25) or low (SD = 10) in different runs. The location of the helicopter usually remained stable but it changed occasionally. The probability of change was zero for the first three trials after a change and 0.125 for the following trials. When the location changed, the new location was sampled from a uniform distribution. Correctly predicting the location of the bag resulted in coins landing in the bucket. These
coins either had positive or neutral value depending on their color, which was randomly assigned for each trial.
Figure 3-1 Overview of the task and theoretical model of belief updating (McGuire et al., 2014)

a Sequence of the task. At the start of each trial, participants predict where a bag will drop from an occluded helicopter by positioning a bucket on the screen. After participants submit their prediction, the bag drops and any rewarded coins that fall in the bucket are added to the participant’s score. The location of the last prediction and the last bag drop are noted on the next trial.

b An example sequence of trials. Each data point represents the location of a bag on each trial (yellow for rewarded coins, gray for neutral coins). The dashed line represents the true generative mean. The mean changes occasionally. The cyan line represents the prediction from a normative model of belief updating. The inset equation shows how the model updates beliefs ($B_t$ = belief, $X_t$ = observed outcome, $\alpha_t$ = learning rate on trial $t$). The vertical dashed line represents the boundary of the noise conditions: high-noise (left) and low-noise condition (right). Noise refers
to the variance of the generative distribution. c Two learning components from the normative model. Change-point probability (CPP) reflects the likelihood that a change-point happens, which is increased when there is an unexpectedly large prediction error. Relative uncertainty (RU) reflects the uncertainty about the generative mean relative to the environmental noise, which is increased after high CPP trials and decays slowly as more precise estimates of the generative mean are possible. The inset formula shows how CPP and RU contribute to single trial estimates of learning rates.
Behavior model

We applied the same normative model described in our previous study (McGuire et al., 2014). An approximation to the ideal observer solution to this task updates beliefs according to a delta learning rule (Fig. 1b):

\[ \delta_t = X_t - B_t \]  
\[ B_{t+1} = B_t + \alpha_t \times \delta_t \]

where \( \delta_t \) is the prediction error, which is the difference between the observed outcome (bag drop location, \( X_t \)) and the prediction (bucket location, \( B_t \)). Beliefs are updated in proportion to the prediction error, and this proportion is determined by \( \alpha_t \), the learning rate. The learning rate is adjusted adaptively on each trial according to two normative factors (Fig. 1c):

\[ \alpha_t = \Omega_t + (1 - \Omega_t) \times \tau_t \]

where \( \Omega_t \) is the change-point probability (CPP) and \( \tau_t \) is the relative uncertainty (RU). The learning rate, CPP and RU are all constrained to be between zero and one, and the learning rate increases when either CPP or RU is high. CPP reflects the likelihood that a change-point has happened (McGuire et al., 2014; Nassar et al., 2010):

\[ \Omega_t = \frac{\nu(X_t | 0, 300)_H}{\nu(X_t | 0, 300)_H + \nu(\mathcal{N}(X_{t+1} | B_{t+1}, \sigma_t^2)^{(1-H)}} \]
where $U(X_t | 0, 300)$ indicates the probability of $X_t$ from a uniform distribution between 0 and 300, $N(X_t | B_t, \sigma^2_t)$ indicates the probability of $X_t$ from a Gaussian distribution with mean of $B_t$ and variance of $\sigma^2_t$. $\sigma^2_t$ is the variance of predictive distribution of the bag location, and $H$ is the average probability of change (0.1) across trials.

RU reflects the uncertainty about the current location of the helicopter relative to the amount of noise in the environment (McGuire et al., 2014):

\[
\tau_{t+1} = \frac{\Omega_t \sigma_N^2 + (1-\Omega_t) \tau_t \sigma_N^2 + \Omega_t (1-\Omega_t) (\delta_t (1-\tau_t))^2}{\Omega_t \sigma_N^2 + (1-\Omega_t) \tau_t \sigma_N^2 + \Omega_t (1-\Omega_t) (\delta_t (1-\tau_t))^2 + \sigma_N^2}
\]  

where $\sigma_N^2$ is the variance of outcome distribution used to generate the location of bag.

There are three terms present in both the numerator and denominator. The first term is the variance of the helicopter distribution conditional on a change-point while the second term is the variance of the helicopter distribution conditional on no change-point. The third term reflects the variance due to the difference in mean between the two conditional distributions. The three terms together capture the uncertainty about the location of the helicopter.

Figure 1c shows an example of the dynamics of CPP and RU. CPP increases when there is an unexpectedly large prediction error. RU increases after CPP increases and decays slowly as more precise estimates of the helicopter location are possible.

As in our previous study, a regression model was applied to investigate how the factors in this normative model, as well as other aspects of the task, influenced participants’ belief updates. We regressed trial-by-trial updates $(B_{t+1} - B_t)$ against the
prediction error \( (\delta_t) \), the interaction between prediction error and the two factors from the normative model, CPP \( (\Omega_t) \) and RU \( (\tau_t) \), as well as the interaction between prediction error and whether the outcome was rewarded or not (McGuire et al., 2014). The form of the regression model can be written as

\[
\text{Update}_t = \beta_0 + \beta_1 \delta_t + \beta_2 \delta_t \Omega_t + \beta_3 \delta_t (1 - \Omega_t) \tau_t + \beta_4 \delta_t \text{Reward}_t + \beta_5 \text{Edge}_t + \epsilon \quad (6)
\]

where Edge is regressor of no interest that captures the tendency to avoid updating toward the edges of the screen \( ((150 - B_{t+1})|150 - B_{t+1}) \). If subjects used a fixed learning rate (Equation 2 alone), \( \beta_2 \) and \( \beta_3 \) will be zero and \( \beta_1 \) will reflect that fixed learning rate. In contrast, if subjects behave exactly in accordance with the normative model (Equation 3), \( \beta_2 \) and \( \beta_3 \) will be one, and \( \beta_1 \) will be zero. Thus, we constructed the regression model so that the weights on \( \beta_2 \) and \( \beta_3 \) reflect the degree to which the two normative factors, CPP and RU, drive dynamic learning rates.

This regression model was fitted separately to each participant’s data to estimate the influence of each factor on each participant’s behavior. We used the residuals of this regression to examine the relationship between subgraph expression and residual updating. To examine the relationship between individual differences in normative learning and functional network dynamics, we used the sum of the regression coefficients on the CPP term \( (\beta_2) \) and the RU term \( (\beta_3) \) as an index of normative learning.

\textit{MRI data acquisition and preprocessing}
MRI data were collected on a 3T Siemens Trio with a 32-channel head coil. Functional data were acquired using gradient-echo echoplanar imaging (EPI) (3 mm isotropic voxels, 64 x 64 matrix, 42 axial slices tilted 30° from the AC-PC plane, TE = 25 ms, flip angle = 75°, TR = 2500 ms). There were four runs with 226 images per run. T1-weighted MPRAGE structural images (0.9375 X 0.9375 X 1 mm voxels, 192 X 256 matrix, 160 axial slices, TI = 1100 ms, TE = 3.11 ms, flip angle = 15°, TR = 1630 ms) and matched fieldmap images (TE = 2.69 and 5.27 ms, flip angle = 60°, TR = 1000 ms) were also collected. Data were preprocessed with FSL (M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, & S. M. Smith, 2012; Stephen M. Smith et al., 2004) and AFNI (Cox, 1996, 2012). Functional data were corrected for slice timing (AFNI’s 3dTshift) and head motion (FSL’s MCFLIRT), attenuated for outliers (AFNI’s 3dDespike), undistorted and warped to MNI space (FSL’s FLIRT and FNIRT), smoothed with 6 mm FWHM Gaussian kernel (FSL’s fslmaths) and intensity-scaled by the grand-mean value per run. Structural images were segmented into gray matter, white matter (WM) and cerebrospinal fluid (CSF) (FSL’s FAST) (Zhang, Brady, & Smith, 2001).

Constructing time-varying functional networks

For each run and each participant, blood-oxygenation-level-dependent (BOLD) time series were obtained from each of 264 regions of interest (ROIs; diameter = 9mm) based on the previously defined parcellation(Power et al., 2011). ROIs that did not have valid BOLD time series for all runs and all participants were removed, resulting in $N = 247$ ROIs. We visualized these ROIs on the brain using the BrainNet Viewer (https://www.nitrc.org/projects/bnv) (Xia, Wang, & He, 2013). For each BOLD time series, a band-pass filter was applied with a cutoff of 0.01-0.08 Hz. This low-frequency
band has been shown to reflect neuronal activation and neural synchronization (Biswal, Zerrin Yetkin, Haughton, & Hyde, 1995; Lu et al., 2007; Zuo et al., 2010). To remove the influence of head motion, a confound regression was implemented to regress out nuisance factors from each BOLD time series. This confound regression included 24 motion parameters (three translation and three rotation motion parameters and their expansion \([R_0 R^2_0 R^2_{t-1} R^2_{t-1}])\) (Friston, Williams, Howard, Frackowiak, & Turner, 1996), as well as average signals from WM and CSF (Fox et al., 2005).

In order to construct dynamic functional networks, we defined sliding time windows and calculated Pearson correlation coefficients between ROI time series in each sliding time window. We assigned these coefficients to the first TR in the time windows. To ensure magnetization equilibrium, the first 6 volumes of each run were removed from the analysis. For the rest of the volumes in each run, a sliding window was defined with a 10-TR (25 seconds) length and 80% overlap across windows. Each run had 106 sliding time windows, leading to \(T = 424\) sliding time windows for each participant. Each participant’s data thus formed a matrix of dynamic functional networks with dimensions \(N \times N \times T\). Then, we took each participant’s \(N \times N\) matrix and unfurled the upper triangle into an \(\frac{N(N-1)}{2}\) vector. By concatenating vectors across all time windows \((T)\), we obtained an \(\frac{N(N-1)}{2} \times T\) matrix. Furthermore, we concatenated matrices from \(S = 32\) participants to form a \(\frac{N(N-1)}{2} \times (T \times S)\) matrix. To ensure that our approach did not give undue preference to either positively or negatively weighted functional edges, we separated this matrix into two thresholded matrices: one composed of positively weighted edges, and one composed of negatively weighted edges. That is, in the matrix of positive functional correlations between ROI time series, the original negative
correlations between ROI time series were set to 0; in the matrix of negative functional correlations between ROI time series, all values were multiplied by -1, and the original positive functional correlations between ROI time series were set to 0. After concatenating the matrix composed of positively weighted edges and the matrix of negatively weighted edges, we had a final \( \frac{N(N-1)}{2} \times (T \times S \times 2) \) matrix \( \mathbf{A} \).

*Clustering functional networks into subgraphs*

We applied an unsupervised machine learning algorithm called non-negative matrix factorization (NMF) (D. D. Lee & Seung, 1999) on \( \mathbf{A} \) to identify subgraphs \( \mathbf{W} \) and the time-dependent expressions of subgraphs \( \mathbf{H} \). The matrix factorization problem \( \mathbf{A} \approx \mathbf{W} \mathbf{H} \) s. t. \( \mathbf{W} \geq 0, \mathbf{H} \geq 0 \) was solved by optimization of the cost function:

\[
\min_{\mathbf{W}, \mathbf{H}} \frac{1}{2} \| \mathbf{A} - \mathbf{W} \mathbf{H} \|_F^2 + \alpha \| \mathbf{W} \|_F^2 + \beta \sum_{t=1}^{T} \| \mathbf{H}(\cdot, t) \|_1^2
\]  

(7)

where \( \mathbf{A} \) is the functional connectivity matrix, \( \mathbf{W} \) is a matrix of subgraph connectivity with size \( \frac{N(N-1)}{2} \times k \), and \( \mathbf{H} \) is a matrix of time-dependent expression coefficients for subgraphs with size \( k \times (T \times S \times 2) \). The parameter \( k \) is the number of subgraphs, \( \alpha \) is a regularization of the connectivity for subgraphs, and \( \beta \) is a penalty that imposes sparsity on the temporal expression coefficients (Kim & Park, 2011). For fast and efficient factorization to solve this equation, we used an alternative non-negative least square with block-pivoting method with 100 iterations (Kim, He, & Park, 2014). The matrices \( \mathbf{W} \) and \( \mathbf{H} \) were initialized with randomized values from a uniform distribution between 0 and 1.
A random sampling procedure was used to find the optimal parameters $k$, $\alpha$, and $\beta$ (Bergstra & Bengio, 2012). In this procedure, the NMF algorithm was re-run 1,000 times with parameter $k$ drawn from $U(2,15)$, parameter $\alpha$ drawn from $U(0.01, 1)$, and parameter $\beta$ drawn from $U(0.01, 1)$. The subgraph learning performance was evaluated through 4-fold cross-validation. In each fold, twenty-four participants were used for training; Eight participants were used for testing and calculating cross-validation error ($\|A - WH\|_F^2$). An optimal parameter set should minimize the cross-validation error. We chose an optimal parameter set ($k = 10$, $\alpha = 0.535$, $\beta = 0.230$) that ensured the cross-validation error in the bottom 25% of the distribution of cross-validation error from our random sampling scheme (Khambhati, Mattar, et al., 2018).

Since the result of NMF is non-deterministic, we implemented consensus clustering to obtain reliable subgraphs (Monti, Tamayo, Mesirov, & Golub, 2003). In this procedure, we (i) used the optimal parameters and ran the NMF 100 times on $A$, (ii) concatenated subgraph matrix $W$ across 100 runs into an aggregate matrix with dimensions $\frac{N(N-1)}{2} \times (k \times 100)$, (iii) applied NMF to this aggregate matrix to obtain a final set of subgraphs $W_{\text{consensus}}$ and expression coefficients $H_{\text{consensus}}$.

**Properties of subgraphs**

Applying NMF yielded a set of subgraphs, or patterns of functional connectivity ($W$), and the expression of these subgraphs over time ($H$). To understand the subgraphs, we first rearranged $W$ into $k$ different $N \times N$ subgraphs. To understand the roles of cognitive systems in each subgraph, we mapped each ROI to 13 putative cognitive systems from the previously defined parcellation: uncertain, sensory, cingulo-
opercular task control, auditory, default mode, memory retrieval, visual, fronto-parietal task control, salience, subcortical, dorsal attention, ventral attention, and cerebellar (Chai et al., 2017; Power et al., 2011). This yielded a 13 x 13 representation of each subgraph. To show which within-system and between-system edges in this representation were strongest, we applied a permutation test. We permuted the system label for ROIs and formed a matrix with system-by-system edges. This process was repeated 10,000 times to determine which strength of system-by-system edges was above the 95% confidence interval threshold after correction for multiple comparisons.

To characterize the connectivity pattern of each subgraph, we ordered them according to the relative strength of within-system edges versus between-system edges. For each subgraph, we calculated the average strength of within-system edges (edges that link two ROIs that both belong to the same system), and the average strength of between-system edges (edges that link an ROI in one system to an ROI in another system). Then, we subtracted the average strength of between-system edges \(E_B\) from the average strength of within-system edges \(E_W\) and divided this difference by the sum of them \(\frac{E_W-E_B}{E_W+E_B}\). We estimated the 95% confidence interval of these measures (average relative strength, average within-system strength or average between-system strength) by implementing bootstrapping 10,000 times.

Next, we investigated the relationship between these connectivity patterns and the temporal expression of each subgraph. As the matrix of functional connectivity was divided in two, with the first half reflecting positive connectivity and the second half reflecting negative connectivity, the temporal expression matrix also had two halves, with the first reflecting positive expression over time and the second reflecting negative
expression over time. As there was a strong negative correlation between positive and negative expression, we did all of our analyses on the relative expression (positive expression minus negative expression) of each subgraph (Khambhati, Medaglia, et al., 2018). Across subgraphs, we calculated Pearson correlation coefficients between the average relative expression and the average within-system strength, average between-system strength, and average relative strength of each subgraph. To determine the significance of the correlation coefficients, we implemented 10,000 permutations of the subgraph labels to form the null distribution of correlation coefficients.

Modulation of subgraph expression by task factors

We investigated how fluctuations in the trial-by-trial relative expression of each subgraph were related to four trial-by-trial task factors: CPP, RU, reward, and residual updating. CPP and RU were estimated based on the normative learning model (McGuire et al., 2014; Nassar et al., 2012; Nassar et al., 2010). Residual updating was derived as the residual of the behavioral regression model described above. We examined the effect of these four trial-by-trial task factors together, including all four in a regression model predicting trial-by-trial relative expression. Since NMF yielded values of temporal expression every 2 TRs (5 seconds), we applied a linear interpolation on the temporal expression values to obtain an expression value aligned with outcome onset on each trial. Regression models were implemented for each participant separately. Regression coefficients were then tested at the group level using two-tailed $t$-tests.

Association of individual learning with subgraph expression

76
Next, we examined the relationship between subgraph expression and individual differences in the extent to which belief updating followed normative principles. As an index of normative learning for each individual, we used the sum of the regression coefficients on the CPP term ($\beta_2$) and the RU term ($\beta_3$) in the behavior model (McGuire et al., 2014). This normative learning index reflected the extent to which a participant's trial-by-trial updates were influenced by the two normative factors CPP and RU. We examined the relationship between this index and two aspects of subgraph expression. First, across subjects, we calculated the Pearson correlation coefficient between normative learning and the dynamic modulation of relative expression by normative factors for each subgraph. This dynamic modulation was indexed as the sum of the regression coefficients for CPP and RU from the regression model predicting trial-by-trial relative expression. That is, dynamic modulation reflected how normative factors were associated with the change in relative expression of the subgraph. Second, across subjects, we calculated the Pearson correlation coefficient between normative learning and the average relative expression of each subgraph. To determine the significance of these correlation coefficients, we permuted the participant labels 10,000 times to form the null distribution.

**Contribution of specific edges**

We evaluated the contributions of different types of edges to the task effects (influence of CPP, RU, reward and residual updating on subgraph expression across time) and individual differences effects (relationship between normative learning and subgraph expression across subjects). We mainly focused on the contribution of within-system edges and between-system edges. For this analysis, we implemented three
types of comparison: Within versus All, Between versus All, and Between versus Within. For Within versus All, we kept within-system edges only and re-estimated task and individual differences effects; then, we compared these effects with the effects estimated using all edges. This comparison showed the change of effects after between-system edges were removed, and thus, this comparison revealed the contribution of between-system edges. For Between versus All, we kept between-system edges only and re-estimated task and individual differences effects. We then compared these effects with the effects estimated using all edges. In this comparison, within-system edges were removed and thus, we examined the contribution of within-system edges. Last, the comparison of Between versus Within is a direct comparison between effects estimated with between-system edges only and effects estimated with within-system edges only. Thus, this comparison examined the different contributions of between-system and within-system edges.

Specifically, for task effects, we examined the change of coefficients in the regression model that investigated the influence of four task factors—CPP, RU, reward and residual updating—on subgraph relative expression. The change was calculated for each participant separately, and the significance of change was then tested at the group level using two-tailed $t$-tests. For individual differences effects, we examined the change of correlation coefficients for two types of relationship: the relationship between individual normative learning and dynamic modulation of subgraph relative expression and the relationship between individual normative learning and average subgraph relative expression. To determine the significance of the change of correlation coefficients, we permuted the labels of participants for individual normative learning 10,000 times to form the null distribution of the change of correlation coefficients.
We also investigated the contribution of different functional systems and the contribution of different system-by-system edges. For the contribution of different functional systems, we compared the effects after removing edges of one functional system with the effects estimated with all edges. For the contribution of different system-by-system edges, we compared the effects after removing one system-by-system edge with the effects estimated with all edges. Statistical testing was conducted with the same procedures described in the previous paragraph.

Relationship between regional activity and connectivity

To investigate the relationship between dynamic functional connectivity and univariate activation, we fit a mass univariate GLM. In this GLM, the regressors were the outcome onset and four modulators of outcome onset: CPP, RU, reward and residual updating. These regressors were convolved with a gamma hemodynamic response function (HRF) as well as the temporal derivative of this function. Six motion parameters were also included as regressors.

To examine what aspects of our functional connectivity results could be accounted for by functional coactivation, we used the regression coefficients from the GLM above (including both the main HRF and its temporal derivative for each regressor) to create a predicted BOLD time series. We then repeated the same sequence of analyses described above on this predicted BOLD time series. This predicted BOLD time series captured all fluctuations in activity in that ROI that could be accounted for by the linear effects of CPP, RU, reward, and residual updating. However, this predicted BOLD time series lacked any statistical dependencies between regions that were present in the actual BOLD time series that could not be explained by task-driven
changes in univariate activation. Thus, any functional connectivity results we observed with this predicted BOLD time series could be fully accounted for by task-driven changes in univariate activation.

Results

*Belief updating is influenced by uncertainty and surprise*

Participants performed a predictive-inference task during fMRI (Fig. 1a). For this task, participants positioned a bucket to catch a bag that dropped from an occluded helicopter. The location of the bag was sampled with noise from a distribution centered on the location of the helicopter. The location of the helicopter usually remained stable but occasionally changed suddenly and unpredictably (with an average probability of change of 0.1 across trials). Additionally, whether the bag (if caught) was rewarded or neutral was assigned randomly on each trial and indicated by color. This task challenged participants to form and update a belief about a latent variable (the location of the helicopter) based on noisy evidence (the location of dropped bags).

We previously described a theoretical model approximating the normative solution for this task (McGuire et al., 2014). This theoretical model takes the form of a delta-rule and approximates the Bayesian ideal observer. Beliefs \((B_{t+1})\) are updated based on the difference between the current outcome location \((X_t)\) and the predicted location \((B_t)\), with the extent of updating controlled by a learning rate \((\alpha_t); \text{Fig. 1b})\). Trial-by-trial learning rates are determined by two factors: (i) change-point probability (CPP), which is the probability that a change-point has happened and represents a form of belief surprise; and (ii) relative uncertainty (RU), which is the reducible uncertainty regarding the current state relative to the irreducible uncertainty that results from
environmental noise and represents a form of belief uncertainty (Fig. 1c). Learning rates are higher when either CPP or RU is higher: \( \alpha_t = \text{CPP} + (1 - \text{CPP})\text{RU} \).

We previously reported how participants’ predictions were influenced by both normative and non-normative factors and how these factors are encoded in univariate and multivariate activity (McGuire et al., 2014; Nassar, McGuire, et al., 2019). Participants updated their beliefs more when the value of CPP or RU was higher, consistent with the normative model. Participants also updated their beliefs more when the outcome was rewarded, however, which is not a feature of the normative model. CPP, RU and reward, as well as residual updating (belief updating not captured by CPP, RU or reward), were all encoded in univariate and multivariate brain activity in distinct regions (McGuire et al., 2014; Nassar, McGuire, et al., 2019). In the current study, we built on these previous findings and investigated how these factors, as well as individual differences in how these factors influence belief updating, are related to the dynamics of whole-brain functional connectivity.

**NMF identified ten subgraphs that varied over time**

We used NMF to decompose whole-brain functional connectivity over time into specific patterns of functional connectivity, called subgraphs, and quantified the expression of these patterns over time. To perform NMF, we first defined regions of interest (ROIs) based on a previously defined parcellation (Power et al., 2011) (Fig. 2a) and extracted blood-oxygenation-level-dependent (BOLD) time series for each ROI (Fig. 2b). For every pair of ROIs, we calculated the Pearson correlation coefficient between the BOLD time series in 10-TR (25 s) time windows, offset by 2 TRs for each time step (and thus 80% overlap between consecutive time windows). This procedure thus yielded
a matrix whose entries represented time-dependent changes in the strengths of these pairwise correlations in the brain during the task. We unfolded each time window from this correlation matrix (Fig. 2c) into a one-column vector, and then concatenated these vectors from all time windows and all participants (Fig. 2d). As required for NMF, we transformed the resulting matrix to have strictly non-negative values: we duplicated the full matrix, set all negative values to zero in the first copy, and set all positive values to zero in the second copy before multiplying all remaining values by negative one. Thus, we divided the final full data matrix into two halves, with one half containing the positive correlation coefficients (zero if the coefficient was negative) and one half containing the absolute values of the negative correlation coefficients (zero if the coefficient was positive) (Khambhati, Medaglia, et al., 2018). This procedure ensured that our approach did not give undue preference to either positive or negative functional connectivity, and that subgraphs were identified based on both positive and negative functional connectivity.

We applied NMF to this matrix (\( \mathbf{A} \)) to identify functional subgraphs and their expression over time. Specifically, we decomposed the full data matrix into a subgraph matrix \( \mathbf{W} \) and an expression matrix \( \mathbf{H} \) (Fig. 2d). The columns of \( \mathbf{W} \) represent different subgraphs and the rows represent different edges (i.e., pairs of regions), with the value in each cell representing the strength of that edge (i.e., the functional connectivity strength for that pair of regions) for that subgraph. The rows of \( \mathbf{H} \) represent different subgraphs, and the columns represent time windows, with the value in each cell representing the degree of expression of that subgraph in that time window. We implemented NMF by minimizing the residual error (\( \| \mathbf{A} - \mathbf{WH} \|^2 \)) via three parameters:
(i) the number of subgraphs \((k)\), (ii) the subgraph regularization \((\alpha)\), and (iii) the expression sparsity \((\beta)\) (Supplementary Fig. 1).

Using NMF, we identified ten subgraphs, which reflected patterns of functional connectivity strengths across every pair of regions in the brain, as well as the expression of these subgraphs over time. The full description of each subgraph specifies the edge strength between every pair of ROIs, corresponding to a 247x247 matrix. We calculated a simpler summary description that specifies the edge strength between every pair of functional systems in the previously defined parcellation, corresponding to a 13x13 matrix (Power et al., 2011). Edges between ROIs were categorized according to the functional system of each ROI. To estimate the diagonal entries in the system-by-system matrix, we averaged the weights of all edges connecting two ROIs within a given system (Fig. 3a). To estimate the off-diagonal entries of the system-by-system matrix, we averaged the weights of all edges linking an ROI in one system with an ROI in another system. In line with common parlance, we refer to the edges within the same system as within-system edges, whereas we refer to the edges between two different systems as between-system edges. For presentation, we ordered and numbered the ten subgraphs according to the strength of within-system edges relative to that of between-system edges (Fig. 3b, Supplementary Fig. 2a-c). Finally, we thresholded the system-by-system matrix to show only edges that passed a permutation test \((p<0.05\) after the Bonferroni correction for multiple comparisons; see Methods). The full data matrix on which we performed NMF was divided into two halves, with the first half corresponding to positive functional connectivity and the second half corresponding to negative functional connectivity. The expression matrix \(H\) was therefore also divided into two halves, with the first half corresponding to positive expression over time and the second half
corresponding to negative expression over time. Positive and negative expression coefficients were highly negatively correlated with each other across time for all the subgraphs (all $r<-0.61$, all $p<0.001$). For the analyses of subgraph expression below, we thus constructed a measure of relative subgraph expression by subtracting the negative expression from the positive expression at each time point (Khambhati, Medaglia, et al., 2018). Across subgraphs, the average relative expression across time was strongly correlated with the relative strength of within- versus between-system edges (Supplementary Fig. 2d-f). That is, higher within-system strength was associated with greater relative expression of the subgraph.
Figure 3-2 Schematic overview of the method

a Regions of interest (ROIs). Functional MRI BOLD signals were extracted from spherical ROIs based on the previously defined parcellation (Power et al., 2011). We only kept 247 ROIs that had usable data from all subjects. Each ROI can be assigned to one of 13 putative functional systems. The brain figure was visualized by the BrainNet Viewer (Xia et al., 2013) under the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

b An example of Pearson correlation coefficients calculated between regional BOLD time series over the course of the experiment. Each BOLD time series was divided into 10-TR (25 seconds) time windows, and consecutive time windows were placed every 2 TRs leading to 80% overlap.
between consecutive time windows. Pairwise Pearson correlation coefficients were calculated between ROI time series in each time window. 

A An example of edge strength over time. In each time window, there were 247*(247-1)/2 edges. 

D Nonnegative matrix factorization (NMF). In each time window, the matrix of edge strengths was unfolded into one column. Then, edges from all time windows in all participants were concatenated into a single matrix. Each row in the full data matrix contained an edge (pairwise correlation coefficients between BOLD time series from two ROIs) and each column contained a time window (across all scans and participants). Correlation values in this matrix were strictly non-negative; the full data matrix was divided into two halves, with one half containing the positive pairwise correlation coefficients (zero if the correlation coefficient was negative) and one half containing the absolute values of negative pairwise correlation coefficients (zero if the correlation coefficient was positive). Thus, subgraphs were identified based on both the similarity of positive functional connectivity and the similarity of negative functional connectivity together. Then, NMF was applied to decompose the concatenated matrix into a matrix $W$, which encoded the strengths of edges for each subgraph, and a matrix $H$, which encoded the time-dependent expression of each subgraph. For example, the strength of edges of the fourth subgraph (the fourth column in the matrix $W$) can be folded into a squared matrix, reflecting the edge strength between every pair of ROIs.
Figure 3-3 Patterns of connectivity in subgraphs

a Converting edges between nodes into edges between systems. First, the edges of each subgraph can be folded into a square matrix, representing the edges between every pair of nodes.
(ROIs). Then, based on the 13 putative functional systems reported by Power et al. (2011), we categorized each edge according to the system(s) to which the two nodes (ROIs) belonged. We calculated the mean strength of edges linking a node in one system to a node in another system, and refer to that value as the between-system edge. Similarly, we calculated the mean strength of edges linking two nodes that both belong to the same system and refer to that value as the within-system edge. Edges between nodes and edges between systems were normalized into the scale between 0 and 1. b Edges between systems in the ten subgraphs identified by NMF. We show only significant edges ($p<0.05$ after the Bonferroni correction for multiple comparisons). For each subgraph, the top matrix shows the significant edges in that subgraph within or between systems. For example, Subgraph 1 has high edge strengths along the diagonal; thus, this subgraph describes functional connectivity that lies predominantly within functional systems. In contrast, subgraph 5 has high edge strengths along a single row and column, corresponding to the visual system; thus, this subgraph describes functional connectivity between the visual system and all other systems. Subgraphs varied in the degree to which they represent interactions within the same system (e.g., subgraph 1) versus interactions between different systems (e.g., subgraph 10). All nodes from systems involved in significant edges are shown on the brain below by the BrainNet Viewer(Xia et al., 2013) under the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).
Normative factors modulated subgraph expression

We investigated how CPP, RU, reward, and residual updating influenced the temporal expression of each subgraph. We identified a particularly strong relationship between the normative factors (CPP, RU and the residuals that reflected the participants’ subjective estimates of those variables) and subgraph 4, whose strongest edges were in the fronto-parietal task-control system, followed by the memory retrieval, salience and dorsal-attention systems (Fig. 4a-b). Specifically, we used multiple regression to estimate the trial-by-trial relationship between these four factors and the relative expression strength of each subgraph. For each subgraph, regression coefficients were fitted separately for each participant and were tested at the group level using t-tests (Supplementary Fig. 3). Among the ten subgraphs, these four factors explained the most variance in the time-dependent relative expression of subgraph 4 (Supplementary Fig. 4), in each case showing positive modulations (CPP: mean±SEM=0.202±0.053, $t_{31}=3.78$, $p<0.001$; RU: 0.392±0.077, $t_{31}=5.11$, $p<0.001$; residual updating: 0.177±0.079, $t_{31}=2.23$, $p=0.033$; Fig 4c). We also evaluated the influence of head motion by including motion, as indexed as the relative root-mean-square of the six motion parameters, in the regression model. Motion was not significant ($p=0.29$) and the effects of CPP, RU and residual updating remained significant and of similar effect size.
Figure 3-4 Temporal expression of subgraph 4 was related to task factors and individual differences

a Summary of the pattern of connectivity in subgraph 4. We summarized the pattern of connectivity as within-system strength (which is the value in the diagonal) and between-system strength (which is the average of values in the off-diagonal) for each system. The fronto-parietal system as well as three other systems (memory retrieval, salience and dorsal attention) showed the strongest contributions to this subgraph in terms of both within-system and between-system strength. The 95% confidence interval of each system was estimated by bootstrapping 10,000 times on the edges of that system. b Nodes for the top four systems with strong within-system and between-system strength. We showed the nodes of fronto-parietal system, memory retrieval system, salience system and dorsal attention system on the brain by the BrainNet Viewer(Xia et al., 2013) under the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

c Modulation of temporal expression of subgraph 4

- Coefficient
  - CPP RU Reward Residual
  - Subgraph 4
  - r=0.448 p=0.004

- Dynamic modulation
  - Subgraph 4
  - Dynamic modulation
  - r=0.332 p=0.029

- Average expression
  - Subgraph 4
  - Average expression
  - r=0.60 p=0.004
4 by task factors. A regression model that included CPP, RU, reward and residual updating as predictors of temporal relative expression (calculated by subtracting negative expression from positive expression) of subgraph 4 was fitted for each participant, and coefficients were tested on the group level by t-tests. The results showed positive effects of CPP, RU and residual updating. Each point represents one participant. Error bars represent one SEM. (*p<0.05, **p<0.001) 

The relationship between individual normative learning and the dynamic modulation of subgraph 4 expression by normative factors. This dynamic modulation was indexed as the sum of the coefficients of CPP and RU in c, and represents the extent to which trial-by-trial expression was influenced by the two normative learning factors. There was a significant positive correlation across participants. Each point represents one participant. The red line represents the regression line and the shaded area represents the 95% confidence interval. 

The relationship between individual normative learning and average relative expression of subgraph 4. There was a significant positive correlation across participants. Each point represents one participant. The red line represents the regression line and the shaded area represents the 95% confidence interval.
Although CPP or RU also modulated the relative expression of some other subgraphs (e.g., subgraph 1, 3 and 7; Supplementary Fig. 3), below we focus on subgraph 4 for several reasons. First, the four factors we investigated explained more variance in the time-dependent relative expression of subgraph 4 than that of any other subgraph. Second, only on subgraph 4 were the effects of CPP and RU strong enough to survive correction for multiple comparisons across ten subgraphs. Third, only on subgraph 4 were the effects of CPP and RU robustly shown across analyses using different sized time windows.

*Individual differences associated with subgraph expression*

The expression of subgraph 4 was not only modulated by task factors that drive normative learning, but also varied across subjects in a manner that reflected individual differences in normative learning. As an index of normative learning, we estimated the influence of CPP and RU on trial-by-trial belief updates using multiple regression and took the sum of the regression coefficients of CPP ($\beta_2$ in Equation 6) and RU ($\beta_3$ in Equation 6) for each participant (McGuire et al., 2014). This sum reflected how much each individual updated their beliefs in response to normative factors. We examined the relationship between individual differences in this normative belief-updating metric and two aspects of subgraph expression.

First, we examined the relationship between normative belief updating and the dynamic modulation of subgraph expression by normative factors (Supplementary Fig. 5). As an index of the dynamic modulation of subgraph expression by normative factors, we used the sum of the regression coefficients of CPP and RU on relative expression from the analyses above (Supplementary Fig. 3). We found a positive correlation
between the dynamic modulation of subgraph 4 expression by normative factors and normative belief updating across participants ($r=0.448$, $p=0.004$; Fig. 4d). Second, We also found a positive correlation between the average relative expression of subgraph 4 and normative belief updating across participants ($r=0.332$, $p=0.029$; Fig. 4e; Supplementary Fig. 6). These effects were still significant when we controlled for the influence of motion on dynamic modulation or average relative expression, whereas the effects of motion itself were not significant (all $p>0.31$). These two results show that participants with the highest average relative expression of subgraph 4, and for whom the normative factors account for the most variance in the relative expression of subgraph 4 across time, tended to update their beliefs in a manner more consistent with the normative model than the other subjects.

**Contribution of specific edges to the identified effects**

Subgraph 4 describes both within- and between-system functional connectivity for multiple functional systems (Figs. 3b and 4a-b, Supplementary Fig. 2a-c). We next examined the contribution of specific edges (i.e., functional connectivity between specific pairs of brain regions) within subgraph 4 to the task and individual difference effects we observed for that subgraph.

The task-related modulations of subgraph 4 involved primarily between-system, not within-system, functional connectivity. Specifically, we re-estimated the effects of CPP, RU, reward, and residual updating on the relative expression of subgraph 4 using only within-system edges (i.e., only the diagonal cells of the system-by-system matrix in Fig. 3b; “Within”) or only between-system edges (i.e., only the off-diagonal cells of the system-by-system matrix in Fig. 3b; “Between”). We compared these effects to our
previous estimates using all edges (Fig. 5a; “All”) through $t$-tests. Removing the between-system edges (Within versus All) reduced the size of the estimated effects of CPP (mean±SEM=-0.155±0.042, $t_{31}$=-3.73, $p<0.001$), RU (-0.300±0.062, $t_{31}$=-4.82, $p<0.001$), and residual updating (-0.140±0.053, $t_{31}$=-2.63, $p=0.013$). In contrast, removing the within-system edges (Between versus All) led to no reliable changes in these effects (all $p>0.21$). Further, in a direct comparison of the reduced subgraphs with only within- or between-system edges, the effects estimated with between-system edges only were stronger for CPP (0.151±0.042, $t_{31}$=3.63, $p<0.001$), RU (0.290±0.063, $t_{31}$=4.63, $p<0.001$), and residual updating (0.139±0.048, $t_{31}$=2.91, $p=0.007$).

The contributions of within- and between-system functional connectivity to the individual difference effects of subgraph 4 were less clear. For the relationship between individual differences in normative learning and average relative expression, the pattern across comparisons was similar to that observed for task effects (Fig. 5b), which would indicate a greater contribution of between-system edges, but none of the comparisons were statistically significant. In contrast, for the relationship between individual differences in normative learning and the dynamic modulation of subgraph 4, within-system edges appeared to be more important, as removing the within-system edges (Between versus All) reduced this correlation (difference = 0.048, $p=0.006$; Fig. 5b).

Supplementary analyses identified contributions of specific functional systems (i.e., one row/column from the system-by-system matrix in Fig. 3b; Supplementary Fig. 7) and of specific system-by-system edges (i.e., one cell from the system-by-system matrix in Fig. 3b; Supplementary Fig. 8) to the task and individual difference effects on subgraph 4.
**Figure 3-5** The contribution of between-system and within-system edges to effects of task factors and individual differences on subgraph 4 expression

**a** The contribution of between-system and within-system edges to the effect of task factors on temporal relative expression of subgraph 4. To determine the relative contribution of between- and within-system edges on time-dependent subgraph 4 expression, we performed three comparisons on the effects estimated by different types of edges using t-tests: within-system edges only (Within), between-system edges only (Between) and all edges (All). First, removing between-system edges (Within versus All) decreased the effect of CPP, RU and residual updating. Second, in contrast, after removing within-system edges (Between vs All), there was no significant change in these coefficients. Third, we directly compared the effects contributed from between-system edges only and from within-system edges only (Between versus Within). For between-system edges, there were stronger positive effects for CPP, RU and residual updating. Error bars represent one SEM. (*p<0.05, **p<0.01, ***p<0.001)

**b** The contribution of between-system and within-system edges to the relationship between normative learning and dynamic modulation and average expression of subgraph 4. We performed the same three comparisons to
determine the relative contribution of between- and within-system edges for each relationship with individual differences. For the effect of dynamic modulation, removing within-system edges (Between versus All) decreased the correlation coefficient. This correlation coefficient was also larger for within-system edges only than between-system edges only, but this effect was not statistically significant. For the effect of average expression, removing between-system edges (Within versus All) decreased the correlation coefficient, and the correlation coefficient was larger for between-system edges only than within-system edges only, though neither of these effects were statistically significant. Error bars represent one SEM. (**p<0.01)
Robust effects across different sized time windows

To determine the sensitivity of our results to the size of this time window, we repeated the entire procedure using shorter (8-TR/20 s window with 6-TR/15 s overlap; Supplementary Fig. 9-12) or longer (12-TR/30 s window with 10-TR/25 s overlap; Supplementary Fig. 13-16) time windows. That is, we shorten or lengthen the time window by the interval of one trial (~5 s). With both shorter and longer time windows, we identified ten subgraphs. There was a high degree of similarity between the ten subgraphs identified in the main analysis and those identified using either shorter (edges between nodes: all $r>0.81$; edges between systems: all $r>0.80$) or longer (edges between nodes: all $r>0.98$; edges between systems: all $r>0.98$) time windows. With longer time windows, the relative expression of subgraph 4 still showed the same relationship to task factors (CPP and RU) and to individual differences in normative learning; with shorter time windows, these effects were also present but weaker.

Relationship between regional activity and connectivity

In our previous report, we described how CPP, RU, reward, and residual updating influenced univariate brain activity. In a final set of analyses, we examined the relationship between these previously reported univariate effects and the changes in dynamic functional connectivity we identified above.

The brain regions that were most strongly represented in subgraph 4 overlapped spatially with the brain regions whose activity was modulated reliably by CPP and RU in our previous report. As a measure of a region’s involvement in subgraph 4, for each ROI, we calculated the mean strength of every edge between that ROI and all other ROIs in subgraph 4, and normalized these mean values between 0 and 1. We then
related this metric to activation from our previous study (McGuire et al., 2014), as measured by the z-statistic of the modulation effect of CPP or RU. This z-statistic indicated the effect size of change of univariate activity in response to CPP or RU across participants. Across all ROIs, there was a positive correlation between edge strength in subgraph 4 and activation for CPP ($r=0.403$, $p<0.0001$; Fig. 6a) and activation for RU ($r=0.704$, $p<0.0001$; Fig. 6b). The Surf Ice software (https://www.nitrc.org/projects/surfice) was used to show the map of normalized mean edge strengths for subgraph 4 alongside the thresholded activation maps for CPP and RU (Fig. 6c). Regions with stronger edge strength in subgraph 4, such as the insula, dorsomedial frontal cortex, dorsolateral prefrontal cortex, posterior parietal cortex, and occipital cortex, also tended to show stronger increases in activation with increases in CPP and RU.

In addition to these strong associations between univariate brain activation and edge strength, effects beyond those captured by univariate task activity also contributed to our dynamic functional connectivity results. To demonstrate this, we estimated functional connectivity from time-series that only contained task-modulated univariate activity, performed NMF on this matrix, and repeated all of our main analyses (Supplementary Fig. 17-20). This analysis again identified a subgraph 4 whose strongest edges were in the fronto-parietal system, but it did not recapitulate all of the relationships between subgraph 4 expression and task factors and individual differences seen in our main analyses. These results implied that the dynamic functional connectivity patterns identified in our main analyses reflect a mixture of coordinated activity across regions (which can be captured by univariate analyses) and other statistical dependencies across regions that require network-based analyses.
Figure 3-6 Relationship between edge strength of subgraph 4 and univariate task activations
a Relationship between the activation for CPP and the edge strength of subgraph 4. We calculated the Pearson correlation coefficient between the z-statistic for CPP from McGuire et al. (2014) and the edge strength across nodes in subgraph 4. Each data point represents an ROI. The edge strength for each ROI was calculated as the column sum of that ROI’s edges to other ROIs, reflecting the summed interactions between that ROI and all others. The edges were normalized into the scale between 0 and 1. A significantly positive correlation was observed. The red line represents the regression line and the shaded area represents the 95% confidence interval. b Relationship between the activation for RU and the edge strength of subgraph 4. We observed a significant positive correlation between the z-statistic for RU from McGuire et al. (2014) and the edge strength across nodes in subgraph 4. The red line represents the regression line and the shaded area represents the 95% confidence interval. c Whole-brain thresholded activation maps for CPP and RU from McGuire et. al (2014) and whole-brain maps for edge strength of subgraph 4 in the current study.
Discussion

We identified a pattern of dynamic functional brain connectivity in human subjects performing a predictive-inference task. This pattern was expressed most strongly during times that demanded faster belief updating and was enhanced in individuals who most effectively used adaptive belief updating to perform the task. To identify this pattern, we used NMF, an unsupervised machine-learning technique that decomposes the full matrix of time-dependent functional connectivity into subgraphs (patterns of functional connectivity), and the time-dependent magnitude of these subgraphs. Among the subgraphs we identified in our data, the expression of one subgraph in particular was modulated reliably by three trial-by-trial factors that influenced the degree of behavioral belief updating: CPP (surprise), RU (uncertainty), and residual updating (updating unaccounted for by surprise or uncertainty). Notably, CPP and RU are factors that normatively promote greater belief updating, scaling the degree to which past observations are discounted relative to the most recent evidence. Residual updating likely captures, at least in part, deviations between the objective values of CPP and RU in the normative model and the individual’s subjective estimates of those factors. Thus, the expression of this subgraph reflects not only normative factors that should influence belief updating but also likely fluctuations in subjective estimates of those factors. In addition to being modulated by these trial-by-trial task factors, expression of this subgraph also varied across individuals in a manner associated with individual differences in belief updating. Participants who tended to update their beliefs in a more normative manner – that is, with a stronger influence of surprise (CPP) and uncertainty
showed stronger dynamic modulation of the expression of this subgraph by
normative factors and showed stronger average expression of this subgraph.

The subgraph modulated by surprise and uncertainty included interactions
between multiple functional systems, most prominently the fronto-parietal task control,
memory retrieval, salience, and dorsal attention systems (Fig. 3b & Fig. 4a). These
systems include multiple regions in the anterior insula, dorsolateral and dorsomedial
frontal cortex, and lateral and medial parietal cortex (Fig. 4b & Fig. 6c). These regions
showed a large degree of overlap with areas that we have previously shown to have
increased univariate activation in response to both surprise and uncertainty (in this same
dataset; Figure 6) (McGuire et al., 2014). A smaller subset of these regions, including
parts of the dorsomedial frontal cortex, anterior insula, inferior frontal cortex, posterior
cingulate cortex, and posterior parietal cortex, was modulated not only by both normative
(surprise and uncertainty) factors, but also by a non-normative one (reward). This
smaller subset includes regions that participate in the fronto-parietal task-control,
memory retrieval, salience, and dorsal attention systems.

Previously, we also reported regions whose univariate activity was modulated by
either surprise or uncertainty alone. Surprise was associated selectively with activation
in occipital cortex, and uncertainty was associated selectively with activation in anterior
prefrontal and parietal cortex (McGuire et al., 2014). We similarly have reported
multivariate activation patterns that were associated selectively with either surprise or
uncertainty alone (Nassar, McGuire, et al., 2019). In the current study, we identified a
key pattern of functional connectivity that was modulated by both surprise and
uncertainty, but we did not identify any other pattern that was modulated reliably by
either surprise or uncertainty alone. One possible explanation for this lack of a positive
result was our need to use relatively long time windows (25 s, corresponding to 4-6 trials) in order to obtain reliable estimates of functional connectivity. These time windows likely included both the surprise elicited by change-points and the uncertainty that follows. Thus, functional connectivity related to surprise and uncertainty may have been difficult to dissociate temporally in our task and analysis design. Using a task that can temporally separate the tracking of surprise and uncertainty (Nassar et al., 2016) might enable the identification of distinct patterns of functional connectivity for each factor.

The identified pattern of whole-brain functional connectivity was also expressed across individuals in a manner that varied with the degree to which they updated their beliefs more in line with the normative model. Thus, individual differences in learning were also reflected in features of individual functional connectomes. In our previous study, we noted a relationship between individual differences in normative learning and the degree to which activity in dorsomedial frontal cortex and anterior insula was modulated by normative factors (surprise and uncertainty) (McGuire et al., 2014). Here we showed that normative learning was also associated with how functional connectivity was modulated dynamically by the same normative factors. These new findings add to previous work showing that brain network dynamics can reflect individual differences in learning in various domains (Bassett et al., 2011; Bassett et al., 2015; Büchel et al., 1999; Gerraty et al., 2018; Sun et al., 2006). Potentially, these differences in individual functional connectomes during learning could reflect individual differences in resting-state (task-independent) functional connectivity (Tompson, Falk, Vettel, & Bassett, 2018), which merits further study.

Functional connectivity captures many different kinds of statistical dependencies between brain regions, including those that result from task-driven co-activation. The
strong association between neural activation and functional connectivity during periods of surprise and uncertainty in our results (Fig. 6), as well as previous studies in other domains (Antzoulatos & Miller, 2014; Bassett et al., 2015; Büchel et al., 1999; Gerraty et al., 2018; Lewis et al., 2009; Sun et al., 2006), raises the possibility that the increases in functional connectivity between brain regions might have arisen because these regions became more tightly synchronized to external task events, without necessarily any increase in communication between them. To refute this possibility, we repeated our analyses on the predicted BOLD time series from univariate GLMs. These predicted time series, which contain only task-driven statistical dependencies between brain regions, could not recapitulate all of the effects that we observed in our actual BOLD time series. Specifically, we found modulations by task (e.g., the modulation of subgraph expression by surprise and residual updating) and individual differences (e.g., the relationship between individual differences in normative learning and the dynamic modulation of subgraph expression by normative factors) that were apparent only in the full, original functional connectivity matrices. Thus, these effects appear to include neural communications that do not simply reflect task-driven co-activation. Even though the changes in functional connectivity that we describe may reflect a mixture of task-driven and endogenous dynamics, the network analysis provides an important higher-level, reduced-dimensionality description of these changes.

A key feature of the brain-wide pattern of functional connectivity that we identified was connectivity involving the fronto-parietal task-control system. We characterized the complex pattern of functional connectivity in the learning-related subgraph by summarizing the connectivity according the putative functional system of each region (Power et al., 2011). Among all the functional systems, the largest proportion of
connectivity in the learning-related subgraph involved the fronto-parietal system. Connectivity associated with the fronto-parietal system has been shown to increase at the beginning of learning and decrease toward the later phases of learning (Bassett et al., 2015; Büchel et al., 1999; Sun et al., 2006). Our result extends this finding by showing that fronto-parietal functional connectivity is modulated dynamically in a trial-by-trial manner according to the need for new learning. That is, the pattern of functional connectivity captured by the learning-related subgraph increased after surprising task changes and then decreased gradually as more information was gained about the current state. The fronto-parietal system is thought of as a control system that is involved in flexible adjustments of behavior (Cole et al., 2014; Power & Petersen, 2013). In particular, connectivity between the fronto-parietal network and other systems has been shown to change in response to different task requirements (Cole et al., 2013). This type of flexible control is critical for learning in a dynamic environment, where people should adjust their degree of belief updating in a context-dependent manner (Behrens et al., 2007b; Nassar et al., 2010).

Although the learning-related subgraph was also characterized by a balanced strength of within-system connectivity and between-system connectivity, the critical features that changed in response to task dynamics involved primarily between-system connectivity. This result implies that faster learning was associated with a greater degree of integration between different functional systems. Several previous studies have shown that complex cognitive tasks are associated with more integration between systems (Bertolero et al., 2015; Cohen & D'Esposito, 2016; Shine et al., 2016; Shine & Poldrack, 2017). Other work has shown that as a task becomes more practiced over time, the interaction between systems decreased while the connections within systems
remained strong (Bassett et al., 2015). Here we demonstrated changes in integration on a fast time scale, as task demands varied from trial to trial. Integration between systems was greater during periods of the task when surprise or uncertainty was high, and therefore there was a need to update one’s beliefs and base them more on the current evidence than on expectations developed from past experience.

In this study, we provided a network-based perspective on the neural substrates of learning in dynamic and uncertain environments. In such environments, people should flexibly adjust between slow and fast learning: beliefs should be updated more strongly when new evidence is most informative, such as when the environment undergoes a surprising change or beliefs are highly uncertain. Here we identified a specific brain-wide pattern of functional connectivity (subgraph) that fluctuated dynamically with changes in surprise and uncertainty. The dynamics and expression of this pattern of functional connectivity also varied across individuals in a manner that reflected differences in learning. This pattern was expressed more strongly and was more strongly modulated by surprise and uncertainty in people who updated their beliefs in a more normative manner, with a stronger influence of surprise and uncertainty. The most important aspect of this learning-related pattern of functional connectivity is functional integration between the fronto-parietal and other functional systems. These results establish a novel link between dynamics adjustments in learning and dynamic, whole-brain changes in functional connectivity.
Supplementary figure 1
Optimal parameters for nonnegative matrix factorization. a Number of subgraphs. We randomly sampled the number of subgraphs from a uniform distribution ($k \in [2, 15]$). The contour plot shows the Kernel density of the bivariate distribution. The darker blue area represents the higher probability mass. We selected the optimal parameter ($k = 10$) by averaging the parameter values that ensured that the cross-validation error was in the bottom 25% of the sampling distribution (orange dashed line). b Subgraph regularization. We randomly sampled values of subgraph regularization from a uniform distribution ($\alpha \in [0.01, 1.0]$) and select the optimal parameter ($\alpha = 0.535$). c Expression sparsity. We randomly sampled values of expression sparsity from a uniform distribution ($\beta \in [0.01, 1.0]$) and select the optimal parameter ($\beta = 0.230$).
Supplementary figure 2

Properties of subgraphs. a Subgraphs differ in the extent of within- versus between-system edge strength. For each subgraph, the strength of within-system edges (edges linking two nodes that both belong to the same system; Fig. 3b) was averaged, and the strength of between-system edges (edges linking one node from one system to another node from another system; Fig. 3b) was averaged. The ten subgraphs are ordered according to the relative strength of within- versus between-system edges. To form a normalized relative strength, we subtracted the average strength of between-system edges from the average strength of within-system edges and then divided this difference by their sum. A high relative strength means that a subgraph has stronger within-system edges than between-system edges (e.g., subgraph 1). The 95% confidence interval of each subgraph was estimated by bootstrapping 10,000 times on the edges of that subgraph. b Subgraphs differ in the extent of within-system strength. For each subgraph, the strength of within-system edges was averaged. For demonstration, the ten subgraphs are ordered according
to within-system strength. The 95% confidence interval of each subgraph was estimated by boostrapping. c Subgraphs differ in the extent of between-system strength. For each subgraph, the strength of between-system edges was averaged. For demonstration, the ten subgraphs are ordered according to the between-system strength. The 95% confidence interval of each subgraph was estimated by boostrapping. d The relationship between relative strength and average expression across subgraphs. Average expression was calculated as the difference between positive expression and negative expression. Each data point represents one subgraph. A significantly positive correlation was observed. The red line represents the regression line and the shaded area represents the 95% confidence interval. e The relationship between within-system strength and average expression across subgraphs. A significant positive correlation was observed. The red line represents the regression line and the shaded area represents the 95% confidence interval. f The relationship between between-system strength and average expression across subgraphs. There was no significant correlation. The red line represents the regression line and the shaded area represents the 95% confidence interval.
Supplementary figure 3

Modulation of temporal expression by learning factors. A regression model that included CPP, RU, reward and residual updating as predictors of trial-by-trial expression was fitted for each participant and subgraph, and then regression coefficients were tested by t-tests on the group level. Each point represents one participant. Error bars represent one SEM. (*p<0.05, **p<0.01, ***p<0.001)
Supplementary figure 4

The relationship between explained variance of task effects and explained variance of individual difference effects across subgraphs. For the task effects, we implemented a regression model that examined the influence of CPP, RU, reward and residual updating on temporal relative expression for each subgraph. Explained variance was indexed as $R^2$ of the regression model. For the individual difference effects, we implemented a regression model to predict individual normative learning for each subgraph. We included two regressors: dynamic modulation of normative factors (CPP and RU) on subgraph expression, and average subgraph expression. We then calculated $R^2$ of the regression model for each subgraph. Among the ten subgraphs, subgraph 4 showed the strongest $R^2$ for both task effects and individual difference effects. We also investigated the relationship between explained variance for task and individual differences effects using Pearson correlation. To determine the significance of this correlation coefficient, we permuted the subgraph labels 10,000 times to form the null distribution of correlation coefficients.
Supplementary figure 5

The relationship between individual normative learning and the dynamic modulation of subgraph expression by normative factors. Normative learning was indexed by the sum of the coefficients of CPP and RU from a behavioral regression model, and represents the extent to which a participant's behavior was influenced by the two normative learning factors. Dynamic modulation was indexed by the sum of coefficients of CPP and RU from the regression model against trial-by-trial expression in Supplementary Fig. 3, and represents the extent to which subgraph expression in that participant was influenced by the two normative learning factors. Each point
represents one participant. The red line represents the regression line and the shaded area represents the 95% confidence interval.
Supplementary figure 6

The relationship between individual normative learning and the average expression of each subgraph. Each point represents one participant. The red line represents the regression line and the shaded area represents the 95% confidence interval.
Supplementary figure 7

Contributions of different functional systems in subgraph 4. **a** Contributions of different functional systems to the effect of learning factors on temporal expression of subgraph 4. We removed all edges from one of the 13 systems and re-estimated the coefficients for CPP, RU, reward and residual updating. Then we compared these coefficients with the original coefficients (including all the edges) to estimate the contribution of each system using t-tests. Error bars represent one SEM. (\(*p<0.05, **p<0.01, ***p<.001\)) **b** Contributions of different functional systems to the relationship between normative learning and dynamic modulation and average expression of subgraph 4. We repeated the same procedure and estimated the change of correlation coefficients for each relationship separately. Error bars represent one SEM. (\(*p<0.05\))
Supplementary figure 8

Contributions of different system edges in subgraph 4. a Contributions of different system edges to the effect of learning factors on temporal expression of subgraph 4. We removed all edges for one of the 91 system-by-system connections and re-estimated the coefficients for CPP, RU, reward and residual updating. Then we compared these coefficients with the original coefficients (including all the edges) to estimate the contribution of each system edge using t-tests. The open circles denote the significant system edges in subgraph 4 (as shown in Fig. 3b). An increase in coefficients is shown in red while a decrease is shown in blue. Lower p values are shown in darker color. A p value around 0.0005 corresponds to a corrected p value of .05 after multiple comparisons (i.e., 0.05/91). b Contributions of different system edges to the relationship between normative learning and dynamic modulation and average expression of subgraph 4. We repeated the same procedure and estimated the change of correlation coefficients for each relationship separately.
Robustness check. Subgraphs identified with smaller sliding time window of 8 TRs, with 6 TRs overlapping between time windows. a Edges between systems in the ten subgraphs identified by NMF, as in Fig. 3b. b Summary of the pattern of connectivity in subgraph 4, as in Fig. 4a, showing the within-system strength and between-system strength of each functional system. The 95% confidence interval of each system was estimated by bootstrapping 10,000 times on the edges of that system.
Supplementary figure 10

Robustness check. Modulation of temporal expression by learning factors in all ten subgraphs identified with smaller sliding time window of 8 TRs (Supplementary Fig. 9). Each point represents one participant. Error bars represent one SEM. (*p<0.05, **p<0.01)
Supplementary figure 11

Robustness check. The relationship between individual normative learning and the dynamic modulation of subgraph expression by normative factors in all ten subgraphs identified with smaller sliding time window of 8 TRs (Supplementary Fig. 10). Each point represents one participant. The red line represents the regression line and the shaded area represents the 95% confidence interval.
Supplementary figure 12

Robustness check. The relationship between individual normative learning and the average subgraph expression in all ten subgraphs identified with smaller sliding time window of 8 TRs (Supplementary Fig. 9). Each point represents one participant. The red line represents the regression line and the shaded area represents the 95% confidence interval.
Supplementary figure 13

Robustness check. Subgraphs identified with larger sliding time window of 12 TRs, with 10 TRs overlapping between time windows. a Edges between systems in the ten subgraphs identified by NMF, as in Fig. 3b. b Summary of the pattern of connectivity in subgraph 4, as in Fig. 4a, showing the within-system strength and between-system strength of each functional system. The 95% confidence interval of each system was estimated by bootstrapping 10,000 times on the edges of that system.
Supplementary figure 14

Robustness check. Modulation of temporal expression by learning factors in all ten subgraphs identified with larger sliding time window of 12 TRs (Supplementary Fig. 13). Each point represents one participant. Error bars represent one SEM. (*p<0.05, ***p<0.001)
Robustness check. The relationship between individual normative learning and the dynamic modulation of subgraph expression by normative factors in all ten subgraphs identified with larger sliding time window of 12 TRs (Supplementary Fig. 14). Each point represents one participant. The red line represents the regression line and the shaded area represents the 95% confidence interval.
Supplementary figure 16

Robustness check. The relationship between individual normative learning and the average subgraph expression in all ten subgraphs identified with larger sliding time window of 12 TRs (Supplementary Fig. 13). Each point represents one participant. The red line represents the regression line and the shaded area represents the 95% confidence interval.
Supplementary figure 17

Robustness check. Subgraphs identified in predicted BOLD signals from univariate GLMs (including predictors for CPP, RU, reward and residual updating). Subgraphs were identified with a sliding time window of 10 TRs, with 8 TRs overlapping between time windows. **a** Edges between systems in the nine subgraphs identified by NMF, as in Fig. 3b. **b** Summary of the pattern of connectivity in subgraph 4, as in Fig. 4a, showing the within-system strength and
between-system strength of each functional system. The 95% confidence interval of each system was estimated by bootstrapping 10,000 times on the edges of that system.
Supplementary figure 18

Robustness check. Modulation of temporal expression by learning factors in all nine subgraphs identified in predicted BOLD signals (Supplementary Fig. 17). Each point represents one participant. Error bars represent one SEM. (*p<0.05, **p<0.01, ***p<0.001)
**Supplementary figure 19**

Robustness check. The relationship between individual normative learning and the dynamic modulation of subgraph expression by normative factors in all nine subgraphs identified in predicted BOLD signals (Supplementary Fig. 18). Each point represents one participant. The red line represents the regression line and the shaded area represents the 95% confidence interval.
Supplementary figure 20

Robustness check. The relationship between individual normative learning and the average subgraph expression in all nine subgraphs identified in predicted BOLD signals (Supplementary Fig. 17). Each point represents one participant. The red line represents the regression line and the shaded area represents the 95% confidence interval.
CHAPTER 4 – Neural encoding of task-dependent errors during adaptive learning

Chang-Hao Kao, Sangil Lee, Joshua I. Gold, Joseph W. Kable. Under revision for resubmitting to *eLife*.

**Abstract**

Effective learning requires using errors in a task-dependent manner, for example adjusting to errors that result from unpredicted environmental changes but ignoring errors that result from environmental stochasticity. Where and how the brain represents errors in a task-dependent manner and uses them to guide behavior are not well understood. We imaged the brains of human participants performing a predictive-inference task with two conditions that had different sources of errors. Their performance was sensitive to this difference, including more choice switches after fundamental changes versus stochastic fluctuations in reward contingencies. Using multi-voxel pattern classification, we identified task-dependent representations of error magnitude and past errors in posterior parietal cortex. These representations were distinct from representations of the resulting behavioral adjustments in dorsomedial frontal, anterior cingulate, and orbitofrontal cortex. The results provide new insights into how the human brain represents errors in a task-dependent manner and guides subsequent adaptive behavior.
Introduction

Errors often drive adaptive adjustments in beliefs that inform behaviors that maximize positive outcomes and minimize negative ones (Sutton & Barto, 1998). A major challenge to error-driven learning in uncertain and dynamic environments is that errors can arise from different sources that have different implications for learning. For example, a bad experience at a restaurant that recently hired a new chef might lead you to update your belief about the quality of the restaurant, whereas a similar experience at a well-known restaurant with a chef that has long been your favorite might be written off as a one-time bad night. That is, the same errors should be interpreted differently in different conditions. In general, errors that represent fundamental changes in the environment or that occur during periods of uncertainty should probably lead you to update your beliefs and change your behavior, whereas those that result from environmental stochasticity are likely better ignored (d’Acremont & Bossaerts, 2016; Li, Nassar, Kable, & Gold, 2019; Nassar, Bruckner, & Frank, 2019; O’Reilly et al., 2013).

Neural representations of key features of these kinds of dynamic, error-driven learning processes have been identified in several brain regions. For example, several studies focused on variables derived from normative models that describe the degree to which individuals should dynamically adjust their beliefs in response to error feedback under different task conditions, including the probability that a fundamental change in the environment just occurred (change-point probability, or CPP, which is a form of surprise) and the reducible uncertainty associated with estimates of environmental features (relative uncertainty, or RU). Correlates of these variables have been identified in dorsomedial frontal (DMFC) and dorsolateral prefrontal (DLPFC) cortex and medial and
lateral posterior parietal cortex (PPC) (Behrens, Woolrich, Walton, & Rushworth, 2007; McGuire, Nassar, Gold, & Kable, 2014; Nassar, McGuire, Ritz, & Kable, 2019). These and other studies also suggest specific roles for these different brain regions in error-driving learning, including representations of surprise induced by either state changes or outliers (irrelevant to state changes) in the PPC that suggest a role in error monitoring (Nassar, Bruckner, et al., 2019; O’Reilly et al., 2013), and representations of variables more closely related to belief and behavior updating in the prefrontal cortex (PFC) (McGuire et al., 2014; O’Reilly et al., 2013). However, these previous studies, which typically used continuous rather than discrete feedback, were not designed to identify neural signals related to a key aspect of flexible learning in uncertain and dynamic environments: responding to the same kinds of errors differently in different conditions.

To identify such task-dependent neural responses to errors, we adapted a paradigm from our previous single-unit recording study (Li et al., 2019). In this paradigm, we generated two different dynamic environments by varying the amount of noise and the frequency that change-points occur (i.e., hazard rate; Behrens et al., 2007; Glaze, Kable, & Gold, 2015; Nassar et al., 2012; Nassar, Wilson, Heasly, & Gold, 2010). In one environment, noise was absent and the hazard rate was high, and thus errors unambiguously signaled a change in state. We refer to this high hazard/low noise condition as the unstable environment, because most errors can be attributed to volatility. In another environment, noise was high and the hazard rate was low, and thus small errors were ambiguous and could indicate either a change in state or noise. We refer to this low hazard/high noise condition as the noisy environment, because most errors can be attributed to stochasticity. Thus, effective learning requires treating errors in the two conditions differently, including adjusting immediately to errors in the unstable
environment but using the size of errors and recent error history as cues to aid interpretation of ambiguous errors in the noisy environment.

In our previous study, we found many single neurons in the anterior cingulate cortex (ACC) or posterior cingulate cortex (PCC) that responded to errors or the current condition, but we found little evidence that single neurons in these regions combined this information in a task-dependent manner to discriminate the source of errors or drive behavior. In the current study, we used whole-brain fMRI and multi-voxel pattern classification to identify task-dependent neural responses to errors and activity predictive of behavioral updating in the human brain. The results show task-dependent encoding of error magnitude and past errors in PPC and encoding of behavioral shifts in frontal regions including ACC, DMFC, DLPFC and orbitofrontal cortex (OFC), which provide new insights into the distinct roles these brain regions play in representing errors in a task-dependent manner and using errors to guide adaptive behavior.

Methods

Participants

All procedures were approved by University of Pennsylvania Internal Review Board. We analyzed data from sixteen participants (9 females, 7 males, mean age = 23.5, SD = 4.3, range = 18–33 years) recruited for the current study. One additional participant was excluded from analyses because of large head movements during MRI scanning (>10% of timepoint-to-timepoint displacements were >0.5 mm). All participants provided informed consent before the experiment. Participants received a participation fee of $15, as well as extra incentives based on their performance (mean = $15.09, SD = $2.26, range = $8.5–17.5).
**Task**

Participants performed a predictive-inference task during MRI scanning. On each trial, participants saw a noisy observation sampled from an unobserved state. The participants’ goal was to predict the location of the noisy observation. To perform this task well, however, they should infer the location of the current state.

In this task (Li et al., 2019), there were 10 targets aligned in a circle on the screen (Figure 1A). At the start of each trial, participants had to fixate a central cross for 0.5 seconds to initialize the trial. After the cross disappeared, participants could choose one of 10 targets (red) by looking at it within 1.5 seconds and keeping fixation on the chosen target for 0.3 seconds. Then, an outcome would be shown for 1 second. During the outcome phase, a green dot indicated the chosen target. A purple or cyan target indicated the rewarded target, with color denoting 10 or 20 points of reward value, respectively. At the end of experiment, every 75 points were converted to $0.25 as participants’ extra incentives.

Participants performed this task in two dynamic conditions separated into two different runs: a high-noise/low-hazard (“noisy”) condition and an low-noise/high-hazard (“unstable”) condition. In the noisy condition, the rewarded target could be one of five targets, given the underlying state (Figure 1B). The rewarded target probabilities for the relative locations ([-2, -1, 0, 1, 2]) of the current state were [0.05, 0.15, 0.6, 0.15, 0.05]. Thus, the location of the current state was most likely rewarded, but nearby targets could also be rewarded. Occasionally, the state would change its location with a hazard rate of 0.02 (Figure 1C). When a change-point happens, the new state would be selected among the ten targets based on a uniform distribution. In the unstable condition, there
was no noise (Figure 1D). That is, the location of the state would be always rewarded.
However, the state was unstable, as the hazard rate in this condition was 0.35 (Figure 1E). There were 300 trials in each run.
**Figure 4-1 Overview of task and experimental design**

(A) Sequence of the task. At the start of the trial, participants look at a cross in the center of the screen and maintain fixation for 0.5 sec to initialize the trial. After the cross disappears, participants choose one of 10 targets (red) by looking at it within 1.5 sec and then holding fixation on the chosen target for 0.3 sec. During the outcome phase (1 sec), a green dot inside the target indicates the participants’ choice. The rewarded target is shown in purple or cyan to indicate the number of earnable points as 10 or 20, respectively. (B) Probability distribution of the rewarded target location in the noisy condition. Target location is relative to the location of the state (generative mean). The rewarded target probabilities for the relative locations of [-2, -1, 0, 1, 2] are [0.05, 0.15, 0.6, 0.15, 0.05]. (C) Example of trials in the noisy condition. The states change occasionally with a hazard rate of 0.02. (D) Probability distribution of the rewarded target location in the unstable condition. Because there is no noise in this condition, the rewarded target is always at the location of the state. (E) Example of trials in the unstable condition. The states change frequently with a hazard rate of 0.35.
Behavior analysis

We investigated how participants used error feedback flexibly across different conditions. Before the behavioral analysis, we removed two different kinds of trials. First, we removed trials in which participants did not make a choice within the time limit (Unstable: median number of trials = 10.5, range = 1–83; Noisy: median = 10, range = 2–88). Second, we also removed trials in which the location of the chosen target was not on the shortest distance between the previously chosen and previously rewarded targets (Unstable: median = 3, range = 0–24; Noisy: median = 17, range = 5–37). All of the belief updating models we tested predict that participants’ choice should be along the shortest distance between the previously chosen target and the previously rewarded target. That is, participants should update in a clockwise direction, if the shortest distance to rewarded target was clockwise of the chosen target. Otherwise, they should update in a counterclockwise direction. We removed trials where participants’ update was in the opposite direction of the rewarded target (which would correspond to a learning rate < 0) and trials where participants’ update was beyond the location of the rewarded target (which would correspond to a learning rate > 1), as this behavior cannot be captured by any of the belief updating models we tested. Further, this behavior might suggest that participants had lost track of the most recently chosen or rewarded targets.

First, we investigated how fast participants learned the location of the current state. For each condition and participant, we binned trials from trial 0 to trial 20 after change-points. Then, we calculated the probability of choosing the location of the current state for each bin.

Second, we examined how different magnitudes of errors lead to shifts in behavior. For each condition and participant, we binned trials based on the current error
magnitude (from 0 to 5). Then, for each bin, we calculated the probability that participants switch their choice to another target on the subsequent trial. We hypothesized participants would have a lower probability of switching after small error magnitudes (1 or 2) in the noisy condition than in the unstable condition since such errors could be due to environment noise in the noisy condition but would signal a state change in the unstable condition.

Third, we further investigated how error history influenced participants' behavioral shifts. Similarly, we binned trials based on the current error magnitude and the error history of the last three trials. Here, we used four bins of error magnitudes (0, 1, 2, 3+). Based on the outcome of correct or error on the last three trials, there were 8 types of error history. For each error magnitude, we calculated the probability of switching for each type of error history. We hypothesized that participants in the noisy condition would tend to switch their choice after small errors more if they had made more errors recently. To test this hypothesis, we ordered the 8 types of error history based on the number of recent errors and calculated the slope of probability of switching against the order of error history. A negative slope means that participants tend to switch as they receive more recent errors.

Behavior modeling

We fit several different computational models to participants' choices to evaluate which ones could best account for their behavior in the task.

Reduced Bayesian (RB) model
Previous studies have shown that a reduced Bayesian model, which approximates the full Bayesian ideal observer, could account well for participants’ behavior in dynamic environments similar to the current task (McGuire et al., 2014; Nassar et al., 2012; Nassar et al., 2010). In this model, belief is updated by a delta rule:

$$
\delta_t = x_t - B_t
$$

$$
B_{t+1} = B_t + \alpha_t \delta_t
$$

where $B_t$ is the current belief and $x_t$ is the current observation. The new belief ($B_{t+1}$) is formed by updating the old belief according to the prediction error ($x_t - B_t$) and a learning rate ($\alpha_t$). The learning rate controls how much a participant revises their belief based on the prediction error. In this model, the learning rate is adjusted on a trial-by-trial basis according to:

$$
\alpha_t = \Omega_t + (1 - \Omega_t) \tau_t
$$

where $\Omega_t$ is the change-point probability and $\tau_t$ is the relative uncertainty. That is, $\alpha_t$ is high as either $\Omega_t$ or $\tau_t$ is high. The change-point probability is the relative likelihood that the new observation represents a change-point as opposed to a sample from the currently inferred state (Nassar et al., 2010):

$$
\Omega_t = \frac{u(x_t|1, 10)H}{u(x_t|1, 10)H + f_p(x_t|y_t, B_t)(1-H)}
$$
where $H$ is the hazard rate, $U(x_t|1,10)$ is the probability of outcome derived from a uniform distribution, and $f_p(x_t|\gamma_t, B_t)$ is the probability of outcome derived from the current predictive distribution. That is, $U(x_t|1,10)$ reflects the probability of outcome when a change-point has occurred while $f_p(x_t|\gamma_t, B_t)$ reflects the probability of outcome when the state has not changed.

The predictive distribution is an integration of the state distribution and the noise distribution:

$$f_p(X|\gamma_t, B_t) = C \times P(X|B_t)^{\gamma_t} \times P(X|B_t)$$  \hspace{1cm} (5)

where $X$ is a random variable determining the locations of target, $P(X|B_t)$ is the noise distribution in the current condition, $P(X|B_t)^{\gamma_t}$ is the state distribution, $\gamma_t$ is the expected run length after the change-point, and $C$ is a normalizing constant to make the sum of probabilities in the predictive distribution equal one. Thus, the uncertainty of this predictive distribution comes from two sources: the uncertainty of the state distribution ($\sigma_s^2$) and the uncertainty of the noise distribution ($\sigma_n^2$). The uncertainty of the state distribution would decrease as the expected run length increases.

The expected run length reflects the expected number of trials that a state remains stable, and thus is updated on each trial based on the change-point probability (Nassar et al., 2010):

$$\gamma_{t+1} = (\gamma_t + 1)(1 - \Omega_t) + \Omega_t$$  \hspace{1cm} (6)
where the expected run length is a weighted average conditional on the change-point probability. If no change-point occurs (i.e., change-point probability is low), the expected run length would increase, leading the uncertainty of the state distribution to decrease. That is, as more observations from the current state are received, participants are more certain about the location of the current state. However, if the change-point probability is high, which signals a likely change in the state, the expected run length would be reset to 1. Thus, the uncertainty of the state distribution becomes large. Participants are more uncertain about the current state after a change-point.

The other factor influencing the learning rate is the relative uncertainty, which is the uncertainty regarding the current state relative to the irreducible uncertainty or noise (McGuire et al., 2014; Nassar et al., 2012):

\[
\tau_{t+1} = \frac{\Omega_t \sigma_t^2 + (1-\Omega_t)\sigma^2 + \Omega_t(1-\Omega_t)[\delta_t(1-\tau_t)]^2}{\Omega_t \sigma_t^2 + (1-\Omega_t)\sigma^2 + \Omega_t(1-\Omega_t)[\delta_t(1-\tau_t)]^2 + \sigma_N^2}
\]  

(7)

The three terms in the numerator contribute to the uncertainty about the current state. The first term reflects the uncertainty conditional on the change-point distribution; the second term reflects the uncertainty conditional on the non-change-point distribution; and the third term reflects the uncertainty due to the difference between the two distributions. The denominator shows the total variance which is the summation of the uncertainty about the current state and the noise. As more precise observations are received in a given state, this relative uncertainty would decrease.

To fit the reduced Bayesian model to behavior, we assumed that participants can depart from the ideal observer by having subjective estimates of the key environmental
variables, hazard rate and noise, that may differ from the true value of these variables. During model fitting, the subjective noise distribution was estimated with the von Mises distribution, which is a circular Gaussian distribution:

\[ P(x_t | B_t, K) = \frac{e^{K \cos(x_t - B_t)}}{\sum_{i=1}^{10} e^{K \cos(x_t - B_i)}} \]  

where \( B_t \) is the location of the current belief, \( x_i \) is the location of target, and \( K \) controls the uncertainty of this distribution. When \( K \) is 0, this is a uniform distribution. As \( K \) increases, the uncertainty decreases. The denominator is used as a normalization term to make sure the sum of all the probabilities equals one. Thus, there are two free parameters in this model: hazard rate (\( H \), in Eq. 4) and noise level (\( K \), in Eq. 8). The range of hazard rate is between 0 and 1 and the noise level is greater than or equal to zero.

**Fixed learning rate (fixedLR) model**

We also consider an alternative model in which participants used a fixed learning rate in each of the two dynamic conditions. That is, the learning rate is the same over all trials in a condition. This model has one free parameter, the fixed learning rate (\( \alpha_{fixed} \)), for each condition (Eq. 2). The fixed learning rate is between 0 and 1.

**Hybrid of RB model and fixedLR model**

Furthermore, we consider a hybrid model, in which the learning rate on each trial is a mixture of the learning rates from the RB model and the fixedLR model:
\[ \alpha_t = w\alpha_{RB} + (1 - w)\alpha_{fixed} \]  

(9)

where \( \alpha_{RB} \) is the learning rate from the RB model and is varied trial by trial according to \( \Omega_t \) and \( \tau_t \), \( \alpha_{fixed} \) is the learning rate from the fixedLR model and \( w \) reflects the weight to integrate these two learning rates. In this model, there are four free parameters: hazard rate, noise level, fixed learning rate and weight. The weight is between 0 and 1.

**Hybrid of RB model and \( P_{stay} \)**

Finally, we consider a hybrid model, which combines the RB model with a fixed tendency to stay on the current target regardless of the current observation. Such a fixed tendency to stay was observed in monkeys in our previous study (Li et al., 2019). Here the belief is updated by:

\[ B_{t+1} = B_t + [(1 - P_{stay}) \times \alpha_t (X_t - B_t) + P_{stay} \times 0] \]  

(10)

where \( P_{stay} \) is the probability that participants stay on the current target. This model has three free parameters: hazard rate, noise level and the probability of stay. The probability of stay is between 0 and 1.

**Model fitting and comparison**
Each model was fitted to data from each participant and within each condition separately. Optimal parameters were estimated by minimizing the mean of the squared error (MSE) between a participant’s prediction and the model prediction.

\[
MSE = \frac{\sum_{t=1}^{n}(B_t - \hat{B}_t)^2}{n}
\]  

(11)

where \( t \) is the trial, \( n \) is the total number of included trials, \( B_t \) is a participant’s prediction on trial \( t \), and \( \hat{B}_t \) is the model prediction on trial \( t \).

Because each model used a different number of parameters and each participant had a different number of included trials, we used Bayesian Information Criterion (BIC) to compare the performance of different models:

\[
BIC = n \ln(MSE) + k \ln(n)
\]  

(12)

where \( n \) is the number of included trials and \( k \) is the number of free parameters in a model. A model with lower BIC performs better.

**MRI Data Acquisition and Preprocessing**

We acquired MRI data on a 3T Siemens Prisma with a 64-channel head coil. Before the task, we acquired a T1-weighted MPRAGE structural image (0.9375 X 0.9375 X 1 mm voxels, 192 X 256 matrix, 160 axial slices, TI = 1,100 ms, TR = 1,810 ms, TE = 3.45 ms, flip angle = 9°). During each run of the task, we acquired functional data using a multiband gradient echo-planar imaging (EPI) sequence (1.9592 X 1.9592 X 2 mm...
voxels, 98 X 98 matrix, 72 axial slices tilted 30° from the AC-PC plane, TR = 1,500 ms, TE = 30 ms, flip angle = 45°, multiband factor = 4). The scanning time (mean = 24.14 minutes, SD = 1.47, range = 21.85-30.00) for each run was dependent on the participants' pace. After the task, fieldmap images (TR = 1,270 ms, TE = 5 ms and 7.46 ms, flip angle = 60°) were acquired.

Data were preprocessed using FMRIB’s Software Library (FSL) (Jenkinson, Beckmann, Behrens, Woolrich, & Smith, 2012; Smith et al., 2004). Functional data were motion corrected using MCFLIRT (Jenkinson, Bannister, Brady, & Smith, 2002), high-pass filtered with a Gaussian-weighted least square straight line fitting of $\sigma = 50 s$, undistorted and warped to MNI space. To map the data to MNI space, boundary-based registration was applied to align the functional data to the structural image (Greve & Fischl, 2009) and fieldmap-based geometric undistortion was also applied. In addition, the structural image was normalized to the MNI space (FLIRT). Then, these two transformations were applied to the functional data.

**fMRI analysis: univariate activity correlated with CPP and RU**

Using similar procedures to our previous study (McGuire et al., 2014), we examined the effects of CPP and RU on univariate activity. Both the current study and the previous study investigate the computational process and neural mechanisms during learning in dynamic environments. The underlying task structures (which involved noisy observations and sudden change-points) are similar between the two studies, but the two studies used very different visual stimuli and motor demands. We specifically focused on the noisy condition in the current study because it was more similar to the
underlying structure, in terms of noisy observations and hazard rate of change-points, to our previous study.

We investigated the factors of CPP, RU, reward values and residual updates. The trial-by-trial CPP and RU were either estimated from the RB model with subjective estimates of hazard rate and noise (as this was the best-fitting model in the current study, analyses presented in Figure 2 – Supplement 3) or from the RB model with true estimates of hazard rate and noise (as this corresponds to how correlates of CPP and RU were identified in our previous study, analyses presented in Figure 2 – Supplement 4). The residual update reflects the difference between the participants’ update and the predicted update, and is estimated from a behavioral regression model in a similar manner as our previous study:

\[
\text{Update}_t = \beta_0 + \beta_1 \delta_t + \beta_2 \delta_t \Omega_t + \beta_3 \delta_t (1 - \Omega_t) \tau_t + \beta_4 \delta_t \text{Reward} + \epsilon
\]  

(13)

where \(\text{Update}_t\) is the difference between \(B_{t+1}\) and \(B_t\), \(\delta_t\) is the error magnitude, both \(\Omega_t\) and \(\tau_t\) were derived from the RB model, and the reward value indicated whether a correct response earned a large or a small value on that trial.

Then, a general linear model using these four factors was implemented on the neural data. Here we further smoothed the preprocessed fMRI data with a 6 mm FWHM Gaussian kernel. We included several trial-by-trial regressors of interest in the GLM: onsets of outcome, CPP, RU, reward value, and residual update. Six motion parameters were also included as confounds. To control false-positive rates (Eklund et al., 2016), statistical testing was implemented through one-sample cluster-mass permutation tests with 5,000 iterations. The cluster-forming threshold was uncorrected voxel \(p<0.01\).
Statistical testing was then based on the corrected cluster $p$ value. For the conjunction analyses, we used the same procedure as the previous study (McGuire et al., 2014). We kept regions that passed the corrected threshold and showed the same sign of effects. For these conjunction tests, we only kept regions that have at least 10 contiguous voxels.

Because the number of participants was fewer in this study ($n=16$) than in the previous study ($n=32$), we might have lower power to detect effects in the whole-brain analyses. Thus, we also implemented ROI analyses. We selected seven ROIs that showed the conjunction effects of CPP, RU and reward value in the previous study (McGuire et al., 2014) and tested the effects of CPP and RU in these ROIs.

We found previously that for a similar task, the presence or absence of reward on a given trial influenced both belief-updating behavior and some aspects of its neural representation (McGuire et al., 2014). To further examine those effects, here we included two different earnable values (10 versus 20 points). However, we did not find any significant effects of earnable values on either belief updating ($\beta_4$ in Eq. 13 was not significantly different than zero) or neural activity (for the contrast of high versus low earnable value). We therefore do not further consider the effects of this manipulation in the current report. We speculate that this lack of an effect contrasts from our earlier finding because here we used high versus low earnable values, whereas in that study we used the presence versus absence of earnable value.

**fMRI analysis: multi-voxel pattern analysis (MVPA)**

We implemented MVPA to understand the neural representation of error signals and subsequent choices. Our analyses focus on the multi-voxel pattern when
participants received an outcome. Before implementing MVPA, we estimated trial-by-trial beta values using the unsmoothed preprocessed fMRI data. We used the general linear model (GLM) to estimate the beta weights for each trial (Mumford, Turner, Ashby, & Poldrack, 2012). In each GLM, the first regressor is the trial of interest and the second combines the rest of trials in the same condition. These two regressors were then convolved with a gamma hemodynamic response function. In addition, six motion parameters were included as control regressors. We repeated this process (one GLM per trial) to estimate trial-by-trial beta values for all the trials in the two conditions. We then used these beta values as observations for MVPA. A whole-brain searchlight was implemented (Kriegeskorte, Goebel, & Bandettini, 2006). In each searchlight, a sphere with the diameter of 5 voxels (10 mm) was formed, and the pattern of activity across the voxels within the sphere were used to run MVPA.

A support vector machine (SVM) with a linear kernel was used to decode different error signals and choices in our whole-brain searchlight analysis. We implemented SVM through the LIBSVM toolbox (Chang & Lin, 2011). To avoid overfitting, we used 3-fold cross-validation, with one fold used as testing data and the other two as training data. Training data were used to train the classifier and then this classifier was used on testing data to examine the classification accuracy. In linear SVM, a free parameter c regularizes the trade-off between decreasing training error and increasing generalization. Thus, during the training of classifier, the training data were further split into 3-folds to select the optimal value of the parameter c through cross-validation. We pick the optimal value for c from \([0.001, 0.01, 0.1, 1, 10, 100, 1000]\) and this optimal parameter should maximize the cross-validation accuracy. Then, we used the optimal parameter c to train the model again based on the entire training data and
calculated the classification accuracy on the testing data. We repeated this procedure with each of the three folds held out as testing data and calculated the average of the classification accuracy. To minimized the influence of different number of trials for each category on the classification accuracy, we used balanced accuracy. For balanced accuracy, we first calculated the classification accuracy within each category, and then averaged the accuracies across all categories. The baseline balanced accuracy was also validated via permutations with 5,000 iterations. For each permutation, each trial was randomly assigned one category with a probability proportional to the number of trials in that category among all the trials. We then used the average of balanced accuracy across these iterations as the baseline accuracy. The baseline accuracy for two categories was 50% and for three categories was 33%.

We first examined how the multi-voxel neural pattern on the current trial could discriminate correct versus error on the current trial or error magnitudes on error trials. For the analysis of error magnitudes, we split trials into three bins of error magnitude: 1, 2, and 3+.

We next examined how the multi-voxel neural pattern on the current trial could discriminate whether the previous trial was an error or not. We also investigated how the classification of past errors differs conditional on the type of the current trial. We classified trial \( t-1 \) as correct or error separately for four different types of current trials: correct trials, error trials, trials with error magnitudes of 0 or 3+ and trials with error magnitudes of 1 or 2. We differentiated between trials with error magnitudes of 0 or 3+ and trials with error magnitudes of 1 or 2 because error magnitudes of 0 or 3+ provide unambiguous evidence regarding a change of state in the noisy condition while error
magnitudes of 1 or 2 provide ambiguous evidence about a change of the state in the noisy condition.

Lastly, we examined how the multi-voxel neural pattern on the current trial could classify the choice on the next trial. In this analysis, we focused only on the trials with error magnitudes of 1 or 2 in the noisy condition, because only under these conditions were participants similarly likely to switch versus stay. For these trials, we examined whether the multi-voxel pattern on the current trial predicted whether the participant stayed or switched on the next trial.

After obtaining the classification accuracy for each participant, we subtracted the baseline accuracy from the classification accuracy. Before conducting a group-level test, we smoothed these individual accuracy maps with a 6 mm FWHM Gaussian kernel. To control false-positive rates (Eklund et al., 2016), statistical testing was implemented through one-sample cluster-mass permutation tests with 5,000 iterations. We used uncorrected voxel $p<0.01$ to form a cluster and estimated the corrected cluster $p$ value for each cluster. For comparison, we report our results using other cluster-forming procedures in supplemental analyses. For the conjunction analyses, we used the same procedure described above.

**Results**

Sixteen human participants performed a predictive-inference task (Figure 1A) while fMRI was used to measure their blood-oxygenation-level-dependent (BOLD) brain activity. The task required them to predict the location of a single rewarded target from a circular array of ten targets. The location of the rewarded target was sampled from a distribution based on the location of the current best target and the noise level in the
current condition. In addition, the location of the best target could change according to a particular, fixed hazard rate ($H$). Two conditions with different noise levels and hazard rates were conducted in separate runs. In the noisy condition (Figure 1B–C), the rewarded target would appear in one of the five locations relative to the location of the current best target, and the hazard rate was low ($H = 0.02$). In the unstable condition (Figure 1D–E), the rewarded target always appeared at the location of the best target, and the hazard rate was high ($H = 0.35$). On each trial, participants made a prediction by looking at a particular target, and then were given explicit, visual feedback about their chosen target and the rewarded target. Effective performance required them to use this feedback in a flexible and task-dependent manner, including typically ignoring small errors in the noisy condition but responding to small errors in the unstable condition by updating their beliefs about the best-target location.

**Behavior**

Nearly all of the participants’ choice patterns were consistent with a flexible, task-dependent learning process (closed symbols in Figure 2). On average, they learned the location of the best target after a change in its location more quickly and reliably in the unstable than the noisy condition (Figure 2A). This flexible learning process had two key signatures. First, target switches (i.e., predicting a different target than on the previous trial) tended to follow errors of any magnitude in the unstable condition but only errors of high magnitude (i.e., when the chosen target was 3, 4, or 5 targets away from the rewarded target) in the noisy condition (sign test for $H_0$: equal probability of switching for the two conditions; error magnitude of 1: median = -0.35, interquartile range (IQR) = [-0.62, -0.25], $p<0.001$; error magnitude of 2: median = -0.30, IQR = [-0.70, -0.11],
p<0.001; Figure 2B–C). Second, target switches depended on error history only for low-magnitude errors (i.e., when the chosen target was 1 or 2 targets away from the rewarded target) in the noisy condition but not otherwise (sign test for H₀: switching was unaffected when recent history contained fewer errors; error magnitude of 1: median = -0.29, IQR = [-0.42, -0.10], p=0.004; error magnitude of 2: median = -0.25, IQR = [-0.38, -0.14], p<0.001; Figure 2D–F).
Figure 4-2 Behavioral results

(A) Probability of choosing the best target after change-points. Symbols and error bars are mean±SEM across subjects (solid symbols) or simulations (open symbols). (B) Relationship between error magnitude and switch probability. Symbols and error bars are as in A. (C) The distribution of switch probabilities for small errors (magnitude of 1 or 2) in both conditions. Each data point represents one participant. Distributions for all error magnitudes are shown in Figure 2 – Figure Supplement 1. (D) Probability of switch as a function of current error magnitude and error history in the unstable condition. Different colors represent different error histories for the past 3 trials. A correct trial is marked as O, and an error trial is marked as X. For example, XOO implies that trial t-1 was an error trial, and trial t-2 and trial t-3 were correct trials. Symbols and error bars are mean±SEM across subjects. (E) Probability of switch as a function of current error magnitude and error history in the noisy condition. Symbols and error bars are as in D. (F) The distribution of the slopes of switch probability against error history for small errors (magnitude of 1
or 2) in both conditions. Each data point represents one participant. Distributions for all error magnitudes are shown in Figure 2 – Figure Supplement 1.
Table 4-1 BIC of behavior models

<table>
<thead>
<tr>
<th>Model</th>
<th>Condition</th>
<th>BIC improvement by RB model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reduced Bayesian model (RB)</td>
<td>Unstable</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td></td>
</tr>
<tr>
<td>Fixed learning rate model (fixedLR)</td>
<td>Unstable</td>
<td>5.06 [3.63, 5.71]**</td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td>-21.05 [-76.63, 0.20]†</td>
</tr>
<tr>
<td>RB + fixedLR</td>
<td>Unstable</td>
<td>-9.83 [-11.20, -8.07]***</td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td>-4.64 [-10.51, 0.89]</td>
</tr>
<tr>
<td>RB + P_stay</td>
<td>Unstable</td>
<td>-5.20 [-5.65, -3.68]**</td>
</tr>
<tr>
<td></td>
<td>Noisy</td>
<td>-5.55 [-5.65, -2.67]*</td>
</tr>
</tbody>
</table>

Values are shown as median [IQR]. A negative value means that the RB model performed better than the alternative model. Significance was tested by a sign test. †p<0.08, **p<0.01, ***p<0.001.
Table 4-2 Parameters of behavior models

<table>
<thead>
<tr>
<th>Model</th>
<th>Parameter</th>
<th>Unstable</th>
<th>Noisy</th>
<th>Unstable &gt; Noisy</th>
</tr>
</thead>
<tbody>
<tr>
<td>RB</td>
<td>H</td>
<td>0.82 [0.64, 0.90]</td>
<td>0.33 [0.11, 0.50]</td>
<td>0.37 [0.24, 0.62]***</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>0.59 [0.03, 2.22]</td>
<td>1.86 [1.22, 2.32]</td>
<td>-0.23 [-1.97, 0.71]</td>
</tr>
<tr>
<td>fixedLR</td>
<td>$\alpha_{\text{fixed}}$</td>
<td>0.96 [0.86, 0.97]</td>
<td>0.63 [0.37, 0.73]</td>
<td>0.33 [0.19, 0.49]***</td>
</tr>
<tr>
<td>RB + fixedLR</td>
<td>H</td>
<td>0.07 [0.00, 0.86]</td>
<td>0.03 [0.00, 0.19]</td>
<td>0.03 [-0.03, 0.77]</td>
</tr>
<tr>
<td></td>
<td>$\alpha_{\text{fixed}}$</td>
<td>0.96 [0.75, 1.00]</td>
<td>0.88 [0.23, 1.00]</td>
<td>0.02 [-0.12, 0.52]</td>
</tr>
<tr>
<td></td>
<td>w</td>
<td>0.38 [0.16, 0.81]</td>
<td>0.71 [0.52, 0.87]</td>
<td>-0.28 [-0.57, 0.22]</td>
</tr>
<tr>
<td>RB + $P_{\text{stay}}$</td>
<td>H</td>
<td>0.73 [0.64, 0.88]</td>
<td>0.31 [0.06, 0.53]</td>
<td>0.27 [0.15, 0.66]**</td>
</tr>
<tr>
<td></td>
<td>K</td>
<td>8.42 [0.73, 30.42]</td>
<td>2.19 [1.62, 9.09]</td>
<td>2.71 [-2.60, 23.94]</td>
</tr>
<tr>
<td></td>
<td>$P_{\text{stay}}$</td>
<td>0.01 [0.00, 0.05]</td>
<td>0.01 [0.00, 0.13]</td>
<td>0.00 [-0.11, 0.03]</td>
</tr>
</tbody>
</table>

Parameter values are shown as median [IQR]. Difference of parameter values between the two conditions was tested by a sign test. **$p<0.01$, ***$p<0.001$. 
We accounted for these behavioral patterns with a reduced Bayesian model that is similar to ones we have used previously to model belief updating in a dynamic environment (open symbols in Figure 2; Tables 1 and 2). This model provides a framework to interpret and use errors differently according to the current task conditions, as defined by hazard rate and noise level. The decision-maker’s trial-by-trial updates are governed by ongoing estimates of the probability that the best target changed (change-point probability, or CPP) and reducible uncertainty about the best target’s location (relative uncertainty, or RU). Both quantities are influenced by the two free parameters in the model, subjective estimates of the task hazard rate and noise level, which were fitted separately in each condition for each participant. As expected, the fitted hazard rates were higher in the unstable condition than in the noisy condition, although both tended to be higher than the objective values, as we have observed previously (Nassar et al., 2010). However, the fitted noise estimates were not reliably different between the noisy and unstable conditions (Table 2). As we observed in our previous study (Li et al., 2019), the subjective estimates of noise level were high in the unstable condition despite the objective absence of noise.

We also tested several alternative models but they did not provide as parsimonious descriptions of the data (Figure 2 – figure supplement 2, and Tables 1 and 2). Notably, an alternative model that assumed a condition-specific fixed learning rate also assumed errors were treated differently in the two conditions but did not include trial-by-trial adjustments of learning rates used by the reduced Bayesian model. Although this model performed better than the reduced Bayesian model in the unstable condition, it cannot capture participants’ behaviors in the noisy condition, where dynamically integrating both current and past errors is required for adapting trial-by-trial
behavior. Other hybrid models performed worse than the reduced Bayesian model in both conditions.

**Neural representation of CPP and RU**

The two key internal quantities in the reduced Bayesian model are CPP and RU, both of which contribute to processing errors in a task-dependent manner (Figure 2 – figure supplement 3). CPP increases as the current error magnitude increases and achieves large values more quickly in the unstable condition because of the higher hazard rate. These dynamics lead to a greater probability of switching targets after smaller errors in the unstable condition. RU increases on the next trial after the participant makes an error and does so more strongly in the noisy condition because of the lower hazard rate. These dynamics lead to a greater probability of target switches when the last trial was an error, which is most prominent for small errors in the noisy condition. Thus, CPP and RU each account for one of the two key signatures of task-dependent learning that we identified in participants' behavior, with CPP driving a task-dependent influence of error magnitude and RU driving a task-dependent influence of error history on target switches.

Though not the main focus of this study, we were able to replicate our previous findings regarding the neural representations of CPP and RU (McGuire et al., 2014). Similar to our previous study, we found activity that was positively correlated with the levels of CPP and RU across DLPFC and PPC (Figure 2 – figure supplement 3 and 4). The regions of DLPFC and PPC that were responsive to both CPP and RU were a subset of those identified as showing this conjunction in our previous study. Because CPP and RU both contribute to responding to errors in a task-dependent manner, the
brain regions that responded to both variables are good candidates for encoding errors in a task-dependent manner. In the following analyses, we aimed to directly identify task-dependent neural representations of error magnitude and error history, as well as activity that predicts subsequent shifts in behavior.

Task-dependent neural representation of errors

We used multi-voxel pattern analysis (MVPA) to identify error-related neural signals that were similar and different for the two task conditions. Given the two key signatures of flexible learning that we identified in behavior, we were especially interested in identifying neural representations of error magnitude and past errors that were stronger in the noisy than the unstable condition.

We found robust, task-dependent representations of the magnitude of the error on the current trial in PPC. Consistent with the task-dependent behavioral effects, this representation of error magnitude was stronger in the noisy than the unstable condition (Figure 3 and Table 3). Specifically, we could classify correct versus error feedback on the current trial across almost the entire cortex, in both the unstable and noisy conditions. However, for error trials, we could classify error magnitude (in three bins: 1, 2, 3+ targets away from the rewarded target) only for the noisy condition and most strongly in the lateral and medial parietal cortex and in the occipital pole. In a parallel set of analyses, we found that univariate activity in PPC also varied in a task-dependent way, responding more strongly to error magnitude in the noisy than the unstable condition (Figure 3 – figure supplement 1).

We also found robust, task-dependent representations of past errors in PPC. These representations also were stronger in the noisy than the unstable condition,
particularly on trials for which past errors had the strongest influence on behavior. Specifically, we could classify correct versus error on the previous trial in PPC for both task conditions (Figure 4). This classification of past errors depended on the outcome of the current trial. We separated trials according to whether the current feedback was correct or an error, or whether the error magnitude provided ambiguous (error magnitudes of 1 or 2) or unambiguous (error magnitudes of 0 or 3+) feedback in the noisy condition (Figure 4). We found reliable classifications of past errors in the lateral and medial parietal cortex in both conditions for correct trials and trials with error magnitudes of 0 or 3+. Moreover, these representations depended on the current condition, and, consistent with behavioral effects of error history, were stronger for error trials and trials with error magnitudes of 1 or 2 in the noisy than in the unstable condition (Table 3). These task-dependent signals for past errors were not clearly present in univariate activity (Figure 4 – figure supplement 1). An additional conjunction analysis across MVPA results showed that PPC uniquely encoded task-dependent error signals for both error magnitude of the current trials and past errors when the current trial’s error magnitude was 1 or 2 (Table 3).
Figure 4-3 Representations of error and error magnitude

For error versus correct analyses, multi-voxel neural patterns were used to classify whether the response on the current trial was correct or an error. For error magnitude analyses, multi-voxel neural patterns were used to classify different error magnitudes (1, 2, 3+) conditional on the current trial being an error. Accuracies were calculated and compared with the baseline accuracy within each subject and then tested at the group level. The representation of current error magnitude is stronger in parietal cortex in the noisy condition than the unstable condition. The cluster-forming threshold was an uncorrected voxel \( p < 0.01 \) \((t=2.6)\), with cluster extent corrected for multiple comparisons using non-parametric permutation tests.
Figure 4-4 Representations of errors on the previous trial conditional on different types of current trials (columns)

Multi-voxel neural patterns were used to classify correct responses versus errors on the previous trial. This analysis was repeated for different types of current trials: all feedback, correct feedback, error feedback, error magnitude of 0 or 3+, and error magnitude of 1 or 2. The representation of past errors is stronger in parietal cortex in the noisy condition than the unstable condition when the current trial is an error or the current error magnitude is 1 or 2. The cluster-forming threshold was an uncorrected voxel $p<0.01$ ($t=2.6$), with cluster extent corrected for multiple comparisons using non-parametric permutation tests.
Table 4-3 Summary of fMRI results: error magnitude and past error

<table>
<thead>
<tr>
<th>Cluster index</th>
<th>#Voxels</th>
<th>Region</th>
<th>Peak t</th>
<th>Peak x</th>
<th>Peak y</th>
<th>Peak z</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Error magnitude: noisy versus unstable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>21032</td>
<td>R Precuneus</td>
<td>5.22</td>
<td>16</td>
<td>-56</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R Angular gyrus</td>
<td>5.17</td>
<td>44</td>
<td>-70</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L Precuneus</td>
<td>5.08</td>
<td>-18</td>
<td>-58</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Occipital pole</td>
<td>5.07</td>
<td>2</td>
<td>-98</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L Superior parietal lobule</td>
<td>4.91</td>
<td>-10</td>
<td>-66</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R Occipital cortex</td>
<td>4.69</td>
<td>26</td>
<td>-76</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L Occipital cortex</td>
<td>4.54</td>
<td>-38</td>
<td>-86</td>
<td>26</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R Superior parietal lobule</td>
<td>4.44</td>
<td>44</td>
<td>-44</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posterior cingulate cortex</td>
<td>4.43</td>
<td>2</td>
<td>-46</td>
<td>20</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Past error on current error magnitude of 1 or 2: noisy versus unstable</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1881</td>
<td>Posterior cingulate cortex</td>
<td>4.79</td>
<td>12</td>
<td>-24</td>
<td>52</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R Superior parietal lobule</td>
<td>4.04</td>
<td>32</td>
<td>-38</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R Precuneus</td>
<td>3.58</td>
<td>6</td>
<td>-54</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L Superior parietal lobule</td>
<td>3.54</td>
<td>-16</td>
<td>-54</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Conjunction: Error magnitude &amp; Past error on current error magnitude of 1 or 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>304</td>
<td>R Superior parietal lobule</td>
<td>3.41</td>
<td>38</td>
<td>-40</td>
<td>52</td>
</tr>
<tr>
<td>2</td>
<td>103</td>
<td>R Precuneus</td>
<td>3.02</td>
<td>2</td>
<td>-58</td>
<td>70</td>
</tr>
<tr>
<td>3</td>
<td>81</td>
<td>L Superior parietal lobule</td>
<td>3.23</td>
<td>-18</td>
<td>-56</td>
<td>72</td>
</tr>
</tbody>
</table>
Neural prediction of subsequent changes in behavior

Although PPC responds to errors in a task-dependent manner that could be used for determining behavioral updates, we did not find that activity in this region was predictive of the participants’ future behavior. Instead, we found such predictive activity more anteriorly in the frontal lobe. Specifically, we investigated whether multi-voxel neural patterns could predict participants’ target switches on the subsequent trial. We focused on the trials with small error magnitudes (1 or 2) in the noisy condition, because these were the only trial types that participants consistently exhibited an intermediate probability of switching (20–80%, Figure 2). We found that activity patterns in large cluster encompassing motor cortex, OFC, ACC, DMFC, and DLPFC could predict subsequent stay/switch decisions (Figure 5, Table 4). We also evaluated this result with different approaches to cluster formation that were more or less spatially specific (Figure 5 – figure supplement 1). We did not find any regions where univariate activity reliably predicted participants’ subsequent behavior (Figure 5 – figure supplement 2).
Figure 4-5 Representations of subsequent behavioral choices (switch versus stay) after ambiguous small errors in the noisy condition

(A) Overlap of results for switch versus stay on the next trial and error magnitude on the current trial. Multi-voxel neural patterns were used to classify whether participants switch their choice to another target or stay on the same target on the next trial. We focused on the most ambiguous errors (error magnitude of 1 or 2 in the noisy condition). Above-chance classification performance was found in a large cluster encompassing the frontal lobe. The cluster-forming threshold was an uncorrected voxel $p<0.01 (t=2.6)$, with cluster extent corrected for multiple comparisons using non-parametric permutation tests. (B) Overlap of results for switch versus stay on the next trial and past error conditional on error magnitude of 1 or 2 on the current trial.
<table>
<thead>
<tr>
<th>Cluster index</th>
<th>#Voxels</th>
<th>Region</th>
<th>Peak t</th>
<th>Peak x</th>
<th>Peak y</th>
<th>Peak z</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12042</td>
<td>Middle cingulate cortex</td>
<td>4.35</td>
<td>14</td>
<td>-8</td>
<td>30</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R Insula</td>
<td>4.33</td>
<td>38</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Medial orbitofrontal cortex</td>
<td>4.24</td>
<td>-4</td>
<td>50</td>
<td>-10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R Frontal pole</td>
<td>4.11</td>
<td>40</td>
<td>46</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R Inferior frontal gyrus</td>
<td>4.11</td>
<td>48</td>
<td>26</td>
<td>10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L Frontal pole</td>
<td>4.01</td>
<td>-24</td>
<td>52</td>
<td>-2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Dorsomedial frontal cortex</td>
<td>3.96</td>
<td>0</td>
<td>26</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Posterior cingulate cortex</td>
<td>3.93</td>
<td>2</td>
<td>-28</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>R Primary motor cortex</td>
<td>3.91</td>
<td>48</td>
<td>-6</td>
<td>50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Anterior cingulate cortex</td>
<td>3.51</td>
<td>0</td>
<td>48</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>3134</td>
<td>L Premotor cortex</td>
<td>4.43</td>
<td>-62</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L Superior temporal gyrus</td>
<td>4.28</td>
<td>-50</td>
<td>-32</td>
<td>12</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L Inferior frontal junction</td>
<td>3.72</td>
<td>-38</td>
<td>4</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td></td>
<td>L Postcentral gyrus</td>
<td>3.61</td>
<td>-50</td>
<td>-26</td>
<td>44</td>
</tr>
</tbody>
</table>
Discussion

We identified task-dependent neural representations of errors in humans performing dynamic learning tasks. Participants were required to learn in two different dynamic environments. In the unstable condition (high hazard rate and low noise), errors unambiguously indicated a change in the state of the environment, and participants reliably updated their behavior in response to errors. In contrast, in the noisy condition (low hazard rate and high noise), small errors were ambiguous, and participants used both the current error magnitude and recent error history to distinguish between those errors that likely signal change-points and those likely arising from environmental noise. Using MVPA, we showed complementary roles of PPC and prefrontal regions (including motor cortex, OFC, ACC, DMFC and DLPFC) in the outcome-monitoring and action-selection processes underlying these flexible, task-dependent behavioral responses to errors. Neural patterns in PPC encoded the magnitude of errors and past errors, more strongly in the noisy than the unstable condition. These task-dependent neural responses to errors in PPC were not reliably linked to subsequent changes in behavior. In contrast, neural patterns in prefrontal regions could predict subsequent changes in behavior (whether participants switch their choice on the next trial or not) in response to ambiguous errors in the noisy condition.

Task-dependent behavior adaptation

Consistent with previous studies of ours and others (d’Acremont & Bossaerts, 2016; McGuire et al., 2014; Nassar, Bruckner, et al., 2019; Nassar et al., 2012; Nassar et al., 2010; O’Reilly et al., 2013; Purcell & Kiani, 2016), human participants adapted
their response to errors differently in different environments. In the unstable condition, participants almost always switched their choice after errors and quickly learned the new state after change-points. In contrast, in the noisy condition, participants ignored many errors and only slowly learned the new state after change-points. In this condition, participants had to distinguish true change-points from environmental noise, and they used error magnitude and recent error history as a cue for whether the state had recently changed or not. These flexible and task-dependent responses to errors could be accounted for by a reduced Bayesian model (McGuire et al., 2014; Nassar et al., 2012; Nassar et al., 2010). This model assumes that participants use approximately optimal inference processes but can have subjective estimates of environmental parameters (hazard rate, noise) that depart from their true values.

**Neural representation of change-point probability and relative uncertainty**

In the reduced Bayesian model, beliefs and behavior are updated dynamically according to two key internal quantities, CPP and RU. Replicating our previous work (McGuire et al., 2014), we identified neural activity correlated with both CPP and RU in PPC and DLPFC. This replication shows the robustness of these neural representations of CPP and RU across experimental designs that differ dramatically in their visual stimuli and motor demands, yet share the need to learn in dynamic environments with similar statistics. We extended those findings to show that some brain regions that encode both CPP, which in the model accounts for task-dependent behavioral responses to error magnitude, and RU, which in the model accounts for task-dependent behavioral responses to recent error history, also encode errors in a task-dependent manner or predict subsequent behavioral updates.
Task-dependent neural representation of errors

Advancing beyond previous work, we identified task-dependent encoding of errors in neural activity in the PPC. Mirroring the task dependence of behavior, the multivariate neural pattern in PPC encoded current error magnitude more strongly in the noisy condition than in the unstable condition and encoded past errors more strongly on trials that provided ambiguous feedback in the noisy condition. These same regions of PPC have been shown previously to represent errors, error magnitudes, surprise and salience (Fischer & Ullsperger, 2013; Gläscher, Daw, Dayan, & O'Doherty, 2010; McGuire et al., 2014; Nassar, Bruckner, et al., 2019; Nassar, McGuire, et al., 2019; O'Reilly et al., 2013; Payzan-LeNestour, Dunne, Bossaerts, & O'Doherty, 2013). In addition, these regions have been shown to integrate recent outcome or stimulus history in human fMRI studies (FitzGerald, Moran, Friston, & Dolan, 2015; Furl & Averbeck, 2011) and in animal single neuron recording studies (Akrami, Kopec, Diamond, & Brody, 2018; Brody & Hanks, 2016; Hanks et al., 2015; Hayden, Nair, McCoy, & Platt, 2008; Hwang, Dahlen, Mukundan, & Komiyama, 2017). Our results extend on these past findings by demonstrating that the neural encoding of error magnitude and error history in PPC is modulated across different conditions in precisely the manner that could drive adaptive behavior.

These whole-brain fMRI results complement our previous results recording from single neurons in ACC and PCC in the same task (Li et al., 2019). In that study, we identified single neurons in both ACC and PCC that encoded information relevant to interpreting errors, such as the magnitude of the error or the current condition. However, we did not find any neurons that combined this information in a manner that could drive...
adaptive behavioral adjustments. Our whole-brain fMRI results suggest that PPC would be a good place to look for task-dependent error representations in single neurons, including a region of medial parietal cortex slightly dorsal to the PCC area we recorded from previously.

**Neural representations of task-dependent behavioral updating**

Also advancing beyond previous work, we identified neural activity predictive of behavioral updates across the frontal cortex, including DLPFC. In the noisy condition, small errors provided ambiguous feedback that could reflect either a change in state or environmental noise. Accordingly, after small errors in the noisy condition, participants exhibited variability across trials in whether they switched from their current choice on the subsequent trial or not. In these ambiguous situations, the multivariate neural pattern in large cluster in frontal cortex, including motor cortex, OFC, ACC, DMFC and DLPFC, predicted whether people switched or stayed on the subsequent trial. These results extend previous findings that the multivariate pattern in frontal cortex, particularly ACC and medial PFC, can decode subsequent switching versus staying in a reversal learning task (Hampton & O'Doherty, 2007). These results suggest a dissociation between PPC regions that monitor error information in a task-dependent manner and frontal regions that may use this information to update beliefs and select subsequent actions.

This ability to decode subsequent choices might arise from different kinds of representations in different areas of frontal cortex. Whereas motor and premotor regions may reflect the change in action plans, other frontal regions might reflect changes in abstract representations of belief states. Medial PFC is involved in performance monitoring, distinguishing errors from different sources such as actions and feedback
(Ullsperger, Danielmeier, & Jocham, 2014), registering a hierarchy of prediction errors from those due to environmental noise to those due to a change in the environmental state (Alexander & Brown, 2015), and interacting with lateral PFC to guide subsequent behavioral adjustments in response to errors (Alexander & Brown, 2015). Consistent with this role, activity in DMFC also reflects the extent of belief updating in dynamic environments (Behrens et al., 2007; Hampton, Bossaerts, & O’Doherty, 2006; McGuire et al., 2014; O’Reilly et al., 2013). OFC and DMFC encode the identity of the current latent state in a mental model of the task environment and neural representations in these regions changes as the state changes (Chan, Niv, & Norman, 2016; Hunt et al., 2018; Karlsson, Tervo, & Karpova, 2012; Nassar, McGuire, et al., 2019; Schuck, Cai, Wilson, & Niv, 2016; Wilson, Takahashi, Schoenbaum, & Niv, 2014). Activity in inferior frontal junction reflects the updating of task representations (Brass & Cramon, 2004; Derrfuss, Brass, Neumann, & von Cramon, 2005). Neural activity in frontopolar cortex (Daw, O’Doherty, Dayan, Seymour, & Dolan, 2006) and DMFC (Blanchard & Gershman, 2018; Kolling, Behrens, Mars, & Rushworth, 2012; Kolling et al., 2016; Muller, Mars, Behrens, & O’Reilly, 2019) increases during exploratory choices, which occur more frequently during periods of uncertainty about the most beneficial option. In a recent study, we identified distinct representations of latent states, uncertainty, and behavioral policy in distinct areas of frontal cortex during learning in a dynamic environment (Nassar et al., 2019). Our results extend these past findings and demonstrate the role of these frontal regions in adjusting behavior in response to ambiguous errors.

**Caveats**
A few caveats should be considered when interpreting our results. First, we had relatively small number of participants in this study (n=16). Although we control the false positive rates through permutation tests that have been validated empirically (Eklund, Nichols, & Knutsson, 2016), it is possible that we lacked the statistical power to detect some effects, and so null results should be interpreted with caution. Second, in this study, we created two qualitatively different task conditions by manipulating both the noise levels and hazard rates. Thus, we cannot attribute any behavioral or neural differences across conditions specifically to changes in either noise levels or hazard rates alone, but rather to how the combinations of these two variables affect the interpretation and use of small errors. Future studies can manipulate hazard rate and noise independently to examine their independent contributions to adaptive learning.

Conclusion

People adapt their behavior in response to errors in a task-dependent manner, distinguishing between errors that indicate change-points in the environment versus noise. Here we used MVPA to identify two distinct kinds of neural signals contributing to these adaptive behavioral adjustments. In PPC, neural patterns encoded error information in a task-dependent manner, depending on error magnitude and past errors only under conditions where these were informative of the source of error. In contrast, activity in frontal cortex could predict subsequent choices that could be based on this information. These findings suggest a broad distinction between outcome monitoring in parietal regions and action selection in frontal regions when learning in dynamic and uncertain environments.
Figure 2 - Figure supplement 1 Distributions of behavior as a function of error magnitude. (A) Distributions of switch probability as a function of error magnitude. Each data point represents one participant. (B) Distributions of slopes of switch probability against error history as a function of error magnitude. Each data point represents one participant.
**Figure 2 - Figure supplement 2** Behavioral data and predictions from different models. (A) Probability of choosing the best target after change-points. (RB: reduced Bayesian; fixedLR: fixed learning rate; Pstay: fixed tendency to stay) (B) The relationship between error magnitude and switch probability. (C) Probability of switch as a function of current error magnitude and error history in the unstable condition. (D) Probability of switch as a function of current error magnitude and error history in the noisy condition. Symbols and colors are as in Figure 2.
Figure 2 - Figure supplement 3 Reduced Bayesian model applied to behavioral and imaging data. (A) Model prediction for CPP. We calculated CPP from the fitted reduced Bayesian model, which incorporates subjective estimates of hazard rate and noise for each condition. The value of CPP increases as the current error magnitude increases in both conditions, but with a stronger dependence on the outcome of the previous trial in the noisy condition. (B) Model prediction for
RU. We calculated RU from the fitted reduced Bayesian model, which incorporates subjective estimates of hazard rate and noise for each condition. The value of RU is minimally affected by the current error magnitude. Instead, a past error tends to increase RU. (C) Model prediction for probability of switching choices. Increasing CPP causes the probability of switching to increase more steeply as the current error magnitude increases in the unstable condition versus in the noisy condition. For small errors (error magnitude of 1 and 2) in the noisy condition, the probability of switching is further influence by RU, which is affected by past errors. (D) Neural representation of CPP and RU. CPP selective effect represents the conjunction of CPP>0 and CPP>RU. RU selective effect represents the conjunction of RU>0 and RU>CPP. The results were thresholded based on uncorrected voxel $p<0.01$ ($t=2.6$). (E) ROI analysis for CPP and RU. These ROIs were selected based on the common regions of CPP, RU, and reward effects in McGuire et al. (2014). Significance was tested by a sign test. *$p<0.05$, **$p<0.01$, ***$p<0.001$. 
Figure 2 - Figure supplement 4 Neural representations of CPP and RU from the approximately ideal observer, which is the reduced Bayesian model with true hazard rate and noise, for direct comparison to analyses in McGuire et al. (2014), which used covariates constructed from the ideal rather than the fitted model. (A) Neural representation of CPP and RU in the current study and in McGuire et al. (2014). CPP selective effect represents the conjunction of CPP>0 and CPP>RU. RU selective effect represents the conjunction of RU>0 and RU>CPP. The results were thresholded based on uncorrected voxel $p<0.01$ ($t=2.6$). (B) ROI analysis for CPP and RU. These ROIs were selected based on the common regions of CPP, RU and reward effects in McGuire et al. (2014). Significance was tested by a sign test. *$p<0.05$, **$p<0.01$, ***$p<0.001$. 

CPP & RU
CPP selective
RU selective

A

B
Figure 3 - Figure supplement 1 Univariate representations of error and error magnitude. A GLM was implemented on the preprocessed fMRI data (smoothed with 6 mm FWHM Gaussian kernel). The trial-by-trial regressors of interest that were included in the GLM were: onset of correct trials, earnable value on correct trials, onset of error trials, error magnitude on error trials, switch or stay on error trials and earnable value on error trials. We focused on the effects of error (which is the difference between the onset of error trials and the onset of correct trials) and error magnitude. Group t-values are shown. For statistical testing, we implemented one-sample cluster-mass permutation tests with 5,000 iterations. The cluster-forming threshold was uncorrected voxel p<0.01 (t=2.6).
Figure 4 - Figure supplement 1 Univariate representations of error on the previous trial conditional on different types of current trials (columns). Several GLMs were implemented on the preprocessed fMRI data (smoothed with 6 mm FWHM Gaussian kernel). First, we examined errors on the previous trial across all trials. The trial-by-trial regressors of interest that were included in the GLM were: onset of trials, error on trial t, error on trial t-1, error on trial t-2, and error on trial t-3. We focused on the effect of error on trial t-1. Second, we separated the analysis of past errors conditional on the current trial being correct or an error. The trial-by-trial regressors of interest that were included in the GLM were: onset of current correct trials, errors on trial t-1, t-2, or t-3 conditional on the current trial being correct, onset of current error trials, errors on trial t-1, t-2, or t-3 conditional on the current trial being an error. We focused on the effects of error on trial t-1 conditional on the current trial being correct or an error. Third, we separated errors conditional on error magnitudes of 0 or 3+ or error magnitudes of 1 or 2. The trial-by-trial
regressors of interest that were included in the GLM were: onset of current trials with error magnitudes of 0 or 3+, errors on trial t-1, t-2 or t-3 conditional on the current trial error magnitude of 0 or 3+, onset of current trials with error magnitudes of 1 or 2, errors on trial t-1, t-2 or t-3 conditional on the current trial error magnitude of 1 or 2. We focused on the effects of errors on trial t-1 conditional on the current trials error magnitude of 0 or 3+ or error magnitude of 1 or 2. Group t-values are shown. For statistical testing, we implemented one-sample cluster-mass permutation tests with 5,000 iterations. The cluster-forming threshold was uncorrected voxel $p<0.01$ ($t=2.6$).
Figure 5 - Figure supplement 1 Representations of subsequent behavioral choices (switch versus stay) thresholded via threshold-free cluster enhancement (TFCE) or with a cluster-forming threshold of $p<0.001$. (A) Overlap of results for switch versus stay on the next trial and error magnitude on the current trial. We implemented two types of cluster-forming approaches: TFCE and uncorrected voxel $p<0.001$. First, significance testing was implemented through permutation tests with threshold-free cluster enhancement (FSL’s randomize), which does not require a pre-defined cluster-forming threshold. The result of switch versus stay showed little spatial specificity. For the purpose of display, the results were thresholded based on uncorrected voxel $p<0.03$ ($t=2$). Second, we used a cluster-forming threshold of uncorrected voxel $p<0.001$ ($t=3.73$) and tested the significance of the formed cluster via one-sample cluster-mass permutation tests with 5,000 iterations. The results showed high spatial specificity and several previously identified...
regions were still significant: middle cingulate cortex [14, -8, 30], right insula [38, 4, 2], medial OFC [-4, 50, -10], left premotor cortex [-62, 2, 24] and left superior temporal gyrus [-50, -32, 12].

(B) Overlap of results for switch versus stay on the next trial and past error conditional on error magnitude of 1 or 2 on the current trial. The two types of cluster-forming approaches are shown.
Figure 5 - Figure supplement 2 Univariate GLM for switch versus stay on small error trials (magnitudes of 1 or 2) in the noisy condition. A GLM was implemented with several trial-by-trial regressors of interest: onset of trials with error magnitude of 0, onset of trials with error magnitude of 3+, onset of trials with error magnitudes of 1 or 2 followed by switching, onset of trials with error magnitudes of 1 or 2 followed by staying. We tested the effects of the difference between switch and stay for small errors. For statistical testing, we implemented one-sample cluster-mass permutation tests with 5,000 iterations. The cluster-forming threshold was uncorrected voxel $p<0.01$ ($t=2.6$). There were no significant clusters. For the demonstration, the results were shown with uncorrected voxel $p<0.01$. 
Summary of studies

In these three studies, I showed how humans adjust learning in dynamic environments and how different neural systems are involved during the adjustment of learning. Humans can evaluate the structure of the current dynamic environment, and adaptively integrate prediction errors to update their belief. Several neural systems were involved in this process. Pupil dilation predicted dynamic learning rates, and its association with learning rates was demonstrated in two ways. It served as a bridge between prediction errors and learning rates, and tracked the variance of learning rates that cannot be accounted for by task factors. In the brain, the integration between fronto-parietal network and other networks was modulated by belief surprise and belief uncertainty, and this modulation was associated with individual behavioral adaptation. Furthermore, frontal and parietal regions played distinct functional roles during learning. Posterior parietal cortex encoded errors in a task-dependent manner while a large array of frontal regions predicted subsequent behavioral change. Through these three studies, we showed the computational mechanisms driving neural activity and human behavior during the adjustment of learning.

Task-dependent influence on learning in dynamic environments

Inferring the state in the dynamic environments requires the understanding of the structure of the environment, which is formed by different task factors, such as environmental noise, drifting rate and hazard rate. Evaluating these factors help people
understand how the state changes and how the observation is generated from the state, and help people adaptively use prediction errors to adjust learning.

In chapter 2, I examined two types of dynamic environment: change-point and drifting. In the change-point environment, the stable state would undergo occasionally sudden changes. Based on the hazard rate and environmental noise, people can evaluate how the current prediction errors indicate the likelihood of the change of the state (i.e., surprise), and evaluate their belief uncertainty about the current state (Krishnamurthy et al., 2017; McGuire et al., 2014; Nassar et al., 2012; Nassar et al., 2010). Such surprise and uncertainty drive the adjustment of learning. In the drifting environment, the current state is generated from the previous state. Based on the drifting rate and environmental noise, people can evaluate their belief uncertainty about the state and adjust their learning using this uncertainty (Daw et al., 2006; S. Lee et al., 2020).

In chapter 4, I compared two change-point environments. One was formed with high noise and low hazard rate while the other one was formed with low noise and high hazard rate. Due to the manipulation of environmental noise and hazard rate, the same errors should be treated differently in different conditions (Y. S. Li et al., 2019). In the unstable condition, a small error always indicated the change of the state, and thus people should change their subsequent behavior. In the noisy condition, however, a small error can indicate either environmental noise or a change of the state, and people integrated past error history to decide whether they should stay or switch their subsequent choice.

In these three studies, I used a reduced Bayesian model to describe behaviors in the change-point environments. In this model, belief surprise and belief uncertainty are
calculated by converting prediction errors based on the task factors such as environmental noise and hazard rate. For example, the same error indicated higher surprise as the hazard rate is high or the environmental noise is low. This model demonstrated how humans evaluate prediction errors in a task-dependent manner, and guide their learning by belief surprise and belief uncertainty.

**Integrated neural systems**

From the three studies, I demonstrated the contribution of different neural systems during learning in dynamic environments.

In chapter 2, I showed that pupil dilation predicted trial-by-trial learning rates. This association was not simply a response to the current prediction errors, which drove the change of learning rates. In addition, pupil dilation can directly modulate learning rates. First, it served as a mediator between the current prediction errors and learning rates. Second, it tracked the variance of learning rates that cannot be accounted for by task factors. Among different measures of physiological arousal (e.g., pupil diameter, skin conductance, heart rate and respiration rate), only pupil diameter showed these relationships with learning rates. This suggests that the learning-related arousal is not general arousal regulated by ANS but potentially reflects activity in the LC-NE system. The role of arousal may be related to the plausible function of neural gain in the LC-NE system (Aston-Jones & Cohen, 2005; Mather et al., 2015). Neural gain means that the activation of a neuron that receives excitatory input would increase but the activation of a neuron that receives inhibitory input decreases. Recent studies showed arousal enhanced attention or performance for salient stimuli and inhibited attention or performance for non-salient stimuli (Eldar et al., 2013; T.-H. Lee et al., 2012; T.-H. Lee
et al., 2014). Prediction errors reflect the deviation of the observation from people’ belief, and may serve as a salient signal. Arousal may amplify the processing of prediction errors, leading to the increase of learning rates. On the other hand, the association between arousal and the variance of learning rates that cannot accounted for by task factors may suggest the function of network resets, which reflect the change of functional connectivity (Bouret & Sara, 2005; Servan-Schreiber et al., 1990). Such network resets may then lead to changes in beliefs (Kao, Kambhati, et al., 2020; Nassar, McGuire, et al., 2019). This possibility of network reset was examined in chapter 3.

In chapter 3 and 4, I showed the functions of frontal and parietal regions during learning in dynamic environments. Past studies have showed the involvement of these regions during adaptive learning (Behrens et al., 2007b; McGuire et al., 2014; Nassar, McGuire, et al., 2019; O’Reilly et al., 2013; Payzan-LeNestour et al., 2013) and I further provided insights about their communication as well as their distinct functional roles. In chapter 3, I demonstrated how the integration of fronto-parietal and other networks changes during the adjustment of learning. Past studies showed that the integration of the fronto-parietal network decreases from the early phase of learning to the later phase of learning (Bassett et al., 2015; Büchel et al., 1999; Sun et al., 2006). I further showed that the integration for fronto-parietal network can be flexibly adjusted. That is, as learning is re-initiated when the state changes, the integration of fronto-parietal network increases again. In chapter 4, I further distinguish the functional roles of frontal and parietal regions. Posterior parietal cortex encoded errors in a task-dependent manner while a large array of frontal cortex encoded subsequent behavioral change in response to ambiguous errors. These results emphasize the contribution of frontal and parietal
regions during the adjustment of learning. These results may be associated with the role of these regions in cognitive control. As the state changes, people should flexibly adjust their learning. Cognitive control is needed when task rules or task requirements change. During this process, the activation in frontal and parietal regions as well as the connectivity in the fronto-parietal network was enhanced (Cole et al., 2014; Power & Petersen, 2013).

These three studies provided insights to establish an integrative neural framework related to the LC-NE system and the fronto-parietal network during learning in dynamic environments. High arousal states are associated with changes in connectivity in regions in fronto-parietal networks (Patanaik et al., 2018; Shine et al., 2016; Wang, Ong, Patanaik, Zhou, & Chee, 2016). Furthermore, atomoxetine, a noradrenergic reuptake inhibitor, increases network integration during task performance (Shine, van den Brink, Hernaus, Nieuwenhuis, & Poldrack, 2018) and modulates participants’ learning and behavior (Jepma et al., 2018; Jepma et al., 2016; Kane et al., 2017). These past findings suggest the coordination between the LC-NE system and fronto-parietal networks during learning.

Future directions

These three studies provided insight about the neural mechanisms for learning in dynamic environments, but indeed, there are still open questions. First, the influence of other task factors on learning should be investigated. For example, in addition to hazard rate (i.e., the frequency of the state change), recent studies have begun evaluating the change of hazard rate during learning (Filipowicz et al., 2020; Glaze, Filipowicz, Kable, Balasubramanian, & Gold, 2018). As the hazard rate
changes, people should adapt different strategies to prediction errors. For example, in
the environment with a low hazard rate, a large prediction error or the change of
observation induced surprise. In the environment with a high hazard rate, instead, a
small prediction error or no change of observation induced surprise. It has been showed
that pupil dilation can adjust its response to prediction errors in the environments with
different hazard rates (Filipowicz et al., 2020). However, there is little evidence about
how the brain activation or the dynamic functional connectivity changes in response to
the change of hazard rates. Future studies can address this issue.

Second, confidence about predictions may also drive the adjustment of learning.
Confidence reflects the subjective probability that the prediction is correct. In contrast to
belief uncertainty, which reflects objective uncertainty about the state, confidence
reflects the subjective uncertainty about the prediction made based on the state.
Confidence has shown to be related with uncertainty or uncertainty-related variables
(Meyniel, Sigman, & Mainen, 2015). Even though confidence is closely linked with
decision performance, they can be dissociated. For example, this dissociation was
shown in a study where obsessive-compulsive disorder (OCD) patients performed a
predictive-inference task in a change-point environment (Vaghi et al., 2017). In addition
to make a prediction, they were also required to make a confidence rating after the
prediction. Similar to healthy controls, patients were able to adjust their confidence in
response to the dynamics of the environment. Confidence decreased immediately after a
change-point and increased gradually over subsequent trials after a change-point.
However, patients were unable to reduce learning rates as the belief uncertainty should
be reduced after a change-point. Patients showed reduced coupling between
performance and confidence and the extent of decoupling was related to their symptom
severity. Although confidence is closely linked with uncertainty, it also provides a unique role in the adjustment of behavior (Hangya, Sanders, & Kepecs, 2016; Sanders, Hangya, & Kepecs, 2016). In perceptual tasks, confidence can be treated as an internal source of feedback to during learning (Guggenmos, Wilbertz, Hebart, & Sterzer, 2016; Zylberberg, Wolpert, & Shadlen, 2018). Moreover, confidence is also associated with pupil dilation (Lempert, Chen, & Fleming, 2015) and activation in frontal and parietal regions (De Martino, Fleming, Garrett, & Dolan, 2013). Future studies should examine the involvement of confidence during learning in dynamic environments and its related neural mechanisms.

Third, little is known about distinct patterns of functional connectivity for surprise and uncertainty. We showed that the same functional network represented both surprise and uncertainty in chapter 4. However, given that high surprise leads to the following high uncertainty in a change-point environment (Nassar et al., 2012; Nassar et al., 2010), it is difficult to identify distinct patterns of functional connectivity for surprise and uncertainty since the reliable functional connectivity needs to be calculated in a long time window. Given that surprise and uncertainty were represented in unique brain regions (McGuire et al., 2014; Nassar, Bruckner, et al., 2019), this implies that they may recruit different functional network during learning. Their distinct functional network may be able to be identified using a design that dissociates the temporal relationship of surprise and uncertainty (Nassar et al., 2016). Future studies should examine this hypothesis.

Fourth, it is not well-established about the relationship between arousal/NE, dynamic functional connectivity and learning. That is, parts of these relationship have been investigated but few studies investigated their relationships together. In chapter 2, I showed the association between pupil dilation and learning, and this association may be
related to the involvement of LC-NE system. In chapter 3, I showed the association between the change of integration of fronto-parietal network and learning. As shown in past studies, using drugs to manipulate the NE level can modulate learning rates (Jepma et al., 2018; Jepma et al., 2016) and drive the change of functional connectivity and pupil dilation (van den Brink et al., 2016; Warren et al., 2016). Little is known whether arousal induced by different factors (e.g., prediction errors, incidental reward and incidental auditory oddball) leads to the changes of different functional networks and how these changes drive different influence on learning. Moreover, well-establishing their relationships may provide insights to causally affect learning rates through modulating arousal or NE level or inducing the change of functional connectivity.

**Conclusion**

Forming a precise belief is important to make an adaptive decision. In dynamic environments, belief updating should be flexibly adjusted in response to changes in the state of the environment. During this process, people should evaluate task factors and adaptively use prediction errors to update their beliefs. In these three studies, I showed how the LC-NE system and fronto-parietal networks contribute to the adjustment of learning. The coordination between these systems is important to achieve adaptive behavior in dynamic environments.


Cannon, W. B. (1929). Bodily changes in pain, hunger, fear and rage.


202


