Between Prediction and Reality: Top-down Propagation, Communication and Modulation.

Yuening Yan

Degree of Ph.D

School of Psychology

College of Science and Engineering

University of Glasgow

June 2022

Abstract

Expectations predict the upcoming visual information, facilitating its disambiguation from the noisy input to speed up behavior. However, the neural mechanisms that support such dynamic feature flow and how they facilitate behavior remain unclear. In this thesis, I will trace the propagation, communication and modulation effect related to the prediction.

In the first study (***Chapter 2***), I initiated a cueing-categorization design and validated its feasibility. I first detected the sensibility of participants in distinguishing auditory pitches (i.e., the cues) by estimating their d-primes. Next, in two phases, I instructed participants to build the coupling relationship between auditory cues and stimuli, then proved that compared with non-informative prediction, informative ones can significantly reduce the reaction time.

Recognizing confounding factors in the first study, I further improved the design by manipulating two specific and separate predicted contents. I used a prediction experiment that cued participants (N = 11) to the spatial location (left vs. right) and spatial frequency (SF, Low, LSF, vs. High, HSF) contents of an upcoming Gabor patch. I reconstructed two networks (prediction network and categorization network) in the following two studies with simultaneous MEG recordings of each participant's neural activity.

In the second study (***Chapter 3***), focusing on the pre-stimulus prediction stage, I answered when, where and how predictions dynamically propagate through a network of brain regions. I traced the dynamic neural representation of predictive cues and reconstructed the communications about predicted contents in a functional network, sequentially from temporal lobe at 90-120ms, to occipital cortex after 200ms, with modulatory supervision of frontal regions at 120-200ms.

In the third study (***Chapter 4***), turning to the post-stimulus stage, I reconstructed the communication network propagating the stimulus feature from occipital-ventral regions (150-250ms) to parietal lobe (250-350ms), finally arriving premotor cortex (>350ms) which modulates behavioral categorization. I found the previous prediction previewed and

then sharpen stimulus representation across the categorization network, leading to a faster reaction time.

I discussed the generalization of the findings to other stimulus features and sensory modalities. Putting forward the plans about developing a series of structured studies on predicting higher-dimensional features, in the future, I aim to understand the neural mechanisms about how prediction tunes perception and to trace the concrete predicted contents in laminar layers with the fusion of E/MEG and fMRI.

Acknowledge

Upon the completion of this thesis, I would like to express my sincere appreciation to many people for supporting me and offering me invaluable help during my four-year Ph.D. study.

Firstly, I would like to foremostly thank my supervisor, **Prof. Philippe G. Schyns** who led me to walk into this intriguing field, raised ideas of great novelty and solid feasibility, inspired me with a broad perspective and diversified knowledge background, and supervised me to form interdisciplinary thoughts, also especially, organized the lab and the projects orderly during the time with the unprecedented challenges.

Secondly, I owe many thanks to **Dr. Jiayu Zhan** who taught me technical measurements and analysis methods from the beginning of my study and, throughout these four years, shared all her suggestions and experiences, always providing generous encouragement whenever I needed.

Thirdly, I would like to express my grateful thanks to **Dr. Robin A. A. Ince** for helping me in the algorithmic and statistical field with his trustworthy expertise, to **Dr. Oliver G. B. Garrod** for giving me practical advice on stimulus design and construction patiently, to **Dr. Kasia Jaworska** who trained me with the experimental skills, to **Prof. Huabing Yin and China Scholarship Council** for providing me with this opportunity to study my Ph.D here.

Fourthly, I would like to extend my heartfelt thanks to **Dr. Bo Wang**, **Mr Yikai Zhang, Miss Yue Yu and Mr Yi Li** for their company during all these years in Glasgow, protecting me from being alone, and leaving me the most cherished and warm memories in my lifetime.

Lastly, thank my family, for caring for me from thousands of miles away, for giving me unconditional love and support in all my pursuits.

Table of Contents

List of Tables

List of Figures

Author's Declaration

I declare that this thesis has been composed by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where states otherwise by reference, the work presented is my own.

1 General Introduction

Visual categorization is one of the most regular but essential tasks people need to complete in daily human life. For example, we categorize weather by observing the sky to decide whether we take an umbrella or not, categorize a friend from crowds to decide whether we greet or not, and categorize the traffic light to decide whether we go or stop. In categorizations, sensory information inputs, after being processed by the brain, drive subjective perceptions, leading to decision making and guiding our behavior. As a bridge linking current status to futural actions, this complex procedure executed by the brain, including how we process, perceive and decide, has become one of the focuses in neuroscience.

Since Helmholtz's "unconscious inferences," vision scientists have worked with the hypothesis that what we perceive is in part influenced by "what we directly see" but also by "what we expect the input might be" (De Lange, Heilbron, & Kok, 2018; Kinchla & Wolfe, 1979). Using similar examples mentioned above, imagine we categorize weather with both forecast and sky, categorize a friend with our memorial impression and the pedestrian's appearance, and categorize the traffic light with our intimate time feeling and the signal change. Plentiful studies over decades have proved that predictions can reduce the disambiguation from noisy information input (Gilbert & Sigman, 2007; P. Kok, J. F. Jehee, & F. P. De Lange, 2012a), facilitating recognition processing such as visual selection (Theeuwes, 2010), visual categorization (Clark, 2013; K. Friston, 2010; Yuille & Kersten, 2006), and speeding up behavior (Bar et al., 2006).

Our brain, generally proposed as multi-layered architecture (Bullmore & Sporns, 2009; K. J. Friston, Daunizeau, & Kiebel, 2009; Grill-Spector & Weiner, 2014; Güçlü & van Gerven, 2015; Kravitz, Saleem, Baker, Ungerleider, & Mishkin, 2013; Mumford, 1992; Van Essen, Anderson, & Felleman, 1992), receives the multiple-modal input from the sensory system (visual, auditory, etc.), transfers the sensory input in feedforward and recognizes the cued information, then sends the predicted contents as feedback. With such a process, a prediction generates and propagates to participate in the upcoming behavior. In recent years, advances in neuroimaging revealed this neural mechanism of prediction to a

large extent. However, the specific pathway and brain regions that propagate such predictions to facilitate behavior remain unclear.

By manipulating the predicted contents, this thesis focuses on tracing the dynamic prediction propagation and understanding its facilitation effect on the categorization of the predicted features. This chapter will start with ***section 1.1***, introducing two fundamental conceptions, bottom-up carrying out sensory input and top-down transferring predictions. ***Section 1.2*** will draw into the variety of prediction modulation, specifically, how different predictions influence visual categorization behavior. ***Section 1.3*** will discuss some theory background, method framework, and analysis approaches in uncovering the neural representation of specific information. This chapter will end with ***section 1.4*** as thesis foci.

1.1 General Overview of Bottom-up and Top-down

Humans process quantities of visual information every second to detect and perceive the world with the eye-opening. In the brain, bottom-up and top-down, two distinct but interactive mechanisms, cooperate to process the vast input selectively and efficiently – bottom-up processes raw sensory input along a low-to-high visual hierarchical pathway, and the top-down predictions generated at the higher hierarchy can propagate down, modulate the activity at lower hierarchical regions, and influence what to process and how to process.

In this section, I will firstly introduce the conception of bottom-up and top-down in ***1.1.1***, then illustrate prediction in a computational view in ***1.1.2*** and the theory of predictive coding in ***1.1.3***.

1.1.1 Bottom-up and Top-down conception

From the very beginning of neuroscience, researchers have been studying the processing and recognition of presented visual stimuli (Linde-Domingo, Treder, Kerrén, & Wimber, 2019). It is well-established that those cognitive procedures require the information contents to be propagated with a flow along visual hierarchical pathways (Contini, Wardle, & Carlson, 2017; Essen, Anderson, & Felleman, 1992). Specifically, the visual information received by the retina with light goes from the early visual areas to intermediate-level and then to higher-level functional areas. Early visual areas (V1)

process the multiple basic features such as shape, orientation, edge, and spatial frequency (Altmann, Bülthoff, & Kourtzi, 2003; Grill-Spector & Malach, 2004). Then, the early visual cortex progressively transfers those visual representations up to intermediate levels (such as V2 and V4) along the ventral and dorsal pathway, where those low-level outputs are merged or integrated to form more complex representations to implement "vision for categorization" and "vision for action" near temporal, parietal or even frontal lobes at higher levels (R. A. A. Ince et al., 2016; Kveraga, Ghuman, & Bar, 2007; Schendan & Stern, 2007; Zhan, Ince, Van Rijsbergen, & Schyns, 2019). From lower propagating to higher visual hierarchy, this information flow suggests a "bottom-up" processing mode (Kinchla & Wolfe, 1979).

Although visual input is one of the most general ways to perceive the surroundings, bottom-up mechanisms alone are not enough to complete the various recognition tasks with innumerable and complicated information from the rapidly changing environment. Many studies indicate that our visual perception is not only influenced by the bottom-up sensory input but also by predictions of what this input might be (De Lange et al., 2018; Kinchla & Wolfe, 1979). Predictions, generated from existing knowledge or received cues, can lead to attention or expectations focusing on the presented or upcoming visual input, impinging on visual selection, processing, or recognition (Bar et al., 2006; Charles D. Gilbert & Wu Li, 2013; Theeuwes, 2010). The propagation of prediction is described as a "top-down" sequential processing, as the representation of prediction from regions with a higher hierarchical cognitive function will subsequently facilitate early visual processing by tuning the neural activity of the primary visual cortex (Gómez-Cuerva & Raymond, 2011; Kok, Jehee, et al., 2012a; Puri, Wojciulik, & Ranganath, 2009; Stokes, Thompson, Nobre, & Duncan, 2009).

Thus, the activity of the visual cortex, as the basis of how we perceive the world, is modulated by two factors in the brain, feedforward bottom-up processing coming from the retina and feedback top-down predictions from high hierarchical regions. Understanding the mechanisms of their independency and interaction is critical in building up vision science. In the following sections, I will introduce how top-down and bottom-up guide decision-making in a computational view and how they interact to achieve the guidance.

1.1.2 Prediction in a computational view

In a real-world environment, both bottom-up flow (i.e., what the present is) and top-down predictions (i.e., what the expected is) are considered to improve the efficiency of information processing and optimize the decision. We can describe this procedure using a computational model.

Consider a mentioned example where we categorize the weather with the present sky and the forecast to decide whether to take an umbrella or not. In a computational view, the potential choices are in dichotomy, including a positive output deciding to take the umbrella and a negative one deciding not. Two factors can influence the final choice: prediction, for example, the forecast saying raining (R) or not ($\neg R$), and the present input stimulus, for example, the color of clouds showing raining (x), using a function of the log of the likelihood ratio (LLR) (Summerfield & de Lange, 2014b):

$$LLR = log\frac{p(R)}{p(\neg R)} + log\frac{p(x|R)}{p(x|\neg R)} \quad (1)$$

where p is the probability of occurring. In this formula, the likelihood of $p(R)$ vs. $p(\neg R)$ means the relative probability of prediction being positive vs. negative; that is, if the first addend, $log\frac{p(R)}{p(\neg R)} > 0$, the positive prediction will be taken, if $log\frac{p(R)}{p(\neg R)} < 0$, the negative prediction will be taken. Similarly, considering the evidence provided by the current situation, $log\frac{p(x|R)}{p(x|\neg R)}$ computes the occurrence probability of x under positive vs. negative conditions. Thus, the outcome will point to a positive choice if $LLR > 0$, and a negative one if $LLR < 0$.

However, sometimes the observers will get multiple information from the present stimulus related to the upcoming decision. In this example, the evidence can be obtained from sky color, cloud shape/color, wind strength, humidity, etc., which highly enrich the content of x. Hence, we can expand the LLR measurement as:

$$LLR = log\frac{p(R)}{p(\neg R)} + log\frac{p(x_1|R)}{p(x_1|\neg R)} + log\frac{p(x_2|R)}{p(x_2|\neg R)} + \cdots + log\frac{p(x_n|R)}{p(x_n|\neg R)} \quad (2)$$

The computation above is named the "sequential sampling model," synthesizing all possibilities of the evidence given the prediction of the upcoming events. With this model, computationally, we see the top-down predictions bias the output LLR (i.e., the later perception or decision) in two ways: (1) By modulating the *LLR* baseline, that is, if the $log\frac{p(R)}{p(\neg R)}$ is positive, observers then need more robust evidence from present circumstances to make a negative decision. Vice versa, if the expectation is negative, it will be harder to make a positive choice; (2) By modulating the reliability of the bottom-up evidence given different predictions, in other words, the decision making depends on the "relative reliability" between bottom-up input and top-down modulation (Mumford, 1992),

1.1.3 Predictive coding

As mentioned, the *LLR* model can explain how humans make a decision with a combination of sensory evidence and prior predictions, as well as the perceptual inference incurred by the predictions. However, with continuous and multiple-modal sensory inputs in the real world, this measure of likelihood will not be accomplished at only one or several times of computations. With the aspect of neural mechanism, predictive coding theory recognizes this processing as a dynamic procedure requiring repeated *LLR* computations.

Specifically, when bottom-up from lower hierarchy and top-down from higher hierarchy flows propagate and converge, neurons in the brain iteratively respond to the two sources of information at the hierarchical processing stages, computing predictive error which is the inconsistency between the observed/heard and the predicted. The neural activity showing high predictive errors in turn feed forwardly to a higher functional hierarchy to refine the estimations to upcoming events, keeping active until the sensory system achieves the highest *LLR*, while the neurons with low predictive errors tend to "explain away" their own neural activity and related information to spread (Mumford, 1992; Murray, Kersten, Olshausen, Schrater, & Woods, 2002; Rao & Ballard, 1999), which is regarded as a strategy to efficiently reduce coding burden (Jehee, Rothkopf, Beck, & Ballard, 2006). In this way, the brain continuously processes the two sources of information, upgrades the recognition, optimizes the explanation of

the sensory input, and finally accomplishes the consistency between what we predict and what we are presented to output a decision.

1.2 Top-down Controls

As discussed, top-down predictions can integrate with bottom-up information, which in turn, modulate the perception or decision. In visual categorization, the contents of predictions involve comprehensive aspects to achieve diverse tasks requiring different recognition processes. For example, we can predict different colors, shapes, objects, contexts, and perceptions according to ongoing or upcoming perceptual demands. Here, in ***1.2.1*** and ***1.2.2***, I will introduce two of the most common forms of top-down control, feature-based attention and feature-based expectation, including their tuning effects and related neural mechanisms, respectively. Then, I will discuss some unsolved questions in the top-down field in ***1.2.3***.

1.2.1 Feature-based attention

Visual attention, generally understood as "focus on something," can determine the targets (e.g., object) or dimensions (e.g., location, color, size, …) for prioritized processing, thus driving a quicker detection of stimulus presentation or change in the attended contents. In the brain, different regions carry out different attention functions. I will first introduce spatial attention, one of the most typical top-down controlled attention, and then some other feature-based attention.

1.2.1.1 Spatial attention

One of the most common top-down modulations is associated with spatial attention (Charles D. Gilbert & Wu Li, 2013), which can allow people to explicitly target specific parts of the visual field, prioritizing the discovery of stimuli there (Posner, Snyder, & Davidson, 1980) and speeding up the processing of stimuli (Summerfield & Egner, 2016). This attentional focus could be modified by either previous knowledge or experience endogenously or given instructions or cues exogenously.

There is a long history that brain imaging researchers cued observers to attend a specific location, aiming to uncover the neural mechanisms of spatial attention controls. In an fMRI study (Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000), an arrow cue

can first activate the occipital regions bilaterally but quickly fade away, dorsally followed by more extended responses from the high-hierarchical regions, including inferior parietal lobe and frontal eye fields, which lasted sustainedly even after the arrow disappeared. Additionally, unlike the bilateral occipital activation, this parietal and frontal activation is contra-lateral towards the attended location. Thus, this sustained and contra-lateral activation suggests an involvement of the high-level regions in the modulation of spatial attention, which has been consistently reported in studies (Corbetta, Kincade, & Shulman, 2002; Hopfinger, Buonocore, & Mangun, 2000).

Despite the short activation when predicting in the occipital cortex, the effect of modulation will be much more profound after the stimulus is physically shown. Quantities of research have proved that attention targeting can enhance the neural responses to the expected visual field as well as suppress the responses outside the focus of attention (Chen et al., 2008; Motter, 1993). This enhancement appears in extensive brain areas, including the primary occipital cortex (V1, V2, V4), medial temporal gyrus and inferior temporal gyrus (Chelazzi, Miller, Duncan, & Desimone, 1993; Li, Piëch, & Gilbert, 2006; Luck, Chelazzi, Hillyard, & Desimone, 1997; Moran & Desimone, 1985; Motter, 1993; Spitzer, Desimone, & Moran, 1988), premotor cortex (Jueptner et al., 1997), parietal lobes and prefrontal cortex (J. Rowe, Friston, Frackowiak, & Passingham, 2002).

1.2.1.2 Other feature-based attention

Besides attending to location, people can also focus on other features. For example, when making visual selection or categorization on shape, color, spatial frequency, or moving direction of the stimulus, or doing visual detection task on object presentation or change, observers can preferentially recognize the goal-directed information and ignore the other aspects of the sensory input.

Similar to the spatial attention, with a visual arrow cueing the moving direction of a set of dots, this specific attention can also activate occipital (middle temporal visual area), parietal (anterior intraparietal sulcus, posterior intraparietal sulcus), and frontal (superior frontal sulcus, precentral sulcus) regions. In contrast, the occipital effect is transient, and the higher-hierarchical effect stays more sustained (Shulman et al., 1999). This contrast suggests that the parietal and frontal regions also get involved in the extensive feature-based attention. Furthermore, different subregions are responsible for different types of

attended features. For example, in the post-cue and pre-stimulus period, cueing motion direction can more strongly activate the intraparietal sulcus and superior parietal lobule compared with cueing color (Shulman, d'Avossa, Tansy, & Corbetta, 2002). Interestingly, the comparison after the stimulus shown indicates the same specificity in the mentioned regions, reflecting the specialization of the activation of predicted contents is consistent with the effect of the physical visual stimulus.

.

1.2.2 Feature-based Expectations

Another common form of top-down control is feature-based expectation. With empirical data, the brain can internally predict the upcoming stimulus features. For instance, people can expect to see the full moon a few days later when looking at a waning moon or predict the following sentence of the lyric when hearing a familiar song. In these cases, the expectations are generated from the same category of sensory input (i.e., vision predicts vision and audio predicts audio). It realizes with a pathway where the lower visual (or auditory) cortex passively receives the related sensory input and propagates to the area at a higher cortical hierarchy, such as the lateral occipital cortex (Kok & de Lange, 2014; Murray et al., 2002) or temporal area (Ekman, Kok, & de Lange, 2017; Muckli, Kohler, Kriegeskorte, & Singer, 2005). Signals from these regions can modulate back the activity of the lower source cortices with the top-down flow containing specific expectations. Moreover, we can also infer the incoming events with cross-modal expectations, such as hearing the horn blaring and expecting a car coming toward us. Though these cross-modal expectations also require the brain to extract critical information as feedforward from and send expectations as feedback, their mechanisms are more complex, as multiple cortical regions, including different origins and destinations, are involved in sensory processing in different modes.

In recent years, studies on understanding neural mechanisms of expectations have kept growing but is still under development. I will introduce the existing theories or studies with the order of expectation propagation (in ***1.2.2.1***), modulation (in ***1.2.2.2***), and perceptual tuning (in ***1.2.2.3***).

1.2.2.1 Neural mechanisms of expectations propagation

As two reversed but integrated procedures, the neural mechanisms between top-down and bottom-up are broadly comparative in studies. It is common to uncover predicted contents by decoding prediction using the representation of its bottom-up feature. For example, using the classifiers trained by the correspondence bottom-up representation to classify top-down signals, fMRI studies revealed that expectations in visual completion trigger a stimulus template in the hippocampal and primary visual cortex (Hindy, Ng, & Turk-Browne, 2016; Kok, Failing, & de Lange, 2014). This finding shows the specific activation in sensory-related regions. In line with this effect, a time-frequency domain reports that the decrease of alpha rhythm power can carry out the modal replays, especially in related processing regions – the decrease during visual replays is sharper in the visual cortex than in the auditory cortex and vice versa (Michelmann, Bowman, & Hanslmayr, 2016).

On the other hand, some studies use MEG to determine the exact time the prediction generates in tracing the temporal processing. In a memory cueing study, replaying the neural representation of a previous face/scene stimulus occurs at ~500ms after the cue (Jafarpour, Fuentemilla, Horner, Penny, & Duzel, 2014). Considering the pre-stimulus stage, a study with the cue-stimulus design successfully decoded the stimulus feature template trained by the Gabor orientation localizer from the prediction in a 40 ms time window before stimulus onset (Kok, Mostert, & De Lange, 2017). Besides the visual template, an auditory study also reported that compared with jitters, the regular rhythms followed by auditory expectations induce higher decoding performance of the pure tone between 100 ms and 80 ms pre-stimulus (Auksztulewicz, Myers, Schnupp, & Nobre, 2019). These studies suggest that a relative bottom-up classifier can detect predicted contents during either post-cue or post-stimulus stages, thus separating the representation of predicted contents induced by the predictive cues from the overall pattern of neural representation and revealing its mechanisms.

Though the sensory septicity of expectations has been well, the propagation pathway to achieve this activation is worth considering. It is conceptually and generally assumed that the high-to-low top-down expectations reverse the low-to-high bottom-up sensory processing. Linde-Domingo's study constructed a stimulus-recall experiment and reported that when seeing an object, observers can recognize the perceptual features more

quickly than conceptual features, while in memory retrieval, recalling the conceptual description of objects is faster than the perceptual. This conclusion is supported by both behavior and ERP results. However, although it provided evidence for the reversed hierarchical reconstruction hypothesis, it is yet to disclose the dynamical spatial-temporal pathway of this high-to-low propagation, in other words, to answer the question about when, where, and how the expectations rapidly flow down into the destination regions in a functional brain network.

1.2.2.2 Neural mechanisms of expectations modulation

The expectation propagation flows from higher hierarchical regions can modulate the later activity of corresponding functional areas at the lower hierarchy. One of the most ubiquitous phenomena is termed "end stopping," or "expectation suppression," that is, cortices carrying out the sensory processing (e.g., early visual cortex or primary auditory cortex) respond more weakly to predictable stimuli compared with unpredictable ones (Summerfield & de Lange, 2014a; Todorovic & de Lange, 2012). Varieties of studies have reported this effect using non-invasive measurements, including magnetoencephalography (MEG) (Garrido, Rowe, Halász, & Mattingley, 2017), electroencephalography (EEG) (Todorovic & de Lange, 2012; Wacongne et al., 2011) and fMRI (Alink, Schwiedrzik, Kohler, Singer, & Muckli, 2010; Egner, Monti, & Summerfield, 2010; Kok, Jehee, et al., 2012a). However, tracking back to its reason raises different voices. There are debates on two likely accounts of the suppression effect, dampening and sharpening (De Lange et al., 2018). Specifically, dampening theory illustrates that the redundant expectations are filtered out, thus weakening the neural response of the predicted contents to achieve more efficient coding (De Lange et al., 2018; Todorovic & de Lange, 2012). On the other hand, sharpening theory thinks the expectations tune the neural response by "explaining away" the predictive errors (K. Friston, 2005), suppressing the inconsistent or unpredicted stimuli representation (Lee & Mumford, 2003a) and enhancing the representation of predicted contents (Kok, Jehee, et al., 2012a).

Over the past years, both explanations have been supported by diverse paradigm designs and neural methodologies. For example, some studies found that macaque's inferior temporal (IT) cortex presents damaged decoding performance for images with fixed presentation sequence compared with a random display (Kaposvari, Kumar, & Vogels, 2018; Kumar, Kaposvari, & Vogels, 2017). Parallelly, Blank's study suggests the

same decreased activation in the posterior Superior Temporal Sulcus with common words input compared with unexpected words (Blank & Davis, 2016). On the contrary, Kok's study reported improved decoding accuracy in the early visual cortex (Kok, Jehee, et al., 2012a) with an informative cueing on upcoming grating stimuli compared with the non-informative, contributing to the demonstration of the theory of sharpening tuning. This debating topic is still open and active at present, without a definitive conclusion on when, where, and how the top-down expectations modulate sensory processing in the primary cortex.

1.2.2.3 Perceptual tuning by expectations

Regardless of the detailed explanation of how feature-based expectations propagate and modulate the neural activity, their aims or passive consequences are tuning the perception and guiding the behavior in various respects. According to different goals, reasonable expectations can facilitate (i.e., speed up or improve performance) the stimuli detection and selection in a visual search task (Chang, Kanai, & Seth, 2015; Stein & Peelen, 2015; Wyart, Nobre, & Summerfield, 2012) or bias the perception (e.g., motion direction) in a discrimination task (Kok, Brouwer, van Gerven, & de Lange, 2013; Sotiropoulos, Seitz, & Seriès, 2011; Sterzer, Frith, & Petrovic, 2008). In addition, by introducing explicit interpretation, expectations can even help observers reconstruct an equivocal image. For example, if one scans Figure 1-1 directly, it is so unsubstantial that the viewer will feel impossible to know the exact meaning. However, with Figure 1-2, a depicted version as an elucidation, the observer can feel this ambiguous image more apparent when returning to the original image.

Figure 1-1 Priming example – Blurred version with only segmented black and white background.

The efficiency of tuning is influenced by both the validity of expectation and the ambiguity of the stimulus. To be exact, when the expectation is solid and the stimulus is noisy, observers rely more on the expectation to reduce disambiguation; in contrast, when the stimulus is clear and stable, while the expectation is not inaccurate, observers prefer to perceive by what they physically see or hear. This trade-off computationally optimizes the recognition process for the highest reliability (Mumford, 1992) and reflects an interaction between bottom-up and top-down.

Figure 1-2 Priming example – Depicted version to prime viewers to reconstruct and understand Figure 1-1.

1.2.3 Problems to solve

Firstly, as discussed, although there are quantities of studies contributing to the understanding of top-down mechanisms (e.g., the downward reactivation or the induction of stimulus-specific template), few of them revealed the dynamics of prediction propagation. That is, to ask when and where the prediction is represented in the brain or reconstruct a dynamic functional pathway/network of prediction communications.

Secondly, the "sharpen vs. dampen" debate continues to discuss the actual reason leading to the end-stopping effect. Most of the studies provided evidence by computing the comparison between a predicted condition and an unpredicted condition, sometimes further reporting a damaged or enhanced decoding performance at regions of interest. However, due to the difficulty of quantifying, little evidence has been reported considering "prediction strength," in other words, whether the solid or weak prediction can influence the later processing of the predicted stimulus.

With the lack of answers to the aforementioned questions, in ***1.3***, I will introduce some background (in ***1.3.1***), structures (in ***1.3.2*** and ***1.3.3***), and measurements (in ***1.3.4*** and ***1.3.5***) of a novel algorithmic method which can trace the information flow of the specific features in the brain, and realize the single-rial decomposition, thus measuring the prediction strength.

1.3 Methods to uncover Neural Coding Mechanism

1.3.1 Feature, Brain and Behavior

As previously discussed, the human brain can process and propagate the input information feed-forwardly, as well as realize top-down modulations reversely in a feedback hierarchy order. This procedure, in a computer vision, can be considered that the information processor (i.e., brain) with multiple hierarchical computations iteratively and efficiently organizes the two functions, and each layer of computation tunes the output according to the input from lower and higher input, finally generating the behavioral commands. Many hierarchical models apply this concept. Predictive coding assumes that the brain repeatedly tunes the neural response to achieve the highest consistency between predictions and perceptions (K. Friston & Kiebel, 2009). The diffusion model deems the

evidence from the sensory input is accumulated continuously until a decision or discrimination can be produced (Bolam, Boyle, Ince, & Delis, 2022). With these models reflecting different input-processor-output outlines, an essential and common question is put forward spontaneously – what are the dynamics of information processing? In other words, **what** information from sensory input is processed by the brain, and **where** and **when** is the information processed to engender later behavior?

To address these questions and uncover the related neural mechanism, three vital factors should be taken into consideration:

(1) Feature, the sensory input from the environment. It could be a visual, auditory, tactile, or olfactory stimulus or anything the researchers are interested in and able to quantify. In investigating top-down/bottom-up processes, the feature could be the cue inducing predictions or the physical predicted contents (i.e., the presented stimulus).

(2) Brain activity, the recording (e.g., EEG, MEG, or fMRI) of brain responses concurrently acquired during feature exposure. Specifically, suppose we collect the neural response in a high temporal resolution (e.g., MEG or EEG). In that case, we can then trace the dynamics of feature representation by computing the relationship between feature and brain activity in the time series of each channel/electrode/source (with source reconstruction).

(3) Behavior response, the outcome of the information input and processing. It could be any behavioral measurements such as categorization decision, reaction time, and accuracy (true or false).

Based on information theory, Schyns provided an algorithmic methodology called Stimulus Information Representation (Schyns, Zhan, Jack, & Ince, 2020) which decomposes all pairwise and triplewise interactions of the data-triple <Features; Brain activity; Behavior>. In the following two sections, I will introduce the computation built on information theory and the diagnostic recognition framework.

1.3.2 Understand Feature, Brain and Behavior using Information theory

To understand what feature is processed in the brain and how the feature is processed to generate behavior, we need to compute the relationship between different pairwise factors. From a statistical view, different methods will quantify the dependence quantification between two variables based on different models. For example, correlation and linear regression assume that the output variable can be calculated from a linear combination of the input(s) (Jaworska, Yan, Van Rijsbergen, Ince, & Schyns, 2022; JIayu Zhan; G. B. Garrod, 2018), or mutual information has no assumption on the relationship (R. A. Ince et al., 2017; R. A. A. Ince et al., 2015; Zhan et al., 2019). Here I will focus on the computation and application of mutual information, the non-parametric measure of the statistical dependence between two random variables (Shannon, 1948).

With setting a null hypothesis that the two sets of variables are independent of each other in advance, mutual information (MI) computation quantifies the statistical power of this null hypothesis, consequently measuring the dependence of the two variables. MI is calculated by the entropy differences between the two variables and can be expressed as following:

$$\begin{aligned} \mathrm{I}(X;Y) &= \mathrm{H}(X) - \mathrm{H}(X \mid Y) \\ &= \mathrm{H}(Y) - \mathrm{H}(Y \mid X) \end{aligned}$$

where $\mathrm{H}(X)$ and $\mathrm{H}(Y)$ are the marginal entropies and $\mathrm{H}(X \mid Y)$ and $\mathrm{H}(Y \mid X)$ are conditional entropies. In neuroscience, supposing X is the brain activity and Y is the target factor (i.e., the stimulus or behavior response), we can understand different equations from different perspectives. The first equation models the reduction of uncertainty of the brain activity itself with the shown stimulus/given response. Vice versa, the second equation is used to model the reduction of uncertainty of the stimulus/response itself given the specific presented neural response (R. A. Ince et al., 2017). As MI computation presupposes no actual model before the calculation, MI avoids the noise brought by false assumptions and widely examines any possible patterns of the connection between the two variables.

In recent years, more and more studies are keen on using information theory (i.e., computing MI) to construct the relationship between neural activity and target factors

(Agarwal, Agrawal, Jain, & Kumar, 2010; Fairhall, Shea-Brown, & Barreiro, 2012; Nelken & Chechik, 2007; Rolls & Treves, 2011; Victor, 2006). Different target factors leads to the answers to different neural coding mechanisms (Borst & Theunissen, 1999). Considering the three vital factors mentioned in 1.3.1 as examples, we can detect two kinds of **neural coding**:

(1) Stimulus-driven. Mutual information can measure the single-trial dependence between stimulus (i.e., feature) type and brain activity, thus delivering direct representational interpretations of brain activity, in other words, **how much** the input stimulus can explain the brain activity and **where** and **when** it is represented in the brain.

(2) Task-directed. Different tasks require different cognitive processes to achieve the goals. In order to understand the neural mechanism of behavior production, we can measure the dependency between behavior measurement (e.g., response choice or response time) and neural activity, and in turn, answer that **when** and **where** the brain organizes and executes the necessary functions leading to the final behavior.

In this thesis, I will use information theory, without any assumptions, on the single-trial relationship between target factors and brain imaging data to detect the unknown neural mechanism of top-down predictions, specifically, to explore that in the brain, how (what, when and where) the predicted contents are (1) induced by an input cue, (2) propagated through the brain and (3) modulate the task-directed neural response and behavioral categorization.

1.3.3 Diagnostic Recognition Framework: Stimulus Information Representation (SIR)

As discussed above, information theory can compute the representation of stimulus-driven/task-directed brain activity. Consider further that the Feature factor is not a united entirety but randomly sampled from the entirety, retaining only partial information. In this case, exposed to different residual inputs on each trial, observers are also likely to form diverse perceptions, judgements, or choices to accomplish the task. Based on this operation, Schyns put forward a framework to identify the diagnostic feature in specific tasks and trace the Stimulus Information Representation (SIR) (Schyns et al., 2020).

In SIR, the diagnostic feature is the residual information that can drive a response or a specific response. Thus, it could be computed by the pairwise MI between the single-trial sampled feature and behavior. Jiayu's study (Zhan et al., 2019) used an ambiguous painting where observers can see both Voltaire's face and the Nuns overapply (*Slave Market with Disappearing Bust of Voltaire*). By showing participants the random sampling of the painting and instructing them to make perception decisions, the researcher found that the visibility of different pixels could contribute to the dichotomous choices (i.e., Voltaire or Nuns), including:

(1) Diagnostic features for Voltaire. When these pixels are visible, observers will likely generate the perception of Voltaire.

(2) Diagnostic features for Nuns. When these pixels are visible, observers will likely generate the perception of Nuns.

(3) Task-irrelevant features. These background pixels will not drive a clear perceptual decision, so keeping them visible or not will not influence the observers' judgements.

Applying information theory similarly to 1.3.2, after input, these decomposed clusters of visible information could activate the neural responses of different regions, detected by the pairwise MI between neural responses and specific feature visibility. In this way, we can uncover the different neural mechanisms of the information processing driving different behavioral perceptions.

Additionally, besides the mentioned pairwise modelling with higher sampling precision, SIR framework also contains the quantification of triple-wise dependency to answer more questions. The triple-wise intersection, expressed as Co-Information <Features; Brain activity; Behavior>, can quantify the co-variation across the three elements, that is, trace where and when the feature variations similarly influence both brain and behavioral responses (mostly, modulate the brain activity in turn influence the behavior responses).

1.3.4 Extensions of Information-theoretic methods

Besides mutual information and the related SIR framework evaluating the dual or triple-wise relationship between variances, I will discuss two more detailed quantifications in tracing specific information propagation (in ***1.3.4.1***) and decomposing the contribution of each individual sample (in ***1.3.4.2***), which are further applied in this thesis.

1.3.4.1 Directed Feature Information (DFI): Tracing specific information propagation

DFI is a novel conception first established in Robin's study in 2015 (R. A. A. Ince et al., 2015). Before understanding the computation of DFI, first, we consider the shortcoming of Directed Information (Massey, 1990), also named Transfer Entropy (TE), a traditional quantification of directed transfer information by computing a conditional MI (CMI):

$$DI = CMI<RA_{x;}\, SA_{x-y}|RA_{x-y}>$$

where RA_x is the activity of receiving processor at time x, SA_{x-y} is the activity of sending processor at time x-y (i.e., with a transfer delay of y), RA_{x-y} is the activity of receiving processor at time x-y as the conditioned-out factor to exclude the history influence from receiving processor itself. Though this is a developed measurement of the communication between two processors, we cannot specify what induces the communication or what the communicated content is.

DFI solves this shortcoming by isolating the feature-related (F) component in DI. Specifically, it is computed by:

$$DFI = DI - DI|F$$

where DI is the entire communication mentioned above from sending processor to receiving processor, DI|F is the DI conditioned on the feature we are interested. Thus, by subtracting DI|F from DI, we quantified the directed communication about the feature information.

Robin's study reconstructed an occipital-temporal-parietal network propagating two kinds of visual features (Voltaire feature and nuns feature mentioned above in (Zhan et al., 2019)), separately leading to different perceptions (R. A. A. Ince et al., 2015). Though, as a newly constructed conception, DFI has not been applied much in neuroimaging, it can indeed quantify the causal transfer of specific contents (e.g., stimulus feature, behavior response, etc.), and in the meantime, consider the timing of the sending and receiving processes.

1.3.4.2 Sample-wise Mutual Information (SMI): Decomposing Single-trial Contribution

The computation of Mutual Information, as discussed above, provides a single number reflecting the overall effect size of the non-parametric dependence between two variances (i.e., without any model hypothesis). Based on MI, sample-wise MI (SMI) computation decomposes the local contribution to the global effect size, generating a single number for each observed sample (i.e., measurement per trial in neuroscience). Positive SMI means that the trial follows the overall dependence, and vice versa, the negative value suggests the failure to follow. Thus, SMI provides a base for analyses considering trial-by-trial variance in neuroscience, which I will use in the following studies to quantify the prediction strength and stimulus discrimination strength and further link the prediction stage and the categorization stage in every single trial.

1.4 Thesis Foci

As discussed above, top-down predictions play an essential role in perception and visual categorization. However, the dynamic neural mechanism of top-down propagation and facilitation remains unclear. In this thesis, focusing on the top-down process, with the design and analysis based on information theory, I uncovered the neural mechanism of the top-down process in a series of three studies: 1) *predictions induce behavioral improvement* (Chapter 2), where I validated the prediction facilitation effect to behavior and initiated the design for later neuroimaging study; 2) *frontal cortex modulates occipital predictions of visual contents in a functional network* (Chapter 3), where I traced the when and where the top-down predictions represent in the brain and how hierarchical regions communicate in this dynamic; 3) *predictions enhance stimulus discrimination and*

modulate behavior in a categorization network (Chapter 4), where I explored the prediction facilitation effect on early visual processing and later decision making.

2 Study 1: Predictions induce behavioural improvement

2.1 Introduction

Human brain can perceive, summarize and learn statistical regularities between an instant input (e.g., a cue or a symbol) and its following event (Seriès & Seitz, 2013). With the learnt correspondence, predictions can be generated from presented cues and prior knowledge to facilitate the subsequent visual processing. Such predictions can shape perception when the stimulus is ambiguous, or improve behavior, including accuracy or reaction time (i.e., the stimulus detection being more precisely or rapidly) (Pinto, van Gaal, de Lange, Lamme, & Seth, 2015; Stein & Peelen, 2015; Weiss, Simoncelli, & Adelson, 2002; Wyart et al., 2012).

Chapter 2, as a preamble of the series of research on top-down predictions, tested this facilitation effect of the prediction of the upcoming stimulus at behavioral level. As discussed, extensive studies have proved that prediction could speed up reaction time in visual selection and categorization. This behavioral study validated this facilitation effect on the specific stimuli set and experiment paradigm, provided the preparation and confirmed the feasibility of the following neuroimaging research.

Specifically, ***experiment 1*** aimed to prepare proper a set of auditory stimuli for later prediction tasks. By measuring D-prime, I validated that all 8 participants could explicitly distinguish the selected auditory pitches by testing the auditory sensitivity on the auditory pitches and their surrounding tones. ***Experiment 2 and experiment 3*** aimed to validate the facilitation effect of prediction on categorization behavior, especially on reaction time. Here I trained 9 participants to couple the prediction relationships between auditory pitches and utilized the coupled predictions in the cueing-categorization task. By analysing the reaction time, I demonstrated the improved behaviour in visual categorization with informative prediction. Further, I took some potential confounding factors (e.g., learning effect) influencing the effect into consideration. Inspired by this study, I will also discuss the likely optimizations for applying this behavior paradigm in the neuroimaging study.

2.2 Experiment 1: Detecting Auditory Sensitivity

2.2.1 Participants

Eight participants participated in this experiment, with mean age = 23.6 years, SD = 3.2 years. All participants reported normal or corrected-to-normal vision and no history or symptoms of neurological or psychiatric illness that affects auditory or visual processing. All participants gave informed consent and received £6 per hour for their participation. I obtained ethical approval from the University of Glasgow College of Science and Engineering Ethics Committee.

2.2.2 Stimuli

The stimuli were 25 pure tones (5 central tones and 20 surrounding tones) with a duration of 200ms and a volume of 75dB. The five central tones were with the frequencies of 587Hz, 880Hz, 1175Hz, 1760Hz and 2218 Hz (with the scientific pitch notation named D5, A5, D6, A6, D7), each surrounded by four surrounding tones (see Table 2-1 for the tone frequency detail). The frequencies of surrounding tones were -80 cents, -40 cents, +40 cents and +80 cents around the central tones. “Cent” is defined as:

$$(F + 1 \text{ cent}) / F = 2^{1/1200},$$

where F is a specific sound frequency.

Table 2-1: Frequencies of central and surrounding tones

Frequencies of central tones (Hz)	Frequencies of surrounding tones (Hz)			
587	560	574	601	615
880	840	860	901	922
1175	1122	1148	1202	1231
1760	1681	1720	1801	1843
2218	2118	2167	2270	2323

2.2.3 Procedure

Each experimental block started with displaying either one tone from the five central tones, instructed to participants as the target tone. On each trial of the block, participants first saw a centrally displayed black fixation for 100ms presentation, 100ms disappearance, and 300ms presentation, which constructed a flicker to help them keep attention. Participants then heard either one tone from five central tones or four surrounding tones around the target with fixation. After the tone displaying, participants pressed the keys to respond to whether the tone heard was the target tone or not, followed by an inter-trial interval (ITI) for 750-1250ms jittered before a new trial began. For each participant, the experiment comprised 450 trials (90 for each central tone as a target), divided into 20 blocks of 27 trials each. Participants sat in a dimly lit room and used a chin rest to maintain the viewing distance at 57cm. I created the interface of the experiment via Psychotoolbox (version: 3.0.14) (Kleiner, Brainard, & Pelli, 2007), setting the background screens as grey with a Gray value of 127.

2.2.4 Analysis and Results

To prepare a distinguishable set of auditory stimuli for later prediction experiments, this task tested the perceptual distance between the chosen auditory pitches by requiring participants to distinguish testing tones from a specific target tone. To test the perceptual distance between the testing tones and the target tone, I computed d-prime (Stanislaw & Todorov, 1999) between 5 central tones and 9 testing tones (4 surrounding tones + the target tone itself + 4 other central tones) with the following 3 steps:

Step 1: For each participant and each target by testing combination, I calculated the Hit rate *H* and False Alarm *F* by:

Hit rate *H* = ratio of "yes" response in trials displaying the target
False Alarm *F* = ratio of "yes" response in trials not displaying the target.

Step 2: For each participant and each target by testing combination, I calculated the d-prime by:

$$D\text{-}prime\ (H, F) = z(H) - z(F)$$

where *z* represents the *z* transforms. Figure 2- 1A showed the 5 (target tones) by 9 (testing tones) d-prime matrix, indicating a clear separation between target-surrounding d-prime and target-other central d-prime at ~1.7 (i.e., Target-Surrounding < 1.7, and Target-Other central > 1.75) which could represent $H = 0.8$ and $F = 0.2$.

Step 3: At group-level, I constructed the significance test by paired-sample t-tests between target-self (target) d-prime and target-not self d-prime, to detect whether the difference in perceptual distance was significant (i.e., whether participants could significantly distinguish the testing tone and the target tone). We applied the Bonferroni correction to the output *p*-value. Figure 2-1 B showed this statistical result, indicating significant distinctions among the five central tones and insignificant distinctions between each central tone and its surrounding tones.

The results showed that, as expected, generally, (1) participants could distinguish each central tone from other central tones; (2) participants could not distinguish each central tone with its surrounding tones very well. The clear distinction among the central tones motivated me to use them as alternative frequencies of the auditory cues in the following experiments.

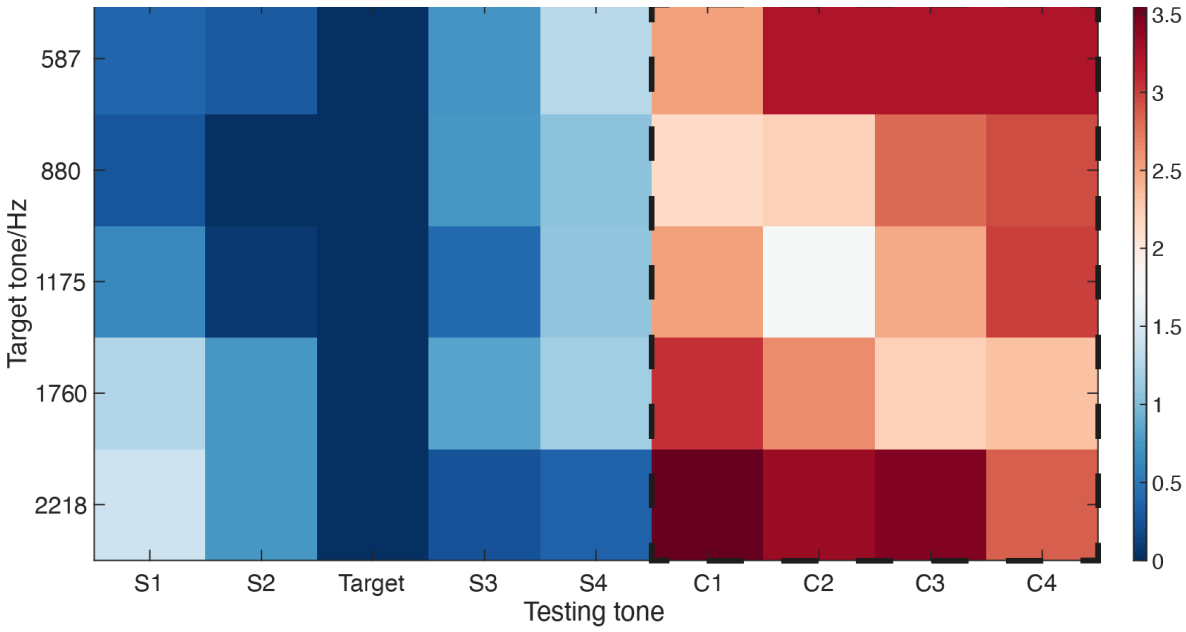


Figure 2-1 Target tone – testing tone d-prime.

The color-coded matrix shows the d-prime between each central tone (Y-axis) and each testing tone (X-axis), including 4 surrounding tones (S1-4), the target tone itself and other

central tones (C1-4). The blue-white-red color scale denotes the increased value from 0-1.75 (in blue, distributed in Targe-Surrounding/Targe d-prime), to 1.75 (the boundary in white), and 1.75-3.5 (in red, distributed). I applied paired-sample t-tests between the Targe-Targa et d-prime and each of 8 Target-Other d-prime. The black frame circles the cells showing significant differences (Bonferroni-corrected). It suggests the significant distinction between each combination of the target tone and the other central tone, and the insignificant distinction between each combination of the target tone and its surrounding tone.

2.3 Experiment 2: Training Couplings between Auditory Cue and Visual Stimuli

2.3.1 Participants

Twelve participants (8 female and 4 male) participated in this experiment, with mean age = 23.6 years, SD = 3.2 years. All participants reported normal or corrected-to-normal vision and no history or symptoms of neurological or psychiatric illness that affects auditory or visual processing. All participants gave informed consent and received £6 per hour for their participation. I obtained ethical approval from the University of Glasgow College of Science and Engineering Ethics Committee.

2.3.2 Stimuli

Auditory cues. I used two pure tones (880 Hz vs. 1760 Hz) as auditory cues with a duration of 200ms.

Visual stimuli. I used two kinds of tilted Gabor patches displayed on a vertical grating background as the visual stimuli, distinguished by the orientation and the location (see Figure 2-2 A Gabors): (1) displayed in 10 degrees anticlockwise (left-tilted) Gabor patch located at the left hemi-visual field; (2) displayed in 10 degrees clockwise (right-tilted) Gabor patch located at the right hemi-visual field. I presented the Gabor patch with a diameter of 3-degree visual angle and eccentricity of 15-degree visual angle at a viewing distance of 57cm. The screen is 23-inch full-HD with a resolution of 1920 × 1080, and a refresh rate of 60 Hz.

Auditory cue – visual stimuli pair. The auditory-visual couplings were counterbalanced across participants. Half of the participants learned the “880 Hz – left

tilted and located Gabor patch" and "1760 Hz – right tilted and located Gabor patch" couplings (see Figure 2-2 A)., and the other half learned the "880 Hz – right tilted and located Gabor patch" and "1760 Hz – left tilted and located Gabor patch" couplings.

2.3.3 Procedure

Before the task, I instructed the couplings between two auditory cues and visual categories. Each experimental block started with learning pages. The participants could familiarize themselves with the couplings by unlimitedly (re)playing the auditory cues and simultaneously seeing the coupled visual stimuli until they felt confident to start the task. Figure 2-2B shows the trial procedure in the training task after the learning page. Each trial started with a fixation with 100ms presentation, 100ms disappearance and 300ms presentation, which constructed a flicker to hint at the upcoming auditory cues. Participants then heard one of the two auditory cues for 200ms, followed by a blank screen for 2000-2500ms jittered. After the cue displaying, participants randomly saw either one visual stimulus from the two kinds mentioned above for 120ms and made a "yes" or "no" judgment with a keypress on whether the visual stimulus coupled with the heard auditory cue. After the response, participants immediately received feedback on the correctness of the judgement ("correct" vs. "incorrect") presented for 250ms and an ITI for 750-1250ms jittered.

Each block contained 20 trials, including 10 trials with matched couplings (i.e., requiring a "YES" response) and 10 trials with mismatched couplings (i.e., requiring a "NO" response). Participants first ran 5 blocks; after these 5 blocks, the training procedure stopped when participants ran 3 successive blocks with 100% accuracy or finished 15 blocks. Participant sat in a dimly lit room and used a chin rest to maintain the viewing distance at 57cm. I created the interface of the experiment via Psychotoolbox (version: 3.0.14) (Kleiner et al., 2007), setting the background screens as grey with a Gray value of 127.

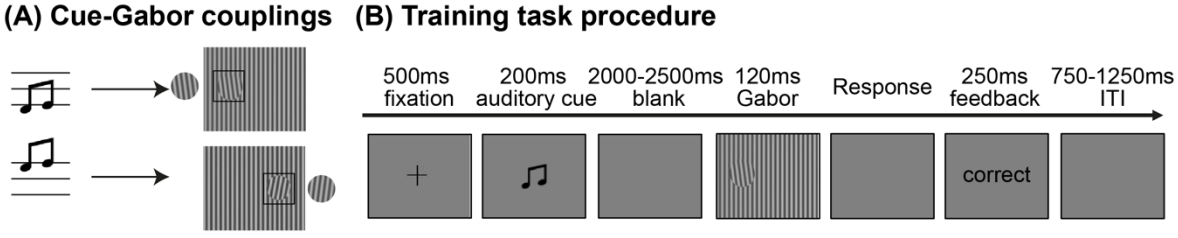


Figure 2-2 Cue-Gabor couplings and task procedure in this training experiment.

(A) Example of cue-Gabor couplings. The low frequency (880Hz) cue coupled with the left-tilted Gabor patch located at the left hemi-visual field, and the high frequency (1760Hz) cue coupled with the right-tilted Gabor patch located at right hemi-visual field. The Gabor patches are highlighted beside the actual stimulus. These couplings were counterbalanced across participants. **(B) Training task procedure.** Each trial started with a fixation with 100ms presentation, 100ms disappearance and 300ms presentation, which constructed a flicker to hint at the upcoming auditory cues. Participants then heard one of the two auditory cues for 200ms, followed by a blank screen for 2000-2500ms jittered. After the cue displaying, participants saw the visual stimulus for 120ms and made a "yes" or "no" judgement on whether the visual stimulus was coupled with the heard auditory cue, followed by a 250 ms feedback and a 750-1250 ms ITI with jitter.

2.3.4 Analysis and Results

The goal of this Experiment 2 is to train participants to thoroughly learn the coupling relationship between the auditory cues and the visual stimuli. After this session, participants are prepared to continue the following Experiment 3, where I explored the prediction effect on visual categorization. As all participants have successfully learned the coupling, I will not report any analysis in this main narrative.

2.4 Experiment 3: Categorizing Visual Stimuli with Auditory Cues

2.4.1 Participants

The participants were the same group that took part in Experiment 2. After performing Experiment 2 and successfully learning auditory cue – Gabor patches couplings, the participants began this Experiment 3 (i.e., this categorization task).

2.4.2 Stimuli

Auditory cues. In addition to the two learned tones (880 Hz and 1760 Hz, in Experiment 2) that could predict the following visual stimuli, I also set a third tone (220 Hz) which participants did not learn and therefore contained no predicted meaning.

Visual stimuli. Experiment 3 contained the two kinds of learned visual stimuli in Experiment 2 (see Figure 2-3 A top row):

(1) left tilted and located Gabor patch
(2) right tilted and located Gabor patch, In Experiment 2, see Figure 2-3 A top row).

I also included three new visual stimuli as fillers (see Figure 2-3 A bottom row):

(3) left-tilted Gabor patch located at the right visual field
(4) right-tilted Gabor patch located at the left visual field
(5) vertical grating background without Gabor patch.

I presented the Gabor patch with a diameter of 3-degree visual angle and eccentricity of 15-degree visual angle at a viewing distance of 57cm.

Auditory cue – visual stimuli couplings. The auditory-visual couplings were counterbalanced across participants. Half of the participants learned the "880 Hz – left tilted and located Gabor patch" and "1760 Hz – right tilted and located Gabor patch" couplings, and the other half learned the "880 Hz – right tilted and located Gabor patch" and "1760 Hz – left tilted and located Gabor patch" couplings. The 220 Hz tone contained no predicted meaning (i.e., coupled with nothing).

2.4.3 Procedure

Before the task, I instructed participants on the 5 types of visual stimuli they would see during the task. Each experimental block started with learning pages as Experiment 2. Participants could replay the two informative auditory cues, simultaneously seeing the coupled visual stimuli to review the learnt couplings, and (re)play the non-informative (NonInfo) auditory cue with a blank screen displayed.

Each trial started with a fixation with 100ms presentation, 100ms disappearance and 300ms presentation, which constructed a flicker to hint at the upcoming auditory cue. Participants then heard one of the three auditory cues for 200ms, followed by a blank screen for 2000-2500ms jittered. After cue displaying, participants saw either one of the five visual stimuli presented for 120ms. Participants responded "left located and tilted" vs. "right located and tilted" vs. "others" with a keypress based on the location and orientation of the Gabor patch (see Table 2-2), followed by an ITI for 750-1250ms jittered.

I blocked the trials according to the types of auditory cues displayed. Each participant performed 10 blocks with 32 trials each, including 5 predictable blocks with the

valid prediction (i.e., the auditory cues displayed were informative) and 5 unpredictable blocks with no prediction (i.e., the auditory cues displayed were non-informative). In predictable blocks, the validity of the auditory cues is 75%. The structure of each predictable block is that:

1. In 75% of trials, the auditory cues correspondingly matched with the following Gabor patches
2. In 6.25% of trials, the auditory cues mismatched with the opposite Gabor patches (the cue predicting the left tilted and located Gabor was followed by the right located and tilted Gabor, and vice versa)
3. In 18.75% of trials, participants heard the auditory cues followed by the other 3 kinds of visual stimuli.

In unpredictable blocks, I displayed the left tilted and located or right tilted and located Gabor patches in 75% of trials, and the other 3 kinds of visual stimuli in 25% of trials. Table 2-3 shows the structure of each type of block.

I instructed participants (1) to respond as accurately and quickly as possible, which encouraged them to utilize the cue to generate prediction; (2) not to ignore the Gabor patches and entirely on the cue because of the existed mismatching and fillers. Participants had at least two minutes for a break to reduce fatigue when one block ended. Participants alternately completed predictive blocks and unpredictive blocks. Participants sat in a dimly lit room and used a chin rest to maintain the viewing distance at 57cm. I created the interface of the experiment via Psychotoolbox (version: 3.0.14) (Kleiner et al., 2007), setting the background screens as grey with a Gray value of 127.

Table 2-2 Visual stimuli and categorization instruction

Orientation (tilted)	Location	Instructed response
Anticlockwise (left)	Left	1 (left tilted and located)
Anticlockwise (left)	Right	3 (“others”)
Clockwise (right)	Left	2 (right tilted and located)
Clockwise (right)	Right	3 (“others”)
Only vertical grating background		3 (“others”)

Table 2-3 The stimuli structure of the two types of blocks.

Here I illustrated using the learned couplings '880 Hz – left located and tilted Gabor' and '1760 Hz – right located and tilted Gabor.'

One predictable block				One unpredictable block		
Condition	**Cue**	**Visual stimuli**	**Percentage (Numbers of trial)**	**Cue**	**Visual stimuli**	**Percentage (Numbers of trials)**
Match	880Hz	Left located and tilted	37.5% (12)	220Hz	Left located and tilted	37.5% (12)
	1760 Hz	Right located and tilted	37.5% (12)	220Hz	Right located and tilted	37.5% (12)
Mismatch	880Hz	Left located and tilted	3.125% (1)	220Hz	Others 1	6.25% (2)
	1760 Hz	Right located and tilted	3.125% (1)	220Hz	Others 2	6.25% (2)
	880Hz	Others 1	3.125% (1)	220Hz	Others 3	12.5% (4)
	1760Hz	Others 1	3.125% (1)			
	880Hz	Others 2	3.125% (1)			
	1760Hz	Others 2	3.125% (1)			
	880Hz	Others 3	3.125% (1)			
	1760Hz	Others 3	3.125% (1)			

2.4.4 Analysis and Results

2.4.4.1 Facilitation of Prediction on Categorization Behaviour

This task required participants to categorize the visual stimuli with the prediction generated by the auditory cues. Same as Experiment 2, I excluded the 3 participants, including one reporting deliberately not using the cue, one reporting not focusing on the task and one showing low ACC (lower than 60%). To measure the prediction effect (i.e., whether the informative cues facilitated the categorization behaviour) for the remaining participants, I proceeded with the following 3 steps:

Step 1: At group-level, I applied 2 (predictable/unpredictable) x 2 (left tilted and located/right tilted and located) ANOVA on ACC and RT. The result indicated no significant effect on ACC, but the informative cues (i.e., in the predictable trials) decreased RT, on average by 120.2 ms, F (1, 34) = 24.5, p = 0.001.

Step 2: I applied the paired-sample *t*-test between median RT of predictable and unpredictable trials of each participant separately for the categorization of "left located and tilted" and "right located and tilted" Gabor patches. Figure 2-3 C shows that the informative auditory cues reduced RT of both categorization of the two kinds of visual stimuli (left located and tilted: 549.7 ms (informative) vs. 670.1 ms (NonInfo), t (8) = 4.3, p = 0.003; right located and tilted: 547.1 ms vs. 665.1 ms, t (8) = 5.5, p = 0.0006).

Step 3: To confirm this group-level effect at each individual level, I applied the 2 by 2 ANOVA and separately for the two kinds of visual stimuli, independent samples t-tests between all single-trial RT under predictable and unpredictable conditions to each participant. The result showed that 8 out of 9 participants significantly facilitated their categorization with the informative auditory cues (see Supplemental Table S2-1).

2.4.4.2 Detecting Potential Learning effect and Other Disturbing Factors

The learning effect in psychology means that a learner shows rapid improvement (e.g., in terms of RT or ACC) in a new task due to increasing familiarity with the task. To understand (1) the potential learning effect during the task and (2) whether the different RT between predictable and unpredictable trials was due to a different improvement speed in the learning effect, I proceeded with the following 3 steps:

Step 1: To see whether facilitated RT of categorization was caused by the learning effect or the cue (as we expected), I computed the relationship between trial order and RT at the group level. Specifically, I extracted the median RT for each trial across participants, and then calculated the Spearman Rank Correlation between trial order and the median RT. Figure 2-3 D shows a significant correlation between trial order and RT of both predictable and unpredictable trials in both two kinds of Gabors (i.e., left located and tilted, right located and tilted) categorization. The result suggested that a learning effect indeed existed, which means the increasing familiarity with the task (probability with the cue or with the stimulus) sped up the behavioral RT.

Step 2: To check the learning effect at the individual level, for each participant, I calculated the Spearman Rank Correlation between trial order and RT per trial, separately for predicted trials and non-predicted trials. Supplemental Table S2-2 indicates the learning

effect quantified by the correlation coefficients, showing evidence that the learning effect existed widely across participants.

Step 3: To see whether the learning effect differs under predicted vs. non-predicted conditions, I applied 2 (predictable/unpredictable) x 2 (left tilted and located/right tilted and located) ANOVA on the obtained correlation coefficients (i.e., R in Supplemental Table S2-2). The result indicated no significant effect induced either by prediction type (F (1, 34) = 1.45, p = 0.26), or by left vs. right-type (F (1, 34) = 0.39, p = 0.55). Thus, although the learning effect was pervasive across participants, it did not differ due to whether an informative cue or a neutral cue was presented.

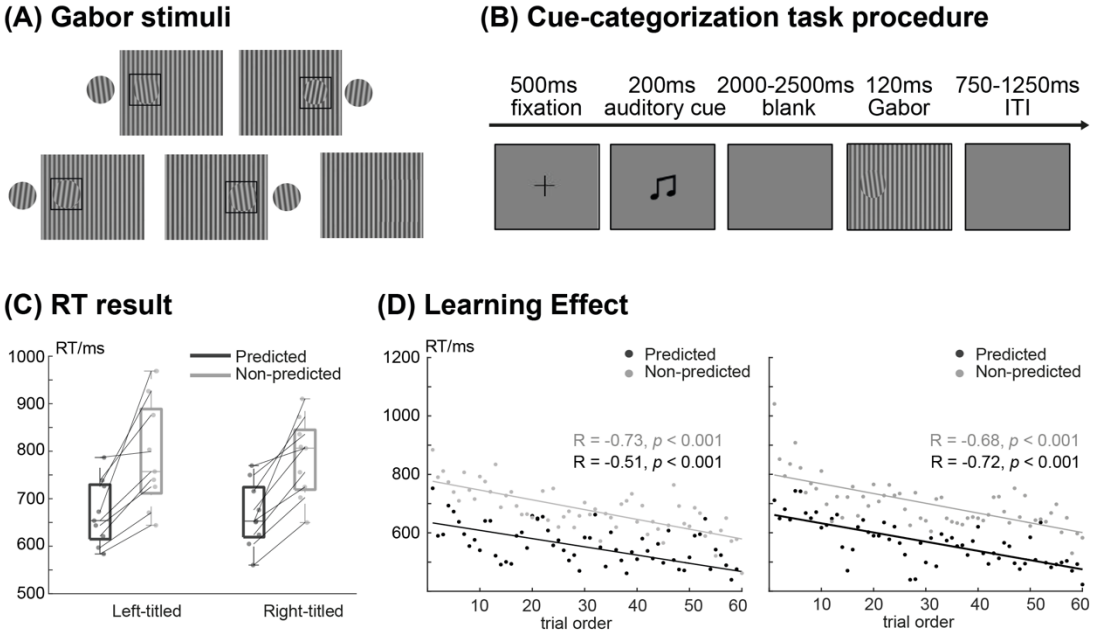


Figure 2-3 Cue-categorization task procedure and behavior-level result.

(A) Gabor stimuli. The top row shows the same two kinds of Gabor stimuli as in Experiment 2 (left-located and tilted, right-located and tilted). In this experiment, participants have learned the couplings between cue and the two Gabors. The bottom row shows the three kinds of filter stimuli (from left to right, left-located and right-tilted, right-located and left-tilted, vertical background). **(B) Cue-categorization task procedure.** Each trial started with a fixation with 100ms presentation, 100ms disappearance and 300ms presentation, constructing a flicker to hint at the upcoming auditory cue. Participants then heard one of the three auditory cues for 200ms, followed by a blank screen for 2000-2500ms jittered. After cue displaying, participants saw one of the five kinds of Gabor stimuli presented for 120ms. Participants responded on "left located and tilted" vs. "right located and tilted" vs. "others", followed by an ITI for 750-1250ms jittered. **(C) RT result.** Boxplots show that the prediction (i.e., informative cueing, black) sped up median Left vs. Right Gabor categorization RTs in each presentation condition, compared with non-prediction (i.e., non-informative cueing, grey). Black dots (vs. light

grey dots) indicate the per-participant median categorization RTs in predicted (vs. non-predicted) trials, indicating non-predicted RT increases in each individual participant replication. (**D) Learning effect.** Each black (vs. grey) dot represents the median RT of predicted (vs. non-predicted) trials, separately for left-tilted Gabor (the left plot) and right-tilted Gabor (the right plot). The black (vs. grey) text shows Spearman correlation coefficient (*R*) and its statistical probability (*p*) between trial order and RT in predicted (vs. non-predicted) trials, indicating the significant group-level learning effect within predicted and non-predicted conditions.

2.5 Discussion

In this study, I tested the distinction of auditory cues and facilitation of informative cues on visual categorization with three behavior experiments. Specifically, using a target detection design (Experiment 1), I demonstrated that participants could clearly distinguish the five selected tones. Thus, these five tones could be the alternative auditory cues in the following experiments and studies. Then, using a coupling training task (Experiment 2) and a cue-categorization task (Experiment 3), I demonstrated that participants could use the cues to generate predicted information after successful learning, facilitating the categorization RT.

It should be noted that, although I have proved the cue facilitation effect, in the design of the following neuroimaging experiment, there are several concerns to be discussed and optimized. In the following paragraphs, I will discuss the potential problems in this study and their corresponding solutions in the later Study 2 and Study 3.

Two concerns exist in the current design of Experiment 2. First, the relatability of cues could be impaired due to the test structure. In the learning test procedure, each block included 50% matched trials and 50% mismatched trials, where participants needed to prove their successful learning by making judgments on match/mismatch. However, given such a high proportion of mismatched trials, the predictive link to low-level perceptions is likely to fade away due to the intensive distractors. To solve or optimize this problem, the physical validity in the learning phase felt by participants should be the same as or higher than in the categorization. In this way, participants can establish the solid predictive couplings by perceptually convincing them with dependable cues rather than conceptually understanding the cue-stimulus mapping.

Second, participants only learnt the informative cues, causing a higher familiarity with the informative cues and lower familiarity with the non-informative cue in the following categorization tasks. As proved, the familiarity difference would not cause different learning effect in categorization task, in turn, would not influence the evidence for the facilitation effect on RT at behavior level. However, the neural mechanisms of the informative and non-informative cues turn to be incomparable when tracing the representations of the cues in the brain. In other words, if replicating this current design in a neuroimaging experiment, it remains a question of whether any difference detected in post-cue MEG signal is due to the prediction vs. non-prediction mechanism, or the low vs. high familiarity of these auditory pitches or even potentially, the slow vs. quick learning under predicted vs. non-predicted conditions. Thus, the familiarity of the three kinds of cues should be consistent in the following MEG study.

To address these two problems, I will optimize the training paradigm by including the learning of non-informative cues to guarantee that the participants' familiarities with the informative and non-informative cues are comparable. For example, when required to learn the '880Hz – left tilted and located Gabor', a participant needs to complete a 3-AFC (left tilted and located Gabor vs. right tilted and located Gabor vs. do not know) after hearing the 880Hz tone. Participants can see the feedback (correct vs. incorrect) after each response. They need to achieve 90% (or more) ACC to ensure that he/she has learned the coupling. Secondly, I will include the learning phase of the non-informative cue to equal the participants' familiarities with the informative and non-informative cues. To this end, before the learning test, I will instruct participants with the non-informative cue. In the learning test (i.e., the 3-AFC task), participants must choose the "do not know" to get the correct feedback. In addition, I will set the same number of the learning trials for the three cues (two informative + one non-informative cue), with the order randomized.

In the current design of Experiment 3, the task instruction required participants to build the prediction link between each cue and its coupled Gabor stimuli. However, as the location and orientation of the predictive Gabor are the same (i.e., predicting the left located & tilted Gabor or the right located & tilted Gabor), this design induces uncertain predicted contents due to the crossover effect. Specifically, participants could only predict location (left vs. right) but not the unique Gabor, thus, only focus on the predicted hemi visual field for the upcoming stimulus. As discussed in the introduction, this modulated

spatial attention (Charles D Gilbert & Wu Li, 2013) as one of the top-down control can enhance/suppress the neural response to the attended/unattended regions (Luck et al., 1997; Motter, 1993) and disambiguate the selection of the target stimulus (Desimone & Duncan, 1995). In this case, the spatial attention modulation would also provide the RT facilitation.

However, our main interest in this study is feature-based expectations where participants generate specific predicted feature contents. In existed study, this feature-based expectations can facilitate detection sensitivity (Summerfield & Egner, 2016), induce sensory template (Kok et al., 2017) in primary visual cortex and cause a reduction of neural activity after stimuli presentation (Alink et al., 2010; Den Ouden, Friston, Daw, McIntosh, & Stephan, 2009; Kok, Rahnev, Jehee, Lau, & De Lange, 2012; Todorovic, van Ede, Maris, & de Lange, 2011), which is built by a different mechanism compared with spatial attention modulation. Although in this Study 1, each of the two potential mechanisms could speed up the categorization RT, I need to exclude the probability of this confounding factors leading to different neural representations when continuing the research with brain activity recordings.

To address this problem, the improved design should specify the predicted contents that participants generated from the input cues, explicitly defining the factors facilitating behavior. In the following study, I will use location cues and auditory cues to respectively manipulate spatial attention and feature expectation, dividing the predicted contents into location and feature prediction for clarity. In this way, we can trace the neural representation of these two kinds of predictions with improved design both separately and integrated.

Another explorable point is the stimuli with location-orientation confounding (i.e., the left-tilted and right-located Gabor and the right-tilted and left-located Gabor) which could potentially induce a Stroop effect (MacLeod, 1991). However, in this design, the number of filler trials with the confounding stimuli was too few to compute the related analysis (4 per block). Here I mentioned it as an inspiring direction for the futural studies.

In conclusion, this study validated the participants could distinguish different cue tones and proved that the predictive cues could reduce the RT in the visual categorization of simple Gabor patches. It is a preamble at behavior level for my following exploration

with measurements in neuroscience. On the other hand, with the problems exposed in this study, I would improve the task paradigm in the following study at the brain level by (1) convincing participants of the reliability of the cue-Gabor couplings unconsciously, (2) combining the non-informative cue into the learning phase and (3) manipulate the specific prediction contents at a different level for clarity.

2.6 Supplemental Material

2.6.1 Supplemental Tables

Supplemental Table S 2-1 Individual-level effect of informative vs. non-informative cues

on the RT of categorization of Gabor patches under 2 (left vs. right-located and tilted) condition, tested with ANOVA and independent samples t-test (p showed before Bonferroni correction).

Participant ID	ANOVA *F*-value: Effect of cue validity	Left condition mean RT reduction (NonInfo-Informative)	Right condition mean RT reduction (NonInfo-Informative)
1	$F(1, 229)=34.7$, $p<0.001$	140.7 ms, $p<0.001$	77.0 ms, $p=0.005$
2	$F(1, 231)=179.0$, $p<0.001$	264.2 ms, $p<0.001$	259.2 ms, $p<0.001$
3	$F(1, 230)=10.2$, $p=0.002$	71.8 ms, $p=0.001$	**46.4 ms, $p=0.05$**
4	**$F(1, 232)=1.34$, $p=0.25$**	**21.6 ms, $p=0.39$**	**18.9 ms, $p=0.44$**
5	$F(1, 233)=28.3$, $p<0.001$	75.2 ms, $p=0.01$	123.0 ms, $p<0.01$
6	$F(1, 230)=128.2$, $p<0.001$	205.5 ms, $p<0.001$	160.7 ms, $p<0.001$
7	$F(1, 231)=59.9$ $p<0.001$	102.2 ms, $p<0.001$	70.2 ms, $p<0.001$
8	$F(1, 233)=82.6$ $p<0.001$	150.0 ms, $p<0.001$	94.4 ms, $p<0.001$
9	$F(1, 223)=18.3$, $p<0.001$	155.3 ms, $p<0.001$	84.9 ms, $p=0.02$

Bold indicates the **insignificant** ($p > 0.05$ after Bonferroni-correction) result.

Supplemental Table S 2-2 Experiment 3 Individual-level learning effect

measured with Spearman correlation between the order and RT of each trial (p showed before Bonferroni correction).

Participant ID	Left Gabors with informative cueing	Left Gabors with NonInfo cueing	Right Gabors with informative cueing	Right Gabors with NonInfo cueing
1	***R* = -0.14, *p* = 0.30**	$R = 0.42$, $p < 0.001$	***R* = -0.27, *p* = 0.04**	$R = -0.35$, $p = 0.006$.
2	***R* = -0.13, *p* = 0.34**	$R = -0.36$, $p = 0.005$	***R* = 0.002 *p* = 0.99**	$R = -0.53$, $p < 0.001$
3	***R* = -0.16, *p* = 0.22**	$R = -0.35$, $p = 0.006$	***R* = -0.11, *p* = 0.39**	$R = -0.38$, $p = 0.003$
4	***R* = -0.33, *p* = 0.025**	***R* = -0.01, *p* = 0.94**	***R* = 0.04, *p* = 0.79**	***R* = -0.23, *p* = 0.08**
5	$R = -0.41$, $p = 0.001$	***R* = -0.17, *p* = 0.20**	$R = -0.48$, $p < 0.001$	***R* = 0.02, *p* = 0.89**
6	***R* = -0.18, *p* = 0.16**	$R = -0.46$, $p < 0.001$	***R* = 0.05, *p* = 0.70**	$R = -0.41$, $p = 0.001$
7	$R = -0.37$, $p < 0.004$	$R = -0.40$, $p = 0.001$	$R = -0.41$, $p = 0.001$	$R = -0.42$, $p < 0.001$
8	$R = -0.68$, $p < 0.001$	$R = -0.44$, $p < 0.001$	$R = -0.52$, $p < 0.001$	$R = -0.47$, $p < 0.001$
9	$R = -0.36$, $p = 0.004$	$R = -0.62$, $p < 0.001$	$R = -0.49$, $p < 0.001$	$R = -0.73$, $p < 0.001$

Bold indicates the **insignificant** ($p > 0.05$ after Bonferroni-correction) result.

3 Study 2: Frontal cortex modulates occipital predictions of visual contents in a functional network

3.1 Introduction

Since Helmholtz' "unconscious inferences," vision scientists have worked with the hypothesis that what we visually perceive is in part influenced by the bottom-up sensory input, but also by top-down expectations of what this input might be (De Lange et al., 2018; Kinchla & Wolfe, 1979). Expectations predict the upcoming visual information (Clark, 2013; K. Friston, 2010; Yuille & Kersten, 2006), thereby facilitating its disambiguation from the noisy input (Gilbert & Sigman, 2007; Kok, Jehee, et al., 2012a) to then speed up behaviour (Bar et al., 2006).

It is revealed that coordinated networks of brain regions process visual information bottom-up (feedforward) and/or top-down (predictive, feedback) (Charles D. Gilbert & Wu Li, 2013; Kinchla & Wolfe, 1979). In bottom-up mode, left and right occipital cortices represent respectively the right and left visual hemifield impinging the retina (Flom, Heath, & Takahashi, 1963). These contra-lateral representations then transfer (including across hemispheres) up to the ventral and dorsal regions (Grill-Spector & Weiner, 2014) where they are merged to implement "vision for categorization" and "vision for action", respectively (R. A. A. Ince et al., 2016; Kveraga et al., 2007; Schendan & Stern, 2007; Zhan et al., 2019). A convergence of formal models (K. Friston, 2010; Rao & Ballard, 1999; Yuille & Kersten, 2006) and empirical evidence (Linde-Domingo et al., 2019) suggest that the top-down mode should reverse the bottom-up mode. Predictions originating from memory would travel from the top of the "vision for categorization" pathway in high hierarchical regions, down to the left vs. right occipital cortices, depending on the visual hemifield where the visual features are expected. In addition, advances in neuroimaging showed that top-down stimulus predictions modulate the neural signal in multiple ways, by inducing patterns of stimulus-specific activation with reduced alpha oscillation power, with enhanced gamma-band activity (Hindy et al., 2016; Kok et al., 2014; Michalareas et al., 2016), or with selective changes of cortical layer activity, as shown with high-field MRI (Margalit et al., 2020). However, the specific dynamic neural mechanisms and brain regions that propagate such visual predictions down the visual hierarchy remain unclear.

Here, to understand how a network of brain regions coordinates to propagate specific predictions, and its relationship with the bottom-up processing, we addressed two key questions:

1) When, where, and how does a prediction dynamically propagate through a network of brain regions?

2) Does the top-down predictive process contain the contents in the bottom-up feedforward processing of a stimulus feature in the visual hierarchy?

To control predictions, we used the three-stage cueing design depicted in Figure 3-1A and 3-1B (see also *Methods, Tasks, Cueing-Categorization*). On each trial, a spatial cue at Stage 1 (a green dot briefly displayed left vs. right of a central fixation cross, cf. Posner cueing (Posner & Petersen, 1990)) predicted the visual hemifield location (left vs. right) of an upcoming Gabor patch (henceforth, Gabor) with 100% validity, followed by a 1-1.5s blank screen. At Stage 2, on informative trials an auditory cue (a 250ms sweeping tone at 196 Hz vs. 2217 Hz) predicted the Spatial Frequency content (SF, Low vs. High) of the upcoming Gabor stimulus with a 90% validity (henceforth, refer as the "predicted" trials), followed by another 1-1.5s blank interval. On uninformative condition, "non-predicted" trials (33% of Stage 2 trials), a 622 Hz neutral auditory cue had no association with LSF or HSF. Finally, at Stage 3, the actual Gabor stimulus appeared in the participant's left vs. right visual hemifield for 100ms. Each participant (N = 11, see *Methods, Participants*) categorized the Gabor SF as quickly and accurately as they possibly could without feedback (i.e. 3-AFC, with responses "LSF" vs. "HSF" vs. "don't know"). We concurrently recorded the participant's dynamic brain activity with MEG and reconstructed it on 12,773 sources, see *Methods, MEG Data Acquisition and Pre-processing*.

In this study, my results will firstly show a reconstructed temporal-frontal-occipital network communicating predictions, secondly validate the frontal cortex mediation in this network, and finally prove that the predicted visual information is indeed represented in a high-to-low hierarchical pathway.

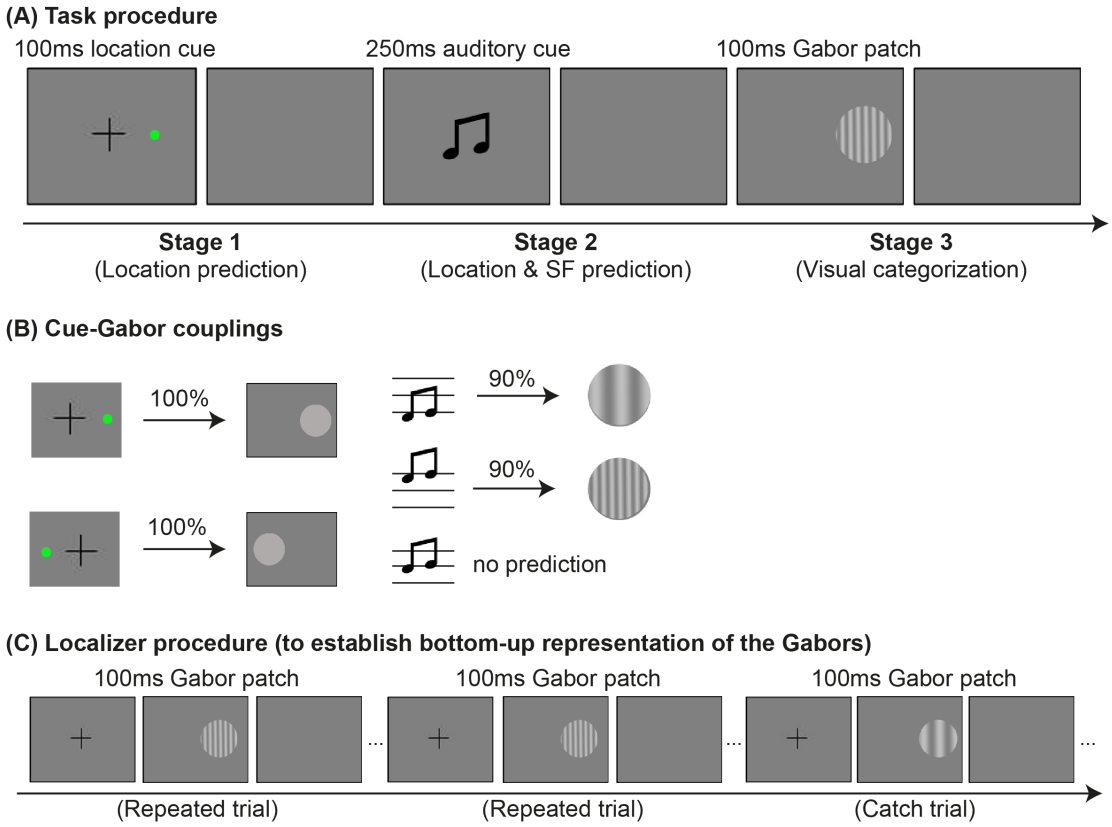


Figure 3-1 Experimental design.

(A) Task procedure. Each trial started with a 500ms fixation, followed by Stage 1 where a 100ms green dot (i.e. location cue) that predicted the location of the upcoming Gabor patch, followed by 1000-1500ms blank screen with jitter. At Stage 2, a 250ms sweeping sound (i.e., SF cue) predicted the LSF vs. HSF content of the upcoming Gabor, followed by a 1-1.5s blank screen with jitter. At Stage 3, the Gabor was presented for 100ms. Participants categorized the SF of the Gabor, followed by a 750-1250ms inter-trial interval (ITI) with jitter. **(B) Cue-Gabor couplings.** At Stage 1, the left vs. right location cue predicted the left vs. right location of the upcoming Gabor with 100% validity. At Stage 2, the 196 Hz vs. 2217 Hz informative auditory cue predicted the Gabor LSF vs. HSF contents with a validity of 90%; a 622 Hz neutral auditory cue served as non-prediction control on 33% of the trials, (i.e. 50% probability of LSF or HSF). **(C) Localizer procedure (to establish bottom-up representation of the Gabors).** Prior to the cueing experiment, we ran each participant in an independent MEG localizer. Each trial started with a 500ms central fixation, followed by a 100 ms Gabor, then a 750ms ITI blank screen. In each block of 20 trials (i.e. one of left-LSF, right-LSF, left-HSF, right-HSF), at the fixed location (left or right presented), participants passively viewed 20 Gabors at the same SF. 1 catch trial presented one different-SF Gabor that we instructed participants to detect in each block.

3.2 Results

3.2.1 Prediction Facilitation on Behavior

Valid predictions (i.e., the 90% valid trials with informative cueing) did indeed improve categorization accuracy compared to non-predicted (neutral cue) trials, on average by 2.58% (96.9% vs. 94.3%), $F(1,86)=22.5$, $p=0.0008$, and sped up Reaction Times (RTs), on average by 87.7ms (454.4ms vs. 542.1ms), $F(1, 86)=20.8$, $p=0.001$. RT improvements applied to each condition of presentation (all condition significant, see Figure 3-2 and Supplemental Table S3-1 and *Methods, Cueing improves Behavior*) and to each individual participant, obtaining Bayesian population prevalence (R. A. Ince, Paton, Kay, & Schyns, 2021) of 11/11 = 1 [0.77 1] (Maximum A Posteriori (MAP) [95%, Highest Posterior Density Interval (HPDI)], see Supplemental Table S3-2).

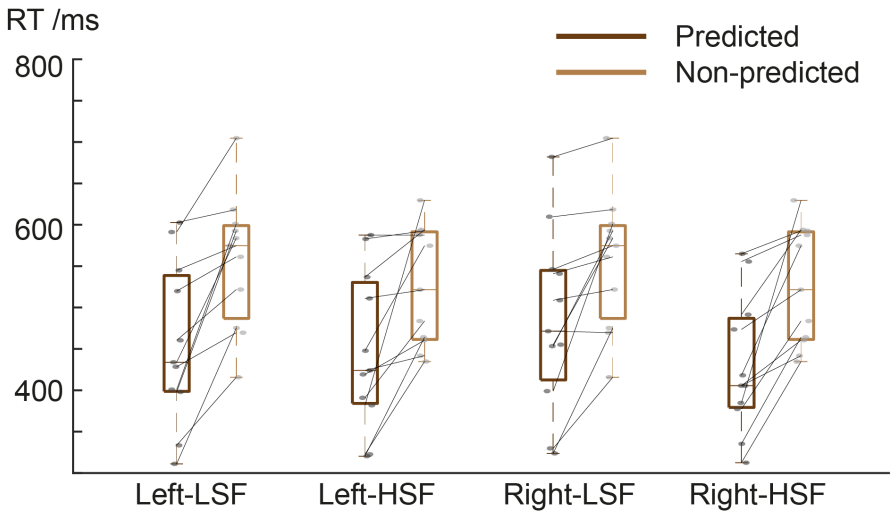


Figure 3-2 Behavioural results.

Boxplots show that the prediction (i.e., informative cueing, dark brown) sped up median LSF vs. HSF Gabor categorization RTs in each presentation condition, compared with non-prediction (i.e., neutral cueing, light brown). Dark grey dots (vs. light grey dots) indicate the per-participant median categorization RTs in predicted (vs. non-predicted) trials, linked to indicate non-predicted RT increases in each individual participant replication.

3.2.2 Prediction Network Nodes

We concurrently measured the participant's dynamic activity with MEG and reconstructed this activity on 12,773 sources across the brain, see *Methods, MEG Data Acquisition and pre-processing*. To address the first question above, of the network of

brain regions that propagate the prediction prior to stimulus onset, we computed when, where and how strongly MEG sources represent the prediction, following Stage 2 cue onset—i.e. computed as the Mutual Information (MI) between <LSF vs. HSF auditory cue; Stage 2 MEG source activity>, on each source and time point (see *Methods, Prediction Network Nodes*). Figure 3 presents these data for each source, color-coded by the timing of prediction representation following Stage 2 cue onset.

Dynamic prediction representation in MEG source activity shows a propagation pattern in three stages, revealed by a clustering analysis, see Supplemental *Methods, Prediction periods clustering* and Supplemental Figure S3-8. We illustrate these three stages in a typical participant (Figure 3A), with an early ([90-120ms], cyan) temporal lobe representational peak post-cue (Period 1), followed by an anterior temporal and frontal cortex peak ([120-200ms], magenta, Period 2, especially in prefrontal cortex), itself followed by a parietal and a final occipital peak ([> 200ms], orange, Period 3). Figure 3-3B generalizes the dynamic propagation of prediction to the group, by locating in each participant (N = 11) the source that maximally represent the prediction (i.e. peak MI), in each of the three time windows, and separately for trials with a left vs. right located prediction (other individual results are shown in Supplemental Figure S3-1). The data suggests that the prediction is maximally represented following a sequential order, starting as expected in the temporal lobe, moving to frontal cortex and then to the occipital cortex contra-lateral to the predicted location. Additionally, this propagation is distinct from an only-temporal pattern induced by the pure bottom-up auditory input, which is tested on the localiser data prior to the cueing task (see Figure 3-1C, Supplemental Figure S3-2 and *Methods, Localiser, Localiser representation*).

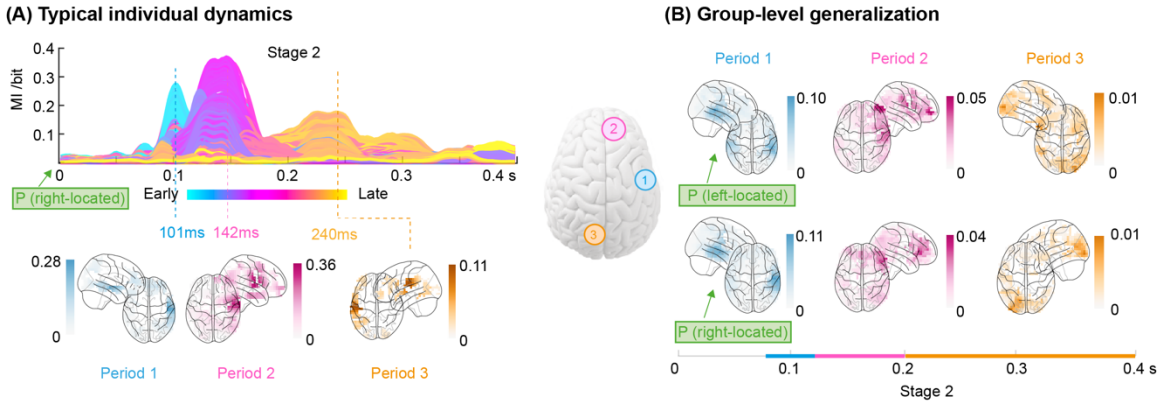


Figure 3-3 Dynamic cue representation across the brain at Stage 2.

(A) Typical individual dynamics. For a typical participant with right located SF prediction, we computed the representation of the prediction (as MI(<LSF vs. HSF cue; MEGt>, Y-axis), between 0 and 0.4s post auditory cue (X-axis), on each source in occipital regions (lingual gyrus, cuneus, inferior occipital gyrus), temporal (fusiform gyrus, inferior temporal gyrus, middle temporal gyrus, superior temporal gyrus), parietal (superior parietal lobe, inferior parietal lobe, angular gyrus, supramarginal gyrus), premotor (precentral gyrus, postcentral gyrus), and frontal (orbitofrontal gyrus, inferior frontal gyrus, middle frontal gyrus, medial frontal gyrus, superior frontal gyrus). The curves show these per-source time courses of prediction representation, color-coded by their (ranked) peak MI time. Glass brains below reveal for this participant the localized sources that peak in each period—i.e. [91-111ms], [136-176ms], [210-305ms]. They show that the representation of the auditory LSF vs. HSF prediction starts as expected in temporal lobe (cyan), then moving to frontal regions (magenta), then through parietal and occipital cortex (orange). The schematic brain in the centre of the figure illustrates these three successive stages of prediction representation. **(B) Group-level generalization.** Using the same method of maximum prediction representation within three time windows, we generalized the temporal sequence across participants using [90-120 ms], [120-200 ms], [>210 ms], color-coded on the X-axis) post SF cue, independently for left- vs. right-located trials. Glass brain shows the cross-participant mean of localized sources in each time window, revealing the temporal-frontal-occipital (contra-laterally to the predicted location) prediction representation sequence.

3.2.3 Prediction Network Reconstruction

The representation dynamics of the predictive cue that it is communicated across a functional network of brain regions. To reconstruct this network, we measured the communications of the cue across the previously identified regions (i.e. temporal, frontal and occipital, as schematized in Figure 3-3), by computing the Directed Feature Information (here, of Prediction P) between these regions, as DFI (P, regionX, regionY), FWER-corrected, $p<0.05$, see *Methods, Prediction Network reconstruction* and Supplemental Figure S3-3. Figure 3-4B shows network communications of the prediction for right-located trials (and left-located trials in Supplemental Figure S3-4). In both cases, Figure 3-4B shows that prediction communications follow a sequence from temporal lobe to frontal cortex and then on from frontal to occipital cortex, both in a typical participant and at the average group level. Figure 3-4C shows replications of these network communications of a prediction, for left-located (dashed lines) and right-located trials, in at least 9/11 participants—i.e., Bayesian population prevalence = 0.81 [0.53 0.96], MAP [95% HPDI]. Supplemental Figure S3-4 shows the individual results.

3.2.4 Prediction Network Modulations of Predictive Cue Communications

Here, we test whether frontal cortex mediates the communications of the predictive cue from temporal to occipital cortex. Figure 3-4D demonstrates this frontal mediation (3-4D.1) in contrast to a direct communication from temporal to occipital cortex without frontal mediation (3-4D.2), see *Methods, Prediction Network Modulations of Predictive Cue Communications*. We show that significant DFI communications of the predictive cue from temporal to occipital (i.e. 1->3 in Figure 3-4D) occur with a 100-150 ms delay (3-4D.1, dashed circled red in the panel, FWER corrected, $p<0.05$, one-tail), but that removing the modulatory effect of frontal source 2 from this computation (in 3-4D.2) also removes communication of the prediction from temporal to occipital cortex (an effect replicated independently for left- and right-located trials in at least 10 participants, Bayesian population prevalence = 0.91 [0.64 0.99] (MAP [95% HPDI], other individual results are shown in Supplemental Figure S3-5). Hence, we conclude that frontal cortex mediates the prediction communication in the network, from temporal to occipital cortex.

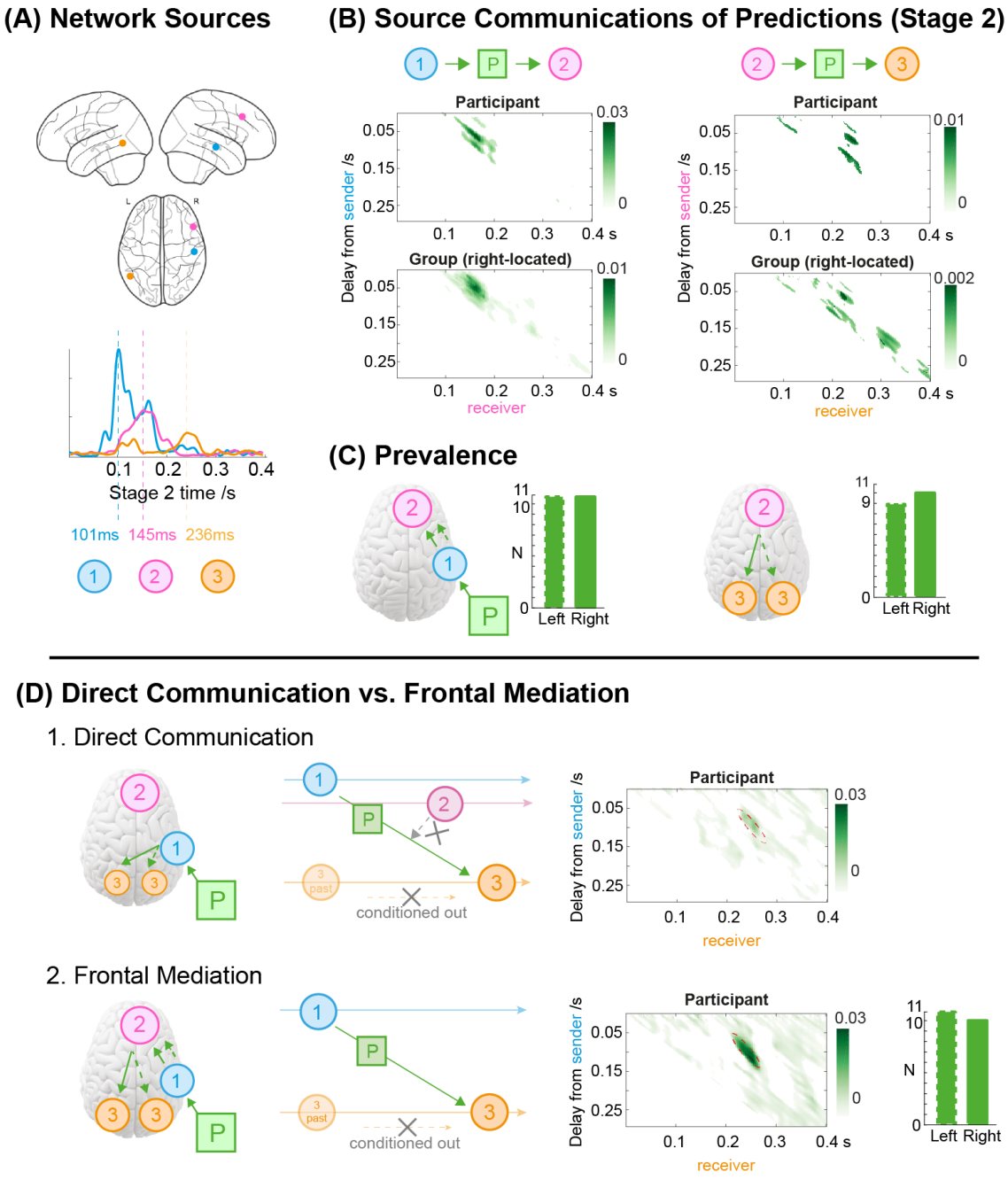


Figure 3-4 Prediction communications in a temporal-frontal-occipital network.

(A) Network sources. We illustrate source selection in one typical participant. We selected the source with highest MI representations (see plot) of the predictive cue following Stage 2 onset time (dashed lines), in the three color-coded time windows, resulting in temporal lobe source 1 (cyan colored, 0.09-0.12s), frontal cortex source 2 (magenta, 0.12-0.2s) and finally occipital cortex source 3 (orange, 0.2 to 0.4s) serving as functional network nodes. **(B) Source communications of predictions (Stage 2).** Using Directed Feature Information, we computed the participant communications across these network nodes (i.e. 1->2 and 2->3). Both participant plots represent these communications in the receiving node time course (X-axis), as communication delays from the sending node (Y-axis). For example, we can see that node 1 sends the predictive cue P to node 2, with a 50-100 ms delay. Group plots averages these communications across participants. **(C) Prevalence.** The brain plots illustrate the network communications of the predictive

cue. We replicated 1->2 for left-located trials (dashed) and right-located trials in 11/11 participants, Bayesian population prevalence = 1 [0.77 1] MAP [95% HPDI]. 2->3 was replicated for left-located trials (dashed) in 9/11 participants, Bayesian population prevalence = 0.81 [0.53 0.96], MAP [95% HPDI] and right-located in 10/11 participants, Bayesian population prevalence = 0.91 [0.64 0.99], MAP [95% HPDI] **D. Mediation of frontal cortex.** *D.1.* Direct communication of the prediction from temporal (source 1) to occipital cortex (source 3).. *D.2.* Mediation of frontal cortex (source 2) to communicate the prediction from temporal (source 1) to occipital cortex (source 3). Illustrated in a typical participant, with a 100-150 ms delay between temporal sender and occipital receiver. This computation conditions out the modulatory role of temporal cortex (source 2) in the prediction communication. The difference between *D.1.* and *D.2.* was replicated for left-located trials (dashed) in 11/11 participants, Bayesian population prevalence = 1 [0.77 1] MAP [95% HPDI], and right-located trials in 10/11 participants, Bayesian population prevalence = 0.91 [0.64 0.99], MAP [95% HPDI]. Frontal cortex mediates network communications of the prediction from temporal to occipital cortex.

3.2.5 Top-down predictive contents along the visual hierarchy

As demonstrated that the predictions start at where the cues initially are processed (i.e., the auditory cortex), mediated by frontal cortex and arriving at occipital cortex contra-laterally to the predicted location. We want to further validate that what propagates along the visual hierarchy indeed contains the visual SF feature information.

Here, we controlled stimulus feature contents (i.e., left vs. right presentation location; LSF vs. HSF) to explicitly test whether the representation of the same stimulus features reversed between their top-down predictions and their bottom-up processing. We did so by using a decoder (a cross-validated supervised classification algorithm) to compare the top-down prediction of each Gabor (at Stage 2 of the cueing experiment, i.e., auditory cueing of LSF vs HSF, cf. Figure 3-1A) to its feedforward bottom-up representation (from a localizer run for each participant prior to the cueing experiment, cf. Figure 3-1C).

Specifically, We first trained decoders to classify the Gabor LSF vs. HSF from the MEG localizer responses (see *Methods, Gabor localizer*), separately for left and right Gabor presentations and every 2 ms between 0 and 500 ms post onset (within-participant Multivariate Pattern Analysis (Treder, 2020), cross-validated, FWER (Nichols & Holmes, 2002), $p<0.05$, one-tailed, see *Methods, Bottom-up Cross-validation*, Supplemental Figure S3-6). Then, we applied these bottom-up decoders to classify the LSF vs. HSF contents of

the top-down predicted Gabor at Stage 2, when the integrated spatial and informative auditory cues enable full prediction of the upcoming stimulus (i.e., its LSF vs. HSF, separately for left and right-located trials, cf. Figure 3-1, see *Methods, Top-down Cross-decoding*).

Figure 3-5 reports such common patterns between bottom-up and top-down Gabor representations, but in a temporally reversed order. Group averaged top-down predictions of LSF vs. HSF Gabor contents at Stage 2 reverse their bottom-up representations in the hierarchical occipito-ventral pathway (FWER (Nichols & Holmes, 2002), $p<0.05$, one-tailed). That is, reversal starts in the deep sources of the right fusiform gyrus, near auditory cortex area A1, where the auditory cue is processed at (~90-160 ms post auditory cue, see Figure 3-1), for both left and right predicted Gabors. Importantly, following this, SF predictions then diverge in the occipital sources contra-lateral to the predicted location of the upcoming Gabor (~160-400ms, Figure 3-2, brain plots). Contra-lateral occipital sources representations of late predictions were significantly different between left and right hemispheres as a function of right and left predicted locations—i.e. group-level ANOVA, 2 (left vs. right prediction) by 2 (right vs. left occipital cortex), $F(1, 34) = 5.65$, $p = 0.02$), an effect replicated in 8/11 participants (Bayesian population prevalence = 0.73 [0.42 0.91] (MAP [95% HPDI], see Supplemental Table S3-4 and Supplemental Figure S3-7).

In sum, the analyses showed that the top-down predictive process reverses the bottom-up processing and contains the same specific stimulus features in the visual hierarchy. Specifically, predictive cues for left vs. right upcoming Gabors at Stage 2 reactivate the LSF vs. HSF contents represented in right fusiform gyrus sources. These contents then flow down the visual hierarchy to the left vs. right occipital cortex location contra-laterally predicted by the visual cue at Stage 1.

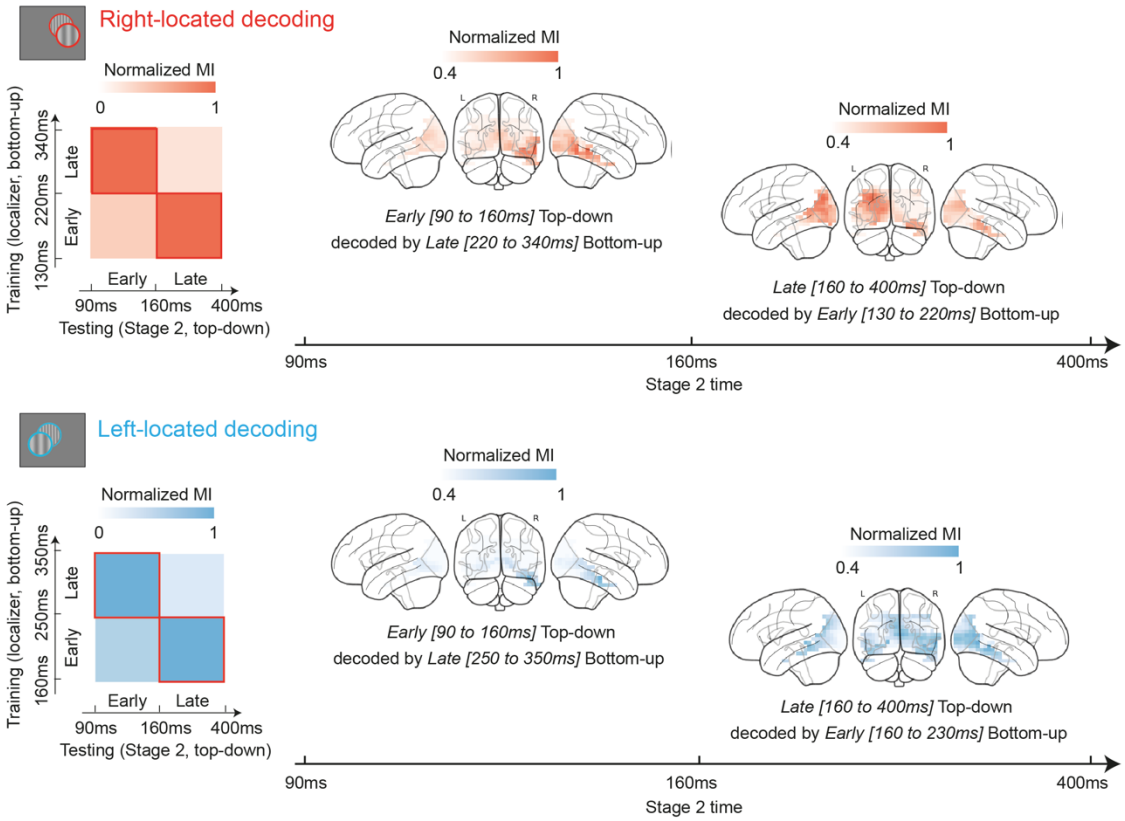


Figure 3-5 Top-down predictions of left vs. right presented, LSF vs. HSF Gabors reverse their bottom-up processing in the visual hierarchy.

For right (color-coded in orange) and left (color-coded in blue) located Gabor trials, the 2 x 2 matrix reveals the predicted (cf. A) group averaged cross-decoding performance between bottom-up representation of Gabors (X axis, as training set) in the MEG localizer and their top-down predictions (Y-axis, as testing set) at Stage 2 of the prediction experiment. We quantified decoding performance as MI(<decoder value (continuous values as criteria to classify decoded LSF vs. HSF); ground truth LSF vs. HSF>) and down-sampled time to a 2 x 2 matrix that averaged MI along the Y-axis (separately for right/left-located Gabor trials, with bottom-up intervals assigned to [~130ms/160ms – 400ms] and [~220ms/250ms – 340ms/350ms] time windows) and X-axis (with top-down intervals assigned to [~90ms – 160ms] and [~160ms – 400ms] time window). The 2 x 2 matrices represent the group-level mean performance of significant decoding (FWER, $p<0.05$, one-tailed), normalized within each decoder (i.e., normalized across each row). This highlights the reversal whereby the late bottom-up classifier better decodes the early top-down responses, and the early bottom-up classifier better decodes the late top-down responses. Brain plots show the sources that contribute to the classification during the top-down stage. To reveal them, we computed MI(<decoder values; Stage 2 MEG>) along the occipito-ventral pathway (i.e. in lingual gyrus, cuneus, inferior occipital gyrus, middle occipital gyrus, superior occipital gyrus and fusiform gyrus). Source color is the dot product of the time × time decoding performance and the time × time MI(<decoder values; Stage 2 MEG). The brain sources show that top-down LSF vs. HSF predictions reverse their bottom-up representations in the visual hierarchy, starting in the right fusiform gyrus [~90 – 160ms] post auditory cue at Stage 2 for both right (orange) and left (blue) predicted Gabors and ending in the contra-lateral occipital cortex [~160 – 400ms].

3.3 Discussion

We asked two key questions pertaining to the dynamic mechanisms of predictive information propagation (i.e. top-down, feedback processes) and their relations to the representation of these same contents from the stimulus itself (i.e. bottom-up, feedforward processes). Our approach used a three-stage predictive experimental design: At Stage 1, a spatial cue predicted the location (left vs. right) of an upcoming Gabor; At Stage 2, an auditory cue predicted the SF contents of the stimulus; At Stage 3, the Gabor stimulus was presented. At behaviour-level, we found that valid predictions of Gabors sped up categorization RTs in all participants and improved accuracy in most. Our analyses reconstructed the functional network that propagates the auditory predictive cue from temporal to occipital cortex. We found a chronological representation of the predictive cue from temporal (90-120ms post Stage 2 cue) to frontal (120-200ms) to occipital (200-400ms) cortex. We showed that frontal cortex mediates the communication of the predictive cue from temporal to the left vs. right occipital cortex, depending on cued location. We further asked whether the top-down predictive process contains the physical visual contents same as the information in bottom-up hierarchical processing of a Gabor. We found that it did so, with a top-down process located deep in rFG sources to propagate down to the predicted contra-lateral OCC. Together, our results build a dynamic prediction communication network from the sensory input of the predictive cue.

Compared with other studies that similarly traced prediction, we promoted the findings in this field in two general aspects. First, we dynamically traced the top-down flow from higher hierarchical regions to the contra-lateral occipital cortex at the source level. The “sensory template” mentioned by Kok (Kok et al., 2014), was proved by decoding analysis on sensor-level MEG signals, while we provided the evidence specifically at the source level. Second, we focus on the dynamics after the cue onset (i.e., Stage 2), addressing how predicted contents propagate. In the previous study conducted by Kok (Kok et al., 2017), their finding is visual-stimulus-locked, and the “stimulus templates” they found are [-40ms, 230ms] around the stimulus onset. Together with these findings, we can build a pathway activating the representation of predicted content from temporal, frontal to occipital, and replaying it after the cue onset and before the predicted stimulus onset.

Functional networks predict and then represent stimulus contents

Methodologically, we reconstructed with DFI a functional network that dynamically propagates a specific auditory prediction of visual contents (LSF vs. HSF) from temporal to occipital cortex, ascribing a mediation role to frontal cortex. That is, without mediation of frontal cortex, there is no such propagation of the prediction. Such connectivity analyses involve individual MEG sources acting as sending and receiving network nodes. Importantly, DFI functional connectivity differs from other signal-to-signal connectivity analyses (such as Granger causality or transfer entropy) because DFI isolates the communication of a specific feature (e.g. the auditory prediction of LSF vs. HSF at Stage 2) as a percentage of the full signal-to-signal connectivity. A similar logic applied to isolating the mediatory role of frontal cortex. Thus, DFI reconstructed the network of regions that dynamically propagate a prediction, addressing the first question schematized in Figure 1-2.

One could object that the transfer of the predicted cue across the temporal-frontal-occipital network is a simple effect of the auditory signal spreading in space x time rather than genuine network communications. We addressed this objection through four aspects. Firstly, by comparing the cue representation between bottom-up and top-down, we found the bottom-up auditory stimuli only indue one explicit activation peak in temporal, and with a top-down modulation, the auditory cue can sequentially activate temporal, frontal, and occipital regions. Secondly, the cue representation finally activates the occipital region contra-laterally to the predicted location, showing a prediction-related specificity. Thirdly, we proved that the frontal cortex is a necessary intermediary to transfer the predictive cue from the temporal lobe to occipital cortex, in contrast to a direct transfer that bypasses the frontal lobe, suggesting a directional top-down propagation rather than the signal spreading pattern. Fourthly, we applied a decoding analysis and directly validated the top-down prediction of a feature contains its contents in bottom-up processing.

The role of high-hierarchal regions

An interesting finding of our functional network is that frontal cortex modulates the temporal to occipital communication of the predictive cue. More precisely, we located the sources with highest representation of the predictive cue in the dorsolateral prefrontal cortex (dlPFC (Sanches et al., 2009)), often related to working memory (D'Esposito et al., 1998; Friedman & Robbins, 2022; J. B. Rowe, Toni, Josephs, Frackowiak, & Passingham,

2000). Frontal cortex could therefore orchestrate the information of the auditory cue (i.e. upcoming LSF vs. HSF) with the memory of the upcoming stimulus location (i.e. left vs. right visual field) and selectively prepare the contra-lateral occipital sources accordingly. Our results are compatible with this hypothesis, because occipital voxels at Stage 2 are indeed contralateral to the predicted field—i.e. left occipital sources for a predicted right visual field stimulus and vice versa. Future work that fuses MEG and high-field fMRI will seek to resolve the specific cortical layer that receives the prediction flow at Stage 2 (e.g. in central laminar layers), and how this prediction then interacts with the cortical layer representation of the feedforward flow when the stimulus is shown at Stage 3 (e.g. in peripheral laminar layers).

The prevalent conception of vision-for-categorization suggests an interplay of two dynamic flows of information within the occipito-ventral pathway. The bottom-up flow progressively reduces the high-dimensional input into a lower-dimensional representation that is compared with memory to produce categorization behavior. The top-down flow predicts category information (i.e. features) from memory that propagates down the same hierarchy to facilitate input processing and behavior. However, the neural mechanisms that support such dynamic feature propagation up and down the visual hierarchy and how they facilitate behavior remain unclear. With the decoding analysis, we show the predicted visual contents flow from right fusiform gyrus to early visual cortex, hierarchically and chronologically reverse to the typical bottom-up pathway (Zhan et al., 2019). In Zhan et al., this bottom-up pathway propagates LSF and HSF features from early visual cortex contra-laterally to the feature location, integrating at rFG to generate the perception and lead to subsequent decision making. In contrast, in our study, with the manipulation of the spatial location, we expect an integration at a higher-level hierarchy (rFG here) between location and SF, although the latter is very fundamental in visual processing.

However, in this study, three limitations should also be put forward. The first is towards the method, where we empirically low-passed the data to 25 Hz (Zhan et al., 2019). Although we remained the alpha and low beta components which have been widely reported related to feedback signalling (Benedek, Bergner, Könen, Fink, & Neubauer, 2011; Haegens, Händel, & Jensen, 2011; Lobier, Palva, & Palva, 2018; Michalareas et al., 2016), we might miss effect at high beta and gamma bands (Michalareas et al., 2016; Siegel, Donner, Oostenveld, Fries, & Engel, 2008). In the future, it is valuable to conduct a

further time-frequency analyses, or even replicate the current analyses with higher frequency bands. Second, we found the frontal cortex modulate the temporal-to-occipital prediction propagation. However, we did not explicitly answer whether this modulation is uniquely provided by frontal cortex or by any other regions like parietal cortex. To solve this, in the future analysis, setting a control region (e.g., parietal cortex) is necessary. Third, when computing MVPA applying bottom-up classifiers to top-down data, we binarized the time window into early and late to emphasize the pattern where early bottom-up better decodes late top-down and late bottom-up better decodes early top-down. We did this due to an inconsistent time window with 2ms step across participants – i.e., the results of different individuals staggered to some extent within the early vs. late time windows. However, the binarization with a low temporal resolution causes an information loss compared with a full matrix, where this loss should have benefited from MEG – the recording with a high temporal resolution. In the future, we need to dig more into the full matrix pattern.

3.4 Conclusions

We sought to understand how a network of brain regions coordinates to propagate specific predictions. We showed when, where, and how predictions of specific contents dynamically propagate through a network from temporal to occipital regions, mediated by frontal cortex. Our novel functional networks in principle generalize to other stimulus features and sensory modalities.

3.5 Methods

3.5.1 Participants

Eleven participants (18-35 years old, mean=26.8 years old, SD=3.0 years old) participated provided informed consent. All had normal or corrected-to-normal vision and reported no history of any psychological, psychiatric, or neurological condition that might affect visual or auditory function. The University of Glasgow College of Science and Engineering Ethics Committee approved the experiment.

3.5.2 Stimuli

We used 2 types of location cues (green dots) at Stage 1, 3 sweeping sounds as SF cues at Stage 2 and 2×2×3 types (location × SF × orientation) of Gabor patches at Stage 3, as the actual stimuli.

3.5.2.1 Location Cues

We presented the left (vs. right) 1 degree of visual angle diameter green dot at a left vs. right 2 degree of visual angle eccentricity. Participants sat at a 182 cm viewing distance from the screen.

3.5.2.2 SF Cues

The three sweeping sounds started with auditory frequency of either 196Hz, 2217Hz or 622Hz, a sweep rate of 0.5 rising octave/second, for a total duration of 250ms. Here, we used the sweeping sounds instead of pure tones to better catch participants' attention.

3.5.2.3 Gabor Stimuli

The left (vs. right)-presented Gabor patch were presented with the diameter of 7.5 visual degree and eccentricity of 12.5 visual degree (left vs. right from the centre), at a 182 cm viewing distance, with LSF (vs. HSF) contents of 0.5 cycle/degree (vs. 1.2 cycle/degree) and one of three randomly chosen orientations (-15 degree, 0 degree, +15 degree). Prior to the task, we calibrated the LSF and HSF Gabor contrast independently for each participant, using an adaptive staircase procedure with target accuracy set at 90%. On each calibration trial, a left (vs. right) green dot presented for 500ms predicted the upcoming left vs. right location of the LSF or (HSF) Gabor patch, itself presented for 100ms. Participant responded "LSF" vs. "HSF" vs. "don't know" without feedback. We adaptively adjusted LSF vs. HSF as follows:

$$Contrast = Contrast - 1*(Correct\ vs.\ Incorrect - target\ accuracy)/Shifting\ Count,$$

where *Shifting Count* counts the number of direction changes (i.e., increasing to decreasing, or decreasing to increasing). The adaptive staircase stopped when the

adjustment step was < 0.01, setting each SF contrast for the Gabor stimuli of this participant in the actual experiment (see Supplemental Table S3-5 provides each participant's contrast values).

3.5.3 Tasks

3.5.3.1 Cueing-Categorization

Each trial started with a central fixation presented for 500ms, followed by three stages (see Figure 3-1A for task procedure):

At Stage 1, a green dot was presented for 100ms, either left or right of the central fixation, predicting the left vs. right location of the upcoming Gabor with validity of 1. This was followed by a blank screen presented for 1000-1500ms with jitter.

At Stage 2, one of three sweeping sounds was presented for 250ms. The 196Hz (vs. 2217Hz) sound predicted the upcoming LSF (vs. HSF) Gabor (both with .9 validity). The 622Hz sound was a neutral cue that did not have any predictive value. It was followed by LSF vs. HSF Gabors with .5 probability, on 33% of trials).

At Stage 3, the LSF vs. HSF Gabor stimulus appeared at one of the three rotations on the left vs. right screen location for 100ms. The Gabor was either LSF or HSF, with one of three randomly chosen orientations, followed by a 750 to 1,250ms inter-trial interval (ITI) with jitter. We instructed participants to respond "LSF" vs. "HSF" vs. "Don't know" as quickly and as accurately as they possibly could. They did not receive feedback.

The experiment comprised blocks of 54 such trials (see Supplemental Table S3-6 for the repetitions on each type of stimuli). Participants performed 10-14 blocks in a single day, with short break between blocks. They completed the total of 38-45 blocks over 3-4 days. Participants completed at least 499 trials for each of the 4 (location × SF combination) Gabor patches presentation. Participants learned the correct relationships between the auditory cues and predicted Gabor SF without explicit instructions within two trial blocks. Thus, We removed the first two blocks from all subsequent analyses.

3.5.3.2 Auditory Localizer

Prior to the experiment, we ran an MEG localizer to model the bottom-up processing of each one of 3 auditory cues; For each cue, each localizer trial started with a blank screen for 500ms, followed by the auditory pitch for 250ms, then a blank screen for 1250ms ITI. In a block of trials, we repeated the same tone for 10 trials in a row. Participants passively listened to the tone but had to detect 2 tones with other SF. Each participant completed 36 such blocks (i.e., 12 blocks per type of tones), with block order of "low frequency", "middle frequency", "high frequency", repeated 12 times.

3.5.3.3 Gabor Localizer

Prior to the experiment, we also ran an MEG localizer to model the bottom-up processing of each one of 2×2×3 Gabor patches (i.e. location × SF × orientation; Figure 3-1C illustrates the localizer task procedure). For each Gabor, each localizer trial started with a 500ms central fixation, followed by the Gabor for 100ms, then a blank screen for 750ms ITI. In a block of trials, we repeated the same Gabor for 20 trials in a row. Participants passively viewed the Gabor but had to detect another Gabor (at the same location and orientation) with other SF. Each participant completed 36 such blocks (i.e., 3 blocks per type of Gabor), with block order of "left-LSF-clockwise", "left-LSF-vertical', "left-LSF-anticlockwise", "left-HSF-clockwise", "left-HSF-vertical', "left-HSF-anticlockwise", "right-LSF-clockwise", "right-LSF-vertical', "right-LSF-anticlockwise", "right-HSF-clockwise", "right-HSF-vertical', "right-HSF-anticlockwise", repeated three times.

3.5.4 MEG Data Acquisition and Pre-processing

We measured participants' MEG activity with a 248-magnetometer whole-head system (MAGNES 3600WH, 4-D Neuroimaging) at a 508Hz sampling rate. We performed the analysis with FieldTrip toolbox (Oostenveld, Fries, Maris, & Schoffelen, 2011) and in-house MATLAB code, according to the recommended guidelines.

3.5.4.1 Cueing-Categorization Task

For each participant, we epoched the raw data into trial windows, separately for each stage: Stage 1, -200ms pre-dot onset to 1,000ms post-dot onset (henceforth [-200ms 1,000ms] around onset); Stage 2: [-200ms 1,000ms] around sweeping sound onset; Stage

3: [-200ms 600ms] around Gabor patch onset. Then, we applied 1Hz high-pass filter (5th order two-pass Butterworth IIR filter) to the epoched data, removed the line noise using discrete Fourier transform and de-noised via a PCA projection of the reference channels. We rejected noisy channels with a visual selection and rejected jump and muscle artifacts with automatic detection. We decomposed the output dataset with ICA, identified and removed the independent components corresponding to artifacts (eye movements, heartbeat; 2-4 components per participant).

We then resampled the output data at 512 Hz, low-pass filtered the data at 25Hz (Zhan et al., 2019) (5th order Butterworth IIR filter), specified the time of interest between 0-500ms (post cue at Stage 2; Gabor stimulus at Stage 3), and performed the Linearly Constrained Minimum Variance Beamforming (LCMV) analysis to reconstruct the time series of sources on a 6mm uniform grid warped to standardized MNI coordinate space. Following the above steps, for each participant we obtained single-trial time series of 12,773 MEG sources at a 512Hz sampling rate between 0 and 500ms for Stages 2 and 3. We used these to analyse dynamics information processing in the brain (see Figure 3-1). Supplemental Table S3-7 reports the numbers of trials that remained after pre-processing and LCMV analysis.

3.5.4.2 Localizer Task

We applied the same pre-processing pipeline to the MEG localizer, using the epoched data [-200ms 500ms] around sound/Gabor patch onset. The LCMV analysis was then applied 0-500ms post sound/Gabor patch, to reconstruct the source representation of the MEG localizer data. Supplemental Table S3-7 summarizes the numbers of trials that remained after pre-processing and LCMV analysis.

3.5.5 Analysis

3.5.5.1 Cueing improves Behavior

At a group-level, we discarded invalid trials and applied a 2 (left vs. right location cues) × 2 (valid informative vs. neutral SF cues) × 2 (LSF vs. HSF Gabor patches) ANOVA on median RTs (excluding incorrect response and outliers) and accuracies (ACC) of valid trials of all participants. We then conducted a paired-sample t-test independently

for each of the 4 experimental conditions (left vs. right locations × low vs. high SFs), and then applied Bonferroni correction.

We analysed the RTs (excluding trials with incorrect responses and outlier trials) and ACC of each individual participant. In each participant, we replicated the significant main effect of SF cue on RTs (tested with the 2 × 2 × 2 ANOVA). In each participant, we also examined the effect of SF cue for experimental conditions (independent-sample t-tests). Supplemental Table S3-2 summarizes these ANOVA and independent-sample t-tests on RT statistics for each participant. Supplemental Table S3-3 summarizes the statistics of 2 × 2 × 2 ANOVA and chi-squared test for the accuracy data.

3.5.5.2 Prediction Network Nodes (supports Figure 3-3 and Supplemental Figure S3-1)

To understand the network of brain regions at Stage 2, that propagate the auditory prediction of LSF/HSF prior to stimulus onset at Stage 3, we computed the dynamics of cue representation across the whole brain, separately for left- and right-located trials.

For each participant, we computed the single-trial MI relationship <LSF vs. HSF auditory cue; Stage 2 MEG source activity>, at each time point from 0 to 400ms following auditory cue onset, on each source in occipital regions (lingual gyrus, cuneus, inferior occipital gyrus), temporal (fusiform gyrus, inferior temporal gyrus, middle temporal gyrus, superior temporal gyrus), parietal (superior parietal lobe, inferior parietal lobe, angular gyrus, supramarginal gyrus), premotor (precentral gyrus, postcentral gyrus), and frontal (orbitofrontal gyrus, inferior frontal gyrus, middle frontal gyrus, medial frontal gyrus, superior frontal gyrus). We then extracted the peak time of the MI for each source. To reveal the dynamics, we localized the source peaking around the first peak in the 90-120ms time window (start), the last peak in 120-200ms (midway) and >200ms (end). Figure 3A illustrates these three peaks in the Stage 2 time series of one typical participant. We computed the group mean of these 3 source localized peaks across participants (see Figure 3B).

3.5.5.3 Localiser Representation (supports Supplemental Figure S3-2)

To compare the pattern induced by the predictive cue and the bottom-up auditory input without predicted contents, we computed the dynamics of the tone representation using localiser data. For each participant, we computed the single-trial MI relationship <low frequency vs. high frequency tone, localiser MEG source activity>, at each time point from 0 to 400ms following tone onset, on each source in occipital, temporal, parietal, premotor and frontal regions. We see that after an explicit temporal activation, no clear stages or peaks occurred later (see Supplemental Figure S3-2).

3.5.5.4 Prediction Network Reconstruction (supports Figure 3-4 A-C)

To reconstruction the functional network that communicates at Stage 2 the predictive cue across the regions identified earlier (i.e., temporal, frontal and occipital), we computed Directed Feature Information (DFI, where F is the auditory cue information that predicts the upcoming LSF vs. HSF Gabor stimulus). In each participant, for each pair of regions (i.e., sender: temporal, receiver: frontal; sender: frontal, receiver: occipital), we computed DFI as follows:

Step 1: Source selection. We selected one source for the sending and one source for the receiving regions. These sources had highest MI(<LSF vs. HSF auditory cue; Stage 2 MEG source activity>) peak value during the time window of interest (temporal: 90-120ms, frontal: 120-200ms, occipital: >200ms). We computed the communications of the predictive cue between these three sources forming a functional network.

Step 2: Directed Information (DI). DI (i.e. Transfer Entropy) quantifies all the information communicated from the sending to the receiving source, removing the information sent from the receiver itself. For the receiver at time x, with a communication delay y from the sender, DI is calculated by Conditional Mutual Information (CMI) between RA_x and SA_{x-y} conditioned on RA_{x-y}:

$$\mathrm{DI} = \mathrm{CMI}\langle RA_x, SA_{x-y} | RA_{x-y} \rangle \quad (1)$$

Here, for each receiver time point between 0 and 400ms post auditory cue onset, and for each communication delay between 0 and 300ms, we computed DI between the sending and receiving sources. We thereby obtained the receiving-time × transfer-delay DI

matrix. The “DI” panel in Supplemental Figure S3-3 shows the computation of DI between the identified regions.

Step 3: DI conditioned on Feature (DI|F). DI|F removes from DI the information communicated about the feature (i.e., here the predictive LSF vs. HSF cue). For each receiving-time × communication-delay, we computed DI conditioned on the LSF vs. HSF predictions. The “DI|F” panel in Supplemental Figure S3-3 shows the computation of DI|F between the identified regions.

Step 4: DFI. The difference between DI and DFI therefore isolates the information communicated that is about the predictive cue. For each receiving-time × communication-delay cell of the matrix, we computed DFI as:

$$DFI = DI - DI|F \quad (2)$$

Figure 4 and the “DFI” panel in Supplemental Figure S3-3 show the computation of DFI between the brain regions.

Step 5: Statistical significance. Was established by recomputing DFI with shuffled feature labels (i.e., the LSF vs. HSF labels), 200 repetitions, using as statistical threshold the 95th percentile of the distribution of 200 maxima (each taken across the DFI matrix of each shuffled repetition, FWER, $p<0.05$, one-tailed).

We applied Step 1-5 to each participant and reconstructed the communication network. Supplemental Figure S3-4 shows the individual DFI result, and Figure 4B shows the group average results.

3.5.5.5 Prediction Network Modulations of Predictive Cue Communications (supports Figure 3-4D and Supplemental Figure S3-5)

We tested whether frontal cortex is a necessary mediator at Stage 2 to communicate the predictive cue from temporal to occipital cortex. We did so by isolating the role of the frontal region in the communication of the predictive cue from the temporal to the occipital regions. We could then compare network communications with and without frontal

mediation. The steps below detail how we computed such frontal mediation in the network of each participant.

Step 1: Frontal Mediation. On the selected temporal and occipital sources, for receiving time points between 0 and 400ms post auditory cue onset and for each delay between 0 and 300ms, we computed the receiving-time × communication-delay of temporal-to-frontal DFI and then frontal-to-occipital DFI (each computed as in *Methods, Network reconstruction*). This quantifies the full communication of predictive cues with frontal mediation (illustrated in Figure 4D.2).

Step 2: Direct Communication. To build the direct communication isolating frontal mediation, on the selected sources, we computed temporal-to-occipital DFI conditioned on the frontal activity. Specifically, at per time point in the combination of (1) receiving time *x* between 0 and 400ms post auditory cue onset (2) communication delay *y* between 0 and 300ms (3) mediation time *z* between receiving time and sending time (i.e., *x* and *x*-*y*), we computed DFI received by occipital at *x*, sent by temporal at *x*-*y*, conditioned on frontal activity at *z*.

In this way, we obtained the 3D receiving-time × communication-delay × mediation-time conditioned DFI matrix. We took the minimum conditioned DFI across the mediation time as the directed communication (i.e., without frontal mediation, Figure 4D.2).

Step 3: Statistical significance. Was established recomputing Steps 1 and 2 and their difference with shuffled feature labels (i.e., the LSF vs. HSF labels), 200 repetitions, using the 95th percentile of 200 maxima as statistical threshold (each maximum taken across the DFI minus DFI|F matrix of each shuffled repetition, FWER, $p<0.05$, one-tailed). This produced the receiving-time × communication-delay cells showing significant enhancement with frontal mediation compared with no frontal mediation. Red circles in Figure 4 denotes significant time points of this mediated communication.

We applied Step 1-3 to each participant. Supplemental Figure S3-5 shows the individual results and Figure 4D their prevalence.

3.5.5.6 Bottom-up Cross-validation (supports Supplementary Figure Figure S3-5)

To model the bottom-up representation of LSF vs. HSF Gabors, we trained decoding classifiers (Linear Discriminant Analysis, LDA) using the MEG localizer, separately for left and right localized Gabors. In each stimulus condition, we segmented the participant's trials into 5 folds based on stratified sampling and performed a 5-fold cross-validation.

In each iteration, we proceeded in 2 steps:

Step 1: we trained a LDA-classifier ((Treder, 2020), MVPA-Light toolbox in Matlab 2015b) to discriminate LSF vs. HSF Gabors, every 2ms between 0 and 500ms post stimulus, using the MEG sensor data from the trials of 4 folds as training set.

Step 2: We tested the trained LDA classifier at Step 1, every 2ms, with the left-out fold trials, leading to a decision value (i.e. "decoder value"), which was the inner product between the classifier weights and the held-out trial activity at this time point.

Following all 5 iterations, we proceeded to Step 3:

Step 3: To quantify decoding performance on a common scale at each time point, we calculated the Mutual Information (MI, (R. A. Ince et al., 2017)) across trials (concatenating test sets across folds) between this classifier decision value and the true "LSF" vs. "HSF" stimulus label. MI quantifies all the LSF vs. HSF information about the stimulus available from the classifier weights, without forcing a specific discrete prediction (Quiroga & Panzeri, 2009).

We repeated Step 1-3 10 times and averaged the resulting 10 MI matrices to quantify Gabor decoding over time. To establish statistical significance, we repeated the decoding procedure described 1,000 times with shuffled SF labels, applying threshold-free cluster enhancement (TFCE, E=0.5, H=0.5, (S. M. Smith & Nichols, 2009)) and using as statistical threshold the 95th percentile of 1,000 maximum values (each taken across the 256 times points per shuffle after TFCE) (i.e., FWER, $p<0.05$, one-tailed, (Nichols & Holmes, 2002)). We repeated the same cross-validation for each individual participant (see

Supplemental Figure S3-6 for significant cross-validation performance for each participant).

3.5.5.7 Top-down Cross-decoding (supports Figure 3-5 and Supplemental Figure S3-7)

To determine whether the top-down prediction at Stage 2 comprises SF contents, we used a cross-decoding approach. Specifically, we used the bottom-up SF classifiers trained with localizer data prior to the cueing task, to decode the SF contents at Stage 2, when the stimulus was not shown yet, but we expected the cues to reactivate the stimulus contents. We proceeded in 3 steps, separately for left and right localized stimuli:

Step 1: We trained the bottom-up classifiers every 2ms between 0 and 500ms post Gabor onset, using all MEG localizer trials in each location condition.

Step 2: We tested each bottom-up classifier on its classification of the MEG sensor data at Stage 2, at its specific time point post auditory cue, when prediction should reactivate the cued LSF vs. HSF into MEG source activity. On each trial, the outcome is a 2D (training time × testing time) decision value matrix. To quantify decoding performance, across trials we computed for each combination of training time and testing time the MI between single-trial classification decision value and the true stimulus label ("LSF" vs. "HSF" prediction). A permutation-based threshold (1,000 repetitions), using TFCE to correct for multiple comparison (FWER over 256 × 256 time points, one-tailed, $p < 0.05$) established statistical significance.

Step 3: We down sampled the decoding performance MI matrix (of dimensions training time × testing time) into a 2x2 matrix that captures the main temporal decoding characteristics. Along the training time axis, we set the early bottom-up time window (i.e., early bottom-up stage) as [147 ± 30ms, 232 ± 42ms] (mean ± standard deviation across participants) and the late bottom-up time window as [232 ± 42ms, 344 ± 48ms], both post localizer Gabor onset. Along testing time axis, we set the early top-down time window (i.e., early top-down stage) as [90ms, 160 ± 20ms] post Stage 2 auditory cue onset, and the late top-down time window as [160 ± 20ms, 400ms]. Note that while these windows were chosen by inspection to visualise the results, the inference corrected FWER over the entire time course at 2ms resolution and is not affected by this choice of bins for visualisation.

We averaged the time × time MI matrix (of statistically significant MI values) within the down-sampled time windows, to derive the early-late bottom-up × early-late top-down decoding performance. Finally, within the early and late bottom-up classifier (i.e., Y-axis), we normalized the 2-by-2 matrix across early and late top-down (i.e., X-axis), respectively.

We repeated the above Steps 1 to 3 for each participant to create their down-sampled 2x2 matrix of decoding performance (see Supplemental Figure S3-7 for each individual result). Figure 3-5 shows the average of these matrices across participants, separately for the predictions of a left and right located Gabor.

In addition, We reasoned that if top-down reversed bottom-up Gabor representations, a common pattern of neural activity would instantiate the LSF vs. HSF contents in both the bottom-up decoders prior to the experiment and in the top-down predictions at Stage 2. Note that this common pattern of neural activity cannot result from the evoked responses to the auditory cues, because these were only presented in the prediction experiment at Stage 2 and not in the localizer task used to train the decoder weights. That is, the decoder's weights could only learn from the visual LSF vs. HSF Gabor presentation, separately for left vs. right presentations.

3.5.5.8 Source Representation of Cross-decoding (supports Figure 3-5 and Supplemental Figure S3-7)

To visualize top-down reactivation of LSF and HSF Gabor representations at MEG source level, We reconstructed the occipito-temporal sources that cross-decode the Gabor predictions in 3 steps, separating left and right located trials.

Step 1: Using the Stage 2 single-trial decision matrix above (see *Methods, Top-Down Cross Validation Step 2*), We computed MI in each cell between the single-trial decision value and Stage 2 single-trial source activity at this specific time point. We confined this analysis to the occipito-temporal pathway involved with vision-for-categorization (Zhan et al., 2019), with regions of interest (ROIs) lingual gyrus, cuneus, inferior occipital gyrus, middle occipital gyrus and superior occipital gyrus and fusiform gyrus. The outcome of Step 1 was therefore a time × time matrix that represented the contribution of occipito-temporal sources to the decoding model.

Step 2: To consider both the decoding strength and localization strength of each source, we computed source representation as the dot product between the time × time decoding performance (after statistical test, see *Methods, Top-Down Cross-decoding Step 2*) and the time × time source localization (see *Step 1*).

Step 3: We further down sampled time × time representation for each source just as we did in *Methods, Top-Down Cross-decoding Step 3* to capture the source representation of the early-late bottom-up × early-late top-down decoding.

Step 4: We applied 2 (left vs. right-located prediction) * 2 (left vs. right hemisphere) ANOVA on the late top-down representation (decoded by early bottom-up) on occipital sources to test the interaction effect (i.e., the contra-lateral effect).

We repeated these four steps for each participant, creating per participant the down-sampled 2-by-2 matrix to show this source representation of early top-down decoded by late bottom-up and late top-down decoded by early bottom-up (see Supplemental Figure S3-7 for each individual result; and Supplemental Table S3-4 for each individual result). Figure 3-5 shows the group average source representation, separately for left and right located predictions.

3.6 Supplemental Material

3.6.1 Supplemental Methods

3.6.1.1 Prediction periods clustering

To compute the number of space x time periods of prediction representations, we applied k-means clustering analysis on all 4413 x 204 (source x time points) dimensional trials as follows:

Step 1: Peak time extraction. First, for each participant, and independently for left- and right-located trials and source, we extracted the peak time MI(<LSF vs. HSF auditory cue; Stage 2 MEG>), 0 and 400ms post auditory cue onset.

Step 2: Matrix computation. Across participants and condition, we counted the numbers of sources per ROI (occipital, temporal, parietal, pre-motor and frontal) that peak during each 10ms-step time window between 0 and 400ms post auditory cue onset (i.e. 39 time windows), producing a ROI x time matrix of MI peaks.

Step 3: Clustering. We k-means clustered (k = 1 to 30, repeating 1,000 times) the matrix from Step 2, using the 39 time windows as samples and selected k as the elbow of the within-cluster sums of point-to-centroid distances metric.

Supplemental Figure S3-8 shows Stage 2, with k = 4 as a good solution, starting with a period 0, before any prediction representation, and then 3 distinct timed periods with temporal, frontal, occipital of peak representations of the prediction.

3.6.2 Supplemental Tables

Supplemental Table S 3-1 Group-level effect of cueing on mean Gabor categorization RTs (for left vs. right located, LSF vs. HSF)

tested with paired samples t-test.

Gabor type	RT (predicted, ms)	RT (non-predicted, ms)	Improvement on RT (ms)	*t* value	*p* value
Left LSF	530.9	456.8	74.1	$t_{(10)} = 3.60$	$p = 0.005$
Left HSF	555.4	447.7	107.7	$t_{(10)} = 5.87$	$p = 0.0002$
Right LSF	556.3	483.6	72.6	$t_{(10)} = 3.37$	$p = 0.007$
Right HSF	525.9	429.5	96.4	$t_{(10)} = 4.82$	$p = 0.0007$

Supplemental Table S 3-2 Individual-level effect of predicted vs. non-predicted on categorization RTs

tested with ANOVA and independent samples t-test (before Bonferroni correction).

Participant ID	ANOVA *F*-value: Effect of valid informative vs. neutral cueing	Left LSF mean RT reduction	Right LSF mean RT reduction	Left HSF mean RT reduction	Right HSF mean RT reduction
1	$F(1, 2174) = 2765.5$, $p<0.001$	71.7ms, $p<0.001$	87.8ms, $p<0.001$	127.4ms, $p<0.001$	121.9ms, $p<0.001$
2	$F(1, 1968) = 90.1$, $p<0.001$	42.9ms, $p<0.001$	30.0ms, $p=0.002$	57.8ms, $p<0.001$	45.0ms, $p<0.001$
3	$F(1, 1888) = 133.5$, $p<0.001$	30.2ms, $p=0.008$	31.8ms, $p<0.001$	96.9ms, $p<0.001$	103.0ms, $p<0.001$
4	$F(1, 2147) = 75.1$, $p<0.001$	**-2.6ms, $p=0.71$**	**1.2ms, $p=0.87$**	65.2ms, $p<0.001$	59.1ms, $p<0.001$
5	$F(1, 1889) = 42.3$, $p<0.001$	**16.8ms, $p=0.12$**	**19.7ms, $p=0.04$**	44.6ms, $p<0.001$	38.5ms, $p<0.001$
6	$F(1, 1751) = 237.6$, $p<0.001$	139.3ms, $p<0.001$	115.2ms, $p<0.001$	102.2ms, $p<0.001$	136.9ms, $p<0.001$
7	$F(1, 1685) = 241.7$, $p<0.001$	81.5ms, $p<0.001$	111.5ms, $p<0.001$	127.6ms, $p<0.001$	93.8ms, $p<0.001$
8	$F(1, 1704) = 1887.1$, $p<0.001$	149.5ms, $p<0.001$	173.1ms, $p<0.001$	234.2ms, $p<0.001$	238.5ms, $p<0.001$
9	$F(1, 1806) = 28.0$, $p<0.001$	**26.2ms, $p=0.046$**	**18.2ms, $p=0.15$**	58.6ms, $p<0.001$	35.8ms, $p=0.007$
10	$F(1, 1510) = 31.0$, $p<0.001$	**15.1ms, $p=0.34$**	**55.9ms, $p=0.016$**	108.2ms, $p<0.001$	**27.5ms, $p=0.10$**
11	$F(1, 1786) = 2424.5$, $p<0.001$	176.3ms, $p<0.001$	162.9ms, $p<0.001$	139.9ms, $p<0.001$	135.1ms, $p<0.001$

Bold indicates **non-significant** result ($p>0.05$, Bonferroni-corrected).

Supplemental Table S 3-3 Individual-level effect of predicted vs. non-predicted on categorization accuracy

tested with ANOVA and chi-squared tests (before Bonferroni correction).

Participant ID	ANOVA *F*-value: Effect of valid informative vs. neutral cueing	Left LSF mean ACC increase	Right LSF mean RT increase	Left HSF mean RT increase	Right HSF mean RT increase
1	***F*(1, 2248) = 7.24, *p*=0.007**	0%	-0.26%, *p*=0.47	**1.7%, *p*=0.011**	1.1%, *p*=0.04
2	*F*(1, 2098) = 0.06, *p*=0.80	0.15%, *p*=0.85	1.5%, *p*=0.31	-2.3%, *p*=0.04	0.08%, *p*=0.94
3	***F*(1, 1998) = 10.2, *p*<0.001**	1.7%, *p*=0.19	**3.7%, *p*=0.006**	2.0%, *p*=0.23	1.6%, *p*=0.20
4	***F*(1, 2348) = 11.8, *p*<0.001**	-1.5%, *p*=0.37	2.6%, *p*=0.27	6.8%, *p*=0.014	**7.2%, *p*<0.001**
5	***F*(1, 2048) = 7.33, *p*=0.007**	-0.23%, *p*=0.91	**7.7%, *p*=0.006**	1.0%, *p*=0.59	**2.7%, *p*=0.011**
6	***F*(1, 1948) = 21.3, *p*<0.001**	4.0%, *p*=0.11	6.2%, *p*=0.013	**7.4%, *p*<0.001**	3.7%, *p*=0.08
7	***F*(1, 1848) = 14.2, *p*<0.001**	-1.1%, *p*=0.60	9.9%, *p*=0.003	**7.5%, *p*<0.001**	1.6%, *p*=0.08
8	***F*(1, 1898) = 13.0, *p*<0.001**	1.6%, *p*=0.31	2.2%, *p*=0.32	**10%, *p*<0.001**	2.1%, *p*=0.30
9	*F*(1, 1948) = 1.09 *p*=0.30	-1.5%, *p*=0.13	0.6%, *p*=0.71	1.7%, *p*<0.001	2.5%, *p*=0.007
10	*F*(1, 1798) = 0.96, *p*=0.32	-1.2%, *p*=0.34	0.83%, *p*=0.016	4.1%, *p*=0.44	-0.02%, *p*=0.08
11	***F*(1, 1898) = 36.6, *p*<0.001**	**3.9%, *p*<0.001**	**3.4%, =0.005**	**3.3%, *p*=0.008**	**2.8%, *p*=0.002**

Bold indicates **significant** ($p<0.05$ after Bonferroni-correction) result.

Supplemental Table S 3-4 Individual-level contra-lateral effect of source representation of LSF vs. HSF prediction

2 (left vs. right-predicted location) x 2 (left vs. right occipital sources), tested with ANOVA.

Participant ID	ANOVA *F*-value: Effect of predicted location	ANOVA *F*-value: Effect of hemisphere	ANOVA *F*-value: Effect of predicted location*hemisphere
1	$F(1, 1308) = 4.31$, $p=0.29$	$F(1, 1308) = 1.00$, $p=0.50$	**$F(1, 1308) = 565.7$, $p<0.001$**
2	$F(1, 1308) = 0.84$, $p=0.53$	$F(1, 1308) = 0.48$, $p=0.62$	**$F(1, 1308) = 181.7$, $p<0.001$**
3	$F(1, 1308) = 1.52$, $p=0.43$	$F(1, 1308) = 0.63$, $p=0.57$	**$F(1, 1308) = 71.4$, $p<0.001$**
4	NaN	NaN	NaN
5	$F(1, 1308) = 2.16$, $p=0.38$	$F(1, 1308) = 1.08$, $p=0.49$	**$F(1, 1308) = 77.9$, $p<0.001$**
6	$F(1, 1308) = 0.25$, $p=0.71$	$F(1, 1308) = 0.72$, $p=0.55$	**$F(1, 1308) = 166.5$, $p<0.001$**
7	NaN	NaN	NaN
8	$F(1, 1308) = 1.35$, $p=0.45$	$F(1, 1308) = 1.00$, $p=0.50$	**$F(1, 1308) = 169.9$, $p<0.001$**
9	$F(1, 1308) = 12902$, $p=0.005$	$F(1, 1308) = 172.8$, $p=0.05$	$F(1, 1308) = 0.05$, $p=0.82$
10	$F(1, 1308) = 2.51$, $p=0.36$	$F(1, 1308) = 1.33$, $p=0.46$	**$F(1, 1308) = 60.7$, $p<0.001$**
11	$F(1, 1308) = 17.6$, $p=0.15$	$F(1, 1308) = 2.5$, $p=0.36$	**$F(1, 1308) = 36.6$, $p<0.001$**

Bold indicates **significant interaction effect.**

Supplemental Table S 3-5 Individual-level contrast and accuracy of LSF vs. HSF Gabor patches in contrast threshold testing.

Participant ID	Left LSF Contrast (ACC)	Right LSF Contrast (ACC)
1	0.2277 (91.5%)	0.1752 (91.9%)
2	0.2152 (91.1%)	0.1752 (91.9%)
3	0.2021 (92.3%)	0.1720 (91.0%)
4	0.2215 (91.3%)	0.2102 (92.5%)
5	0.1633 (93.0%)	0.1815 (91.4%)
6	0.2019 (92.9%)	0.1852 (91.7%)
7	0.2219 (91.5%)	0.1952 (91.5%)
8	0.2367 (83.8%)	0.1967 (94.6%)
9	0.2300 (92.6%)	0.2152 (92.0%)
10	0.2356 (90.5%)	0.2252 (92.7%)
11	0.2152 (91.1%)	0.1668 (92.7%)

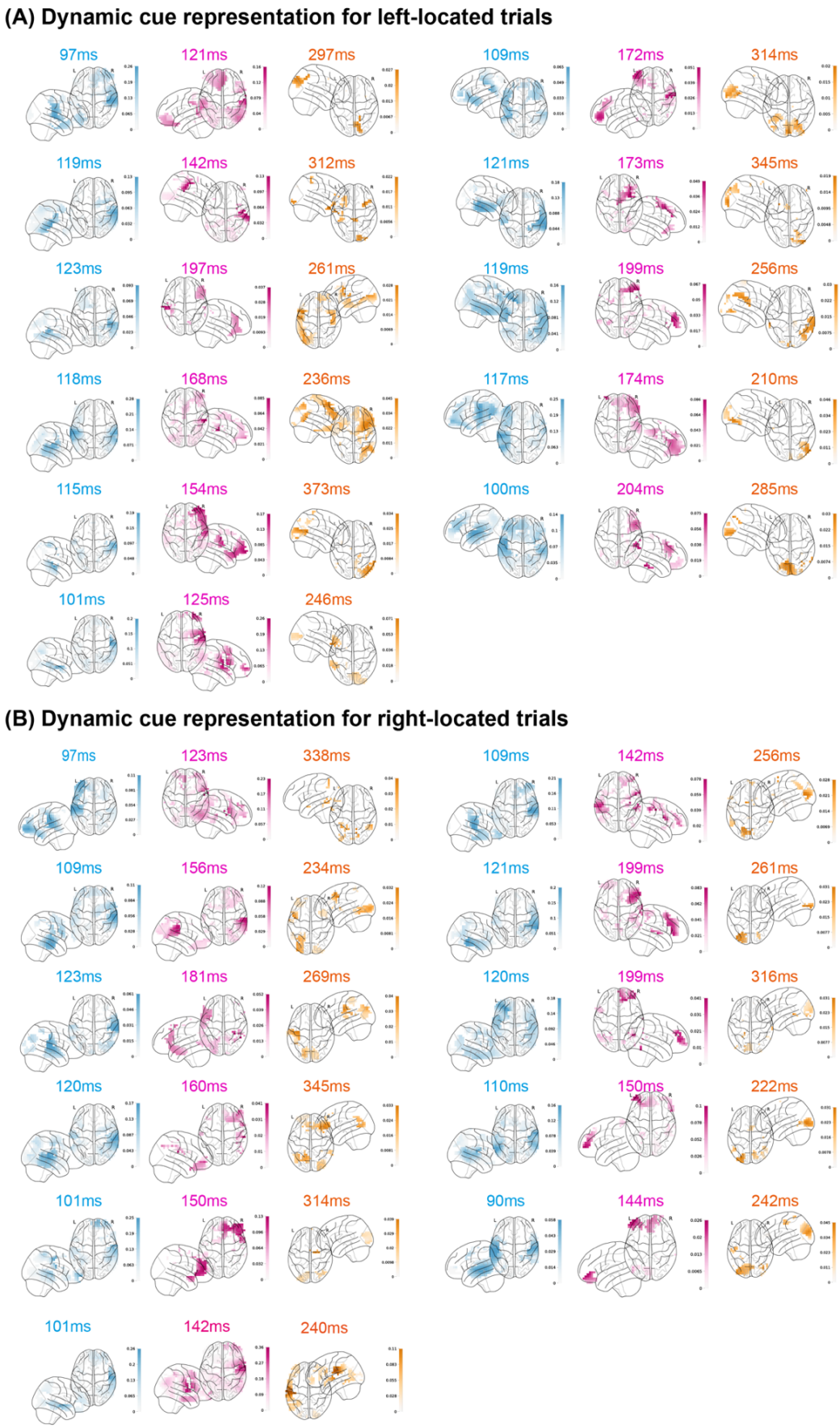
Supplemental Table S 3-6 Stimulus repetition in one cueing-categorization block

	Location cue	SF cue	Visual stimuli (random from 3 orientations)
Repetitions/type	27 left cues	9 LSF cues	8 left-LSF + 1 left-HSF
		9 HSF cues	8 left-HSF + 1 left-LSF
		9 neutral cues	9 left-random LSF/HSF
	27 right cues	9 LSF cues	8 right-LSF + 1 right-HSF
		9 HSF cues	8 right-HSF + 1 right-LSF
		9 neutral cues	9 right-random LSF/HSF
Sum	54		

Supplemental Table S 3-7 Trials remaining following pre-processing and LCMV analysis

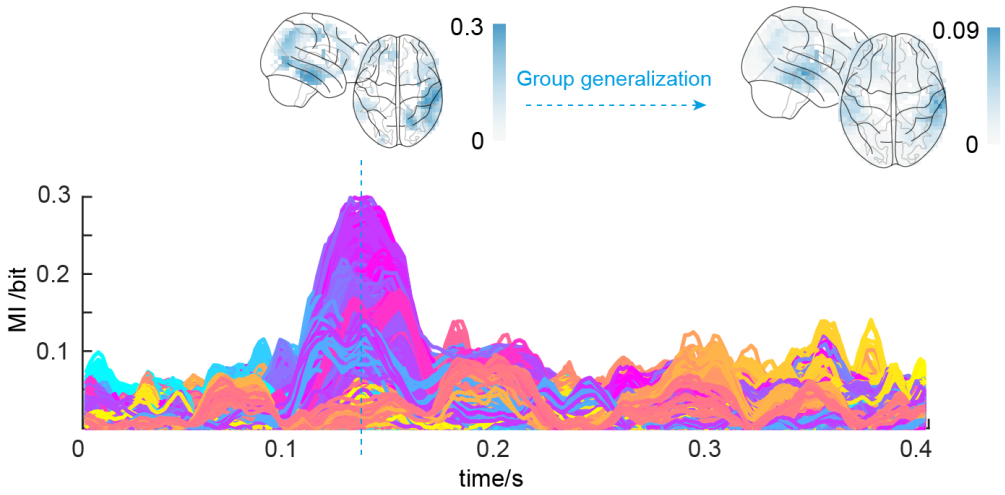
Stage	Mean trial number/participant	Range
Stage 1	1609	1492-1759
Stage 2	1604	1505-1751
Stage 3	1620	1507-1743
Auditory localizer	218	130-325
Gabor localizer	592	336-699

3.6.3 Supplemental Figures



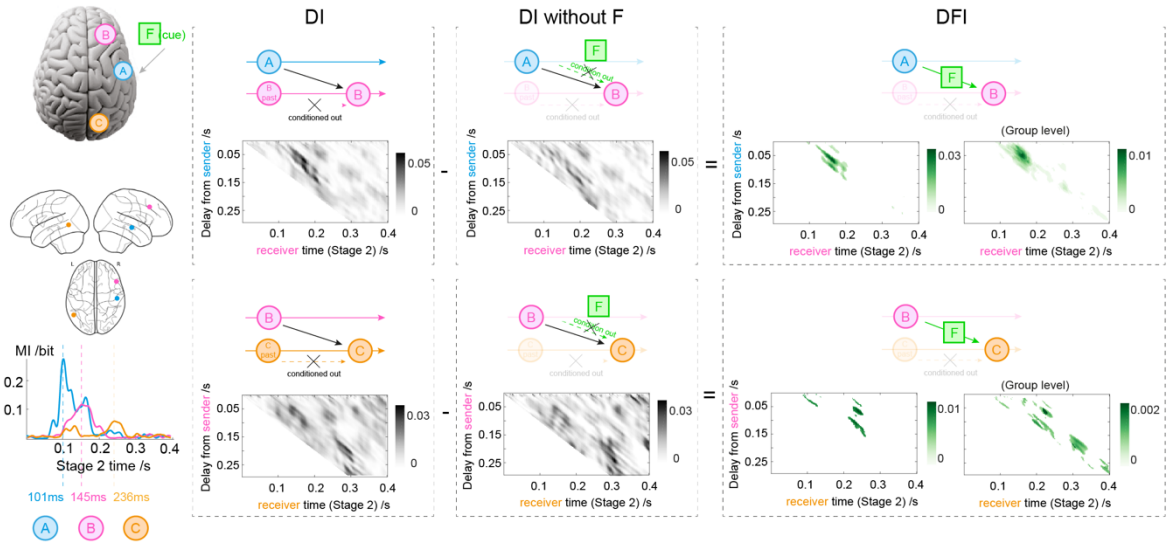
Supplemental Figure S 3-1 Individual prediction network node.

For each participant, we computed the single-trial MI relationship <LSF vs. HSF auditory cue; Stage 2 MEG source activity>, at each time point from 0 to 400ms following auditory cue onset, on each source in occipital regions (lingual gyrus, cuneus, inferior occipital gyrus), temporal (fusiform gyrus, inferior temporal gyrus, middle temporal gyrus, superior temporal gyrus), parietal (superior parietal lobe, inferior parietal lobe, angular gyrus, supramarginal gyrus), premotor (precentral gyrus, postcentral gyrus), and frontal (orbitofrontal gyrus, inferior frontal gyrus, middle frontal gyrus, medial frontal gyrus, superior frontal gyrus). We then extracted the peak time of the MI for each source. To reveal the dynamics, we localized the source peaking around the first peak in the 90-120ms time window (start, in blue), the last peak in 120-200ms (midway, in pink) and >200ms (end, in orange), separately for **left-located trials (A) and right-located trials (B)**. The specific time point of each pattern is denoted above the brain plot.



Supplemental Figure S 3-2 Dynamic sound representation across the brain at auditory localization.

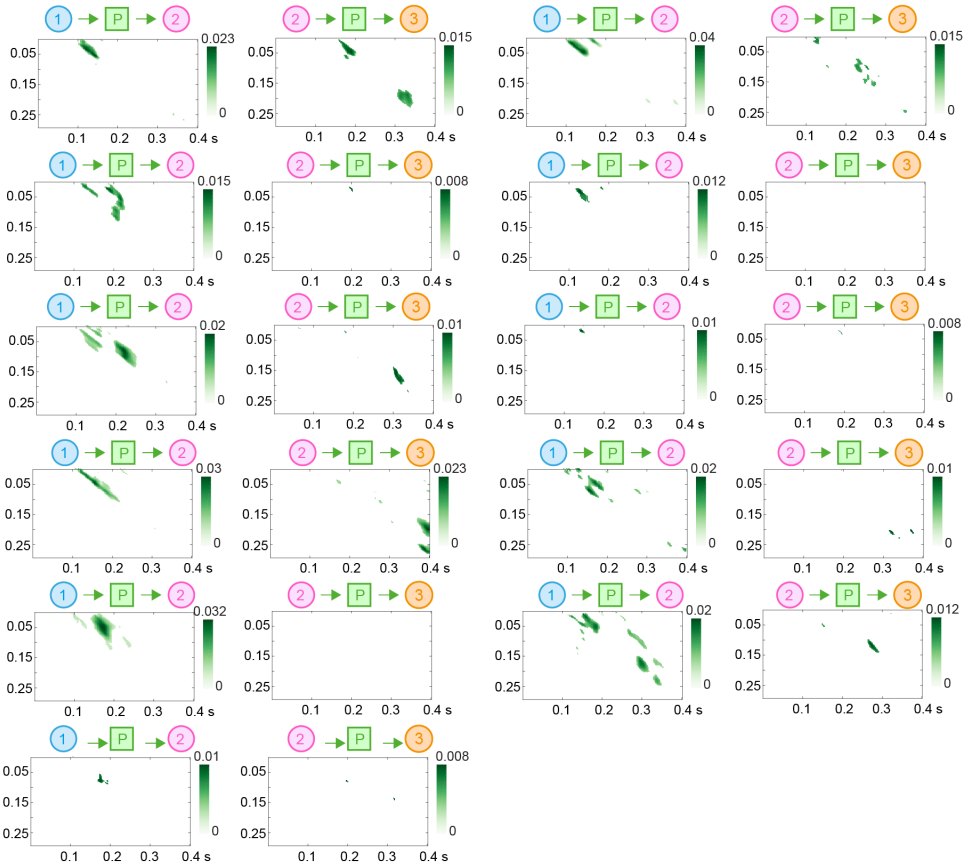
For a typical participant, we computed the representation of the prediction (as MI(<LSF vs. HSF cue; MEGt>, Y-axis), between 0 and 0.4s post sound (X-axis) in auditory localization, on each source in temporal (fusiform gyrus, inferior temporal gyrus, middle temporal gyrus, superior temporal gyrus), frontal (orbitofrontal gyrus, inferior frontal gyrus, middle frontal gyrus, medial frontal gyrus, superior frontal gyrus) and occipital regions (lingual gyrus, cuneus, inferior occipital gyrus). The curves show these per-source time courses of sound representation, color-coded by their (ranked) peak MI time. Glass brains below reveal for this participant the localized sources that peak in the only period—i.e. [122-142ms]. It shows that the representation of the pure sound without prediction only activates temporal lobe (cyan). The right-sided glass brains show the consistent result of group-level generalization.



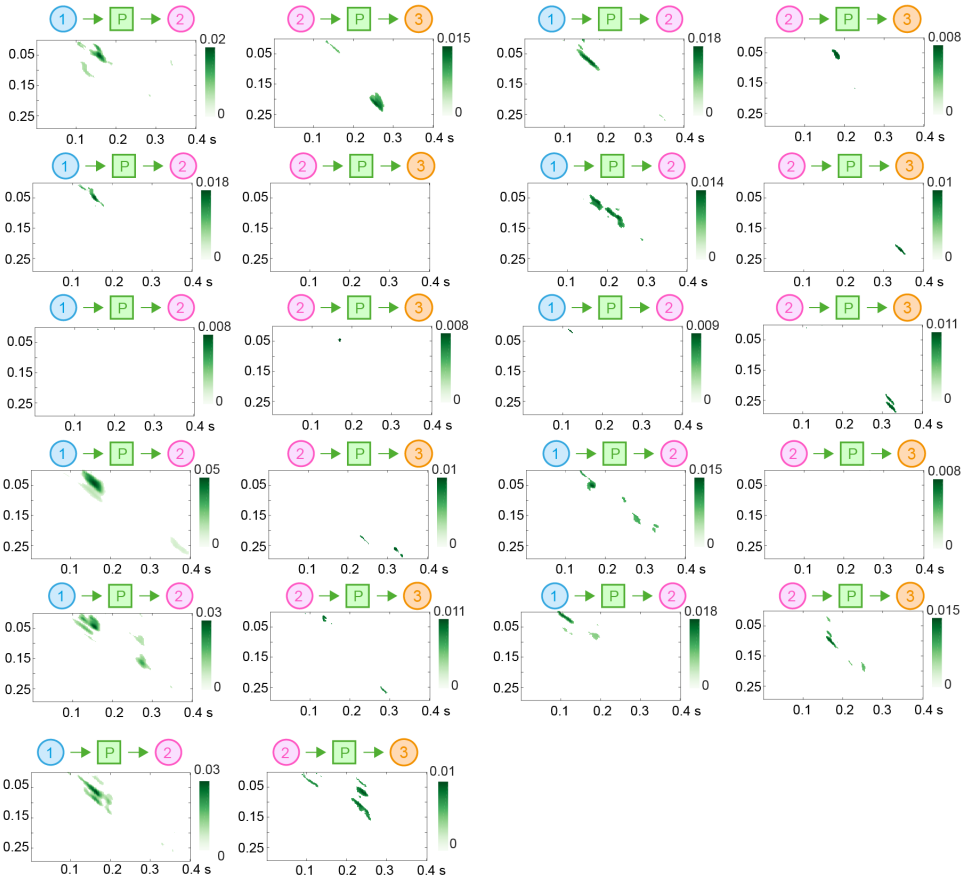
Supplemental Figure S 3-3 DFI computation.

Brain plots and curves During time bin 1, time bin 2 and 3, time bin 4 (cf. Figure 2), we selected three representative sources showing the highest MI (Stage 2 MEG; LSF vs. HSF cue) respectively in temporal lobe (top brain node A, blue), frontal cortex (top brain node B, pink) and occipital cortex (top brain node C, orange). The middle brain plots show the specific location of each source. The bottom curves show the time courses of MI (Stage 2 MEG; LSF vs. HSF cue) on the three sources with their peak time denoted. **DI; DI|F; DFI sketch maps and matrice plots** illustrate the way to obtain DFI transferred from the sender source to the receiver source (top panel: from node A, temporal to node B, frontal; bottom panel: from node B, frontal to node C, occipital). DI quantifies the total transfer of signal information from the sender node to the receiver node, conditioning out the early activity of the receiver. DI|F quantifies the transferred information conditioned on (i.e., independent from) the feature (i.e., LSF vs. HSF cue). DFI, as the difference between DI and DI|F, quantifying the transferred feature information. The matrices show the raw DI, raw DI|F from a typical participant and the significant DFI (including the result from the typical participant and the group mean) for each delayed time × receiving time, with X-axis representing Stage 2 time between 0 and 0.4s from auditory cue onset on the receiver source and Y-axis representing the delayed time between the sender source and the receiver source. The result indicates that A sends cueing information between 100 to 150ms to B which receives it between 140ms to 200ms – i.e., with a delay of 50ms to 100ms, and B sends cueing information between 140ms to 200ms to C which receives it between 200 to 250ms.

(A) DFI results for left-located trials



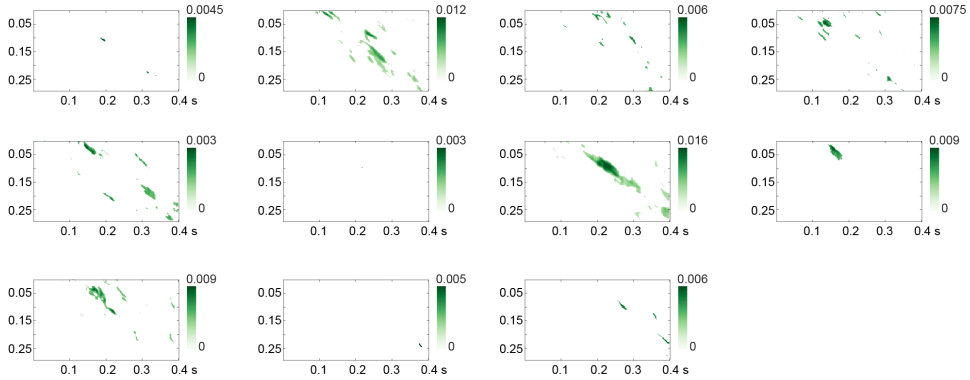
(B) DFI results for right-located trials



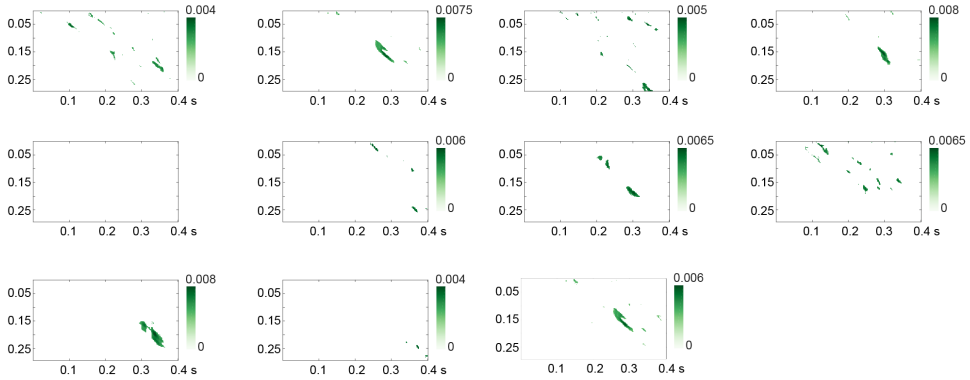
Supplemental Figure S 3-4 DFI individual results.

Using Directed Feature Information, we computed the participant communications across these network nodes (i.e. 1->2 and 2->3) for each individual participant, separately for **left-located trials (A) and right-located trials (B)**.These plots represent the significant (FWER-corrected, $p<0.05$) communications in the receiving node time course (X-axis), as communication delays from the sending node (Y-axis).

(A) Significant difference between direct communication and frontal mediation for left-located trials



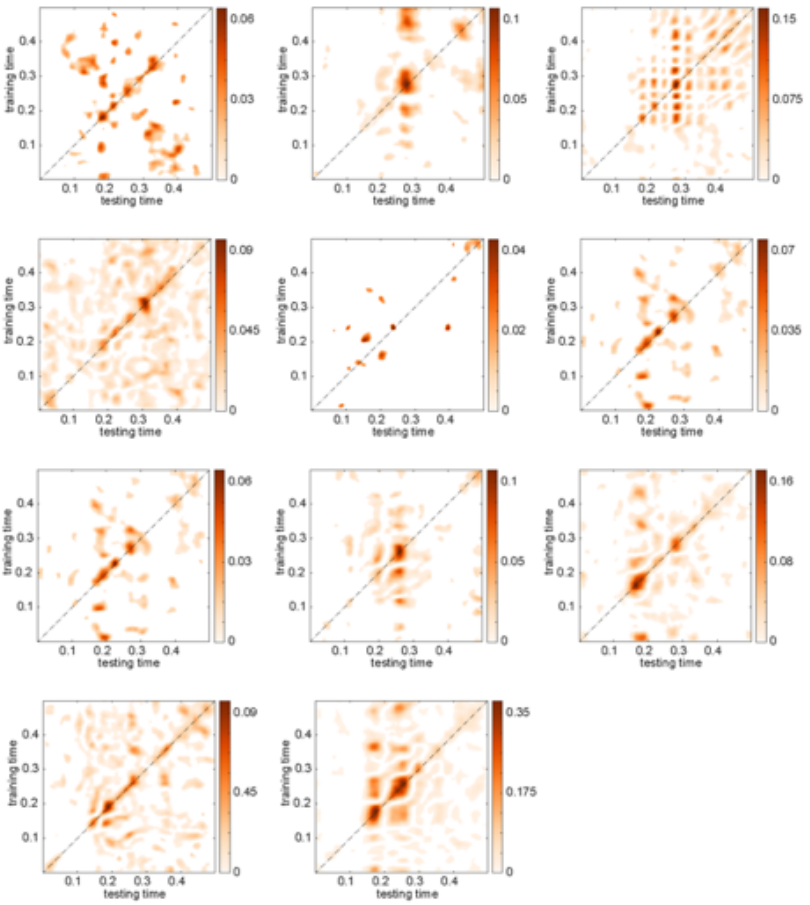
(B) Significant difference between direct communication and frontal mediation for right-located trials



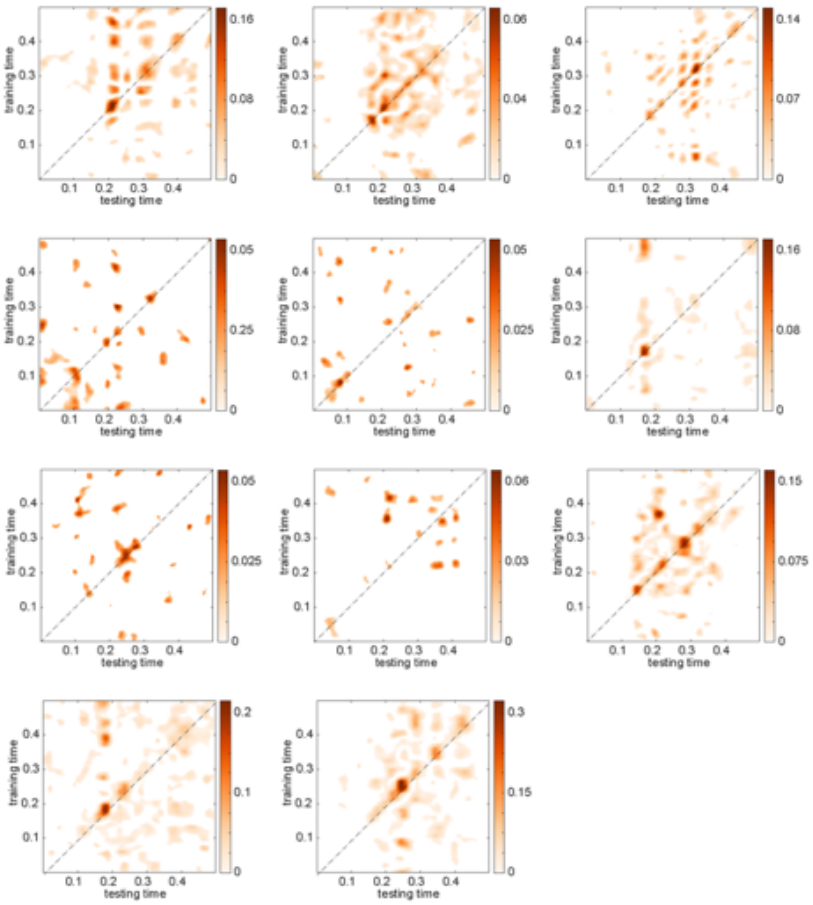
Supplemental Figure S 3-5 Individual mediation effect.

For each participant, We show here the significant difference (FWER-corrected, $p<0.05$, one-tailed) between Mediation of frontal cortex (source 2) to communicate the prediction from temporal (source 1) to occipital cortex (source 3, see Figure 4D.1) and Direct communication of the prediction from temporal (source 1) to occipital cortex (source 3, see Figure 4D.2). This computation conditions out the modulatory role of temporal cortex (source 2) in the prediction communication, separately **for left-located trials (A) and right-located trials (B)**. Frontal cortex mediates network communications of the prediction from temporal to occipital cortex.

(A) Bottom-up Cross-validation (right-located)

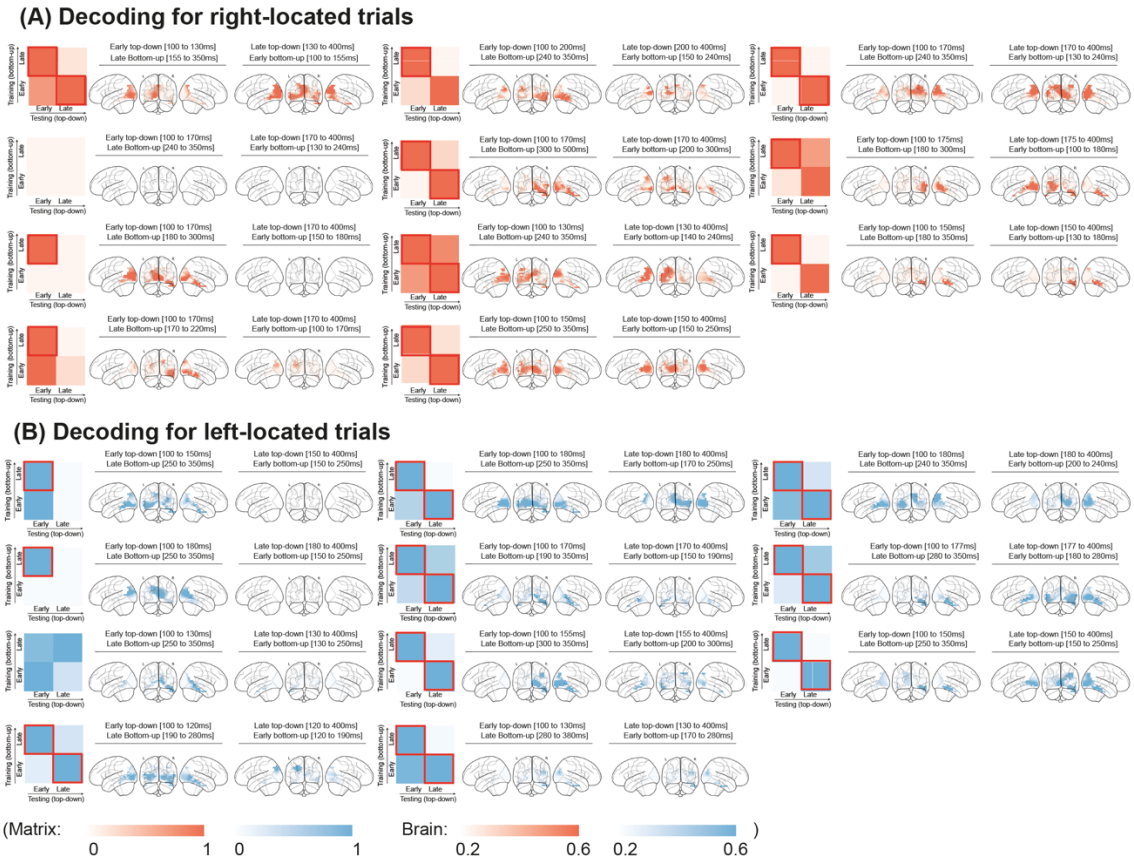


(B) Bottom-up Cross-validation (left-located)



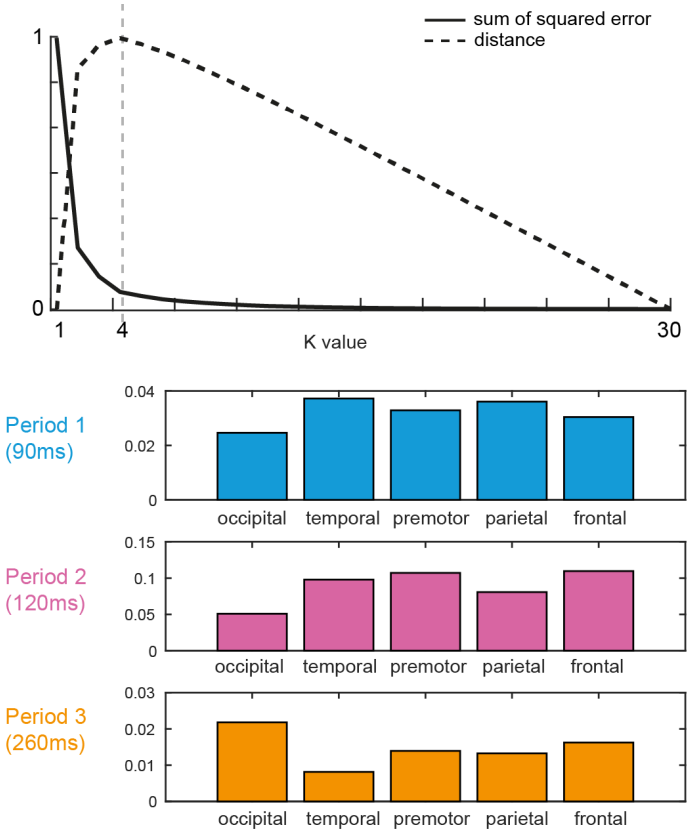
Supplemental Figure S 3-6 Bottom-up cross-validation performance at individual-level for (A) right-located and (B) left-located trials.

We computed and tested Gabor LSF vs. HSF classifiers on the localizer dataset by applying a 5-fold cross-validation with 10 repetitions and quantified the cross-validation performance by computing MI between decision values and true “LSF” vs. “HSF” labels. Each color-indexed matrix represents significant cross-validation performance (MI/bit, FWER, $p < 0.05$, after TFCE, one-tailed) in each participant, demonstrating that bottom-up SF classifiers decoded top-down predicted LSF vs. HSF contents.



Supplemental Figure S 3-7 Decoding performance and source representation at Individual-level for (A) right-located and (B) left-located trials.

Matrices. For right (color-coded in orange) and left (color-coded in blue) located Gabor trials, each matrix illustrates the decoding performance obtained in each participant. We used the bottom-up SF classifier (learned from MEG localizer) to decode the LSF vs. HSF contents of Stage 2 top-down predictions (MEG sensor data). We quantified decoding performance as MI(<decoder value; ground truth LSF vs. HSF>). We down-sampled the significant decoding performance (FWER, $p<0.05$, one-tailed) into a 2-by-2 matrix that averaged each MI matrix along the Y-axis that represents the bottom-up time window (i.e., early – late localizer; training dataset) and X-axis that represents the top-down time window (i.e., early – late prediction; testing dataset), with the specific time windows indicated above the brain plots. We normalized within each classifier (i.e., normalized across each row) to highlight that the early bottom-up classifier better decodes the late top-down responses, and the late bottom-up classifier better decodes the early top-down responses. Brain plots show the Stage 2 source representation of significant Gabor LSF vs. HSF decoding. We first calculated MI(<decoder value; Stage 2 MEG) along the occipito-ventral pathway (i.e. lingual gyrus, cuneus, inferior occipital gyrus, middle occipital gyrus, superior occipital gyrus and fusiform gyrus), and then computed source representation as the dot product between the time × time decoding performance and the time × time MI(decoder value; Stage 2 MEG). Similarly, we down-sampled the source representation with the early-late top-down × early-late bottom-up time window. We showed the source representation of early top-down decoded by late bottom-up (left-side brain plots) and late top-down decoded by early bottom-up (right-side brain plots), normalized across classifiers, revealing the top-down flow from right fusiform gyrus down to the primary occipital cortex contra-lateral to the predicted location of the upcoming Gabor.



Supplemental Figure S3-8. Clustering of representational time windows at Stage2 prediction representation.

K-means analysis, with k = 1..30, with sum of squared error of each k (plain curve) and distance of the plain curve from straightline (dashed curve). We selected k with furthest distance (elbow method). Following a baseline period (Period 0), the bar plot shows the percentage of sources peaking at the center time of each period, at occipital, temporal, premotor, parietal, frontal respectively. Stage 2 comprises 3 periods of predictive cue representations, dominated by temporal, frontal and occipital.

4 Study 3: Predictions enhance stimulus discrimination and modulate behavior in a categorization network

4.1 Introduction

Top-down modulations reflect brain mechanisms attune to the incoming stimulus information, facilitating its processing and subsequent behavior. For example, neural predictions could sharpen representations of the predicted stimulus contents, by enhancing a specific neural response (Kok, Rahnev, et al., 2012), while suppressing irrelevant contents (Lee & Mumford, 2003a). Conversely, neural predictions could dampen predicted contents (De Lange et al., 2018; Todorovic & de Lange, 2012), thereby enhancing selectivity to the unpredicted contents (Blank & Davis, 2016; Kumar et al., 2017). The key to addressing the mechanisms of prediction is therefore first to understand where, when and how networks of specific brain regions coordinate to predict a specific content from memory, and then understand how predictions in turn facilitate the processing of the upcoming stimulus and modulate behavior.

Having reconstructed at Stage 2 the temporal-frontal-occipital functional network propagating prediction contra-laterally to the predicted location, we now turn to another two questions related to Stage 3 (cf. Figure 3-1) that:

(1) When, where, and how does a stimulus feature dynamically propagate through a Categorization Network of brain regions? This is to reconstruct when shown at Stage 3, a functional network that categorizes the stimulus.

(2) How does the prediction interact with stimulus representation, finally modulating behavior? This is important to understand the specific effects of valid predictions (i.e., valid informative cueing, as opposed to non-predictive) on early visual processing and later decision making.

This study is still based on the same experiment as Study 2, while focusing more on Stage 3 about the facilitation provided by prediction on visual categorization. We will briefly recap the three-stage cueing design of the experiment which have been depicted in detail in Figure 3-1 (see also *Chapter 3 Methods, Cue-categorization*). To recap, on each trial, a spatial cue at Stage 1 (a green dot briefly displayed left vs. right of a central fixation

cross, cf. Posner cueing (Posner & Petersen, 1990)) predicted the visual hemifield location (left vs. right) of an upcoming Gabor patch (henceforth, Gabor) with 100% validity, followed by a 1-1.5s blank screen. At Stage 2, on informative trials, an auditory cue (a 250ms sweeping tone at 196 Hz vs. 2217 Hz) predicted the Spatial Frequency content (SF, Low vs. High) of the upcoming Gabor stimulus with a 90% validity (henceforth, referred as the "predicted" trials), followed by another 1-1.5s blank interval. In uninformative, "non-predicted" trials (33% of Stage 2 trials), a 622 Hz neutral auditory cue had no association with LSF or HSF. Finally, at Stage 3, the actual Gabor stimulus appeared in the participant's left vs. right visual hemifield for 100ms. Each participant (N = 11, see *Chapter 3 Methods, Participants*) categorized the Gabor SF as quickly and accurately as they possibly could without feedback (i.e. 3-AFC, with responses "LSF" vs. "HSF" vs. "do not know"). We concurrently recorded the participant's dynamic brain activity with MEG and reconstructed it on 12,773 sources, see *Chapter 3 Methods, MEG Data Acquisition and Pre-processing*.

4.2 Results

4.2.1 Categorization Network Nodes

We have reconstructed a functional network that communicates a prediction of visual contents, at Stage 2 before the stimulus is shown, from auditory cortex to occipital cortex, with frontal mediation. Here we show that these predicted contents then modulate stimulus representation when shown at Stage 3, in the functional network that categorizes the stimulus.

To do so, we first computed the dynamic representation of the Gabor stimulus at Stage 3—i.e. as the Mutual Information (MI) between <LSF vs. HSF Gabor; Stage 3 MEG source activity>, on each source and time point (see *Methods, Categorization Network Nodes*). Figure 4 presents these representational data for each source. Stimulus representation reveals again three sequential periods illustrated in a typical participant (Figure 4-1 A), starting with an early ([150-250ms], orange, Period 4) occipital representational peak contra-lateral to the presented Gabor stimulus, followed by a parietal lobe peak ([250-350ms], red, Period 5) and a pre-motor cortex peak ([> 350ms], brown, Period 6). Figure 4-1 B generalizes this three-period representational dynamics to the group, locating in each participant (N = 11) the source that maximally represent the

stimulus (i.e. peak MI), in each time window, separately for left vs. right located trials (see Supplemental Figure S4-1 all individual results, and besides, all participants show a simultaneous or slightly later ventral pathway activation in Period 4.

In a second step, we reconstructed with DFI the functional network of regions that communicates the stimulus contents over time (see *Methods, Categorization Network Reconstruction*). Figure 4-1 C and 4-1 D show that Gabor LSF vs. HSF communications proceed sequentially from contra-lateral occipital cortex to parietal cortex and then on from parietal to pre-motor cortex at the average group level. We replicated this functional network independently, for left- (dashed lines, Figure 6) and right-located trials, in ≥ 9/11 participants—i.e. Bayesian population prevalence = 0.81 [0.53 0.96], MAP [95% HPDI], individual results in Supplemental Figure S4-2.

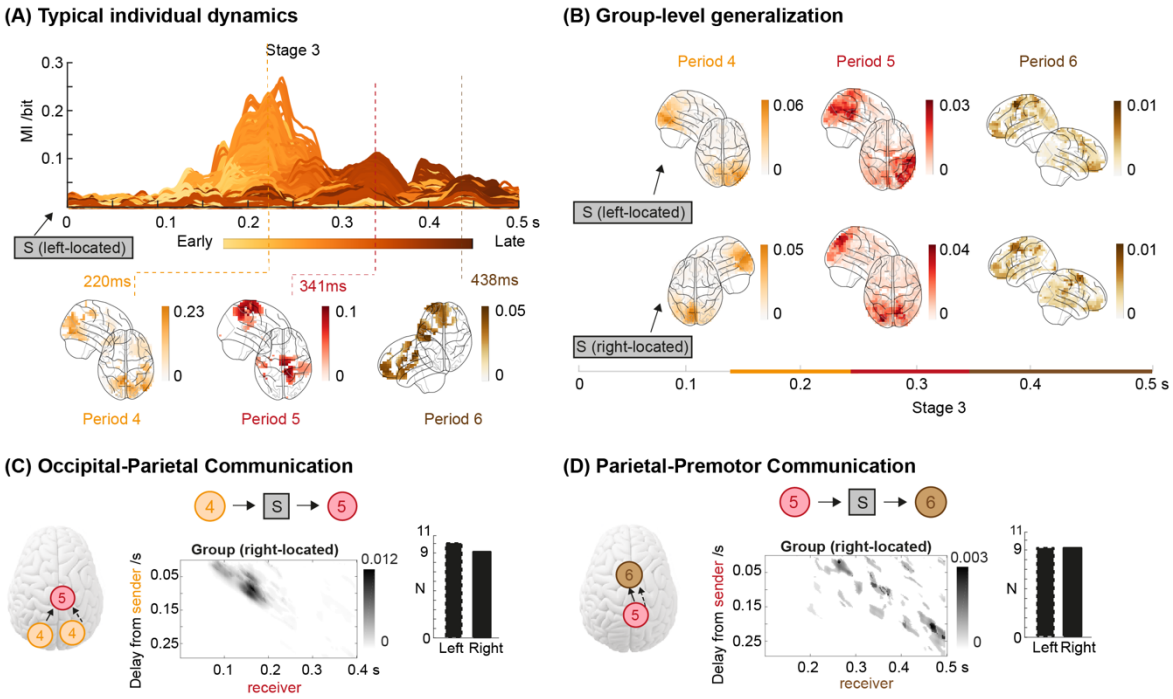


Figure 4-1 Dynamic communications of stimulus in the categorization network at Stage 3.

(A) Typical individual dynamics. For a typical participant with right located Gabor, we computed the representation of the stimulus (as MI(<LSF vs. HSF Gabor; MEGt>, Y-axis), between 0 and 0.5s post Gabor (X-axis), on each temporal, frontal and occipital source. The curves show these per-source Gabor representation time courses, color-coded by their (ranked) peak MI time. Glass brains below reveal for this participant the sources that maximally represent the Gabor in each period—i.e. [210-230ms], [330-350ms], [428-448ms], showing that LSF vs. HSF representation starts in contra-lateral occipital cortex (orange), moving to parietal regions (red), and finally to premotor and frontal cortex (brown). The schematic brain illustrates the successive stages of stimulus representation.
(B) Group-level generalization. We generalized the temporal sequence post Gabor

presentation across participants, using [150-250 ms], [250-350 ms], [>350 ms] time windows color-coded on the X-axis, independently for left- vs. right-located trials. Glass brain shows the cross-participant mean of the sources that maximally represented the stimulus in each time window, revealing the occipital-parietal-premotor stimulus representation sequence for stimulus categorization. **(C) Occipital-parietal communication and (D) Parietal-premotor communication.** We computed the communications across Stage 3 network nodes (i.e. 4->5 and 5->6). Both plots represent for right-located trials these group-average communications in the receiving node time course (X-axis), as communication delays from the sending node (Y-axis). For example, we can see that node 4 sends the stimulus S to node 5, with a 75-150 ms delay. We replicated 4->5 for left-located trials (dashed) in 10/11 participants, Bayesian population prevalence = 0.91 [0.64 0.99], MAP [95% HPDI], and right-located trials in 9/11 participants, Bayesian population prevalence = 0.81 [0.53 0.96], MAP [95% HPDI]. 5->6 was replicated for left-located trials (dashed) and right-located trials in 9/11 participants, Bayesian population prevalence = 0.81 [0.53 0.96], MAP [95% HPDI].

4.2.2 Influences of Stage 2 Prediction on Stage 3 Stimulus Representation

We have reconstructed the Stage 2 functional network that predicts the upcoming LSF vs. HSF stimulus contents, and the Stage 3 functional network that categorizes the LSF vs. HSF of the left and right presented Gabor stimulus. Now, we turn to understand the mechanisms whereby how Stage 2 valid predictions influence Stage 3 stimulus categorizations. Specifically, we address three key questions:

(1) Does Stage 2 occipital node prediction preview (i.e. replay) Stage 3 physical stimulus representation?
(2) Does valid prediction change Stage 3 brain activity to speed up behavior?
(3) Does Stage 2 prediction sharpen Stage 3 stimulus representation?

4.2.2.1 Does Stage 2 occipital node prediction preview (i.e., replay) Stage 3 physical stimulus representation?

Here, we show that valid Stage 2 cueing LSF vs. HSF improves LSF vs. HSF representation of the Gabor stimulus when it is shown at Stage 3. To do so, we proceed in three aspects. In summary, first, we validated the representational overlap between the predicted SF contents (Stage 2) and the real SF feature (Stage 3); secondly, we compared the SF representation between predicted condition and unpredicted condition; thirdly, within predicted condition, considering trial-by-trial variance, we related the single-trial

prediction strength with the single-trial representation strength of the visual SF feature. With these three steps, we answered the question by "sharpen". The results are depicted as following.

To address this, we selected per participant the occipital source with maximal MI representation of the Stage 2 prediction, and that with maximal MI representation of the Stage 3 Gabor. Then, we computed how much the representation of LSF vs. HSF prediction (Stage 2, Period 3) on source 1 overlaps with LSF vs. HSF Gabor stimulus representation on source 2 (Stage 3, Period 4)—i.e. using Co-Information(<LSF vs. HSF; Stage 2 occipital MEG1; Stage 3 occipital MEG2>), FWER-corrected, $p<0.05$, see *Methods, Prediction overlaps with stimulus representation* (Bell, 2003; McGill, 1954). Orange curves in Figure 4-2A plots the cross-participant average prediction-to-categorization representational overlap magnitude in occipital cortex, with a peak ~200ms post-stimulus—independently replicated for left (dashed lines) and right-located trials, in 9/11 participants, Bayesian population prevalence = 0.81 [0.53 0.96], MAP [95% HPDI], individual results in Supplemental Figure S4-3 A-B. Hence, we can conclude that Stage 2 predictions in occipital cortex preview/replay representation of the physical stimulus at Stage 3.

Besides, to more strongly conclude the representation overlap not limited to the single source, we searchlight-computed this per-trial representational overlap (i.e., Co-Information (<LSF vs. HSF; Stage 2 MEG sources; Stage 3 MEG sources>), using rFG and contra-lateral OCC sources at Stage 2 and the sources of 6 main ROIs of the occipito-ventral pathway primarily involved with vision-for-categorization (Grill-Spector & Weiner, 2014) (spanning contra-lateral occipital cortex, right fusiform gyrus, temporal lobe, parietal lobe, frontal lobe and pre-motor cortex) at Stage 3. We found that Gabor SF representations overlapped mostly on contra-lateral occipital sources between Stages 2 (when LSF vs. HSF is predicted) and 3 (when LSF vs. HSF is shown). Supplemental Methods presents the analysis and Supplemental Figure S4-5 presents this result.

4.2.2.2 Does Stage 2 prediction change Stage 3 brain activity to speed up RT?

Here, we computed in each participant the representational overlap of predicted vs. non predicted trials between Stage 3 pre-motor cortex source activity and behavioral

RTs—i.e. using Co-I(<predicted vs. non-predicted; Stage 3 MEG; RT>), selecting the pre-motor cortex source with highest MI(<predicted vs. non-predicted; Stage 3 MEG>), separately for left vs. right-presented trials, FWER-corrected, $p<0.05$, see *Methods, Prediction modulates Source Activity and RT*. The brown curve in Figure 4-2A plots this representational overlap averaged across participants, with a peak ~400ms post Gabor for left- (dashed) and right-located trials—replicated in all participants for left- (dashed) and right-located trials, Bayesian population prevalence = 1 [0.77 1] MAP [95% HPDI], see individual results in Supplemental Figure S4-3 C-D. Thus, Stage 2 prediction does indeed change Stage 3 pre-motor cortex neural activity to speed up RTs.

Similarly, we searchlight-computed the overlap of representation between predicted vs. non-predicted trials at Stage 3 neural activity of the sources of the 6 ROIs used earlier and behavioural RTs, separately for left vs. right-presented trials (FWER-corrected, $p<0.05$). We found that this RT modulation is mostly on premotor sources. Supplemental Figure S4-5 presents this result.

4.2.2.3 Does Stage 2 predictions sharpen Stage stimulus representation?

Having shown that Stage 2 prediction previews Stage 3 stimulus representation, leading to faster RTs, we test the mechanistic hypothesis that valid Stage 2 predictions sharpen Stage 3 stimulus representation, throughout the categorization network. Remember that Stage 2 comprises both informative and neutral trials (cf. Figure 3-1B, “no prediction”). In each participant, we can therefore compare how prediction vs. no prediction changes source representation of the same Gabor LSF vs. HSF stimulus—computed as MI (<LSF vs. HSF; Stage 3 MEG>), see *Methods, Prediction enhances Stimulus Representation*. Figure 4-2B shows Stage 2 prediction did indeed significantly sharpen Stage 3 discrimination of the LSF vs. HSF Gabors in contra-lateral occipital, parietal, and premotor cortices over the temporal sequence of the categorization process (FWER, $p<0.05$, two-tailed).

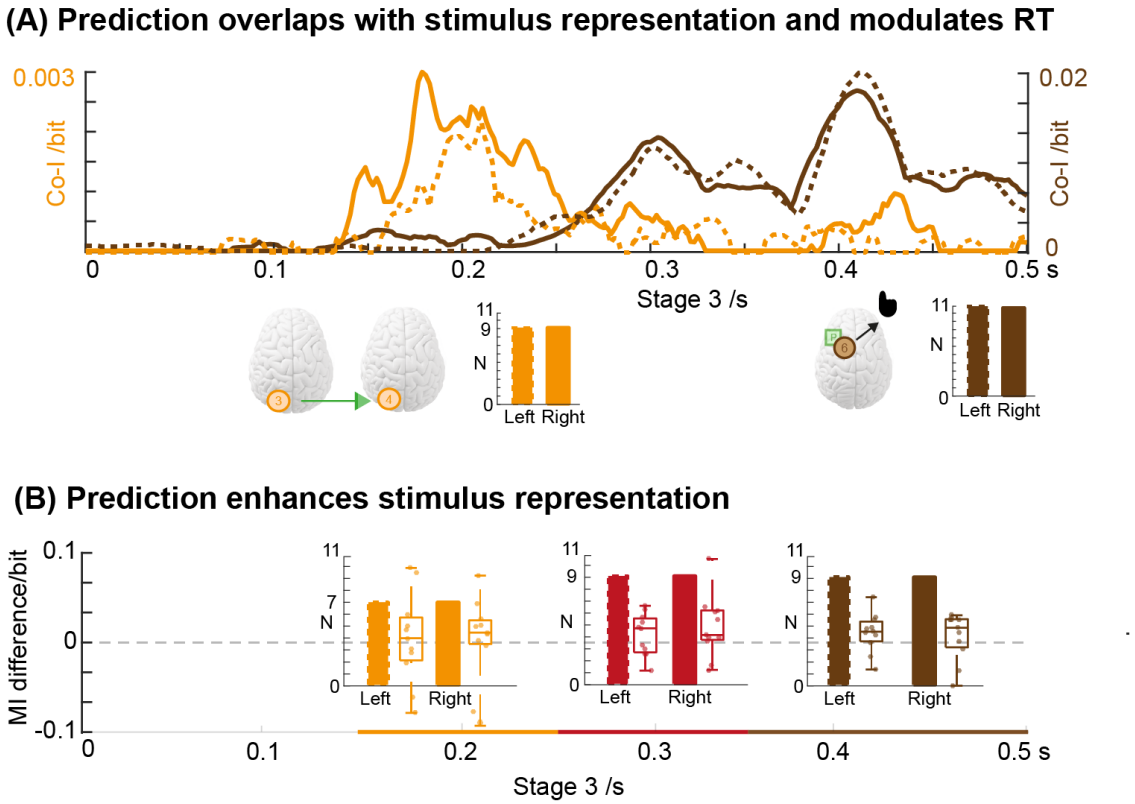


Figure 4-2 Stage 2 prediction influences Stage 3 categorization.

(A) Prediction previews physical stimulus representation and modulates RT. Orange curves. Prediction-categorization overlap. We computed Co-I(<LSF vs. HSF; Stage 2 MEG1; Stage 3 MEG2>), selecting two occipital sources per participant: that which maximally represents (MI) the LSF vs. HSF prediction at Stage 2; that which maximally represent (MI) the LSF vs. HSF Gabor stimulus at Stage 3. The plot presents the cross-participants average Co-I along Stage 3 occipital source time (X-axis), separately for left-located (dashed) and right-located (solid) trials, peaking ~200ms post Gabor onset, replicated for left-located trials (dashed) and right-located trials in 9/11 participants, Bayesian population prevalence = 0.81 [0.53 0.96], MAP [95% HPDI]. Brown curves. Prediction modulates premotor activity and RT. To test that prediction modulates both premotor cortex activity and RT, we computed Co-I(<Predicted vs. Non-predicted trials; Stage 3 premotor MEG; RT>), using the source with highest MI(<Predicted vs. Non-predicted trials; Stage 3 premotor MEG>). The plot represents the group-average positive Co-I along Stage 3 premotor source time (X-axis), and separately for left- (dashed) and right-located (solid) trials, with peaks ~400ms post Gabor onset, replicated individually in all 11 participants, Bayesian population prevalence = 1 [0.77 1] MAP [95% HPDI]. **(B) Prediction sharpens LSF vs. HSF Gabor discrimination across the categorization network**. We compared representations of the same Gabor LSF vs. HSF discrimination–computed with MI(<LSF vs. HSF Gabor; Stage 3 MEG>)–in predicted vs. non-predicted trials, separately for left- and right-located. On each of occipital-ventral, parietal and pre-motor cortex source, valid predictions lead to significantly stronger representations of the LSF vs. HSF discrimination (i.e. peak MI, FWER, p<0.05, two-tailed) in the critical time window of the sequence (i.e., contra-lateral occipital-ventral: 150-250ms; parietal: 250-250ms; pre-motor: >350ms). Boxplots show peak MI difference between valid prediction vs. without prediction against the null hypothesis of no difference. Scatters show individual

results. Sharpening is replicated for left- (left boxplot and prevalence bar) and right-located trials (in right boxplot and prevalence bar), in 7/11 participants in contra-lateral occipital cortex, Bayesian population prevalence = 0.64 [0.33 0.85] (MAP [95% HPDI]), in 9/11 participants in parietal lobe and premotor cortex, Bayesian population prevalence = 0.81 [0.53 0.96], MAP [95% HPDI].

4.2.3 Stronger Predictions lead to Stronger Discrimination in Occipital Cortex

Having proved that valid predictions, compared with non-prediction, can enhance the initial representation of the predicted contents, in turn speed up RT. We then turn to understand this effect within the predictive condition, considering trial-by-trial variance. In Study 2, with two predictive cues (i.e., LSF and HSF cues), we have reconstructed a functional network that communicates the prediction of visual contents, at Stage 2 before the stimulus is shown, from auditory cortex to occipital cortex, with frontal mediation. Here, we further proved the facilitation effect of the two informative cues at single trial level (cf. Figure 3-1).

To do so, in each participant we compared the strength of the prediction representation on the occipital sources at the end of Stage 2 with the representation of the stimulus when shown at Stage 3. Specifically, we isolated the contribution of each trial to the overall representation of the predictive cue on each receiving occipital network source at the end of Stage 2 (computed with sample-wise MI, SMI, see *Methods, Stronger Prediction leads to Stronger Categorization*). We reiterated the same single-trial decomposition, this time following Stage 3 stimulus onset, when LSF vs. HSF is represented on the same source. In each participant, we can therefore compare (trial by trial, with Pearson correlation) how Stage 2 strength of prediction on the receiving source (when the stimulus is absent) coincides with Stage 3 strength of stimulus representation on the same source (when the stimulus is shown). Figure 4-3 shows the group mean of significant correlation coefficients across participants (no-correction, $p<0.05$), thus localizes the occipital sources where this correspondence is significant. Prevalence bar shows that for left-located (dashed line) and right-located trials, this enhancement effect is both replicated in all 11 participants (i.e. Bayesian population prevalence = 1 [0.77 1], MAP [95% HPDI]. However, this effect does not remain after FWER multiple comparison correction.

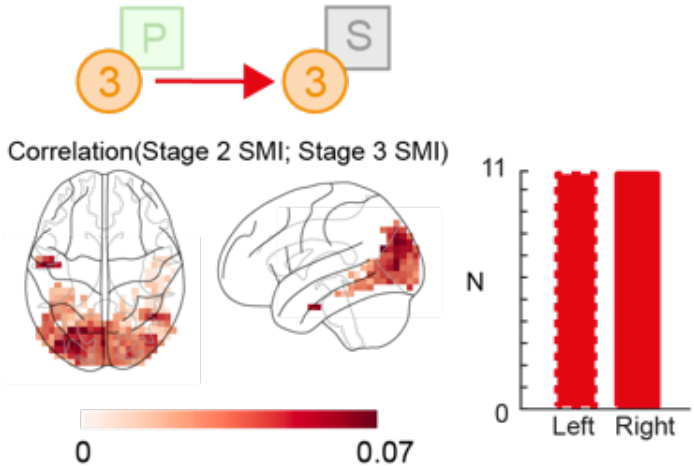


Figure 4-3 Single-trial prediction sharpens stimulus contents representations.

With SMI, we decomposed MI(<LSF vs. HSF auditory cue; Stage 2 MEG source activity>), to quantify the contribution of each trial to the representational strength of the prediction on occipital sources at the end of Stage 2. We similarly used SMI to decompose MI(<LSF vs. HSF stimulus; Stage 3 MEG source activity>) and isolate the contribution of each trial to the representation of LSF vs. HSF contents when the stimulus is shown at Stage 3. For each source, we then computed the single-trial correlation between Stage 2 SMI (i.e. maximium >200ms post-cue) and Stage 3 SMI (maximum in [0.15 to 0.25 s] post Gabor onset), and averaged them across participants when significant. We replicated such sharpening effect of prediction on Gabor representations in occipital cortex for left- (dashed) and right-located trials in 11/11 participants, Bayesian population prevalence = 1 [0.77 1], MAP [95% HPDI].

4.2.4 Stronger Categorization in Premotor Cortex leads to Faster RT

To understand how single-trial modulations of premotor cortex sources in turn modulate single-trial behavioral reaction times (RT), in each participant we computed the strength of the relationship between single-trial LSF vs. HSF representation on each pre-motor cortex source and corresponding RTs (Pearson correlation, no correction, $p<0.05$, see *Methods, Premotor Cortex modulates RT*). Figure 4-4 shows the group mean of the significant correlation coefficients across participants, thus localizes which source positively and negatively related to RT, suggesting that higher SF discrimination in left premotor cortex (i.e., contra-laterally to the instructed right-hand response) and lower SF discrimination in right premotor cortex (i.e., contra-laterally to the instructed right-hand response) produced faster behavioral SF categorizations of the Gabors (i.e., lower RT). Prevalence bar shows that for left-located (dashed line) and right-located trials, this effect is both replicated in at least 8/11 participants—i.e. Bayesian population prevalence = 0.73

[0.42 0.91], MAP [95% HPDI]. After FWER multiple comparison correction, only left premotor cortex effect remains in 5/11 participants.

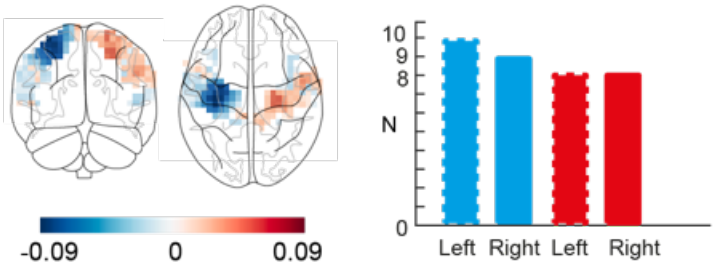


Figure 4-4 Single-trial premotor activity modulates RT.

For each premotor cortex source, we correlated between the maximum Stage 3 SMI (> 0.3s post Gabor onset) and the corresponding behavioral RT. The group averages of significant correlations show that higher discrimination of SF in left premotor cortex sources (contra-lateral to response hand) and lower discrimination in right premotor cortex (ipsilateral to response hand) lead to faster behavioral RT. We replicated the left premotor cortex effect for left-located trials (dashed) in 10/11 participants, Bayesian population prevalence = 0.91 [0.64 0.99], MAP [95% HPDI], and right-located trials in 9/11 participants, Bayesian population prevalence = 0.81 [0.53 0.96], MAP [95% HPDI]. We replicated the right premotor cortex effect for left-located trials (dashed) and right-located trials in 8/11 participants, Bayesian population prevalence = 0.73 [0.42 0.91], MAP [95% HPDI].

In sum, we showed that Stage 2 LSF vs. HSF predictions on contra-lateral occipital sources at Stage 2 preview/replay LSF vs. HSF representations when the stimulus is shown at Stage 3, speeding up RT by modulating premotor cortex activity. We showed that the underlying mechanisms of sharpened the representation of the LSF vs. HSF Gabor stimulus across the categorization network, from its onset in occipital cortex representation until categorization behavior.

4.3 Discussion

To address the question of how prediction propagation influences the processing of the stimulus in a categorization network, When the Gabor stimulus is finally shown at Stage 3, we reconstructed a functional network post-stimulus that propagates the actual feature from occipital cortex (150-250ms post Stage 3 Gabor), parietal lobe (250-350ms

post Stage 3 Gabor) and premotor cortex (>350ms post Stage 3 Gabor). We found this the actual feature representation replays the predictions, enhanced by valid predictions across the network, leading to faster categorization behavior.

Particularly, the earliest node in the categorization network activates at 150ms-250ms at occipital-ventral regions, we should not ignore the missing of a common early component before 100ms (Sharon, Hämäläinen, Tootell, Halgren, & Belliveau, 2007). This missing is due to a weak effect size of the early activation, which is regarded as insignificant with the strict FWER statistical test. However, we indeed see the occipital activation during 50ms-100ms from the Gabor onset. This inspires us to ask, what the cognitive difference is between the early activation and the later (i.e., 150ms-250ms) activation in network communication, and whether it is sharpened, dampened, or not influenced by the pre-stimulus prediction.

Beyond that we have proved behaviorally that the prediction could speed up the categorization significantly, it is indeed valuable to trace where the difference is from. In other words, whether this facilitation is manipulated by one or some specific brain regions. Here we showed that the end node of the network – the premotor cortex contributes to this speeding up most significantly. However, Supplemental Figure S4-5 also indicates that searching the whole brain, occipital cortex shows a much weaker but existing relevance with RT, compared with the parietal and premotor cortices showing the highest effect size. Thus, it is suggested that the enhanced representations contribute to the brain through the whole network, with an accumulated effect size along the network communication. This inspires us to put forward another question – what the function of parietal cortex is in the network communication about visual categorization? Previous studies have found that the parietal cortex is deeply related to decision-making, driving a perceptual response to noisy or ambiguous displays (Andersen & Cui, 2009). Particularly, in this cognitive process, the posterior parietal cortex presents a signature of evidence accumulation and encoded a graded value of this accumulation (Hanks et al., 2015). Linking to our study, this reported function is essential to answer what cognition function of parietal cortex is to realize the strong manipulation of RT with prediction.

An enduring debate questions whether prediction sharpens or dampens stimulus representation (De Lange et al., 2018), typically supported by enhanced or impaired

decoding performance of the predicted stimulus in the regions of interest (Blank & Davis, 2016; P. Kok, Janneke F. M. Jehee, & Floris P. de Lange, 2012b; Kumar et al., 2017; Lee & Mumford, 2003a). Here, we showed that predictions enhance the representation of LSF vs. HSF stimulus contents, meantime answering when and where this enhancement happens. Although most of our participants (7/11) show sharpened feature discrimination with prediction in occipital cortex, we do not want to ignore the other 3 participants suggesting a contrary impaired representation. This distinction could be explained by the individual strategies, that with a prediction, different participants attend more vs. less to the physical visual stimulus, causing a stronger vs, weaker discrimination. On the other hand, we found the prediction enhancement on parietal and premotor cortex is more prevalent across participants, leading to a faster behavioral categorization.

Further, using SMI decomposition, we found a tendency that predictions enhance stimulus content representations on single trials. That is, in trials when an occipital source represents more strongly the LSF vs. HSF prediction at Stage 2, we found that the same source also better represents the LSF vs. HSF contents of the subsequently presented stimulus at Stage 3. Although this effect fades away after multiple comparison corrections, we can still consider such a per-trial tendency as a direction to apply further research. On the other hand, we should not deny the confounding caused by trial-by-trial fluctuations in attention or even simply, in onset baseline. Thus, we need to be cautious to use the trial-by-trial correlation and draw specific conclusions with this method.

However, there is still an unsolved issue about whether the enhancement of representation is related to the reduction of RTs. Through the analysis, although some participants have shown a negative correlation between single-trial occipital representation and RT, it is not a general pattern across participants. Thus, we cannot provide direct evidence to link the initial discrimination and the final speed of response.

4.4 Methods

4.4.1 Analysis

4.4.1.1 Categorization Network Nodes (supports Figure 4-1A and B)

To understand the network of brain regions at Stage 3, that propagate the stimulus feature of LSF/HSF post Gabor onset, we computed the dynamics of stimulus representation across the whole brain, separately for left- and right-located trials.

For each participant, we computed the single-trial MI relationship <LSF vs. HSF Gabor; Stage 3 MEG source activity>, at each time point from 0 to 500ms following the Gabor onset, on each source in occipital regions (lingual gyrus, cuneus, inferior occipital gyrus), temporal (fusiform gyrus, inferior temporal gyrus, middle temporal gyrus, superior temporal gyrus), parietal (superior parietal lobe, inferior parietal lobe, angular gyrus, supramarginal gyrus), premotor (precentral gyrus, postcentral gyrus), and frontal (orbitofrontal gyrus, inferior frontal gyrus, middle frontal gyrus, medial frontal gyrus, superior frontal gyrus). We then extracted the peak time of the MI for each source. A clustering analysis revealed again three sequential space x time periods of stimulus representation (see *Methods, Categorization periods clustering* and Supplemental Figure S4-4). To reveal the dynamics, we localized the source peaking around the first peak in the 150-250ms time window (start), the last peak in 250-350ms (midway) and >350ms (end). Figure 4-1A illustrates these three peaks in the Stage 3 time series of one typical participant. We computed the group mean of these 3 source localized peaks across participants (see Figure 4-1B for group mean and Supplemental Figure S4-1 for each individual result).

4.4.1.2 Categorization Network Reconstruction (supports Figure 4-1 C and D)

To reconstruction the functional network that communicates at Stage 3 the Gabor SF across the regions identified earlier (i.e., occipital, parietal, premotor), we computed DFI where F is the Gabor SF information that decides the final response. In each participant, for each pair of regions (i.e., sender: occipital, receiver: parietal; sender: parietal, receiver: premotor), we computed DFI as follows:

Step 1: Source selection. We selected one source for the sending and one source for the receiving regions. These sources had the highest MI (<LSF vs. HSF Gabor; Stage 3 MEG source activity>) peak value during the time window of interest (occipital: 150-250ms, parietal: 250-350ms, premotor: >350ms). We computed the communications of the predictive cue between these three sources forming a functional network.

Step 2: DFI. On the selected sources, for receiver time points between 0 and 500ms post Gabor stimulus onset, and for delays between 0 and 300ms, we computed the receiving-time × communication-delay of LSF vs. HSF stimulus representation with DFI (see the computation methods in Stage 2: Network reconstruction in prediction).

Step 3: Statistical significance was established by recomputing DFI with shuffled feature labels (i.e., the LSF vs. HSF labels), 200 repetitions, using as statistical threshold the 95th percentile of 200 maxima (each taken across the DFI matrix of each shuffled repetition, FWER, $p<0.05$, one-tailed).

We applied Step 1-3 in each participant, reconstructing the occipital-to-parietal and parietal-to-premotor network that communicates the LSF vs. HSF stimulus contents (Figure 4-1C and D shows the group average; Supplementary Figure S4-2 shows Stage 3 DFI results per participant).

4.4.1.3 Prediction overlaps with Stimulus Representation (supports Figure 4-2A orange curve)

To determine whether trial-per-trial Gabor predictions at Stage 2 comprise the same SF contents as when the Gabor stimuli are actually processed at Stage 3, we quantified the overlap of Gabor LSF vs. HSF representations between Stages 2 and 3. We did so with information theoretic redundancy quantified with co-information—i.e. Co-I(<LSF vs. HSF, Stage 2 Period 3 MEG; Stage 3 Period 4 MEG>) on trials with valid predictions and separately for left and right localized stimuli. This quantifies the information about LSF vs HSF which is shared between or common to the Stage 2 and Stage 3 MEG responses.

Step 1: Source selection. We selected one source for the Stage 2 Period 3 and one source for the Stage 3 Period 4 in contra-lateral occipital cortex. These sources had the

highest MI (<LSF vs. HSF cue/Gabor; Stage 2/3 MEG source activity>) peak value during the time window of interest (Stage 2: >200ms, Stage 3: 150-250ms).

Step 2: Co-Information. To trace the per-trial representational overlap between Gabor LSF vs. HSF predictions at Stage 2 and actual LSF vs. HSF Gabor representations at Stage 3, we then computed the single-trial information theoretic redundancy—i.e. the Co-Information Co-I(<Stage 2 MEG; Stage 3 MEG; LSF vs. HSF>)

$$\text{Co-I} = \text{MI}(\text{Stage 2 MEG; LSF vs. HSF}) + \text{MI}(\text{Stage 3 MEG; LSF vs. HSF}) - \text{MI}(\text{Stage 2 MEG, Stage 3 MEG; LSF vs. HSF}),$$

where Stage 2 MEG refers to the single-trial activity of the Stage 2 Period 3 occipital source chosen at Step 1 above; Stage 3 MEG refers to the Stage 3 Period 4 occipital source chosen. We computed Co-I on the full combinatorics of Stage 2 MEG and Stage 3 MEG, every 2ms between 0 and 400ms post auditory cue onset and every 2ms between 0 and 500ms post Gabor onset, using valid informative cueing trials when the single-trial LSF vs. HSF label is consistent between Stages 2 and 3. This computation produced a 2D matrix (Stage 2 time × Stage 3 time).

Step 3: Statistical significance was established by recomputing the Co-I with shuffled feature labels (i.e., the LSF vs. HSF labels), 200 repetitions, using as statistical threshold the 95th percentile of 200 maxima (each taken across the 2D Co-I matrix of each shuffled repetition, FWER, $p<0.05$, one-tailed). To reflect when and how strongly the stimulus representation at Stage 3 overlaps with the previous prediction, for each time point at Stage 3, we took the maximum Co-I value across Stage 2 time.

We repeated the above Steps 1-3 for each participant. The orange curve in Figure 4-2A shows the cross-participants mean Co-I from Stage 3 Gabor onset. Supplementary Figure S4-3 shows the individual results.

4.4.1.4 Prediction modulates Source Activity and RT (supports Figure 4-2A brown curve)

To detect how valid predictions modulate MEG activity and facilitate behavior, we compared the effect of Stage 2 prediction (i.e., predicted vs. non-predicted) jointly on Stage 3 brain activity and RT.

Step 1: Source selection. We selected one source for the Stage 3 Period 6 in premotor cortex. This source had the highest MI (<LSF vs. HSF Gabor; Stage 3 MEG source activity>) peak value after 350ms post Gabor onset.

Step 2: Co-Information. We computed single-trial information theoretic redundancy—i.e. positive Co-I(<predicted vs. non-predicted; Stage 3 MEG; RT>)—as follows:

Co-I = MI (predicted vs. non-predicted; Stage 3 MEG) + MI (predicted vs. non-predicted; RT) – MI (predicted vs. non-predicted; Stage 3 MEG, RT).

where Stage 3 MEG refers to the single-trial activity of the Stage 3 Period 6 premotor source chosen at Step 1 above, every 2ms between 0 and 500ms post Gabor onset. Co-I computations produced a vector along Stage 3 time.

Step 3: Statistical significance was established by recomputing the Co-I with shuffled feature labels (i.e., the LSF vs. HSF labels), 200 repetitions, using as statistical threshold the 95th percentile of 200 maxima (each taken across the 2D Co-I matrix of each shuffled repetition, FWER, $p<0.05$, one-tailed).

We applied Step 1-3 to each participant. The brown curve in Figure 4-2A shows the cross-participants mean Co-I from Stage 3 Gabor onset. Supplementary Figure S4-3 shows the individual results.

4.4.1.5 Prediction enhances Stimulus Representation (supports Figure 4-2B)

To understand how Stage 2 predictions facilitate Stage 3 visual categorizations, we compared representation of Gabor LSF vs. HSF between predicted and non-predicted trials

at Stage 3. Specifically, we computed MI between MEG source and LSF vs. HSF with predicted (informative cued) vs. non-predicted (neutral cued) trials across the network nodes. Separately for left and right located trials, in each participant, for each region (i.e., contra-lateral occipital, parietal, premotor), we computed MI as follows:

Step 1: Source selection. We selected one source at Stage 3. This source had the highest MI (<LSF vs. HSF Gabor; Stage 3 MEG source activity>) peak value during the time window of interest (occipital: 150-250ms, parietal: 250-350ms, premotor: >350ms).

Step 2: MI computation. At the selected source, we computed source-by-time MI (LSF vs. HSF Gabor; Stage 3 MEG), every 2ms between 0 and 500ms post Gabor onset, separately for predicted and non-predicted trials, matching trial numbers with random selection for this computation and averaging 5 random trial selections.

Step 3: MI comparisons.. Statistical significance was established by recomputing the source-by-time MI with shuffled predicted and unpredicted trials (repeated 200 times), calculating the difference of peak between recomputed predicted and non-predicted MI during the time window of interest, using as statistical threshold the 95th percentile of 200 maxima, each taken across the source-by-time difference of each shuffled repetition (FWER, $p<0.05$, two-tailed).

We repeated above Step 1-3 for each participant. Figure 4-2B shows the result of MI difference.

4.4.1.6 Stronger Prediction leads to Stronger Categorization (supports Figure 4-3)

To determine whether prediction sharpens or dampens representation of stimulus when it is shown, we computed the single-trial relationship between the strength of the stimulus prediction before it is shown and its actual representation when it is shown. We did so per participant, separately for left-located and right-located trials, as follows:

Step 1: Stage 2 single-trial decomposition of prediction. For each occipital source, we isolated the contribution of each trial to the overall discrimination of the LSF/HSF predictive cue at Stage 2, by computing sample-wise MI (SMI), in the time window > 200

ms following predictive cue onset, selecting the source with maximum SMI on this trial, within this time window. Across trials, we obtained a trial × source matrix of prediction strength > 200 ms at Stage 2.

Step 2: Stage 3 representation. To determine the Stage 3 time of interest, for each occipital source, we computed the single-trial MI relationship <LSF vs. HSF Gabor; Stage 3 MEG source activity>, at each time point from 0 to 500ms following Gabor onset. We extracted peak time for each occipital source and took the median across sources during 150-250ms post Gabor.

Step 3: Stage 3 single-trial decomposition of representation. We proceeded similarly for each occipital source, this time following Gabor stimulus presentation at Stage 3. We decomposed SMI to quantify the single-trial representational strength of LSF vs. HSF, selecting the source with maximum SMI between 150ms and 250ms post Gabor stimulus. Across trials, this produced the trial × source matrix of Gabor SF representation strength.

Step 4: Correlation between prediction and representation. For each occipital cortex source, we Pearson correlated single-trial SMI values between the LSF vs. HSF prediction at Stage 2 (Step 1) and LSF vs. HSF representation at Stage 3 (Step 2). We then localized occipital sources with significant correlations.

Steps 1 to 4 localized the participant's occipital sources where predicted SF contents at Stage 2 led to enhances SF contents representations at Stage 3. Across participants, we averaged these significant coefficients (no correction, $p<0.05$) per source, shown in Figure 4-3.

4.4.1.7 Premotor Cortex modulates RT (supports Figure 4-4)

To understand how the premotor cortex receiving network node in turn modulates behavioral RTs, in each participant, separately for left- and right-located trials, we computed with SMI the single-trial relationship between the strength of LSF vs. HSF representation and the corresponding RT as follows:

Step 1: Stage 3 representation. To determine the Stage 3 time of interest, for each premotor source, we computed the single-trial MI relationship <LSF vs. HSF Gabor; Stage 3 MEG source activity>, at each time point from 0 to 500ms following Gabor onset. We extracted peak time for each premotor source and took the median across sources after >350ms post Gabor.

Step 2: Stage 3 single-trial decomposition. For each premotor source, we quantified with SMI the single-trial LSF vs. HSF Gabor representation, taking the maximum SMI after 300ms post Gabor per source, to produce the trial × source strength matrix of stimulus representation strength.

Step 3: Correlation. For each premotor cortex source, we Pearson correlated Step 1 stimulus representation SMI and corresponding RT, localizing the sources with significant correlation between stimulus representation and RT.

Across participant, we averaged per source the significant correlations (no correction, $p<0.05$) to visualize the effect of premotor modulation of stimulus representation on RT in Figure 4-4.

4.5 Supplemental Material

4.5.1 Supplemental Methods

4.5.1.1 Categorization periods clustering

To compute the number of space x time stimulus representations period, we applied again k-means cross-trials clustering analysis on all 4,413 sources x 256 time points as follows:

Step 1: Peak time extraction. First, for each participant, and independently for left- and right-located trials and source, we extracted the peak LSF vs. HSF representation MI in 50 10-ms time windows spanning 0-500ms post Gabor.

Step 2: Matrix computation. Across participants and conditions, we counted the numbers of sources per ROI (occipital, temporal, parietal, pre-motor and frontal) and time window the number of sources that peak, producing a ROI x time matrix of MI peaks.

Step 3: Clustering. We k-means clustered (k = 1 to 30, repeating 1,000 times) the matrix from Step 2, using the 50 time windows as samples and selected k as the elbow of the within-cluster sums of point-to-centroid distances metric.

Stage 3 comprised k = 4 clusters (see Supplemental Figure S4-4). A first period with no LSF vs. HSF stimulus representation, followed an occipital-ventral (150-250ms, start), parietal (250-350ms), and premotor-frontal (>350 ms) periods of stimulus representation.

4.5.1.2 Representational overlap across the brain

To reveal the representational overlap across the brain, we replicated the Co-I (<LSF vs. HSF, Stage 2 Period 3 MEG; Stage 3 MEG>) across the whole brain at Stage 3, by applying the following steps:

Step 1: To simplify calculations, we chose one source at Stage 2: one for the contra-lateral occipital cortex and one for the right fusiform gyrus. In each region, we chose that

source that maximizes MI(<LSF vs. HSF prediction; Stage 2 MEG>) computed every 2ms over the [0 - 500ms] interval post auditory cue.

Step 2: To trace the per-trial representational overlap between Gabor LSF vs. HSF predictions at Stage 2 and actual LSF vs. HSF Gabor representations at Stage 3, we used 6 ROIs—i.e. (1) contra-lateral occipital cortex, (2) right fusiform gyrus, (3) temporal lobe (inferior temporal gyrus, middle temporal gyrus, superior temporal gyrus), (4) parietal lobe (inferior parietal lobe, superior parietal lobe, supramarginal gyrus and angular gyrus), (5) frontal lobe (orbitofrontal cortex, inferior frontal gyrus, middle frontal gyrus, medial frontal gyrus and superior frontal gyrus), (6) pre-motor cortex (precentral gyrus and postcentral gyrus).

Step 3: We then computed the single-trial information theoretic redundancy—i.e. the Co-Information Co-I (<Stage 2 MEG; Stage 3 MEG; LSF vs. HSF>)

$$\text{Co-I} = \text{MI}(\text{Stage 2 MEG; LSF vs. HSF}) + \text{MI}(\text{Stage 3 MEG; LSF vs. HSF}) - \text{MI}(\text{Stage 2 MEG, Stage 3 MEG; LSF vs. HSF}),$$

where Stage 2 MEG refers to the single-trial activity of the 2 sources chosen at *Step 1* above; Stage 3 MEG refers to each source of the 6 ROIs defined in *Step 2* above. We computed Co-I on the full combinatorics of Stage 2 MEG and Stage 3 MEG, every 2ms between 0 and 500ms post Gabor onset, using valid informative cueing trials when the single-trial LSF vs. HSF label is consistent between Stages 2 and 3. This computation produced a 4D matrix (Stage 2 sources × Stage 3 sources × Stage 2 time × Stage 3 time).

Step 4: When the Gabor appears at Stage 3, we quantified how its representation in each one of the 6 ROIs overlaps with its prediction at Stage 2. In each ROI, we selected the top 25% of the maximum Co-I values (taken for each source, at peak Co-I time at Stage 3) and averaged these values per ROI. This produced a 2×6 matrix for Co-I peak amplitudes and peak times (i.e. 2 sources selected at Stage 2 × the top 25% Co-I in each ROI at Stage 3).

We repeated the above Steps 1-4 for each participant. Round and diamond scatters above plots A and B in Figure 3 show the individual Co-I peak amplitude and peak time at Stage 3 in contra-lateral occipital cortex, separately for right and left-located trials.

Supplemental Figure S4-5 Y-axis shows the individual-level and group-level Co-I strength for all 6 ROIs at Stage 3.

4.5.1.3 RT modulation across the brain

To reveal the RT modulation across the brain, we replicated the Co-I (<predicted vs. non-predicted, RT; Stage 3 MEG>) across the whole brain at Stage 3, by applying the following steps:

Step 1: We computed single-trial information theoretic redundancy—i.e. positive Co-I (<predicted vs. non-predicted; Stage 3 MEG; RT>)—as follows:

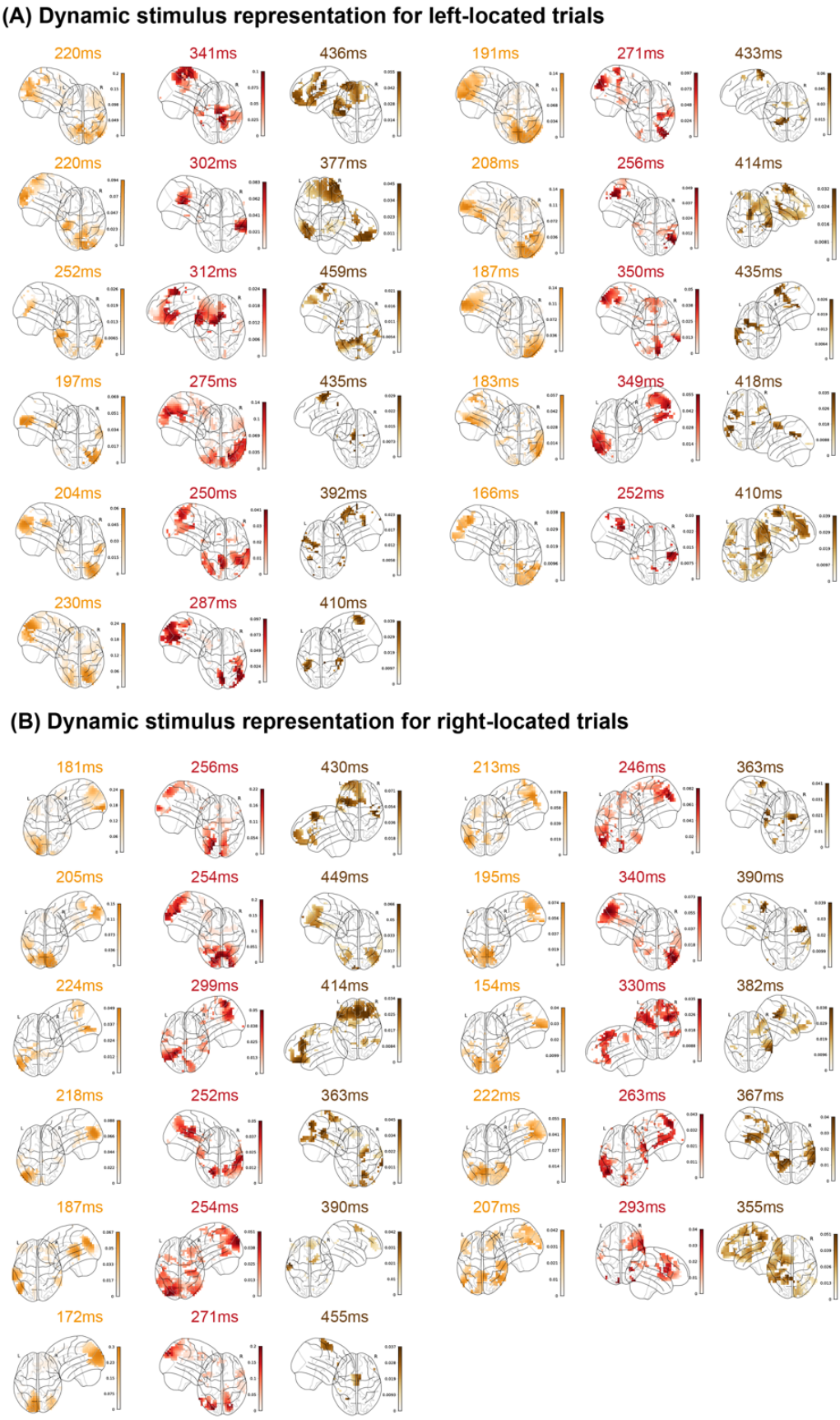
Co-I = MI (predicted vs. non-predicted; Stage 3 MEG) + MI (predicted vs. non-predicted; RT) – MI (predicted vs. non-predicted; Stage 3 MEG, RT)

for all sources of the 6 ROIs at Stage 3, every 2ms between 0 and 500ms post Gabor onset. Co-I computations produced a source-by-time redundancy matrix. We established statistical significance by recomputing the matrix with shuffled cueing labels (repeated 1,000 times), using as statistical threshold the 95th percentile of 1,000 maxima (each taken across the redundancy matrix of each shuffled repetition, FWER, $p<0.05$, one-tailed).

Step 2: In each ROI, we quantified how the valid informative cueing modulated brain activity and RT. We averaged the top 25% of maximum Co-I values in each ROI (taken for each source, across time, after statistical test) at their peak time. This produced a vector of 6-ROI redundancy peak amplitudes and peak times.

We applied Step 1 and 2 to each participant, to derive peak values of redundancy strength and peak time for each of 6 ROIs. The square scatters above plots C and D in Figure 3 show the Co-I peak time and amplitude in pre-motor cortex, separately for right and left located trials. Supplemental Figure S4-5 X-axis shows the individual-level and group-level redundancy strength in all 6 ROIs at Stage 3.

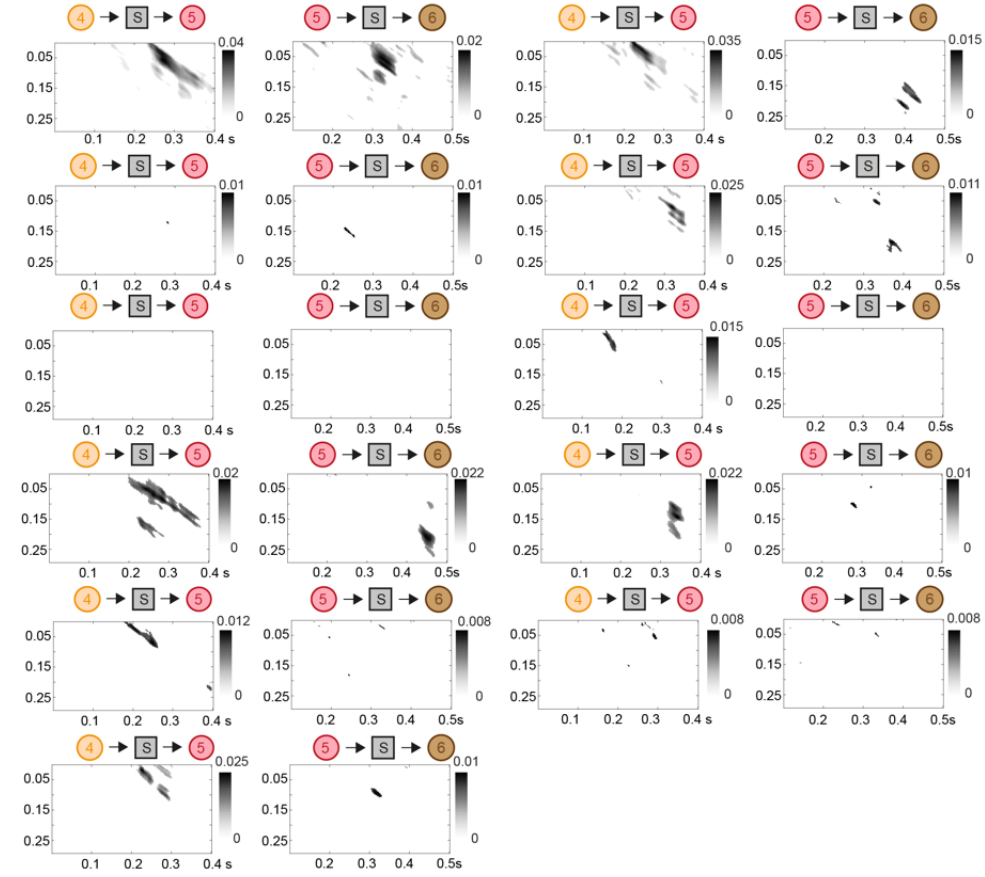
4.5.2 Supplemental Figures



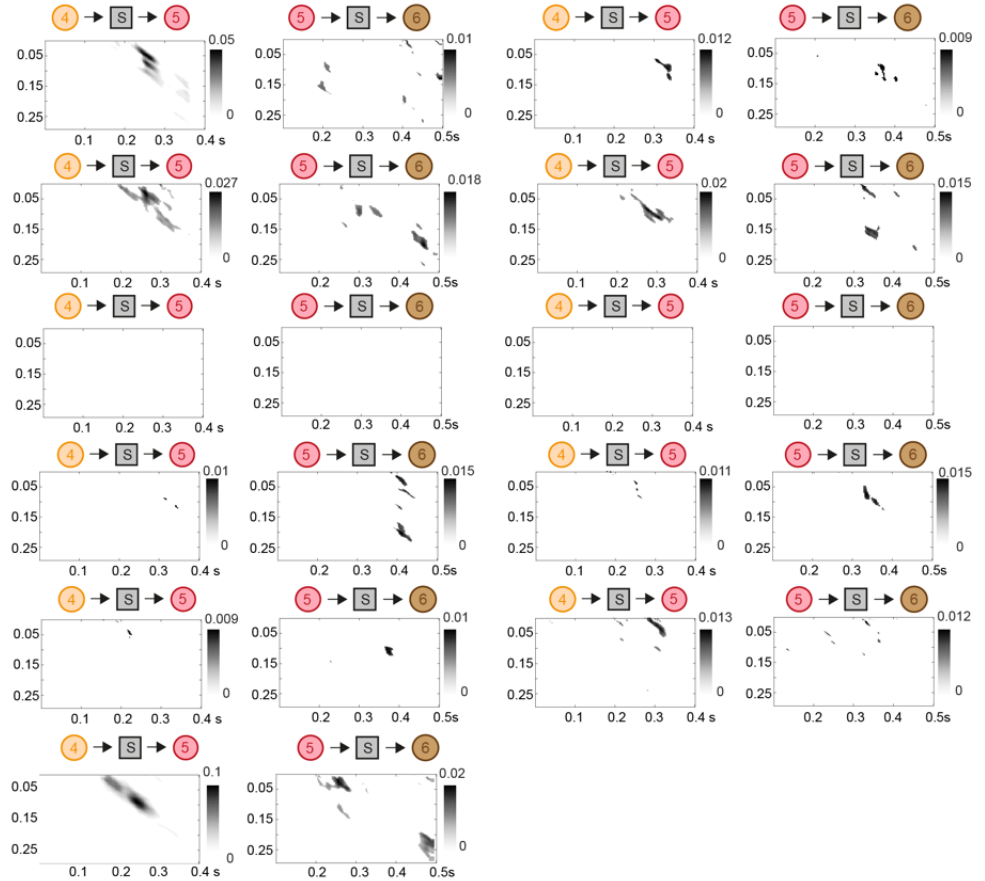
Supplemental Figure S 4-1 Individual dynamic representation of stimulus.

For each participant, separately **for (A) left-located trials and (B) right-located trials**, we computed the single-trial MI relationship <LSF vs. HSF Gabor; Stage 3 MEG source activity>, at each time point from 0 to 400ms following auditory cue onset, on each source in temporal, frontal and occipital regions. We then extracted the peak time of the MI for each source. To reveal the dynamics, we localized the source peaking around the first peak in the 150-250ms time window (start, in orange), the last peak in 250-350ms (midway, in red) and >350ms (end, in brown). The specific time point of each pattern is denoted above the brain plot.

(A) DFI results for left-located trials (Stage 3)

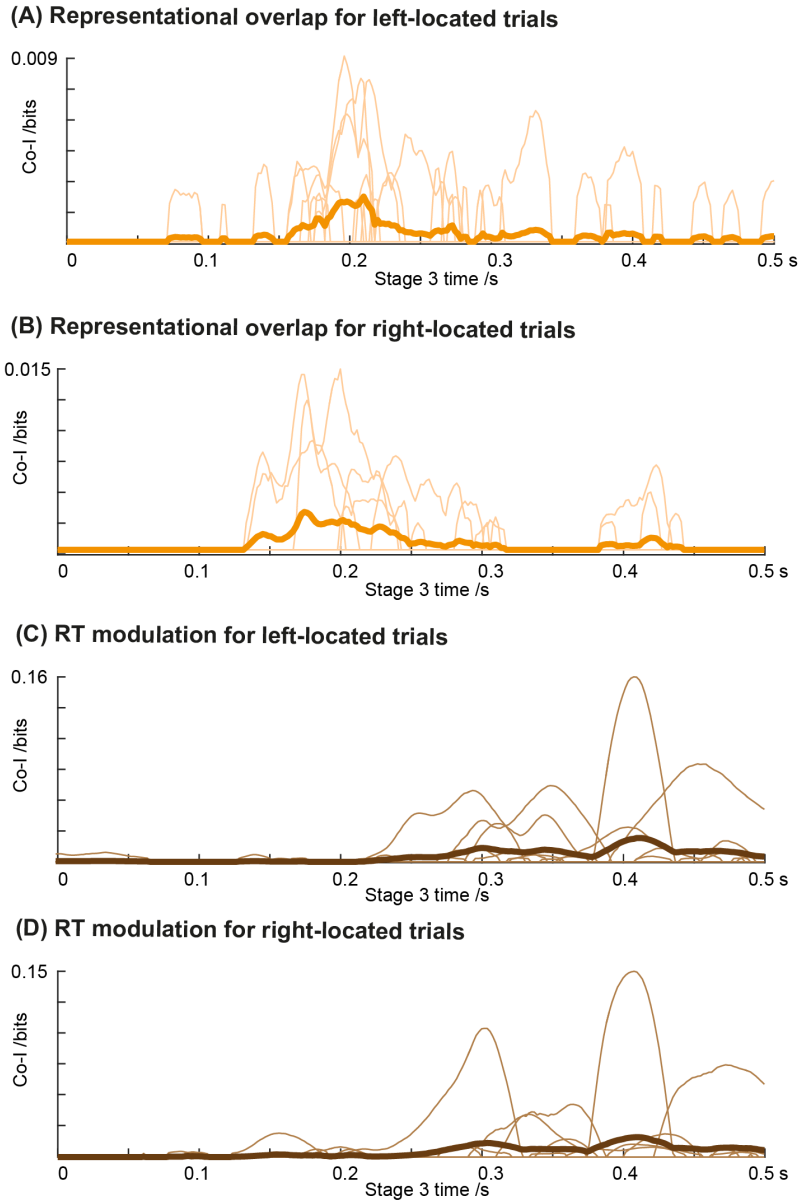


(B) DFI results for right-located trials (Stage 3)



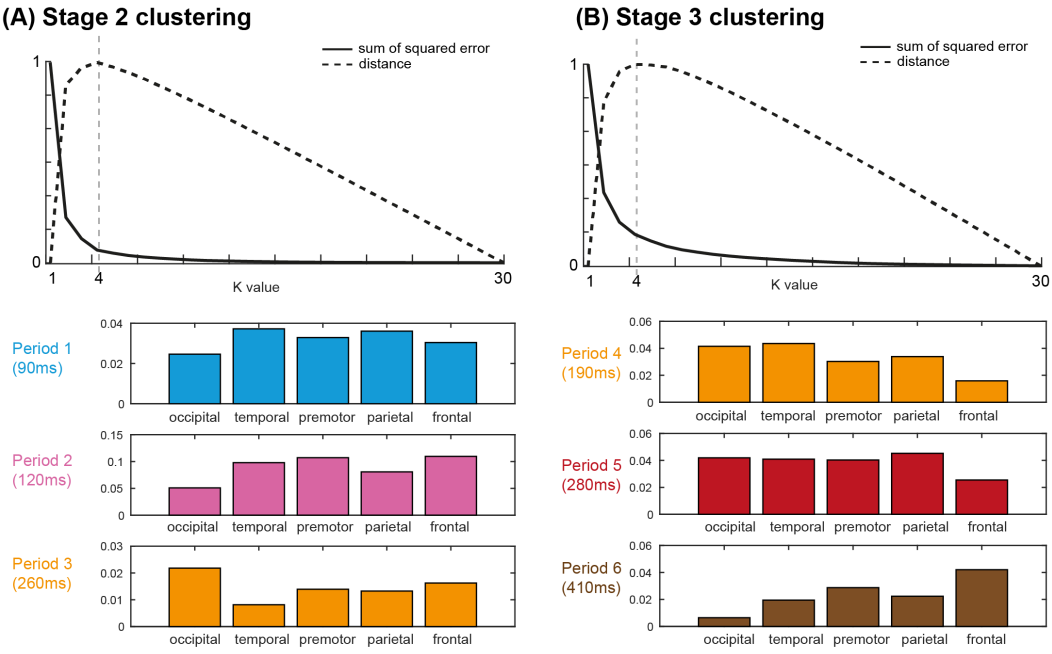
Supplemental Figure S 4-2 Individual DFI results in categorization network.

Using Directed Feature Information, separately for **(A) left-located trials and (B) right-located** trials, we computed the participant communications across Stage 3 nodes (i.e. occipital 4 -> parietal 5 and parietal 5 -> premotor 6) for each individual participant. These plots represent the significant (FWER-corrected, $p<0.05$) communications in the receiving node time course (X-axis), as communication delays from the sending node (Y-axis).



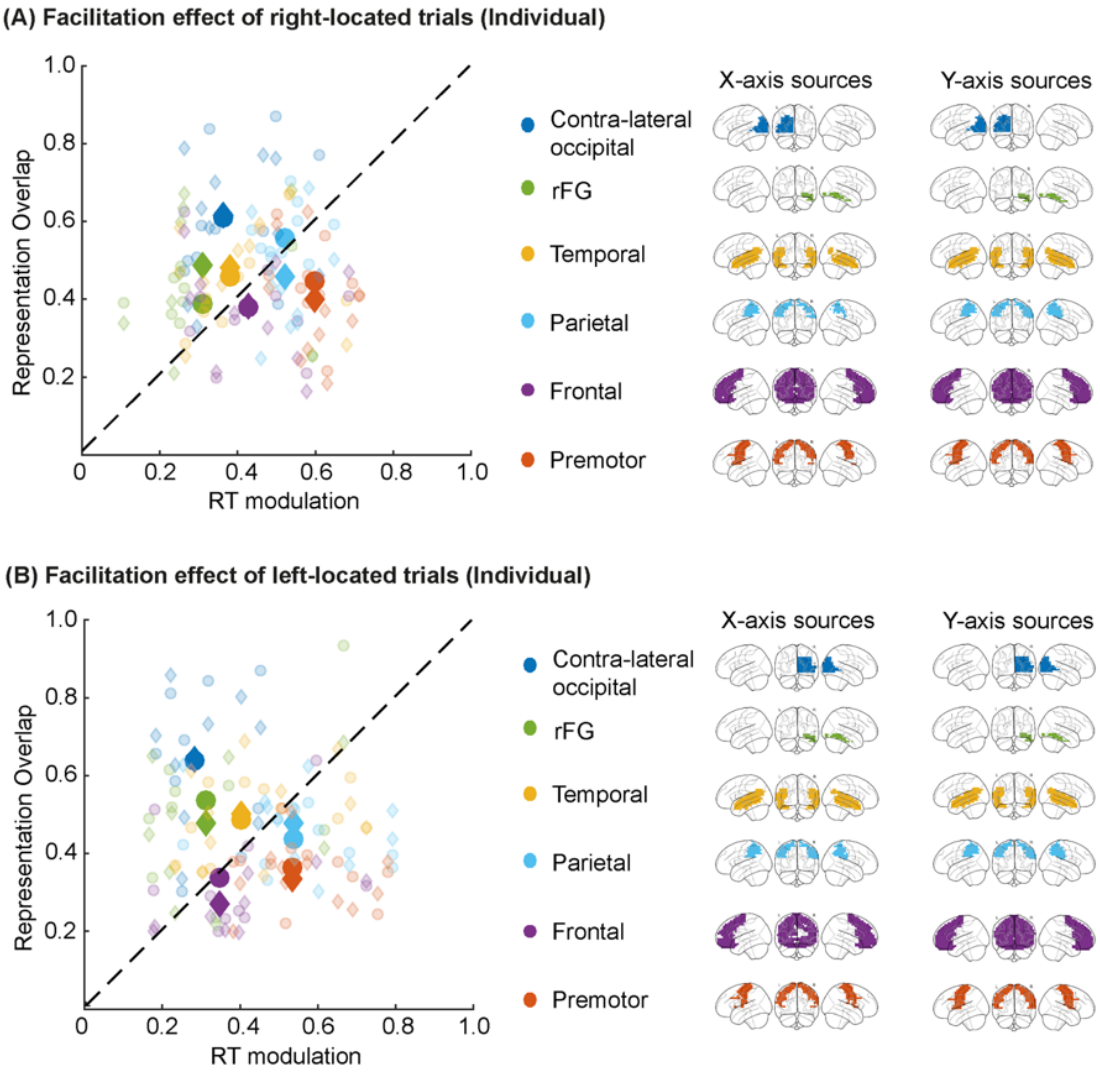
Supplemental Figure S 4-3 Individual results of representational overlap and RT modulation.

(A) and (B) Representational overlap for left-located and right-located trials. We computed the representational overlap between prediction of Gabor LSF vs. HSF (in Stage 2 occipital node) and physical Gabor LSF vs. HSF (in Stage 3 occipital node) using Co-Information – i.e., Co-I (LSF vs. HSF; Period 3 MEG; Period 4 MEG). We took the maximum Co-I across Stage 2 time. The light orange curves represent individua results and the dark orange curve represents the group mean. **(C) and (D) RT modulation for left-located and right-located trials.** We tested the prediction modulation on RT and premotor cortex activity —i.e., computed as Co-I (Predicted vs. Non-predicted; Stage 3 premotor MEG; RT). The light brown curves represent individua results and the dark brown curve represents the group mean.



Supplemental Figure S4-4. Clustering of representational time windows at Stage 3 (stimulus representation).

K-means analysis, with k = 1..30, with sum of squared error of each k (plain curve) and distance of the plain curve from straightline (dashed curve). We selected k with furthest distance (elbow method). Following a baseline period (Period 0), the bar plot shows the percentage of sources peaking at the center time of each period, at occipital, temporal, premotor, parietal, frontal respectively. Stage 3 contains 3 periods of stimulus representations, dominated by occipital-temporal, parietal and premotor-frontal.



Supplemental Figure S 4-5 SF redundancy – RT redundancy distribution across the whole brain for right-located (A) and left-located (B) trials.

Left panel shows 3.1 Co-I(LSF vs. HSF; Stage 2 MEG; Stage 3 MEG) on the X-axis, 3.2 Co-I(predicted vs. non-predicted; RT; Stage 3 MEG) on the Y-axis. Color-indexed scatters show SF redundancy (X-axis, circles, redundancy with Stage 2 contra-lateral occipital source, diamonds, redundancy with Stage 2 right fusiform gyrus source), RT redundancy (Y-axis) in contra-lateral occipital cortex (blue), right fusiform gyrus (green), temporal lobe (yellow, including inferior temporal gyrus, middle temporal gyrus and superior temporal gyrus), parietal lobe (light blue, including inferior parietal lobe, superior parietal lobe, supramarginal gyrus and angular gyrus), frontal lobe (purple, including orbitofrontal cortex, inferior frontal gyrus, middle frontal gyrus, medial frontal gyrus and superior frontal gyrus), and premotor cortex (orange, including precentral gyrus and postcentral gyrus). Each t scatter represents the mean of top 25% of maximum Co-I (LSF vs. HSF; Stage 2 MEG; Stage 3 MEG) value (taken for each Stage 3 source, across Stage 2 times x Stage 3 times), the mean value of top 25% of maximum Co-I (predicted vs. non-predicted; RT; Stage 3 MEG) value (taken for each Stage 3 source, across Stage 3 times). Each opaque scatter represents the group-level mean result. Right panel shows the location of color-coded ROI sources that have significant Co-I in the top 25% of maximum Co-I for at least one of the participants.

5 General Discussion

To understand how a network of brain regions coordinates to propagate specific predictions and to analyze how these predictions in turn change stimulus processing, we addressed three fundamental questions (illustrated in Figure 5-1):

(1) When, where, and how does a prediction dynamically propagate through a Prediction Network of brain regions?
(2) When, where, and how does a stimulus feature dynamically propagate through a Categorization Network of brain regions?
(3) How does the prediction interact with stimulus representation, finally modulating behavior?

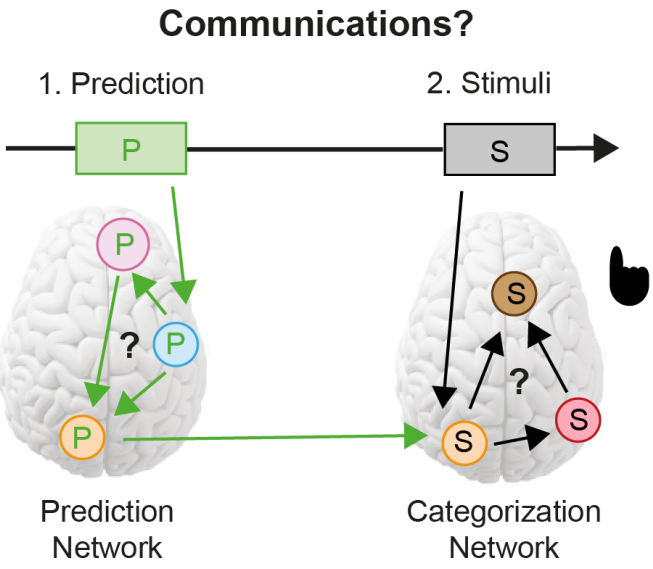


Figure 5-1 Three key questions about prediction.

(1) When, where and how does a prediction dynamically propagate throughout in a Prediction Network of brain regions? Pre-stimulus onset, an auditory cue (P, the prediction color-coded in green) signals an upcoming visual stimulus feature. Our first question is about that network of regions that communicate (see green arrows) this prediction from temporal to occipital cortex. **(2) When, where and how does a stimulus feature dynamically propagate throughout a Categorization Network of brain regions?** Post-stimulus onset, our second question is about the network of regions that communicate the stimulus feature information. **(3) How does the prediction interact with stimulus representation, finally modulating behavior?**

In this thesis, before answering these questions at the brain level, I first validated the feasibility of the neuroimaging cueing design and the prediction effect using a series of behavior experiments as ***Study 1***. Sequentially, I detected the auditory sensitivity to pick

up differentiable cue stimuli, trained participants to couple the auditory cues and visual stimuli and finally instructed participants to categorize stimuli in parallel with predictions induced by cues and based on the learnt relationship. Though the results of this preamble behavior study were, to some extent, confounded between attention and expectations, it still delivered the successful manipulation in the protocol of the learning-categorization paradigm and inspired to improve the further neuroimaging experiment (e.g., by specifying the predicted contents).

Moving to the neuroimaging stage, I addressed the three questions mentioned in two subsequent studies via a three-stage experimental design that involved a spatial cue predicting the left vs. right location of an upcoming Gabor stimulus at Stage, followed by an auditory cue predicting the Gabor's SF contents at Stage 2, and finally the Gabor stimulus was presented at Stage 3.

In ***study 2***, my analyses reconstructed the functional network propagating the auditory predictive cue from the temporal to the occipital cortex. We found a chronological representation of the predictive cue from the temporal (90-120ms post Stage 2 cue) to the frontal (120-200ms) to the occipital (200-400ms) cortex. We showed that the frontal cortex mediates the communication of the predictive cue from the temporal to the left vs. right occipital cortex, depending on cued location. Additionally, using a multivariate decoding analysis, we proved the information propagating into the occipital cortex contra-lateral to the predicted location indeed contains the visual information the same as bottom-up processing.

In ***study 3***, we first reconstructed a functional network communicating stimulus feature, tracing the stimulus communications from occipital-ventral (150-250ms post Stage 3 Gabor) to parietal (250-350ms) and then to premotor (>350ms). With this network, we further showed that Stage 2 LSF vs. HSF predictions on contra-lateral occipital sources at Stage 2 preview/replay LSF vs. HSF representations when the stimulus is shown at Stage 3, speeding up RT by modulating premotor cortex activity. We showed that the underlying mechanisms of sharpened the representation of the LSF vs. HSF Gabor stimulus across the categorization network, from its onset in occipital cortex representation until categorization behavior.

In this ***Chapter 5***, I will discuss the potential source of prediction in ***5.1***, the contribution of our result in the debate of sharpening and dampening in ***5.2***, the generalization to the prediction of other kinds of or higher-dimensional contents in ***5.3***, in other sensory modalities in ***5.4***, and futural plans for tracing the prediction in laminar layers in ***5.5***.

5.1 Where is prediction from?

One of our prominent findings is the prediction propagation from temporal to occipital with the modulation of frontal cortex in a functional network. However, the question, of where the prediction is from, is still waiting for its answer. Hippocampus can be one of the most probable regions according to previous research, with a function to associate stimuli of which relationship is newly learnt without previous interference (Wallenstein, Hasselmo, & Eichenbaum, 1998), similarly to the auditory cues and Gabor spatial frequency in our design. In recent years, fMRI studies have come forth to report the predictive signals detected in hippocampus (Buckner, 2010; Davachi & DuBrow, 2015; Hanks et al., 2015). Limitedly, our MEG data did not tell a clear activity in hippocampus, which might be due to the source reconstruction of MEG sensor data being limited by the depth of hippocampus. We should conduct further experiments with deeper MEG source reconstruction or fMRI to examine this hypothesis.

Another candidate could be middle temporal gyrus showing anticipatory responses of learning and early imagery (Kreiman, Koch, & Fried, 2000; Reddy et al., 2015). This is also in line with our earliest prediction network node in temporal regions. However, in the current design, the temporal gyrus activation largely contains the auditory effect. Thus, we cannot distinguish the effect caused by the auditory pitch and potentially caused by the prediction generation in temporal lobe. In the future, one highly interesting direction is to separate the auditory effect and the prediction effect using advanced methods.

Besides, it is widely reported that different frequency bands distinctively work for the feedforward (bottom-up) and the feedback (top-down) signals. Specifically, lower frequency bands such as alpha and beta are responsible for feedback signalling (Benedek et al., 2011; Haegens et al., 2011; Lobier et al., 2018) and the gamma bands propagate information feedforward (Michalareas et al., 2016). Thus, additionally, the interesting

questions would be which frequency bands are responsible for the communications between temporal and frontal, and between frontal and occipital in our reported network.

5.2 Sharpening the visual representations with predictions

For decades, researchers have reported the enhancement of bottom-up sensory signals congruent with predictions compared with incongruent conditions, which is against another group of studies showing suppression. Our study is in line with both empirical and theoretical evidence provided by previous studies supporting the enhancement. Empirically, by conducting neuroscience experiments on humans and other primates, researchers have proved the sharpening effect by the higher decoding ability of Gabor orientation with its expectation compared with unexpected orientation (Kok, Jehee, et al., 2012b), stronger responses with matching stimuli compared with mismatch ones (Desimone, 1996) and the enhanced synaptic plasticity with perceptual priming. Theoretically, our result supports the predictive coding models, in which with Bayesian inference, the specific neurons integrate the current sensory input with prior estimations to this incoming input, filtering the predictive errors and achieving the highest congruency (Lee & Mumford, 2003b).

As a contribution to the progression of this field compared with the existing studies, our result further indicates that the sharpening effect happens not only in the visual cortex, but through the network communication about visual categorization at specific time windows. However, we still want to ask why this debate exists and both two opinions have been supported by lots of studies. A bold assumption is that the debate generates due to the different objects of observations. The studies supporting sharpening mostly focus on a cluster of voxels or neurons showing a particularly strong response to the stimulus, in other words, processing the stimulus contents specifically. This is also what we have done in the current study. In contrast, the “dampening” evidence usually tells a general pattern across early visual cortex, V1/V2 (Summerfield & de Lange, 2014a; Todorovic & de Lange, 2012), or even higher regions such as inferior temporal cortex (Kaposvari et al., 2018; Kumar et al., 2017) and posterior superior temporal sulcus (Blank & Davis, 2016). With this assumption, the next question for us in this field could be whether there is also a dampening pattern according to our data, potentially in a larger range within early visual cortex, or even in a different time window (e.g., the 50-100ms time window).

In addition, the explanation of sharpening theory is the predictions can explain away the predictive errors, in turn suppressing the incongruent stimuli representation and enhancing the congruent ones. Although with our design, to guarantee the feasibility where participants can implicitly learn the coupling relationship between cues and stimuli, the mispredicted (i.e., invalidly predicted) trials are relatively few to conduct solid analysis. However, by changing the structure of our design to get enough trials with a mismatch between the prediction and the subsequent stimulus (e.g., LSF prediction followed by HSF Gabor), we can then compare the stimulus representation in predicted trials, non-predicted trials, and mispredicted trials to further uncover the influence of prediction, uncovering the mechanisms around predictive errors.

5.3 Generalizing from Gabors to other stimulus features in face, object, body and scene categorizations.

Our contribution is methodological, providing the means to isolate the specific dynamic processing of a predicted stimulus feature into the visual hierarchy. Generalizing from our case study with Gabor stimuli to more naturalistic face, object and scene categorization tasks will incur several challenges that we can frame in the context of generally studying the details of information processing (Schyns, Gosselin, & Smith, 2009; Schyns et al., 2020). A key challenge is that the features of real-world faces, objects and scenes, may differ across behaviours and participants, rather than being fixed as they were here to Gabor LSF vs. HSF. Such task- or participant-specific features remain to be considered in studies of complex neural representation, processing and categorization, though see (R. A. A. Ince et al., 2016; Schyns, Petro, & Smith, 2007; Zhan et al., 2019).

A key constraint of studying information processing is that we need to characterize the information processed as precisely as possible. In complex categorizations, this implies understanding which specific image features are task-relevant and therefore predicted by each participant that performs the same behavioural task, but possibly with different stimulus features (Daube et al., 2021; Frédéric Gosselin & Schyns, 2003; M. L. Smith, Gosselin, & Schyns, 2006, 2012). Furthermore, we also need to know the behavioural outcomes of the process, which the typical passive viewing or one-back designs of neuroimaging experiments do not control. For example, a participant will predict a smiling

mouth feature to verify that a given incoming face is “happy” but the forehead wrinkles to predict “45 years of age” from the same face. Relatedly, experts will use different features than novices to classify their expert categories. Such relative perceptual expertise and associated features generally characterize the relationship between visual cognition and outside world stimuli (Gauthier, Tarr, Anderson, Skudlarski, & Gore, 1999; Malcolm, Nuthmann, & Schyns, 2014).

A methodological challenge will be to study visual predictions along the hierarchical decomposition of features, from their integrated representation in higher visual hierarchy (e.g., right fusiform gyrus), to their simplest Gabor representation in V1. This will likely require integration of brain measures that finely tap into the laminar layers of these regions (Charles D Gilbert & Wu Li, 2013), with measures that tap into the dynamic flow (R. A. A. Ince et al., 2015). In sum, to understand the information processing subtending dynamic predictions in the brain, we need to understand the task behaviour that the participant engages with and the stimulus features of the task (F Gosselin & Schyns, 2001; Schyns, Bonnar, & Gosselin, 2002) and then trace predictions of these features and their decompositions across the different hierarchical stages along the ventral pathway (Yuille & Kersten, 2006).

5.4 Generalizing from visual features to features in other sensory modalities.

Predictions of stimulus information can in principle occur in any sensory modality. The methodological framework proposed here within the visual modality could straightforwardly extend to other sensory modalities—e.g. auditory—by adapting the design proposed in Figure 1. For example, auditory stimuli can also be decomposed in low- (Daube, Ince, & Gross, 2019) or high-level (Burred, Ponsot, Goupil, Liuni, & Aucouturier, 2019) stimulus properties, for which localizer bottom-up dynamic signatures could be compared with the top-down reactivations of the properties, as we did with our left- and right-presented SF Gabors. From thereon, one could also examine the time-course of predicted multi-modal information in each modality.

To conclude, with a three-stage predictive design, we have traced the dynamic top-down flow of a predicted visual content (left vs. right, LSF vs. HSF Gabor path). We

found that bottom-up-reversed predictions improve early visual cortex representation of the shown stimulus and its subsequent categorization. As discussed, our methodology in principle generalizes to predictions in other sensory modalities and from simple to more complex stimulus features and categorizations.

5.5 Futural work.

With the findings from predicted SF contents as Step 1, I aim to expand the futural research to two aspects. In 5.3.1, I will discuss the Step 2 plan to endow the general and basic concepts of spatial frequencies with specific perception meaning, and further disclose how expectations shape perceptions. In 5.3.2, I will put forward a blueprint for Step 3, aiming to trace the different prediction contents in laminar layers with the fusion of MEG and fMRI.

5.5.1 Applying perceptual specificity to general SF conception

Though Step 1 study has answered some mechanisms about prediction, it is constructed on a fundamental feature – SF. However, according to the comments from participants, they are not very familiar with the conception of SF. Instead, the participants tend to understand SF by "narrow vs. wide bands". This confusion could inspire the futural study to provide specific meaning to understand the mechanism of concrete feature prediction.

In fact, SF is widely used in visual categorization. Suppose when seeing a blossom, we approach and recognize it as a rose. In this procedure, the LSF information at a long distance leads to the categorization of "blossom", and the HSF feature of the flower at a short range decides the categorization of "rose". Though there are many combinations between LSF and HSF guided categorization, most of them have a relationship of affiliation (e.g., rose belongs to flower), which can cause the crossover confounding in the experimental manipulation. Thus, I will choose a stimulus with separate senses between LSF and HSF components.

Under this circumstance, I aim to construct stimuli set using a painting by Dali, which is mutually exclusive in SF but spatially overlapped (Bonnar, Gosselin, & Schyns, 2002). Specifically, in the raw ambiguous figure (Figure 5-2A), LSF features initiate the

perception of Voltaire, and HSF features produce the perception of Nuns. In addition, it has also been proved that the two perceptions derived from the different SF components can activate solid and detached representation in the brain (Zhan et al., 2019), which is also a reason for the choice.

In this plan, we will use two independent components driving respective perceptions: Nuns (Figure 5-2B) vs. Voltaire (Figure 5-2C) components, obtained by separately Gabor sampling the raw picture within pixel-level masks that significantly drive nuns vs. Voltaire perceptual decision, at frequencies of 8, 16 and 32 cycles per degree. Using a similar cueing-categorization design, I will train participants to couple cues with the two components, then in the task to categorize the hybrid (Figure 5-2D) and make a perceptual decision.

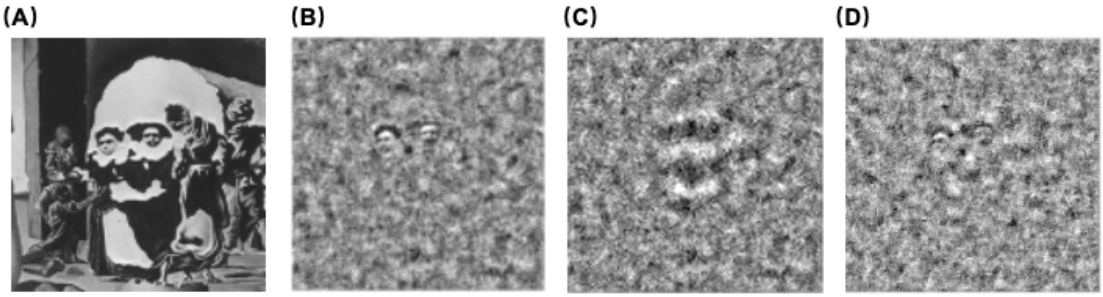


Figure 5-2 Stimuli construction.

(A) Original painting. Part of Dali's painting: Slave Market with the Disappearing Bust of Voltaire **(B) Nuns component.** I obtained the independent Nuns component by Gabor sampling the raw painting within the area driving nuns perception (Bonnar et al., 2002) at frequencies of 8 and 16 cycles per degree and adding the noisy background. **(C) Independent Voltaire component.** I obtained the independent Nuns component by Gabor sampling the raw painting within the area driving nuns perception (Bonnar et al., 2002) at frequencies of 16 and 32 cycles per degree and adding the noisy background. **(D) Hybrid example.** This is constructed by overlaying 50% Nuns (i.e., randomly shuffle the location of 50% pixels and remain the other 50%) with 50% Voltaire.

Besides generalizing findings from the two built studies mentioned in this thesis, this design is further trying to address two further questions:

(1) What is the predicted feature content, and when and where are the concrete predicted contents represented in the brain?

(2) how do the different predictive cues change the representation of stimulus features (i.e., Nuns or Voltaire component) and in turn influence Nuns and Voltaire perceptions by modulating neural activity?

5.5.2 Tracing prediction in laminar layers

As we already built the fundamental SF prediction and have the Step 2 plan to generalize the findings to specific perception prediction, in Step 3, I will model how brain selectively processes pre-stimulus predicted contents. In step 2, I exclude the affiliation relationship in the stimulus design to avoid confounding. However, in Step 3, I will utilize the affiliation relationship to trace the hierarchical prediction. I will ask whether different levels of prediction (e.g., “flower” vs. “rose”) will finally represent in the different laminar layers related to the specific bottom-up processing and if so, how high hierarchical region guides this directional propagation.

Addressing this question requires the fusion of brain measures such as high-field fMRI that finely tap into the laminar layers of these regions (Charles D Gilbert & Wu Li, 2013), with dynamic measures such as E/MEG (R. A. A. Ince et al., 2015). Thus, to generally understand the dynamics of different predictions and representations in the brain, we need to understand the categorization task (e.g. “city vs. New York”), the hierarchical composition of features that represent each category in the participant’s memory, trace their predictions across the visual hierarchy in the feedback flow (Yuille & Kersten, 2006) and their subsequent representation in the feedforward flow when the stimulus is shown. This raises the thorny question of the granularity of visual features that can be predicted from memory at the different levels of the visual hierarchy.

5. Concluding Remarks

"Top-down prediction", named by the implication of a high-to-low hierarchical pathway, can modulate our neural activity, facilitate our behavior or even shape our perception. Starting from an enduring and essential topic, I first discussed the conception, the built theories, the uncovered issues, and the debates on the neural mechanism of top-down, which motivated me to establish three studies to reveal the unsolved mysteries of prediction propagation and facilitation. In my first study, I validated the basic effect and tested the feasibility of a cueing paradigm guiding the following neuroimaging experiment. In my second study, I traced the dynamics of prediction and reconstructed the functional network of prediction communication. In my third study, I further developed the communication network after the predicted contents are shown, and understood the mechanism of the modulation on the stimulus processing provided by the previous prediction. These approaches and results shown in the three studies uncovered the prediction-behavior process in the brain, methodologically provided an open view to study specific information/feature, and contributed to the prediction field of neuroscience.

List of References

Agarwal, M., Agrawal, H., Jain, N., & Kumar, M. (2010). *Face recognition using principle component analysis, eigenface and neural network.* Paper presented at the 2010 International conference on signal acquisition and processing.

Alink, A., Schwiedrzik, C. M., Kohler, A., Singer, W., & Muckli, L. (2010). Stimulus predictability reduces responses in primary visual cortex. *Journal of neuroscience, 30*(8), 2960-2966.

Altmann, C. F., Bülthoff, H. H., & Kourtzi, Z. (2003). Perceptual Organization of Local Elements into Global Shapes in the Human Visual Cortex. *Current Biology, 13*(4), 342-349. doi:10.1016/S0960-9822(03)00052-6

Andersen, R. A., & Cui, H. (2009). Intention, action planning, and decision making in parietal-frontal circuits. *Neuron, 63*(5), 568-583.

Auksztulewicz, R., Myers, N. E., Schnupp, J. W., & Nobre, A. C. (2019). Rhythmic Temporal Expectation Boosts Neural Activity by Increasing Neural Gain. *The Journal of Neuroscience, 39*(49), 9806-9817. doi:10.1523/jneurosci.0925-19.2019

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., . . . Rosen, B. R. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences, 103*(2), 449-454.

Bell, A. J. (2003). *The co-information lattice.* Paper presented at the Proceedings of the Fifth International Workshop on Independent Component Analysis and Blind Signal Separation: ICA.

Benedek, M., Bergner, S., Könen, T., Fink, A., & Neubauer, A. C. (2011). EEG alpha synchronization is related to top-down processing in convergent and divergent thinking. *Neuropsychologia, 49*(12), 3505-3511.

Blank, H., & Davis, M. H. (2016). Prediction errors but not sharpened signals simulate multivoxel fMRI patterns during speech perception. *PLOS Biology, 14*(11), e1002577.

Bolam, J., Boyle, S. C., Ince, R. A., & Delis, I. (2022). Neurocomputational mechanisms underlying cross-modal associations and their influence on perceptual decisions. *NeuroImage, 247*, 118841.

Bonnar, L., Gosselin, F., & Schyns, P. G. (2002). Understanding Dali's Slave Market with the Disappearing Bust of Voltaire: A case study in the scale information driving perception. *Perception, 31*(6), 683-691.

Borst, A., & Theunissen, F. E. (1999). Information theory and neural coding. *Nature neuroscience, 2*(11), 947-957.

Buckner, R. L. (2010). The role of the hippocampus in prediction and imagination. *Annual Review of Psychology, 61*, 27-48.

Bullmore, E., & Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience, 10*(3), 186-198.

Burred, J. J., Ponsot, E., Goupil, L., Liuni, M., & Aucouturier, J.-J. (2019). CLEESE: An open-source audio-transformation toolbox for data-driven experiments in speech and music cognition. *Plos one, 14*(4), e0205943. doi:10.1371/journal.pone.0205943

Chang, A. Y. C., Kanai, R., & Seth, A. K. (2015). Cross-modal prediction changes the timing of conscious access during the motion-induced blindness. *Consciousness and Cognition, 31*, 139-147. doi:https://doi.org/10.1016/j.concog.2014.11.005

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature, 363*(6427), 345-347.

Chen, Y., Martinez-Conde, S., Macknik, S. L., Bereshpolova, Y., Swadlow, H. A., & Alonso, J.-M. (2008). Task difficulty modulates the activity of specific neuronal populations in primary visual cortex. *Nature neuroscience, 11*(8), 974-982.

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and brain Sciences, 36*(3), 181-204.

Contini, E. W., Wardle, S. G., & Carlson, T. A. (2017). Decoding the time-course of object recognition in the human brain: From visual features to categorical decisions. *Neuropsychologia, 105*, 165-176. doi:https://doi.org/10.1016/j.neuropsychologia.2017.02.013

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature neuroscience, 3*(3), 292-297.

Corbetta, M., Kincade, J. M., & Shulman, G. L. (2002). Neural systems for visual orienting and their relationships to spatial working memory. *Journal of cognitive neuroscience, 14*(3), 508-523.

D'Esposito, M., Aguirre, G. K., Zarahn, E., Ballard, D., Shin, R. K., & Lease, J. (1998). Functional MRI studies of spatial and nonspatial working memory. *Cognitive Brain Research, 7*(1), 1-13. doi:https://doi.org/10.1016/S0926-6410(98)00004-4

Daube, C., Ince, R. A., & Gross, J. (2019). Simple acoustic features can explain phoneme-based predictions of cortical responses to speech. *Current Biology, 29*(12), 1924-1937. e1929.

Daube, C., Xu, T., Zhan, J., Webb, A., Ince, R. A., Garrod, O. G., & Schyns, P. G. (2021). Grounding deep neural network predictions of human categorization behavior in understandable functional features: The case of face identity. *Patterns*, 100348.

Davachi, L., & DuBrow, S. (2015). How the hippocampus preserves order: the role of prediction and context. *Trends in Cognitive Sciences, 19*(2), 92-99.

De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception? *Trends in Cognitive Sciences, 22*(9), 764-779.

Den Ouden, H. E., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cerebral Cortex, 19*(5), 1175-1185.

Desimone, R. (1996). Neural mechanisms for visual memory and their role in attention. *Proceedings of the National Academy of Sciences, 93*(24), 13494-13499.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual review of neuroscience, 18*(1), 193-222.

Egner, T., Monti, J. M., & Summerfield, C. (2010). Expectation and surprise determine neural population responses in the ventral visual stream. *Journal of neuroscience, 30*(49), 16601-16608.

Ekman, M., Kok, P., & de Lange, F. P. (2017). Time-compressed preplay of anticipated events in human primary visual cortex. *Nature communications, 8*(1), 1-9.

Essen, D. C. V., Anderson, C. H., & Felleman, D. J. (1992). Information Processing in the Primate Visual System: An Integrated Systems Perspective. *Science, 255*(5043), 419-423. doi:doi:10.1126/science.1734518

Fairhall, A., Shea-Brown, E., & Barreiro, A. (2012). Information theoretic approaches to understanding circuit function. *Current opinion in neurobiology, 22*(4), 653-659.

Flom, M. C., Heath, G. G., & Takahashi, E. (1963). Contour interaction and visual resolution: Contralateral effects. *Science, 142*(3594), 979-980.

Friedman, N. P., & Robbins, T. W. (2022). The role of prefrontal cortex in cognitive control and executive function. *Neuropsychopharmacology, 47*(1), 72-89. doi:10.1038/s41386-021-01132-0

Friston, K. (2005). A theory of cortical responses. *Philosophical transactions of the Royal Society B: Biological sciences, 360*(1456), 815-836. doi:doi:10.1098/rstb.2005.1622

Friston, K. (2010). The free-energy principle: a unified brain theory? *Nature Reviews Neuroscience, 11*(2), 127-138.

Friston, K., & Kiebel, S. (2009). Predictive coding under the free-energy principle. *Philosophical transactions of the Royal Society B: Biological sciences, 364*(1521), 1211-1221.

Friston, K. J., Daunizeau, J., & Kiebel, S. J. (2009). Reinforcement learning or active inference? *Plos one, 4*(7), e6421.

Garrido, M. I., Rowe, E. G., Halász, V., & Mattingley, J. B. (2017). Bayesian Mapping Reveals That Attention Boosts Neural Responses to Predicted and Unpredicted Stimuli. *Cerebral Cortex, 28*(5), 1771-1782. doi:10.1093/cercor/bhx087

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

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience, 14*(5), 350-363.

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature Reviews Neuroscience, 14*, 350. doi:10.1038/nrn3476

Gilbert, C. D., & Sigman, M. (2007). Brain States: Top-Down Influences in Sensory Processing. *Neuron, 54*(5), 677-696. doi:https://doi.org/10.1016/j.neuron.2007.05.019

Gómez-Cuerva, J., & Raymond, J. E. (2011). Perception of facial expression depends on prior attention. *Psychonomic bulletin & review, 18*(6), 1057-1063.

Gosselin, F., & Schyns, P. (2001). Bubbles: A new technique to reveal the use of information in recognition tasks. *Journal of vision, 1*(3), 333-333.

Gosselin, F., & Schyns, P. G. (2003). Superstitious perceptions reveal properties of internal representations. *Psychological Science, 14*(5), 505-509.

Grill-Spector, K., & Malach, R. (2004). The human visual cortex. *Annu. Rev. Neurosci., 27*, 649-677.

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience, 15*(8), 536-548.

Güçlü, U., & van Gerven, M. A. (2015). Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream. *Journal of neuroscience, 35*(27), 10005-10014.

Haegens, S., Händel, B. F., & Jensen, O. (2011). Top-down controlled alpha band activity in somatosensory areas determines behavioral performance in a discrimination task. *Journal of neuroscience, 31*(14), 5197-5204.

Hanks, T. D., Kopec, C. D., Brunton, B. W., Duan, C. A., Erlich, J. C., & Brody, C. D. (2015). Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature, 520*(7546), 220-223.

Hindy, N. C., Ng, F. Y., & Turk-Browne, N. B. (2016). Linking pattern completion in the hippocampus to predictive coding in visual cortex. *Nature neuroscience, 19*(5), 665-667.

Hopfinger, J. B., Buonocore, M. H., & Mangun, G. R. (2000). The neural mechanisms of top-down attentional control. *Nature neuroscience, 3*(3), 284-291.

Ince, R. A., Giordano, B. L., Kayser, C., Rousselet, G. A., Gross, J., & Schyns, P. G. (2017). A statistical framework for neuroimaging data analysis based on mutual information estimated via a gaussian copula. *Human brain mapping, 38*(3), 1541-1573.

Ince, R. A., Paton, A. T., Kay, J. W., & Schyns, P. G. (2021). Bayesian inference of population prevalence. *Elife, 10*, e62461.

Ince, R. A. A., Jaworska, K., Gross, J., Panzeri, S., van Rijsbergen, N. J., Rousselet, G. A., & Schyns, P. G. (2016). The Deceptively Simple N170 Reflects Network Information Processing Mechanisms Involving Visual Feature Coding and Transfer Across Hemispheres. *Cerebral Cortex, 26*(11), 4123-4135. doi:10.1093/cercor/bhw196

Ince, R. A. A., van Rijsbergen, N. J., Thut, G., Rousselet, G. A., Gross, J., Panzeri, S., & Schyns, P. G. (2015). Tracing the Flow of Perceptual Features in an Algorithmic Brain Network. *Scientific Reports, 5*, 17681. doi:10.1038/srep17681 https://www.nature.com/articles/srep17681#supplementary-information

Jafarpour, A., Fuentemilla, L., Horner, A. J., Penny, W., & Duzel, E. (2014). Replay of very early encoding representations during recollection. *Journal of neuroscience, 34*(1), 242-248.

Jaworska, K., Yan, Y., Van Rijsbergen, N. J., Ince, R. A., & Schyns, P. G. (2022). Different computations over the same inputs produce selective behavior in algorithmic brain networks. *Elife, 11*, e73651.

Jehee, J. F., Rothkopf, C., Beck, J. M., & Ballard, D. H. (2006). Learning receptive fields using predictive feedback. *Journal of Physiology-Paris, 100*(1-3), 125-132.

JIayu Zhan; G. B. Garrod, O. V. R., Nicola; Philippe. (2018). Someone like you? Modelling Face Memory Reveals Task-generalizable Representations.

Jueptner, M., Stephan, K. M., Frith, C. D., Brooks, D. J., Frackowiak, R. S., & Passingham, R. E. (1997). Anatomy of motor learning. I. Frontal cortex and attention to action. *Journal of neurophysiology, 77*(3), 1313-1324.

Kaposvari, P., Kumar, S., & Vogels, R. (2018). Statistical learning signals in macaque inferior temporal cortex. *Cerebral Cortex, 28*(1), 250-266.

Kinchla, R. A., & Wolfe, J. M. (1979). The order of visual processing:“Top-down,”“bottom-up,” or “middle-out”. *Perception & psychophysics, 25*(3), 225-231.

Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3?

Kok, P., Brouwer, G. J., van Gerven, M. A., & de Lange, F. P. (2013). Prior expectations bias sensory representations in visual cortex. *Journal of neuroscience, 33*(41), 16275-16284.

Kok, P., & de Lange, Floris P. (2014). Shape Perception Simultaneously Up- and Downregulates Neural Activity in the Primary Visual Cortex. *Current Biology, 24*(13), 1531-1535. doi:https://doi.org/10.1016/j.cub.2014.05.042

Kok, P., Failing, M. F., & de Lange, F. P. (2014). Prior expectations evoke stimulus templates in the primary visual cortex. *Journal of cognitive neuroscience, 26*(7), 1546-1554.

Kok, P., Jehee, J. F., & De Lange, F. P. (2012a). Less is more: expectation sharpens representations in the primary visual cortex. *Neuron, 75*(2), 265-270.

Kok, P., Jehee, Janneke F. M., & de Lange, Floris P. (2012b). Less Is More: Expectation Sharpens Representations in the Primary Visual Cortex. *Neuron, 75*(2), 265-270. doi:https://doi.org/10.1016/j.neuron.2012.04.034

Kok, P., Mostert, P., & De Lange, F. P. (2017). Prior expectations induce prestimulus sensory templates. *Proceedings of the National Academy of Sciences, 114*(39), 10473-10478.

Kok, P., Rahnev, D., Jehee, J. F., Lau, H. C., & De Lange, F. P. (2012). Attention reverses the effect of prediction in silencing sensory signals. *Cerebral Cortex, 22*(9), 2197-2206.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. *Trends in Cognitive Sciences, 17*(1), 26-49.

Kreiman, G., Koch, C., & Fried, I. (2000). Imagery neurons in the human brain. *Nature, 408*(6810), 357-361. doi:10.1038/35042575

Kumar, S., Kaposvari, P., & Vogels, R. (2017). Encoding of Predictable and Unpredictable Stimuli by Inferior Temporal Cortical Neurons. *Journal of cognitive neuroscience, 29*(8), 1445-1454. doi:10.1162/jocn_a_01135

Kveraga, K., Ghuman, A. S., & Bar, M. (2007). Top-down predictions in the cognitive brain. *Brain and Cognition, 65*(2), 145-168. doi:https://doi.org/10.1016/j.bandc.2007.06.007

Lee, T. S., & Mumford, D. (2003a). Hierarchical Bayesian inference in the visual cortex. *JOSA A, 20*(7), 1434-1448.

Lee, T. S., & Mumford, D. (2003b). Hierarchical Bayesian inference in the visual cortex. *Journal of the Optical Society of America A, 20*(7), 1434-1448. doi:10.1364/JOSAA.20.001434

Li, W., Piëch, V., & Gilbert, C. D. (2006). Contour saliency in primary visual cortex. *Neuron, 50*(6), 951-962.

Linde-Domingo, J., Treder, M. S., Kerrén, C., & Wimber, M. (2019). Evidence that neural information flow is reversed between object perception and object reconstruction from memory. *Nature communications, 10*(1), 1-13.

Lobier, M., Palva, J. M., & Palva, S. (2018). High-alpha band synchronization across frontal, parietal and visual cortex mediates behavioral and neuronal effects of visuospatial attention. *NeuroImage, 165*, 222-237.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *Journal of neurophysiology, 77*(1), 24-42.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological bulletin, 109*(2), 163.

Malcolm, G. L., Nuthmann, A., & Schyns, P. G. (2014). Beyond gist: Strategic and incremental information accumulation for scene categorization. *Psychological Science, 25*(5), 1087-1097.

Margalit, E., Jamison, K. W., Weiner, K. S., Vizioli, L., Zhang, R.-Y., Kay, K. N., & Grill-Spector, K. (2020). Ultra-high-resolution fMRI of human ventral temporal cortex reveals differential representation of categories and domains. *Journal of neuroscience, 40*(15), 3008-3024.

Massey, J. (1990). *Causality, feedback and directed information.* Paper presented at the Proc. Int. Symp. Inf. Theory Applic.(ISITA-90).

McGill, W. (1954). Multivariate information transmission. *Transactions of the IRE Professional Group on Information Theory, 4*(4), 93-111.

Michalareas, G., Vezoli, J., Van Pelt, S., Schoffelen, J.-M., Kennedy, H., & Fries, P. (2016). Alpha-beta and gamma rhythms subserve feedback and feedforward influences among human visual cortical areas. *Neuron, 89*(2), 384-397.

Michelmann, S., Bowman, H., & Hanslmayr, S. (2016). The Temporal Signature of Memories: Identification of a General Mechanism for Dynamic Memory Replay in Humans. *PLOS Biology, 14*(8), e1002528. doi:10.1371/journal.pbio.1002528

Moran, J., & Desimone, R. (1985). Selective attention gates visual processing in the extrastriate cortex. *Science, 229*(4715), 782-784.

Motter, B. C. (1993). Focal attention produces spatially selective processing in visual cortical areas V1, V2, and V4 in the presence of competing stimuli. *Journal of neurophysiology, 70*(3), 909-919.

Muckli, L., Kohler, A., Kriegeskorte, N., & Singer, W. (2005). Primary Visual Cortex Activity along the Apparent-Motion Trace Reflects Illusory Perception. *PLOS Biology, 3*(8), e265. doi:10.1371/journal.pbio.0030265

Mumford, D. (1992). On the computational architecture of the neocortex. *Biological cybernetics, 66*(3), 241-251.

Murray, S. O., Kersten, D., Olshausen, B. A., Schrater, P., & Woods, D. L. (2002). Shape perception reduces activity in human primary visual cortex. *Proceedings of the National Academy of Sciences, 99*(23), 15164-15169.

Nelken, I., & Chechik, G. (2007). Information theory in auditory research. *Hearing research, 229*(1-2), 94-105.

Nichols, T. E., & Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping, 15*(1), 1-25.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J.-M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience, 2011*.

Pinto, Y., van Gaal, S., de Lange, F. P., Lamme, V. A. F., & Seth, A. K. (2015). Expectations accelerate entry of visual stimuli into awareness. *Journal of vision, 15*(8), 13-13. doi:10.1167/15.8.13

Posner, M. I., & Petersen, S. E. (1990). The attention system of the human brain. *Annual review of neuroscience, 13*(1), 25-42.

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of experimental psychology: general, 109*(2), 160.

Puri, A. M., Wojciulik, E., & Ranganath, C. (2009). Category expectation modulates baseline and stimulus-evoked activity in human inferotemporal cortex. *Brain Research, 1301*, 89-99.

Quiroga, R. Q., & Panzeri, S. (2009). Extracting information from neuronal populations: information theory and decoding approaches. *Nature Reviews Neuroscience, 10*(3), 173-185.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience, 2*(1), 79-87.

Reddy, L., Poncet, M., Self, M. W., Peters, J. C., Douw, L., Van Dellen, E., . . . Roelfsema, P. R. (2015). Learning of anticipatory responses in single neurons of the human medial temporal lobe. *Nature communications, 6*(1), 1-8.

Rolls, E. T., & Treves, A. (2011). The neuronal encoding of information in the brain. *Progress in neurobiology, 95*(3), 448-490.

Rowe, J., Friston, K., Frackowiak, R., & Passingham, R. (2002). Attention to action: specific modulation of corticocortical interactions in humans. *NeuroImage, 17*(2), 988-998.

Rowe, J. B., Toni, I., Josephs, O., Frackowiak, R. S. J., & Passingham, R. E. (2000). The prefrontal cortex: response selection or maintenance within working memory? *Science, 288*(5471), 1656-1660.

Sanches, M., Caetano, S., Nicoletti, M., Monkul, E. S., Chen, H. H., Hatch, J. P., . . . Soares, J. C. (2009). An MRI-based approach for the measurement of the dorsolateral prefrontal cortex in humans. *Psychiatry research, 173*(2), 150-154. doi:10.1016/j.pscychresns.2009.02.007

Schendan, H. E., & Stern, C. E. (2007). Mental rotation and object categorization share a common network of prefrontal and dorsal and ventral regions of posterior cortex. *NeuroImage, 35*(3), 1264-1277.

Schyns, P. G., Bonnar, L., & Gosselin, F. (2002). Show me the features! Understanding recognition from the use of visual information. *Psychological Science, 13*(5), 402-409.

Schyns, P. G., Gosselin, F., & Smith, M. L. (2009). Information processing algorithms in the brain. *Trends in Cognitive Sciences, 13*(1), 20-26.

Schyns, P. G., Petro, L. S., & Smith, M. L. (2007). Dynamics of visual information integration in the brain for categorizing facial expressions. *Current Biology, 17*(18), 1580-1585.

Schyns, P. G., Zhan, J., Jack, R. E., & Ince, R. A. (2020). Revealing the information contents of memory within the stimulus information representation framework. *Philosophical Transactions of the Royal Society B, 375*(1799), 20190705.

Seriès, P., & Seitz, A. (2013). Learning what to expect (in visual perception). *Frontiers in human neuroscience, 7*, 668.

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal, 27*(3), 379-423.

Sharon, D., Hämäläinen, M. S., Tootell, R. B., Halgren, E., & Belliveau, J. W. (2007). The advantage of combining MEG and EEG: comparison to fMRI in focally stimulated visual cortex. *NeuroImage, 36*(4), 1225-1235.

Shulman, G. L., d'Avossa, G., Tansy, A. P., & Corbetta, M. (2002). Two attentional processes in the parietal lobe. *Cerebral Cortex, 12*(11), 1124-1131.

Shulman, G. L., Ollinger, J. M., Akbudak, E., Conturo, T. E., Snyder, A. Z., Petersen, S. E., & Corbetta, M. (1999). Areas involved in encoding and applying directional expectations to moving objects. *Journal of neuroscience, 19*(21), 9480-9496.

Siegel, M., Donner, T. H., Oostenveld, R., Fries, P., & Engel, A. K. (2008). Neuronal Synchronization along the Dorsal Visual Pathway Reflects the Focus of Spatial Attention. *Neuron, 60*(4), 709-719. doi:https://doi.org/10.1016/j.neuron.2008.09.010

Smith, M. L., Gosselin, F., & Schyns, P. G. (2006). Perceptual moments of conscious visual experience inferred from oscillatory brain activity. *Proceedings of the National Academy of Sciences, 103*(14), 5626-5631.

Smith, M. L., Gosselin, F., & Schyns, P. G. (2012). Measuring internal representations from behavioral and brain data. *Current Biology, 22*(3), 191-196.

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *NeuroImage, 44*(1), 83-98.

Sotiropoulos, G., Seitz, A. R., & Seriès, P. (2011). Changing expectations about speed alters perceived motion direction. *Current Biology, 21*(21), R883-R884. doi:https://doi.org/10.1016/j.cub.2011.09.013

Spitzer, H., Desimone, R., & Moran, J. (1988). Increased attention enhances both behavioral and neuronal performance. *Science, 240*(4850), 338-340.

Stanislaw, H., & Todorov, N. (1999). Calculation of signal detection theory measures. *Behavior research methods, instruments, & computers, 31*(1), 137-149.

Stein, T., & Peelen, M. V. (2015). Content-specific expectations enhance stimulus detectability by increasing perceptual sensitivity. *Journal of experimental psychology: general, 144*(6), 1089.

Sterzer, P., Frith, C., & Petrovic, P. (2008). Believing is seeing: expectations alter visual awareness. *Current Biology, 18*(16), R697-R698. doi:https://doi.org/10.1016/j.cub.2008.06.021

Stokes, M., Thompson, R., Nobre, A. C., & Duncan, J. (2009). Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences, 106*(46), 19569-19574.

Summerfield, C., & de Lange, F. P. (2014a). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience, 15*(11), 745-756. doi:10.1038/nrn3838

Summerfield, C., & de Lange, F. P. (2014b). Expectation in perceptual decision making: neural and computational mechanisms. *Nature Reviews Neuroscience, 15*, 745. doi:10.1038/nrn3838

Summerfield, C., & Egner, T. (2016). Feature-based attention and feature-based expectation. *Trends in Cognitive Sciences, 20*(6), 401-404.

Theeuwes, J. (2010). Top–down and bottom–up control of visual selection. *Acta psychologica, 135*(2), 77-99. doi:https://doi.org/10.1016/j.actpsy.2010.02.006

Todorovic, A., & de Lange, F. P. (2012). Repetition suppression and expectation suppression are dissociable in time in early auditory evoked fields. *Journal of neuroscience, 32*(39), 13389-13395.

Todorovic, A., van Ede, F., Maris, E., & de Lange, F. P. (2011). Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *Journal of neuroscience, 31*(25), 9118-9123.

Treder, M. S. (2020). MVPA-Light: a classification and regression toolbox for multi-dimensional data. *Frontiers in Neuroscience, 14*, 289.

Van Essen, D. C., Anderson, C. H., & Felleman, D. J. (1992). Information processing in the primate visual system: an integrated systems perspective. *Science, 255*(5043), 419-423.

Victor, J. D. (2006). Approaches to information-theoretic analysis of neural activity. *Biological theory, 1*(3), 302-316.

Wacongne, C., Labyt, E., van Wassenhove, V., Bekinschtein, T., Naccache, L., & Dehaene, S. (2011). Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Sciences, 108*(51), 20754-20759.

Wallenstein, G. V., Hasselmo, M. E., & Eichenbaum, H. (1998). The hippocampus as an associator of discontiguous events. *Trends in neurosciences, 21*(8), 317-323.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature neuroscience, 5*(6), 598-604.

Wyart, V., Nobre, A. C., & Summerfield, C. (2012). Dissociable prior influences of signal probability and relevance on visual contrast sensitivity. *Proceedings of the National Academy of Sciences, 109*(9), 3593-3598.

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *Trends in Cognitive Sciences, 10*(7), 301-308.

Zhan, J., Ince, R. A., Van Rijsbergen, N., & Schyns, P. G. (2019). Dynamic construction of reduced representations in the brain for perceptual decision behavior. *Current Biology, 29*(2), 319-326. e314.