Neural mechanisms underlying human cognitive control and working memory

A dissertation presented

by

Yuchen Xiao

to

The Division of Medical Sciences

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biological and Biomedical Sciences

Harvard University

Cambridge, Massachusetts

July 2022

Dissertation Advisor: Gabriel Kreiman Yuchen Xiao

Neural mechanisms underlying human cognitive control and working memory

Abstract

Cognitive control involves flexibly combining multiple sensory inputs with task-dependent goals during decision making. Several tasks involving distinct sensory inputs and motor outputs have been proposed to examine cognitive control, including the Stroop, the Eriksen-flanker, and the multi-source interference task. It remains unclear whether neural signals during cognitive control extrapolate across different conditions within each task. Furthermore, because these tasks have mostly been studied independently, it remains elusive whether the neural signatures of cognitive control reflect an abstract control mechanism or specific combinations of sensory and behavioral aspects of each task. To address these questions, we recorded invasive neurophysiological signals from 16 subjects and directly compared neural responses within and across tasks. Neural activity patterns in the theta (4-8 Hz) and high-gamma (70-120 Hz) frequency bands differed between incongruent and congruent conditions, revealing strong modulation by conflict. These neural signals were specific to each task, generalizing within a task but not across tasks. These results highlight the complex interplay between sensory inputs, motor outputs, and task demands underlying cognitive control processes.

Working memory is an essential cognitive function that is important for every aspect of life. In this study we implemented a well-known memory matching game to investigate complex

human working memory behaviors under rich contexts. Subjects were initially presented with covered images on a board and were instructed to reveal any two images each time, until all matching pairs of images were located. We recorded intracranial field potentials from 20 pharmacologically-intractable epilepsy patients while they were playing the game. Leveraging generalized linear models to assess the relative contribution of multiple parameters on the neural responses simultaneously, we found that neural activities in the gamma band (30-150 Hz) captured a wide array of working memory status including novelty, familiarity, or recency. The ability to represent the recency of pair existed only during successful associative recall, which was forecasted by distinctive gamma responses.

Contents

Figures

Tables

Acknowledgements

Nothing in this dissertation could have been achieved without the help and support from the following people. I would like to thank my thesis advisor Gabriel Kreiman for providing enormous support throughout my Ph.D. study. He is undoubtedly both a great scientist and a mentor. I would like to thank Professors Joseph Arboleda-Velasquez and Mark Andermann, who served on my dissertation advisory committee, for providing thoughtful inputs on my scientific projects as well as encouragement and guidance throughout my Ph.D. journey.

I would like to thank my colleague Paula Sanchez, a master student in our lab, who worked with me together on the working memory project and made great contributions. I'd like to thank Ruijie Wu, my colleague and friend in Beijing, who collected plenty of intracranial EEG data.

I would like to thank neurosurgeons and neurologists who made data collection from epilepsy patients possible: Dr. Joseph R. Madsen and Dr. Scellig Stone at Boston Children's Hospital (BCH); Dr. Chien-Chen Chou and Dr. Hsiang-Yu Yu at Taipei Veterans General Hospital (TVGH); Dr. William Anderson and Dr. Nathan Crone at Hopkins; Dr. Daniel Weisholtz and Dr. Garth Rees Cosgrove at Brigham and Women's Hospital (BWH); and Dr. Guoguang Zhao and Dr. Yongzhi Shan at Xuanwu Hospital in Beijing. I'd also like to thank Eun-Hyoung Park at BCH, Brannon Cha, and Dustine Reich at BWH, Yen-Cheng Shih at TVGH, Penghu Wei at Xuanwu Hospital, Ian Reucroft, Yousef Salimpour, and Yujing Wang at Hopkins, who also assisted in data collection in many ways.

I would like to thank my dissertation examination committee: Professors Joseph

Arboleda-Velasquez, Alexander Rotenberg, Mark Richardson, and Ueli Rutishauser for their dedication in reading my thesis.

I would like to thank my fellow lab members for sharing their valuable ideas and thoughts, helping solve technical issues, and creating a delightful working environment. They include but not limited to: Jie Zheng, Hanlin Tang, Pranav Misra, Mengmi Zhang, Katarina Bendtz, Jerry Wang, Jiye Kim, Joseph Olson, Shane Shang, David Mazumder, and Will Xiao.

I would like to thank Harvard University and the BBS (Biological and Biomedical Sciences) program for making my education experience enjoyable. I cannot be more proud to have been studying in here.

Last but not least, I would like to thank my parents for their unconditional support in all walks of life. I am forever in your debt.

Introduction

Cognitive neuroscience is the field of study focusing on the neural substrates of mental processes. Common human mental processes, or cognitive functions, include perception, memory, emotion, thinking, and reasoning. Mental phenomena are carried out by neurons which communicate through action potentials, the electrical pulse that is sent from one neuron to another. Researchers have developed diverse methods to measure neuronal activities at different levels to probe the neural basis of cognition and consciousness, from single-unit recording that can measure action potentials (Fried et al., 2014), to sEEG (stereo-electroencephalography) and ECoG (electrocorticography) that record the local field potentials (LFP) or intracranial field potentials (IFP) which reflect the electrical activities from populations of neurons, to functional MRI (fMRI) that images brain dynamics through measuring the BOLD (blood-oxygen-level-dependent) signal, and other technologies like scalp-EEG, PET (positron emission tomography), and MEG (magnetoencephalography), etc. Single-unit recording, sEEG, and ECoG are invasive techniques and exhibit the highest spatiotemporal resolution and signal-to-noise ratio that are the keys to the study of human cognition (Dubey and Ray, 2019; Johnson and Knight, 2015; Mukamel and Fried, 2012). SEEG and ECoG were originally developed for localizing epileptogenic foci for pharmacologically-intractable epilepsy patients and are currently widely used for addressing neuroscience questions. Here we used sEEG and ECoG to record intracranial field potentials in order to investigate the neural mechanisms underlying human cognitive control and working memory.

Cognitive control refers to the ability to selectively process information and guide

behavior in line with overriding goals, while tune down behaviors toward goal-irrelevant but often more salient stimuli (Botvinick et al., 2004). Compared with most animals, human beings can make sense of substantially more information from the surrounding environment, which yields tremendous behavioral options. However, richer information tends to introduce greater confusion, interference, and anxiety (Bawden and Robinson, 2009; Miller and Cohen, 2001), and thus requires effective control processes to ensure appropriate actions toward achieving internally generated goals and plans. Failure of cognitive control can be troublesome, frustrating, and harmful, from minor instances like driving to a routine but wrong destination to severe conditions like environmental dependency syndrome (Lhermitte, 1986) and various forms of addictions (Groman and Jentsch, 2012; Hyman, 2007).

Neuroscientists have been studying the mechanisms of cognitive control using experimental tasks that contain conflicting information and require subjects to select and respond to goal-relevant but usually less salient component(s) of the stimuli. The Stroop task is the most widely used paradigm both in clinical (Assef et al., 2007) and research settings (MacLeod, 1991) for assessing the capacity of cognitive control. In its original version (Stroop, 1935), subjects were instructed to read lists of color words like “red” and “green” as well as name the ink color of words. Stroop found that incongruent ink color did not interfere with reading but with naming the color of the words. For example, the word “red” written in green ink took longer reaction time to name its color (green) than read the word (red). Such “conflict effect” might be due to the prepotent tendency of reading or mentally reading words, a response that needs to be suppressed in order to achieve the task goal. The conflict effect then became the center of cognitive control

studies where researchers have been taking efforts for identifying its neural substrates.

A key neural signature of cognitive control is the discrepancy in neural responses between congruent and incongruent conditions, which is usually manifested by higher activation in specific brain regions when there is a presence of conflict (Barch et al., 2001; Bunge et al., 2002; Bush and Shin, 2006; Caruana et al., 2014; Egner and Hirsch, 2005b; Fan et al., 2003; Fu et al., 2022; Milham and Banich, 2005; Parris et al., 2019; Sheth et al., 2012; Tang et al., 2016). Such neural effects are typically interpreted as domain-general or abstract cognitive control processes. A reasonable concern, if not flaw, is that ascribing neural conflict effects discovered in one task to domain-general cognitive control processes, which by definition should be invariant to whatever task used. By studying only one task, it is impossible to draw any conclusion about domain-generality. Although one can combine multiple studies that used various tasks, this attempt cannot address the issues of subject variability and fine somatotopic organization within a single brain area (Paus et al., 1993). The anterior cingulate cortex (ACC) is proposed to monitor the occurrence of conflict and direct strategic adjustments in cognitive control (Botvinick et al., 2004; Carter et al., 1998), which is widely assumed to be a domain-general function. However, several studies expressed disagreement on the domain-general function of ACC. For example, (Turken and Swick, 1999) reported that focal right ACC lesion led to selective impairment in performing executive tasks (including a Stroop-like paradigm) that required manual but not vocal responses, demonstrating that different response modalities may recruit distinct ACC subregions to monitor or process conflict. A recent human single-neuron study (Ebitz et al., 2020) using the multi-source interference task reported that ACC neurons address conflict by magnifying task-specific characteristics

rather than serving a domain-general monitoring role. Other studies that suggest a non-domain-general or even nonessential role of ACC in cognitive control include (Cole et al., 2009; Heilbronner and Hayden, 2016; Nakamura et al., 2005; Parris et al., 2019; van Veen and Carter, 2005). It is important to note that the ACC is certainly not the only region that orchestrates cognitive control; nevertheless, cognitive control has been demonstrated to recruit widely distributed areas including the prefrontal cortex with a higher emphasis on the dorsolateral portion (Botvinick et al., 2001; Bunge et al., 2002; Bush and Shin, 2006; Coulthard et al., 2008; Egner and Hirsch, 2005b; Liston et al., 2006; Menon and D'Esposito, 2022; Miller, 2000; Miller and Cohen, 2001; Shenhav et al., 2013; Tang et al., 2016; Widge et al., 2019), the superior frontal lobe (Egner and Hirsch, 2005b; Fan et al., 2003; Milham and Banich, 2005; Tully et al., 2014), the middle frontal lobe (Bush and Shin, 2006; Egner and Hirsch, 2005a, b; Fan et al., 2003; Janssens et al., 2018; Milham and Banich, 2005; Ridderinkhof et al., 2004; Tang et al., 2016), the inferior frontal lobe (Banich, 2019; Egner and Hirsch, 2005a; Milham and Banich, 2005; Parris et al., 2019; van Veen et al., 2001), the orbitofrontal cortex (Fan et al., 2003; Kuusinen et al., 2018; Tang et al., 2016), the motor, premotor, and supplementary motor cortices (Bunge et al., 2002; Bush and Shin, 2006; Caruana et al., 2014; Fan et al., 2003; Fu et al., 2022), the parietal lobe (Bunge et al., 2002; Coulthard et al., 2008; Egner and Hirsch, 2005b; Fan et al., 2003; Liston et al., 2006; Milham and Banich, 2005; Milham et al., 2001; Parris et al., 2019; van Veen and Carter, 2005), the temporal lobe (Bush and Shin, 2006; Egner and Hirsch, 2005a; Milham and Banich, 2005; Sani et al., 2021), and the occipital lobe (Fan et al., 2003; Janssens et al., 2018; Milham and Banich, 2005). To note, there might be overlap among these regions due to different nomenclatures or parcellations used in

different studies. Unfortunately, there is even less discussion of domain-generality or domain-specificity regarding non-ACC regions. By recording intracranial field potentials from 1,877 widely located depth electrodes in 16 pharmacologically-intractable epilepsy patients who completed three cognitive control tasks in one session, we directly compared the neurophysiological responses under distinct response modalities and processing domains. Our results indicate that cognitive control processes are largely task-specific and suggest complex interaction between input/output modes and task demands.

Working memory refers to the ability to temporarily maintain information in mind (Baddeley, 1992), which is quickly forgotten when attention is directed elsewhere. It is an essential cognitive faculty that is constantly functioning. During every second of awake time, one holds in mind what he or she is doing and what the purpose of that action is. It is not difficult to perceive how loss of working memory can impact life; imagine if one always losses track of the internal goals or cannot remember what went on 5 seconds ago during a conversation. The earliest attempts to "localize" working memory functions in the human brain relied on the observations from patients with brain lesions, who concurrently exhibited impairments in working memory capacity. Several lesion studies have reported the critical role of the prefrontal cortex (PFC) in working memory (Baddeley, 1986; Courtney et al., 1998; Muller et al., 2002; Owen et al., 1990; Warren et al., 1964). In general, these studies observed deficits in performing working memory tasks consequent to PFC lesions. PFC's involvement in working memory function has also been clearly demonstrated with evidence from intracranial recordings, fMRI, and PET studies (Barbey et al., 2011; Bergmann et al., 2012a; Courtney et al., 1998; Fuster and Alexander, 1971; Jiang et al., 2000; Lara and Wallis, 2015; Miller et al., 1996; Ranganath et al., 2004;

Smith et al., 1995). In addition to the prefrontal cortex, there is growing evidence suggesting the roles of the ventral visual stream (Bergmann et al., 2012a; Kucewicz et al., 2014; Ranganath et al., 2004; Rutishauser et al., 2021) and the medial temporal lobe (MTL) (Bergmann et al., 2012a; Crane and Milner, 2005; Duncan et al., 2009; Giovanello et al., 2003; Johnson and Knight, 2015; Murray et al., 2014; Paller and McCarthy, 2002; Ranganath et al., 2004; Yonelinas, 2013) in various working memory processes, despite that the MTL was historically believed to serve long-term memory only (Cave and Squire, 1992; Scoville and Milner, 1957; Squire and Zola-Morgan, 1991).

The ability to judge a scene, a person, or an object as novel or familiar is important for appropriate memory behaviors. Differentiating novelty versus familiarity helps encode new information (Knight, 1996) and quickly identify previously encountered entities. Human and monkey single unit studies have identified neurons in the medial temporal lobe that are tuned to novelty or familiarity (Fried et al., 1997; Rutishauser et al., 2006; Rutishauser et al., 2021; Rutishauser et al., 2008; Viskontas et al., 2006; Xiang and Brown, 1998). Responses to novelty or familiarity have also been reported by intracranial EEG (Baudena et al., 1995; Murray et al., 2014; Wang et al., 2005; Zaehle et al., 2013), scalp-EEG (Courchesne et al., 1975; Duzel et al., 2004; Friedman et al., 2001; Knight, 1996; Opitz et al., 1999), and fMRI studies (Daselaar et al., 2006; de Chastelaine et al., 2017; Duzel et al., 2004; Kohler et al., 2005; Opitz et al., 1999), which collectively demonstrate that neural responses to novelty or familiarity might be more distributed than localized to the MTL. Most studies investigating novelty or familiarity used long-term recognition or episodic memory paradigms, leaving their counterparts in short-term or working memory less examined. Furthermore, in real world, the level of familiarity with

an entity may constantly change, as new information takes over attention and memory decays. How the human brain keeps track of the dynamics of familiarity is poorly understood.

Associative memory refers to the ability to associate or link two or more usually unrelated items. Neural mechanisms of associative memory have several characteristics: (1) neurons are able to learn associations and become selective to distinct components within an association (Duzel et al., 2004; Ison et al., 2015; Ranganath et al., 2004; Rutishauser et al., 2021; Sakai and Miyashita, 1991; Zhou et al., 2007); (2) the presence of the cue itself is sufficient to activate the mental representation of its pairing information and thereby leads to successful associative memory retrieval (Bergmann et al., 2012a; Bunge et al., 2004; Ranganath et al., 2004; Staresina et al., 2019; Zhou et al., 2007); 3) successful versus unsuccessful retrieval, as well as strong vs. weak association, can be manifested by differential neural activities (Bunge et al., 2004; Staresina et al., 2019).

Building upon these important findings, several questions remain to be answered. First, since associative memory has often been studied with rather simple task paradigms where human or animal subjects were instructed or trained to choose an object's associate or pair from a highly limited range of options (Bunge et al., 2004; Sakai and Miyashita, 1991; Staresina et al., 2019; Zhou et al., 2007) or simply report whether the presented association was right or wrong or whether it matched the sample (Bergmann et al., 2012a; Courtney et al., 1998; Duncan et al., 2009; Duzel et al., 2004; Ranganath et al., 2004), it is difficult to assess how one retrieves associated information based on the cue from a large repertoire, which reflects more natural scenarios.

Second, it is unknown whether there are distinct brain structures that support

memory recall and those that orchestrate familiarity-based processes in the context of associative working memory. This question points to the long-standing controversy of whether recollection and familiarity are neurally distinct. The dual-process theory states that recollection and familiarity are independent processes and are operated by distinct brain structures (Bowles et al., 2007; Daselaar et al., 2006; Jacoby and Dallas, 1981; Yonelinas, 2001), especially that the hippocampus is involved in recollection but not familiarity (Diana et al., 2007; Eldridge et al., 2000; Yonelinas, 2001, 2013). The opposing view deems that recollection and familiarity are not completely dissociable but can very well be highly intertwined with each other (Merkow et al., 2015; Rutishauser et al., 2008; Wais et al., 2010; Wais et al., 2006; Wixted and Squire, 2011). Although familiarity and recollection are often discussed in the long-term memory settings, these processes do serve short-term or working memory (Danker et al., 2008; Gothe and Oberauer, 2008; Oberauer, 2005; Yonelinas, 2013). Nevertheless, how recollection and familiarity operate under associative working memory has not been elucidated.

To address the abovementioned questions, we recorded intracranial field potentials from 20 epilepsy patients implanted with sEEG of ECoG electrodes when they were playing a well-known memory matching game to examine both non-associative and associative working memory. Subjects were initially presented with covered images and were instructed to locate all pairs of images by revealing two tiles each time. The working memory status of each tile was constantly changing throughout the game, creating a highly complex and dynamic setting. In most blocks, subjects needed to recall the location from a large number of possibilities. We built generalized linear models to estimate the relative contribution of different memory-related parameters to the neural responses. Our

results indicate that neural responses in the gamma (30-150 Hz) band capture novelty and different grades of familiarity as well as signal successful associative memory retrieval. Neural responses during successful, but not unsuccessful associative memory retrieval, also contain information about the level of familiarity, suggesting that recollection and familiarity may not be neurally dissociable during associative working memory processes.

Chapter 1

Cross-task specificity and within-task invariance of cognitive control processes

Introduction

The ability to flexibly route information is central to daily activities, especially when faced with a complex and conflicting interplay of sensory information, choices, and goals. Cognitive control refers to the ability to regulate actions toward achieving overriding goals. Cognitive control is mentally effortful due to the necessity to suppress autonomous responses toward goal-irrelevant but usually salient stimulus attributes (Gratton et al., 1992; Miller and Cohen, 2001). Such costs are unavoidable for successful adaptation to various environments (Diamond, 2013). Impairment in cognitive control is associated with a wide range of mental disorders, including addiction, depression, and schizophrenia (Goschke, 2014; Lesh et al., 2011; Zilverstand et al., 2018). An essential component of cognitive control is conflict resolution, which entails mental operations involving conflict detection and monitoring (Botvinick et al., 2001), response selection and inhibition (Goghari and MacDonald, 2009), performance monitoring and evaluation (Ridderinkhof et al., 2004), and error-detection (Fu et al., 2019; Ridderinkhof et al., 2004; Tang et al., 2016).

Many experimental tasks have been used to study cognitive control during conflict resolution. Paradigmatic examples include the Stroop task (Stroop, 1935), the Eriksen-flanker task (referred to as "Flanker" throughout the text, (Eriksen and Eriksen, 1974), and the multi-source interference task (MSIT, referred to as "Number" throughout the text, (Bush and Shin, 2006)). Common to all these tasks is the comparison between congruent condition and incongruent

condition (**Figure 1**). In the Stroop task, subjects name the font color of a color word (e.g., "red", "green", "blue") when the semantic meaning of the word agrees (congruent condition) or disagrees (incongruent condition) with its font color. In the Flanker task, subjects have to recognize a symbol such as a letter or an arrow, embedded among the same symbols (congruent condition) or different symbols (incongruent condition) (Davelaar and Stevens, 2009; Eriksen and Eriksen, 1974; Mayr et al., 2003). The multi-source interference task (Bush and Shin, 2006) combines multiple dimensions of cognitive interference from the Stroop, Flanker, and Simon (Simon and Berbaum, 1990) tasks. Each MSIT stimulus consists of three numbers (chosen from 0, 1, 2, or 3) in which one number (target) is always different from the other two numbers (distractors). Subjects are instructed to say the location (“one,” “two,” “three”) of the target number under conditions where it is congruent (e.g., 100) or incongruent (e.g., 313) with its position.

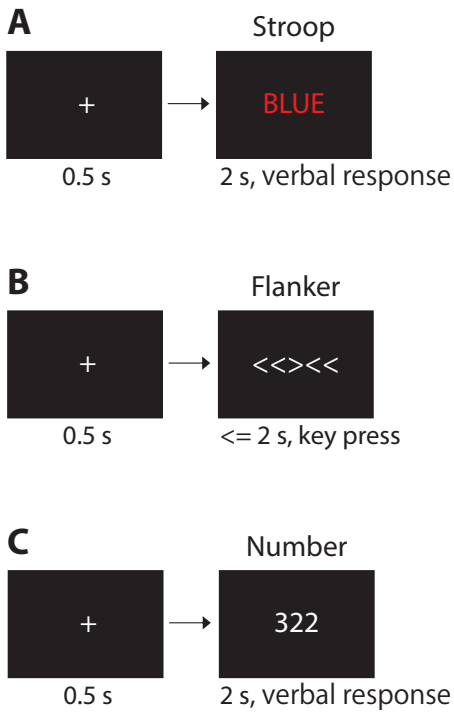

Figure 1. Experimental paradigms. Subjects performed the Stroop (**A**), Flanker (**B**), and Number (**C**) tasks in one session during intracranial neurophysiological recordings with depth electrodes. A standard session contained 18 blocks and each block comprised 30 trials of one task. **A**. The Stroop task required subjects to say the font color. In the congruent condition, the semantic meaning coincided with the font color, while the two conflicted in the incongruent condition. Each stimulus remained on the screen for 2 seconds. **B**. The Flanker task required subjects to press the left or the right key to indicate the direction of the central arrow. In the congruent condition, all the arrows pointed in the same direction while in incongruent condition, the arrow in the middle pointed oppositely from the others (flankers). Each stimulus was immediately off if subjects responded within 2 seconds or off after 2 seconds. **C**. The Number task required subjects to say the position ("one", "two", or "three") where the unique number was located. In the congruent condition, the target number and its position were the same while in the incongruent condition these were different. Each stimulus remained on the screen for 2 seconds. All trials in this figure show incongruent conditions. The order of blocks was 6 repetitions of Stroop, Flanker, and then Number.

The behavioral signature of this family of tasks is longer reaction time (RT) for incongruent stimuli (containing conflict) compared with congruent stimuli (conflict-free). For example, in the Stroop task, subjects take longer to name the font color of the word red when its font color is not red. The increase in reaction time during incongruent conditions is due to interference from irrelevant but conflicting information and the selection among competing motor plans (Goghari and MacDonald, 2009; Miller and Cohen, 2001; Stroop, 1935).

Multiple studies have examined brain signals associated with each one of these cognitive control tasks, including measurements derived from human neuroimaging (Barch et al., 2001; Bunge et al., 2002; Bush and Shin, 2006; Fan et al., 2003; Parris et al., 2019; Robertson et al., 2014; Sani et al., 2021), human scalp electroencephalography (Hanslmayr et al., 2008; Janssens et al., 2018; Robertson et al., 2014), human invasive neurophysiology (Caruana et al., 2014; Koga et al., 2011; Oehrn et al., 2014; Sheth et al., 2012; Tang et al., 2016), and monkey neurophysiology (Blackman et al., 2016; Cole et al., 2009; Li et al., 2019; Nakamura et al., 2005). These studies have described an extensive network of frontal and parietal regions, and to a lesser extent temporal and other regions, that demonstrate distinct activation patterns between congruent and incongruent trials.

Here we evaluate whether there are shared mechanisms involved in conflict monitoring and resolution that are common across different sensory inputs and motor outputs. We focus on how conflict is represented in the brain by directly comparing neurophysiological responses during three cognitive control tasks, analyzing intracranial field potentials from 694 electrodes implanted in patients with pharmacologically-intractable epilepsy. Our first hypothesis is that conflict-related responses should show invariance to the stimulus properties within each task (within-task invariance). For example, in the Stroop task, we would expect that neural responses would distinguish congruent (RED/red, GREEN/green, or BLUE/blue) from incongruent (RED/green, RED/blue, GREEN/red, GREEN/blue, BLUE/red, or BLUE/green) conditions, irrespective of the specific color/semantic combination. Extending this hypothesis of within-task invariance to the comparison across tasks, the assumption of an abstract notion of conflict led to our second hypothesis, that neural responses would distinguish conflict irrespective of whether incongruency is dictated by color, shape, or number stimuli, and also regardless of the specific response

modalities involved (cross-task invariance). Our results are consistent with the first hypothesis; neural signals that show modulation between incongruent and congruent trials are invariant to stimulus attributes within a task. In contrast, our results are inconsistent with the second hypothesis; the majority of the neural responses demonstrate robust modulation between incongruent and congruent trials that is task-specific and does not generalize across tasks.

Results

We recorded intracranial field potentials (IFPs) from 16 epilepsy patients implanted with depth electrodes (**Table 1**). Subjects performed three cognitive control tasks: Stroop, Flanker, and Number (**Methods**, **Figure 1**). Importantly, subjects performed the three tasks during the same session, therefore enabling direct comparisons among the tasks. Each task began with a fixation cross shown for 500 ms at the center of the screen. The Stroop task stimulus consisted of color words ("RED," "GREEN," "BLUE," or the corresponding traditional Chinese characters for patients in Taipei, **Methods**) shown in red, green, or blue font. Subjects were instructed to name the font color (**Figure 1A**). Conflict arose when the font color did not match the meaning of the word. The Flanker task stimulus consisted of five arrows in a horizontal row, and subjects were asked to press the left or the right key to indicate the direction of the central arrow (**Figure 1B**). Conflict arose when the central arrow pointed in the opposite direction to the other four arrows. The Number task required subjects to say the position of the unique number ("one", "two", or "three") among three numbers shown in a horizontal row (**Figure 1C**). Conflict arose when the position of the unique number did not match the actual number (e.g., number "3" in position "1" in the stimulus "322"). For all the tasks, congruent and incongruent conditions, as well as the

stimulus dimensions (word, color, arrow direction, number identity), were randomly interleaved and counterbalanced.

Table 1. Subject information. Information about each participant and number of blocks completed. Bolded entries indicate subjects that performed fewer or more than the default target number of blocks (6 blocks for each task).

Subject number	Age	Gender	Hospital	Number of blocks (Stroop, Flanker, Number)	Number of electrodes
1	13	male	BCH	6, 6, 6	145
2	12	female	BCH	6, 6, 6	168
3	14	female	BCH	6, 6, 6	194
4	13	male	BCH	**3, 3, 3**	215
5	41	female	BWH	**7, 7, 5**	72
6	58	female	BWH	6, 6, 6	119
7	62	male	JHMH	**7, 6, 6**	99
8	41	male	JHMH	6, 6, 6	48
9	26	male	TVGH	6, 6, 6	92
10	27	female	TVGH	6, 6, 6	104
11	29	female	TVGH	6, 6, 6	101
12	29	male	TVGH	6, 6, 6	126
13	25	male	TVGH	6, 6, 6	102
14	20	female	TVGH	**6, 6, 4**	98
15	12	female	TVGH	6, 6, 6	90
16	24	male	TVGH	6, 6, 6	104

Subjects showed behavioral evidence of conflict in the three tasks

Subjects showed high accuracy in all three tasks (**Figure 2A-D**): Stroop (congruent) = 96.8 ± 0.9%; Stroop (incongruent) = 90.1 ± 2.2%; Flanker (congruent) = 96.3 ± 2.8%; Flanker (incongruent) = 90.2 ± 2.9%; Number (congruent) = 96.6 ± 1.7%; Number (incongruent) = 90.4 ± 2.3% (mean±SEM). On average, performance was significantly higher in the congruent condition compared to the incongruent condition in all three tasks; this difference reached statistical significance in the Stroop task (p=0.007, two-sided permutation test; 10,000 iterations), but not in the Flanker (p=0.33) or Number (p=0.39) tasks. These observations are consistent with previous work (Bush and Shin, 2006; Davelaar and Stevens, 2009; Eriksen and Eriksen, 1974; MacLeod, 1991; Sheth et al., 2012; Stroop, 1935; Tang et al., 2016), and are mostly ascribed to a ceiling

effect (Carter and van Veen, 2007). Since accuracy was high in the three tasks, there were not enough error trials to have sufficient power to distinguish incongruent from congruent trials statistically. We focused exclusively on correct trials for the remainder of the study.

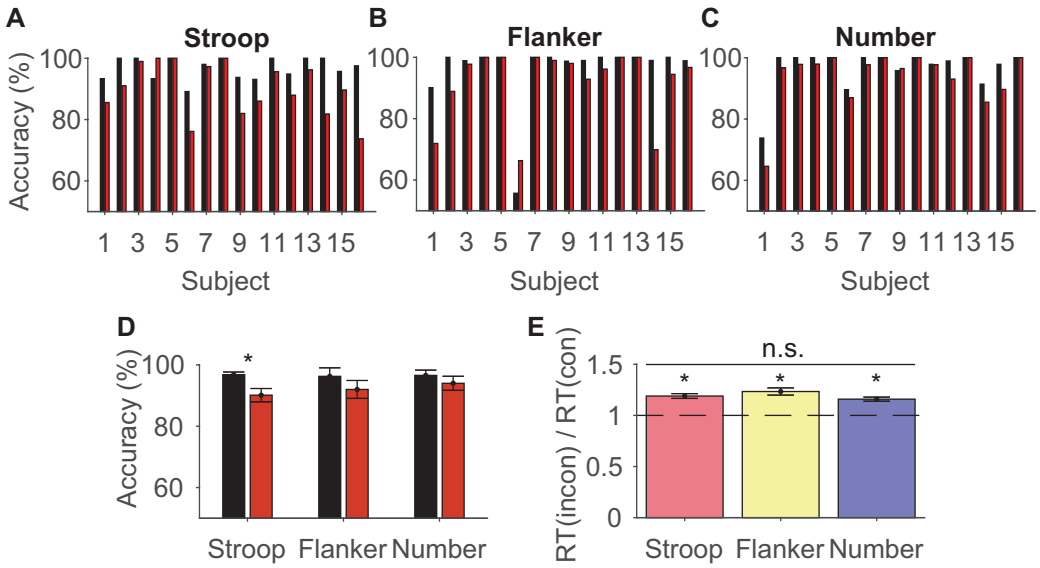

Figure 2. Accuracy and difficulty of tasks. A-C. Accuracy of each task by each subject. Black bars indicate congruent trials and red bars incongruent trials. **D**. Average accuracy across all subjects. Error bars indicate s.e.m. Asterisk indicates statistically significant difference between congruent and incongruent conditions (n=16, permutation test, 10,000 iterations, p=0.007). **E**. There was no difference among the three tasks in task difficulty calculated as the ratio of reaction times of incongruent to congruent trials (one-way ANOVA, p=0.16). Asterisks denote significant differences in each task with respect to the null hypothesis corresponding to a ratio of 1 (two-sided permutation test, 10,000 iterations, α=0.05).

A hallmark of conflict in cognitive control tasks is the longer reaction time associated with incongruent trials (**Figure 3**). As demonstrated in previous work (Davelaar and Stevens, 2009; Sheth et al., 2012; Tang et al., 2016), reaction times were longer during incongruent trials for all three tasks (Stroop: 1,122±8 ms vs. 953±7 ms, $p<0.001$; Flanker: 875±11 ms vs. 722±9 ms, $p<0.001$; Number: 1,110±8 ms vs. 972±8 ms, $p<0.001$; mean±SEM, two-sided permutation test, 10,000 iterations). The longer reaction times during incongruent trials were also statistically significant at the individual subject level in the majority of cases (Stroop: 16/16 subjects; Flanker: 14/16 subjects; Number: 15/16 subjects). Subject number 4 showed no significant difference in the Flanker and Number tasks, but this subject completed only half of a standard session. Absolute

reaction times differed across tasks because of the distinct response modalities (verbal or keypress), because of the different number of response options (2 or 3), and because of the different processing modalities (language, symbol, and number). Therefore, to assess the difficulty of each task, we computed the ratio of reaction times in incongruent versus congruent trials. There was no significant difference in difficulty among the three tasks (**Figure 2E**, p=0.16, non-parametric one-way ANOVA). In sum, behavioral results were consistent with previous work and demonstrated almost ceiling accuracy and longer reaction times associated with incongruent than congruent trials across all the three tasks.

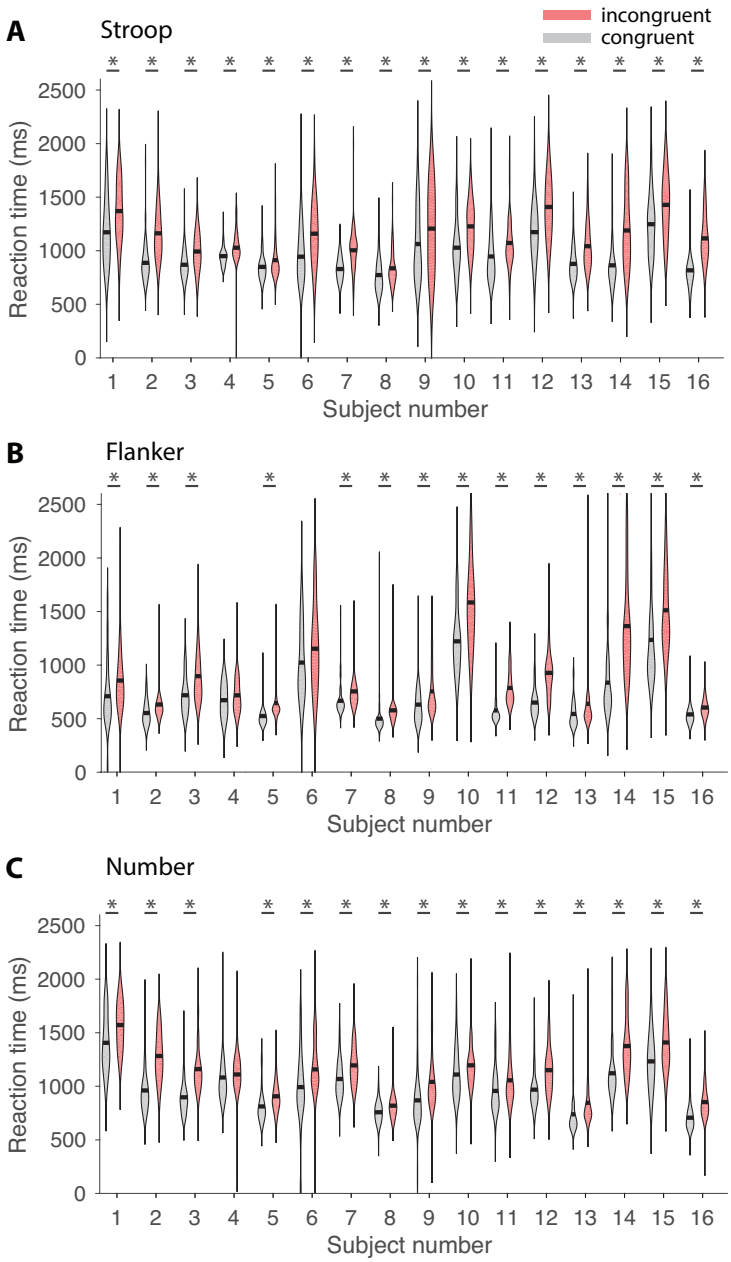

Figure 3. Subjects were slower in incongruent trials in the three tasks. Violin plots showing distribution of reaction times for each subject for congruent trials (gray) and incongruent trials (red) during the Stroop (**A**), Flanker (**B**), and Number (**C**) task. Only correct trials are shown. Black bars indicate mean reaction time. The asterisks denote statistically significant differences between congruent RT and incongruent RT (two-sided permutation test, 10,000 iterations, α=0.05).

Neural responses were modulated by conflict

We recorded intracranial field potentials from 1,877 electrodes (**Table 1** reports the number of electrodes in each subject). We analyzed the activities from 694 bipolarly-referenced electrodes

that were not in the white matter (**Methods**); **Figure 4** and **Table 2** report the distribution of electrode locations. We focused on the neural activities in the theta band (4-8 Hz) because it constitutes a key component of cognitive control (Cavanagh and Frank, 2014; Gratton et al., 2018; Helfrich and Knight, 2016; Widge et al., 2019) and also on the high-gamma band (70-120 Hz) given its significance in sensory, motor, control, and other cognitive functions (Crone et al., 1998; Liu et al., 2009; Norman et al., 2019; Oehrn et al., 2014; Tang et al., 2016). Additional results in other frequency bands (alpha, beta, low-gamma) were also examined. In our previous work (Tang et al., 2016), we reported that multiple electrodes showed activities in the high-gamma band that was modulated by the presence of conflict during the Stroop task. Consistently, **Figure 5** (left) depicts the high-gamma activities during the Stroop task of an electrode, located in the left orbitofrontal cortex, that showed enhanced responses during incongruent trials compared to congruent trials when aligning the neural signals to the behavioral response. The differences between incongruent and congruent trials were highly robust and could even be discerned in individual trials (compare **Figure 5B**, left versus **Figure 5C**, left). Notably, the enhancement associated with conflict was also evident when the neural responses were aligned to stimulus onset (**Figure 5D**, left).

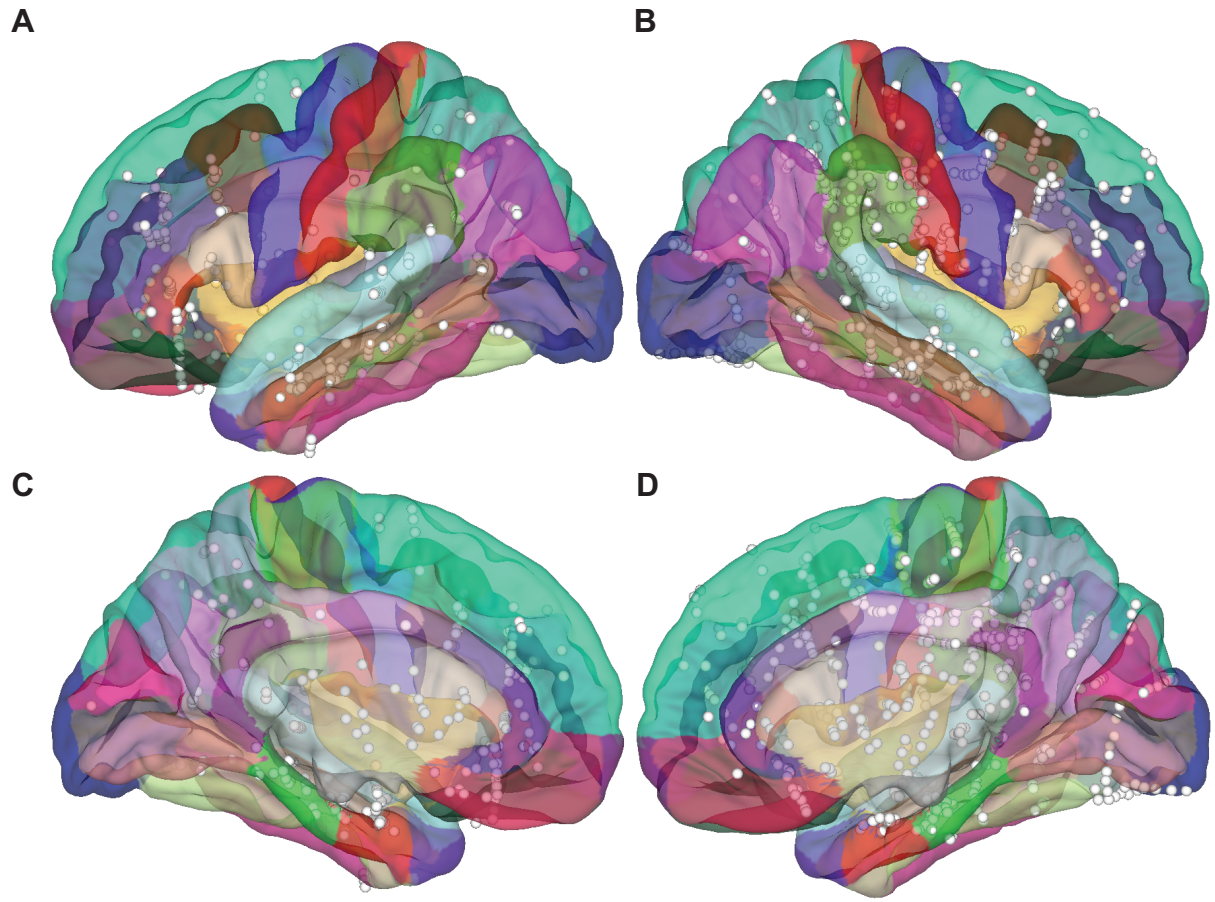

Figure 4. Electrode locations. Each sphere reflects one of each pair of nearby electrodes that were bipolarly referenced (n=694), overlayed on the Desikan-Killiany Atlas with different views: **A**: left lateral; **B**: right lateral; **C**: left medial; **D**: right medial.

Table 2. Distribution of electrode locations. Number of electrodes in each location; "ctx" = cortex; "lh" = left hemisphere; "rh" = right hemisphere; 'bankssts" = banks of superior temporal sulcus.

Location	Count
'left-amygdala'	16
'left-hippocampus'	16
'left-putamen'	9
'right-amygdala'	12
'right-hippocampus'	11
'ctx-lh-bankssts'	2
'ctx-lh-caudalanteriorcingulate'	2
'ctx-lh-caudalmiddlefrontal'	1
'ctx-lh-fusiform'	5
'ctx-lh-inferiorparietal'	22
'ctx-lh-inferiortemporal'	10
'ctx-lh-insula'	14
'ctx-lh-lateraloccipital'	1
'ctx-lh-lateralorbitofrontal'	14
'ctx-lh-lingual'	3
'ctx-lh-medialorbitofrontal'	6
'ctx-lh-middletemporal'	15
'ctx-lh-parahippocampal'	4

Table 2 (Continued).

‘ctx-lh-parsopercularis’	3
‘ctx-lh-parsorbitalis’	1
‘ctx-lh-parstriangularis’	8
‘ctx-lh-postcentral’	2
‘ctx-lh-posteriorcingulate’	2
‘ctx-lh-precentral’	1
‘ctx-lh-rostralanteriorcingulate’	3
‘ctx-lh-rostralmiddlefrontal’	30
‘ctx-lh-superiorfrontal’	15
‘ctx-lh-superiorparietal’	4
‘ctx-lh-superiortemporal’	14
‘ctx-lh-supramarginal’	6
‘ctx-rh-bankssts’	5
‘ctx-rh-caudalanteriorcingulate’	3
‘ctx-rh-caudalmiddlefrontal’	22
‘ctx-rh-cuneus’	3
‘ctx-rh-entorhinal’	1
‘ctx-rh-frontalpole’	1
‘ctx-rh-fusiform’	19
‘ctx-rh-inferiorparietal’	21
‘ctx-rh-inferiortemporal’	18
‘ctx-rh-insula’	27
‘ctx-rh-isthmuscingulate’	8
‘ctx-rh-lateraloccipital’	7
‘ctx-rh-lateralorbitofrontal’	11
‘ctx-rh-lingual’	7
‘ctx-rh-medialorbitofrontal’	5
‘ctx-rh-middletemporal’	23
‘ctx-rh-paracentral’	9
‘ctx-rh-parahippocampal’	6
‘ctx-rh-parsopercularis’	9
‘ctx-rh-parsorbitalis’	3
‘ctx-rh-parstriangularis’	5
‘ctx-rh-pericalcarine’	6
‘ctx-rh-postcentral’	29
‘ctx-rh-posteriorcingulate’	8
‘ctx-rh-precentral’	43
‘ctx-rh-precuneus’	13
‘ctx-rh-rostralanteriorcingulate’	3
‘ctx-rh-rostralmiddlefrontal’	15
‘ctx-rh-superiorfrontal’	26
‘ctx-rh-superiorparietal’	33
‘ctx-rh-superiortemporal’	21
‘ctx-rh-supramarginal’	29
‘ctx-rh-temporalpole’	1
‘ctx-rh-transversetemporal’	2
Total	**694**

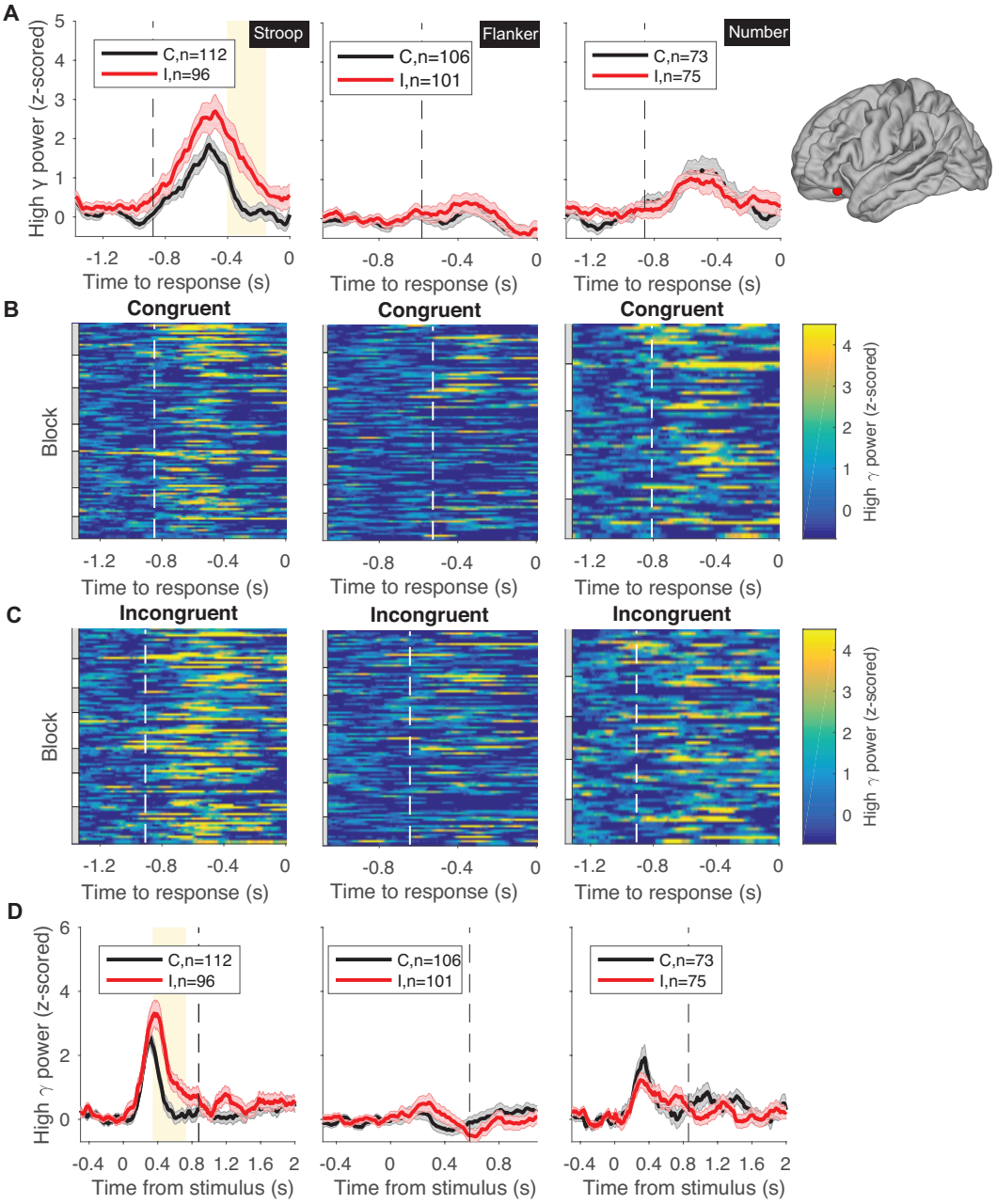

Figure 5. Example electrode in the left orbitofrontal cortex showing conflict modulation in the high-gamma band during the Stroop task only. A. The traces show the mean±SEM z-scored high-gamma power aligned to behavioral response time for incongruent trials (red) and congruent trials (black) for each of the three tasks (Column 1: Stroop; Column 2: Flanker; Column 3: Number). The vertical dashed lines denote the average stimulus onsets. Yellow background indicates statistically significant power differences between congruent and incongruent trials (permutation test, 5,000 iterations, α=0.05, Methods). Legend shows the number of congruent (C) and incongruent (I) trials. The electrode location is shown on the right. **B**-**C**. Raster plots showing the neural signals in individual trials (see color scale on the right) for congruent (**B**) and incongruent (**C**) trials. The white dashed lines show the average stimulus onsets. These lines are shifted to the left in **C** compared to **B**, reflecting the longer reaction times during incongruent trials (see **Figure 3**). Gray and white bars on the left represent different blocks. **D**. Z-scored high-gamma power (mean±SEM) aligned to stimulus onset. Vertical dashed lines denote the average behavioral response times. Yellow background indicates statistically significant power difference between congruent and incongruent trials (permutation test, 5,000 iterations, α=0.05, Methods).

An electrode was considered to be conflict-modulated if the band-filtered power during incongruent conditions was significantly different from that during congruent conditions for at least 150 consecutive milliseconds (permutation test, 5,000 iterations, $\alpha = 0.05$) both when responses were aligned to the behavioral response (**Figure 5A**, left) and to the stimulus onset (**Figure 5D**, left, **Methods**). These strict selection criteria using both alignment to behavior and stimulus were implemented in order to exclude potential false positives. For example, signals from a visually responsive electrode could be confused for conflict modulation when aligning the neural responses to behavior due to the different reaction times between congruent and incongruent trials (see **Figure 3**). An example of such a visually responsive electrode located in the right lateral occipital cortex is shown in **Figure 6A-B**. Even though there seemed to be a difference between incongruent and congruent conditions when neural signals were aligned to the behavioral response (**Figure 6A**), this difference was completely absent when the neural signals were aligned to the stimulus onset (**Figure 6B**). Therefore, we did not consider this type of response to reveal any conflict modulation. Conversely, a motor responsive electrode could also be confused for conflict modulation when aligning the neural signals to stimulus onset for the same reasons (**Figure 6C-D**). Thus, the evaluation criteria for conflict modulation excluded purely sensory and purely motor responses.

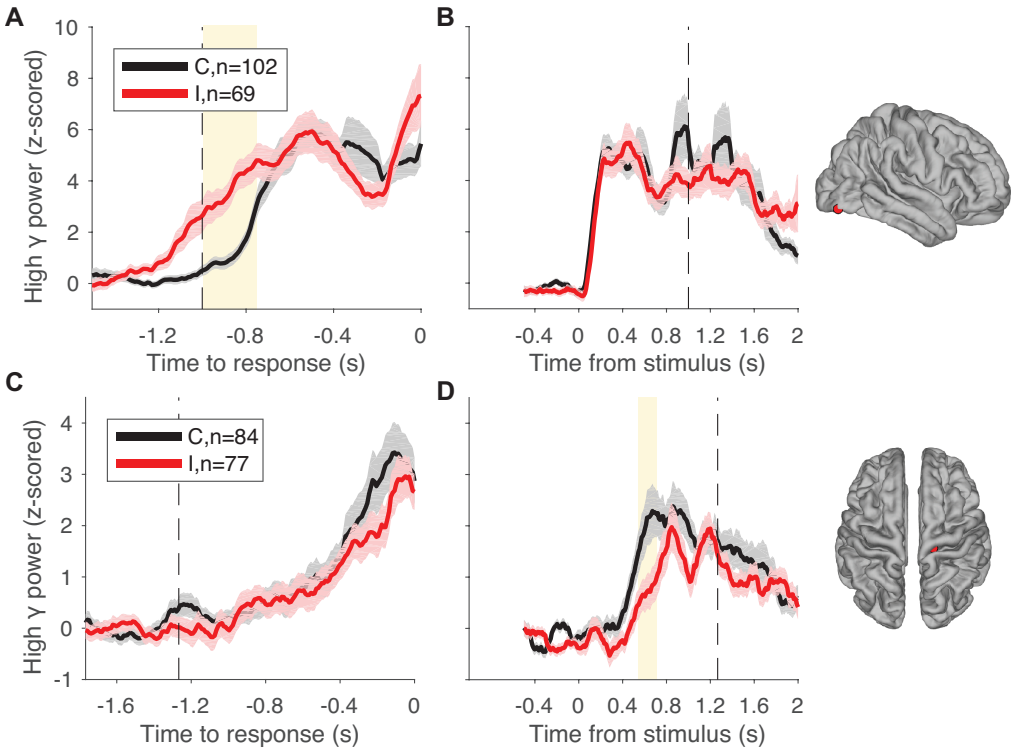

Figure 6. Alignment to stimulus and behavioral response is critical to interpret conflict modulation signals. An electrode in the right lateral occipital cortex showed conflict modulation when the high gamma power (mean±SEM, black for congruent and red for incongruent) was aligned to behavioral response (**A**) but no such effect emerged when aligned to stimulus onset (**B**). Conversely, an electrode at the right precentral gyrus showed conflict modulation when the high gamma power (mean±SEM) was aligned to stimulus onset (**D**) but not behavioral response (**C**). These electrodes reflect either purely visual response (**B**) or purely motor response (**C**).

Figure 5 shows an example electrode that revealed conflict modulation in the high-gamma band during the Stroop task. Electrodes demonstrating robust conflict modulation were also observed during the Flanker and Number tasks. **Figure 7A** (middle) depicts the responses of an electrode in the right superior parietal lobule that showed enhanced activity during incongruent trials in the Flanker task. As described for the Stroop task, conflict modulation was observed in single trials (**Figure 8A**, middle) and also when aligning the responses to stimulus onset (**Figure 8B**, middle). **Figure 7B** (right) depicts the responses of an electrode in the right precuneus that showed enhanced activity during incongruent trials in the Number task. **Figure 9A** (right) shows conflict modulation for this electrode during single trials and **Figure 9B** confirms this conflict modulation even when aligning neural activities to the stimulus onset. Similar results were

observed when considering the theta frequency band. **Figure 10** shows an example electrode in the right pars triangularis that demonstrated conflict modulation in the theta band during the Stroop task. Such modulation can be appreciated both in response-aligned signals (**Figure 10A**) and stimulus-aligned signals (**Figure 10C**) signals, as well as in individual trials (**Figure 10B**). We also found electrodes that exhibited conflict modulation in the theta band for the Flanker and Number tasks, respectively. These results strongly demonstrated that neural incongruency effect in one task does not necessarily equate with an abstract conflict signal.

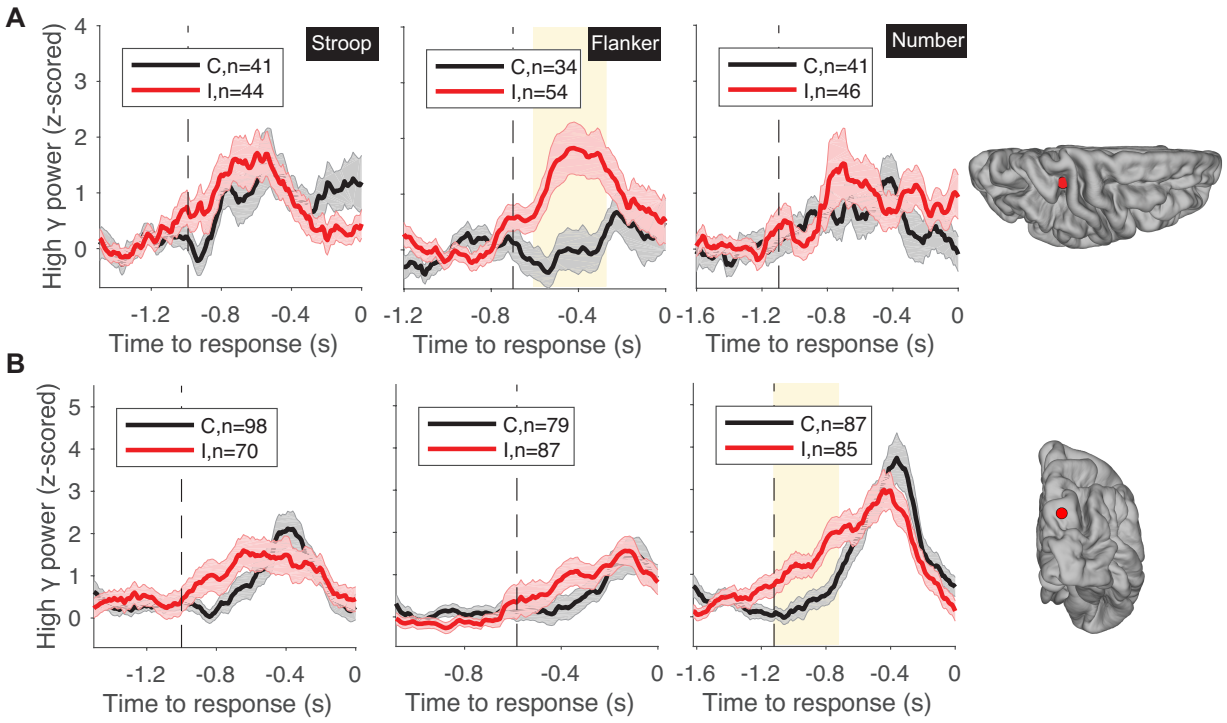

Figure 7. Example Flanker-specific (A) and Number-specific (B) electrodes in the high gamma band. The traces show mean±SEM z-scored high-gamma power aligned to behavioral response time for incongruent trials (red) and congruent trials (black) for each of the three tasks (Column 1: Stroop; Column 2: Flanker; Column 3: Number). Legend denotes number of congruent (**C**) and incongruent (**I**) trials. The vertical dashed lines denote the average stimulus onset. Yellow background indicates statistically significant differences between congruent and incongruent trials (permutation test, 5000 iterations, α=0.05, Methods). Electrode locations are shown on the right (**A**: right superior parietal; **B**: right precuneus).

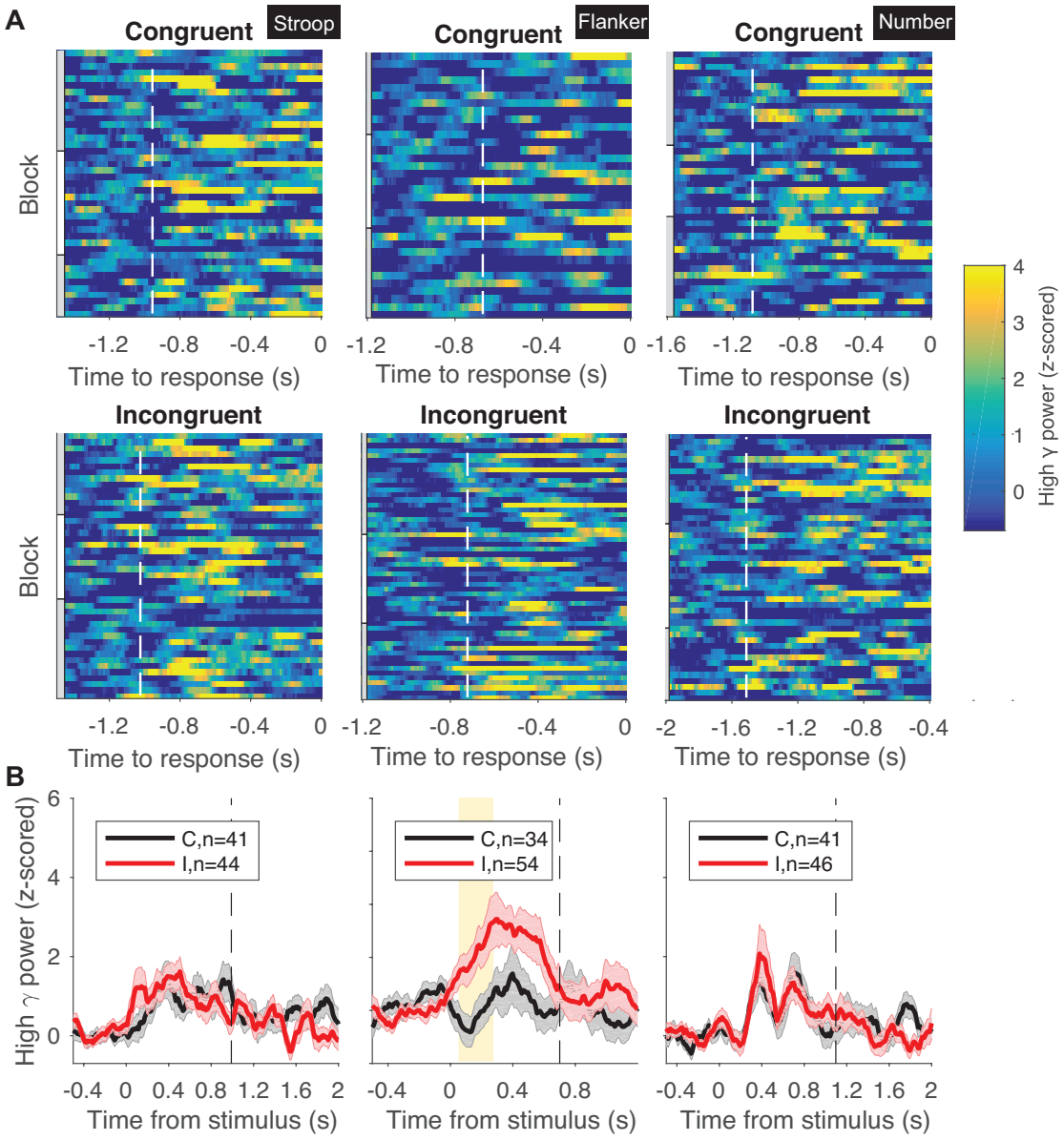

Figure 8. Example Flanker-specific electrode in the high gamma band (same electrode as in Figure 7A). An electrode located in the right superior parietal cortex exhibited conflict modulation in the Flanker task only. **A**. Raster plots showing the neural signals in individual trials (see color scale on the right) for congruent and incongruent trials. The white dashed lines show the average stimulus onsets. Gray and white bars on the left represent different blocks. **B**. Z-scored high gamma power (mean±SEM, black for congruent and red for incongruent) aligned to stimulus onset. Vertical dashed lines denote the average behavioral response times. Yellow background indicates statistically significant power difference between congruent and incongruent trials (permutation test, 5,000 iterations, α=0.05, Methods).

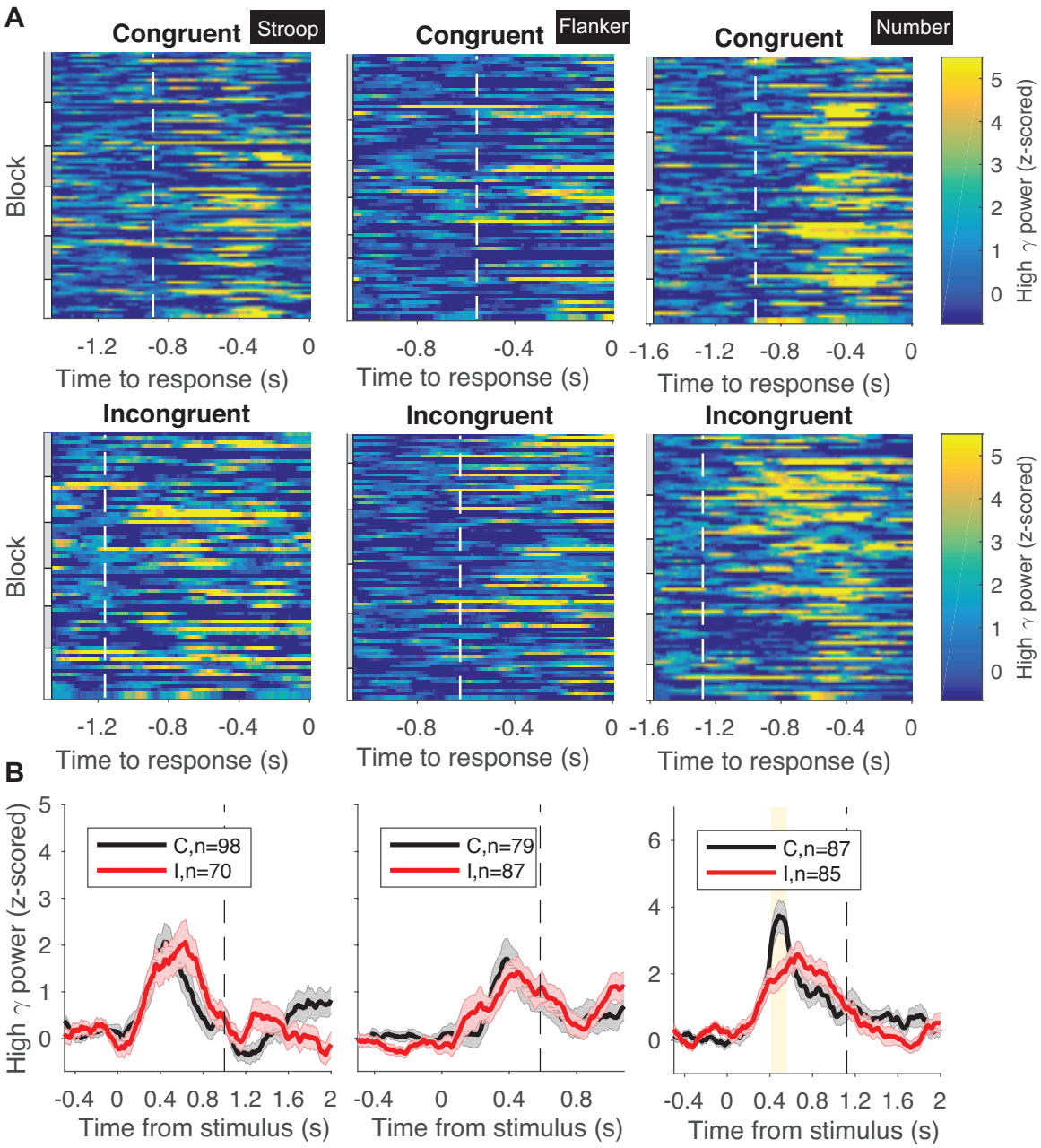

Figure 9. Example Number-specific electrode in the high gamma band (same as in Figure 7B). An electrode located in the right precuneus exhibited conflict modulation in the Number task only. **A**. Raster plots showing the neural signals in individual trials (see color scale on the right) for congruent and incongruent trials. The white dashed lines the average stimulus onsets Gray and white bars on the left represent different blocks. **B**. Z-scored high gamma power (mean±SEM, black for congruent and red for incongruent) aligned to stimulus onset. Vertical dashed lines denote the average behavioral response times. Yellow background indicates statistically significant power difference between congruent and incongruent trials (permutation test, 5,000 iterations, α=0.05, Methods).

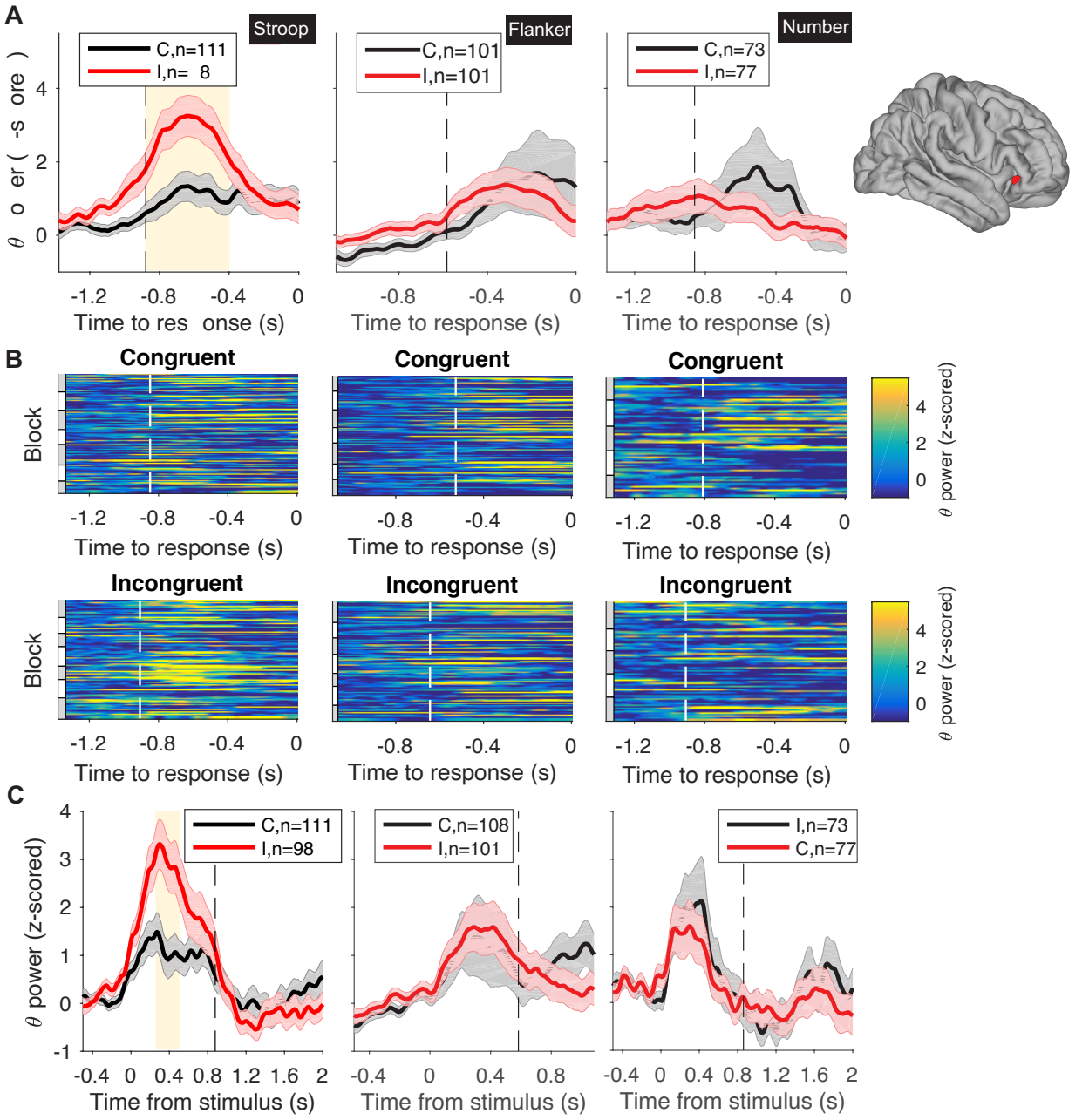

Figure 10. Example Stroop-specific electrode in the theta band (right pars triangularis). A. The traces show the mean±SEM z-scored theta (4-8 Hz) power aligned to behavioral response time for incongruent trials (red) and congruent trials (black) for each of the three tasks (Column 1: Stroop; Column 2: Flanker; Column 3: Number). The vertical dashed lines denote the average stimulus onsets. Yellow background indicates statistically significant power difference between congruent and incongruent trials (permutation test, 5,000 iterations, α=0.05, Methods). Legend shows the number of congruent (C) and incongruent (I) trials. The electrode location is shown on the right. **B**. Raster plots showing the neural signals in individual trials (see color scale on the right) for congruent and incongruent trials. The white dashed lines show the average stimulus onsets. Gray and white bars on the left represent different blocks. **C**. Z-scored theta power aligned to stimulus onset. Vertical dashed lines denote the average behavioral response times.

Out of the total of 694 electrodes, we identified 134 electrodes (19%) that exhibited conflict modulation in at least one task in the high-gamma band (**Table 3**) and 109 electrodes (16%) when considering the theta band (**Table 4**). In most cases, conflict modulation was characterized by enhanced high-gamma-band power in the incongruent condition compared to the congruent condition. A few electrodes exhibited the reverse modulation direction where the congruent response was higher than the incongruent one (**Figure 17A**, middle). **Figure 11** shows the distribution of locations of electrodes revealing conflict modulation for each task. In sum, using strict criteria, we found electrodes that demonstrate robust conflict modulation in each of the three tasks, considering both high-gamma and theta band signals, evident in both behavior- and stimulus-aligned responses, and even in single trials.

Table 3. Location and specificity of conflict-modulated electrodes (high-gamma band). For each location, the table reports the number of electrodes that show conflict modulation in one task only, in two tasks, or in all three tasks. S=Stroop, F=Flanker, N=Number.

Location	One task only			Two tasks			All tasks	Sum
	S	F	N	S+F	S+N	F+N	S+F+N	
amygdala		1	1					2
ctx-caudalmiddlefrontal	1	1						2
ctx-entorhinal			1					1
ctx-fusiform		2	4					6
ctx-inferiorparietal	3	5	2	1				11
ctx-inferiortemporal		2	6			1		9
ctx-insula	1	3	3			2		9
ctx-lateraloccipital			3					3
ctx-lateralorbitofrontal	4	2						6
ctx-middletemporal	2	1	1					4
ctx-paracentral	1							1
ctx-parahippocampal	1		1					2
ctx-parstriangularis		1				1		2
ctx-pericalcarine			1					1
ctx-postcentral	2		2					4
ctx-posteriorcingulate		1	1					2
ctx-precentral	4	5	1			1		11
ctx-precuneus			2					2
ctx-rostralanteriorcingulate			1					1
ctx-rostralmiddlefrontal	6	2			2			10
ctx-superiorfrontal	3	5	1		1	4		14
ctx-superiorparietal	1	2	10			2		15
ctx-superiortemporal	1	1						2
ctx-supramarginal	4	3	2		1			10
hippocampus	1		1					2
putamen	2							2
Total	**37**	**37**	**44**	**1**	**4**	**11**	**0**	**134**
	118			**16**			**0**	**134**

Table 4. Location and specificity of conflict-modulated electrodes (theta band). For each location, the table reports the number of electrodes that show conflict modulation in one task only, in two tasks, or in all three tasks. S=Stroop, F=Flanker, N=Number.

Location	One task only			Two tasks			All tasks	Sum
	S	F	N	S+F	S+N	F+N	S+F+N	
amygdala	1							1
ctx-bankssts			1					1
ctx-caudalmiddlefrontal	1	3	1					5
ctx-cuneus			1					1
ctx-fusiform	1	1	5		1	1		9
ctx-inferiorparietal		3	3	1				7
ctx-inferiortemporal		2	5			1		8
ctx-insula	1	2	1					4
ctx-isthmuscingulate	1	1		1				3
ctx-lateraloccipital			4	1	2			7
ctx-lateralorbitofrontal	1			1				2
ctx-lingual		1	2	1	1			5
ctx-medialorbitofrontal		1	1					2
ctx-middletemporal		1	1					2
ctx-parahippocampal		1	1					2
ctx-parsopercularis		1	2					3
ctx-parstriangularis	2		1					3
ctx-pericalcarine	2							2
ctx-postcentral	1	1						2
ctx-precentral		3						3
ctx-precuneus	1		1		1			3
ctx-rostralmiddlefrontal		2	3					5
ctx-superiorparietal		4	7		5			16
ctx-superiortemporal	2	1	1					4
ctx-supramarginal	2	2						4
hippocampus	1	1						2
putamen			3					3
TOTAL	**17**	**31**	**44**	**5**	**10**	**2**	**0**	**109**
		92			**17**		**0**	**109**

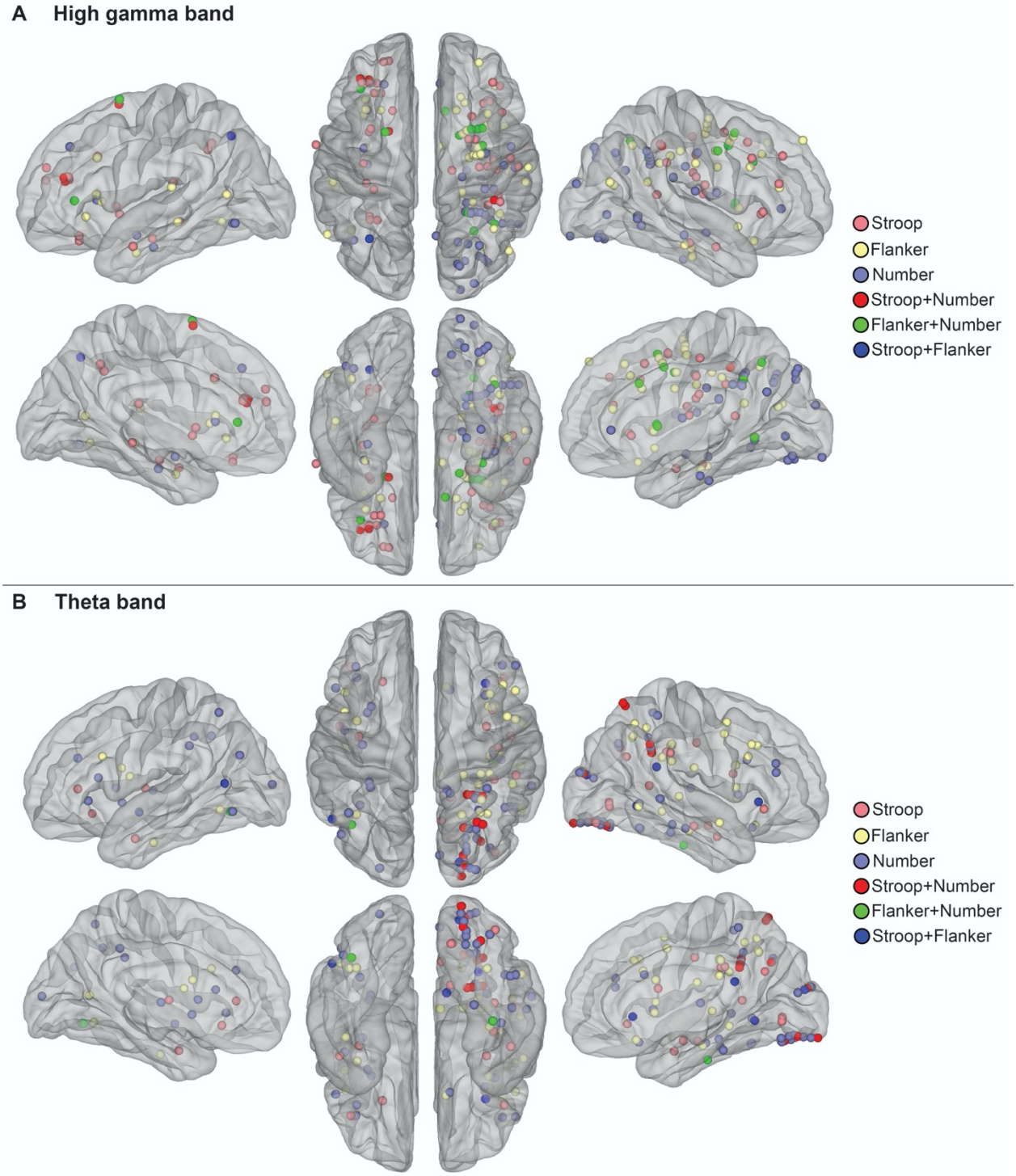

Figure 11. Electrodes exhibiting conflict modulation in high gamma (A) and theta (B) band. Specific locations of these electrodes can be found in **Table 3** and **Table 4**. We didn't find any electrode that was conflict-selective for all the three tasks in all the frequency bands analyzed.

Neural signals in the high-gamma band during incongruent trials correlated with reaction times

Next we examined whether the neural signals were correlated with behavior. For each of the conflict-modulated electrodes, we plotted the mean high-gamma band power from stimulus onset to the behavioral response time as a function of the RT in each trial (**Methods**). **Figure 12A** shows an example electrode located in the left rostral middle frontal cortex that was modulated by conflict during the Stroop task. The mean high-gamma power was not correlated with reaction times during congruent trials (**Figure 12A**, left, p=0.3), but there was a significant correlation during incongruent trials (**Figure 12A**, right, p=0.03). Similarly, **Figure 12B** shows an example electrode in the right superior frontal gyrus that showed a correlation with reaction times during the Flanker task and **Figure 12C** shows an example electrode in the right inferior temporal cortex that showed a correlation with reaction times during the Number task. In all, 8.3%, 12.2%, and 10.2% of the conflict modulated electrodes showed a correlation with reaction time during incongruent trials, but not congruent trials, for the Stroop, Flanker, and Number task, respectively.

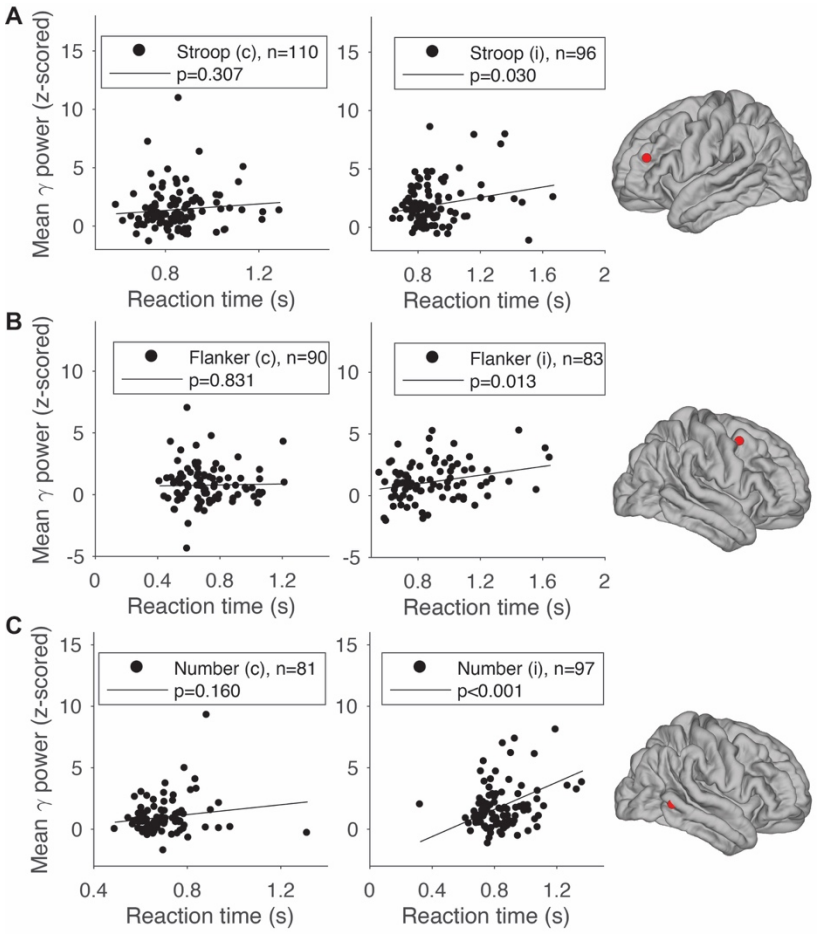

Figure 12. Example conflict modulated responses showing correlation with reaction time. The plots show the mean high-gamma power (z-scored) in each trial as a function of the reaction time for 3 example electrodes during the (**A**) Stroop (left rostral middle frontal cortex), (**B**) Flanker (right superior frontal cortex), and (**C**) Number task (right inferior temporal cortex). Each point shows one trial. The number of trials is shown in each subplot. Electrode locations are shown on the right. The solid lines indicate the linear fits. Correlations were statistically significant for incongruent trials (right) but not congruent trials (left) (see p values in legend).

These observations did not extend to the theta band. Signals in the theta band showed a much weaker correlation with reaction times. We found only 4 conflict-modulated electrodes, 2 in the Flanker task, 2 in the Number task, and none in the Stroop task, that demonstrated a statistically significant correlation between theta band power and reaction times.

Conflict representation exhibited within-task invariance

In each task, there were different stimuli that define conflict. For example, in the Stroop task, there were six different word/color combinations that were incongruent and three that were congruent (**Figure 13A-I**). Our first hypothesis states that conflict modulation is invariant to the different stimuli defining incongruence within a task. To test this hypothesis, we evaluated whether the modulation of neural signals underlying conflict was evident only in some specific incongruent stimuli but not others, or evident in all incongruent stimuli.

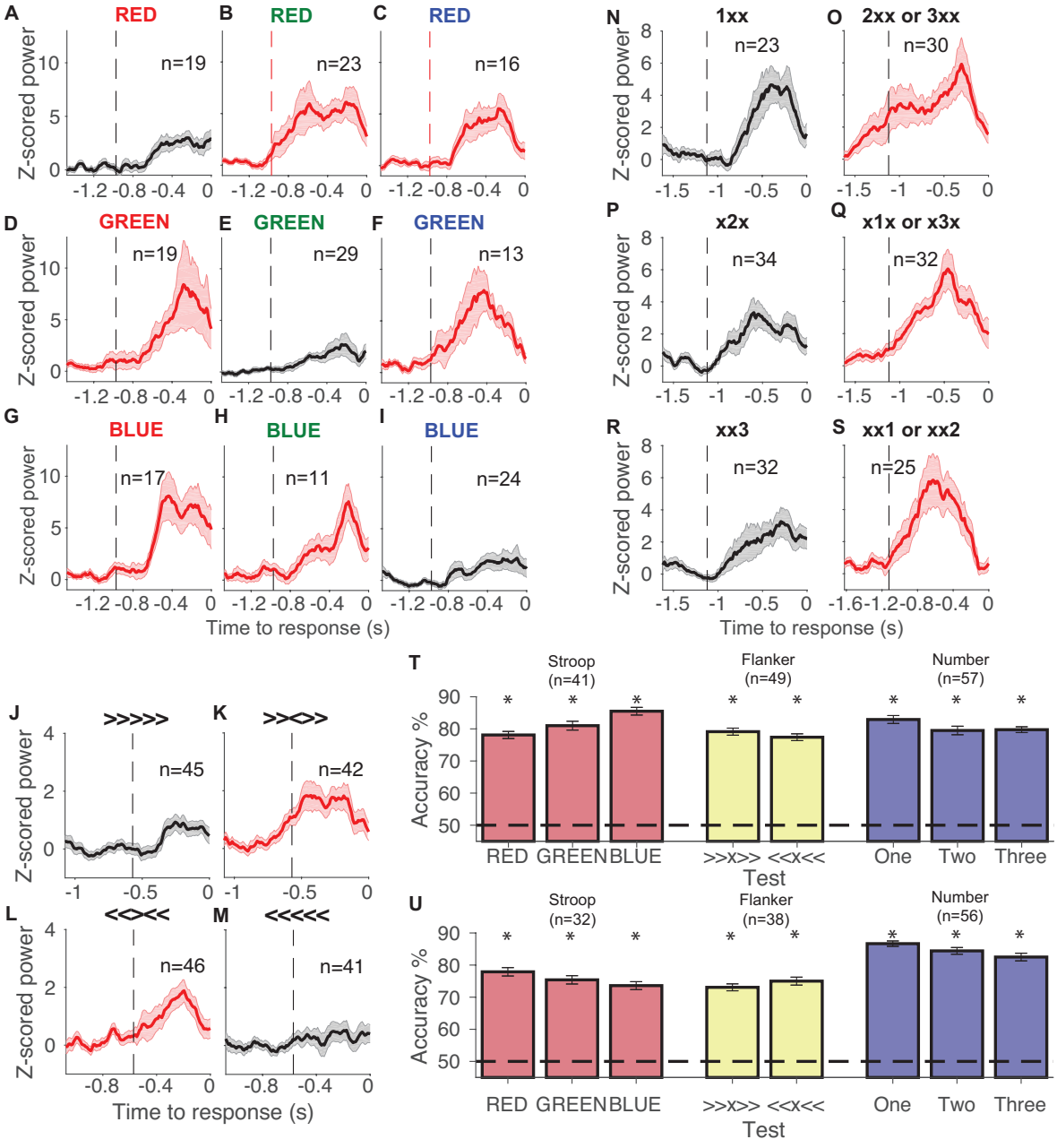

Figure 13 (Continued).

Figure 13. Neural signals showed within-task invariance. A-S. Example task-specific electrodes demonstrating within-task invariance (Stroop: **A-I**, left inferior parietal; Flanker: **J-M**, right superior frontal; Number: **N-S**, right superior parietal). Z-scored high-gamma power (mean±SEM) aligned to behavioral response time (black for congruent and red for incongruent). Vertical dashed lines denote the average stimulus onsets. Subplot titles indicate specific stimulus types. Conflict modulation occurred in all incongruent color/word (Stroop), target/flanker (Flanker), and target/distractor (Number) combinations. **T-U**. Accuracy of support vector machine (SVM) classifier in discriminating incongruent from congruent trials extrapolating across conditions within each task (within-task invariance) using high-gamma (**T**) and theta (**U**) band power. Stroop task: The first bar labeled "RED" was trained using the "GREEN" trials (as in panels **DEF**) and "BLUE" trials (as in panels **GHI**), and tested on "RED" trials (as in panels **ABC**). A similar procedure was followed for the other combinations. In bar 2, the SVM was trained using "RED" and "BLUE" trials, and tested on "GREEN" trials. In bar 3, the SVM was trained using "RED" and "GREEN" trials, and tested on "BLUE" trials. Flanker task: in the 4th bar, the SVM was trained on "<<<<<" and "<<><<", and tested on ">>>>>" and ">><>>". In the 5th bar, training and testing data were reversed. Number task: in the 6th bar, the SVM was trained on trials where the correct answer was "two" (as in panels **PQ**) or "three" (as in panels **RS**), and tested on trials where the correct answer was "one" (as in panels **NO**). Similarly in bar 7, the SVM was trained on "one" and "three", and tested on "two". In bar 8, the SVM was trained on "one" and "two", and tested on trials whose target answer was "three". For each task, the training and testing data for each condition were randomly subsampled to contain an equal number of congruent and incongruent trials. Electrodes that had very few correct trials in any condition were removed from this analysis. Error bars indicate s.e.m. over 50 sessions. The dashed line indicates chance performance (50%). Asterisks denote higher than chance accuracy (permutation test with 10,000 iterations, $p<0.001$ for all bars).

One might expect stimulus specificity given the extensive documentation of selective responses to different sensory inputs (e.g., (Liu et al., 2009), among many others). Indeed, consistent with previous work, we found multiple visually selective electrodes (Stroop: 15 electrodes; Flanker: 8 electrodes; Number: 0 electrodes; total = 23 electrodes; **Methods, Table 5**). Similarly, we found 36 motor selective electrodes (verbal response: 26 electrodes; keypress response: 10 electrodes; total = 36 electrodes; **Methods, Table 6**). Among these 23+36=59 electrodes, there were only 5 electrodes (3 visually-selective and 2 motor-selective) that showed both visual or motor selectivity and conflict modulation in the same task. These 5 electrodes constituted 8% of the visually/motor selective electrodes and 4% of all the electrodes that showed conflict modulation. Thus, the majority of electrodes that showed conflict modulation were not visually or motor selective.

Table 5: Location and number of visually-selective electrodes.

Location	Stroop	Flanker	Number	Sum
ctx-fusiform	2			2
ctx-inferiorparietal	1	1		2
ctx-inferiortemporal	3			3
ctx-insula	1	1		2
ctx-lateralorbitofrontal	1			1
ctx-lingual		2		2
ctx-postcentral		1		1
ctx-precuneus		1		1
ctx-rostralmiddlefrontal	2			2
ctx-superiorparietal	4	2		6
ctx-supramarginal	1			1
Total	**15**	**8**	**0**	**23**

Table 6: Location and number of motor-selective electrodes.

Location	Verbal	Keypress	Sum
ctx-caudalanteriorcingulate		1	1
ctx-caudalmiddlefrontal	1		1
ctx-inferiorparietal		2	2
ctx-insula	1		1
ctx-parahippocampal		1	1
ctx-parsopercularis	1		1
ctx-parstriangularis	1		1
ctx-precentral	7	1	8
ctx-postcentral	11	2	13
ctx-superiorfrontal	1		1
ctx-superiorparietal		1	1
ctx-superiortemporal	2		2
ctx-supramarginal	1	2	3
Total	**26**	**10**	**36**

To further investigate whether conflict modulation generalized across the different sensory inputs, we directly compared the responses to all possible stimuli within each task. **Figure 13A-I** describes the responses of an electrode in the left inferior parietal cortex for every word/color combination during the Stroop task. Conflict modulation cannot be ascribed to responses to specific word/color combinations; that is, conflict modulation showed within-task invariance with enhanced responses during incongruent trials for the six different possible incongruent word and

font color combinations compared to the three different possible congruent word and font color combinations. An example examining the theta band is shown in **Figure 14**. Similarly, **Figure 13J-M** describes the responses of an electrode in the right superior frontal gyrus for every combination of central and peripheral arrow directions during the Flanker task. Conflict modulation in the Flanker task was also invariant within the task; that is, there was higher activity during the two incongruent target/flanker combinations compared to the two congruent combinations. An example from the theta band is shown in **Figure 15**. **Figure 13N-S** describes the responses of an electrode in the right superior parietal lobule, showing that conflict modulation was evident for all the different incongruent conditions in the Number task. An example from the theta band is shown in **Figure 16**.

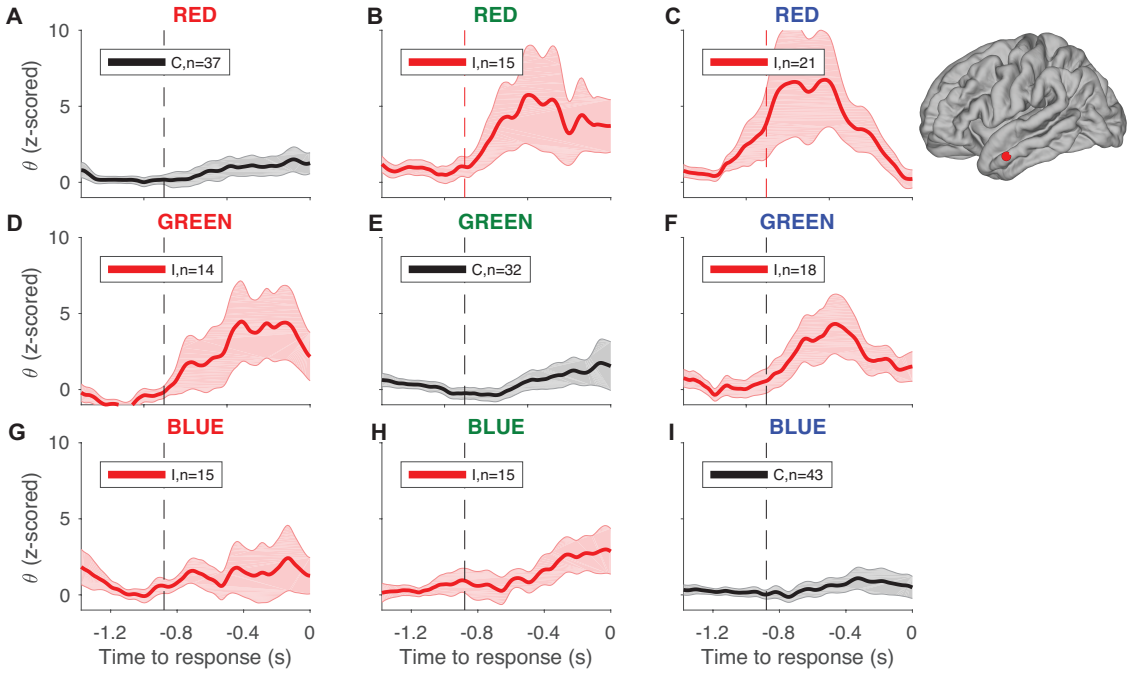

Figure 14. Example Stroop-specific electrode showing within-task invariance in the theta band (left superior temporal). Z-scored theta power (mean±SEM) aligned to behavioral response time (black for congruent and red for incongruent). Vertical dashed lines denote the average stimulus onsets. Subplot titles indicate specific stimulus types. Conflict modulation occurred in all incongruent color/word combinations.

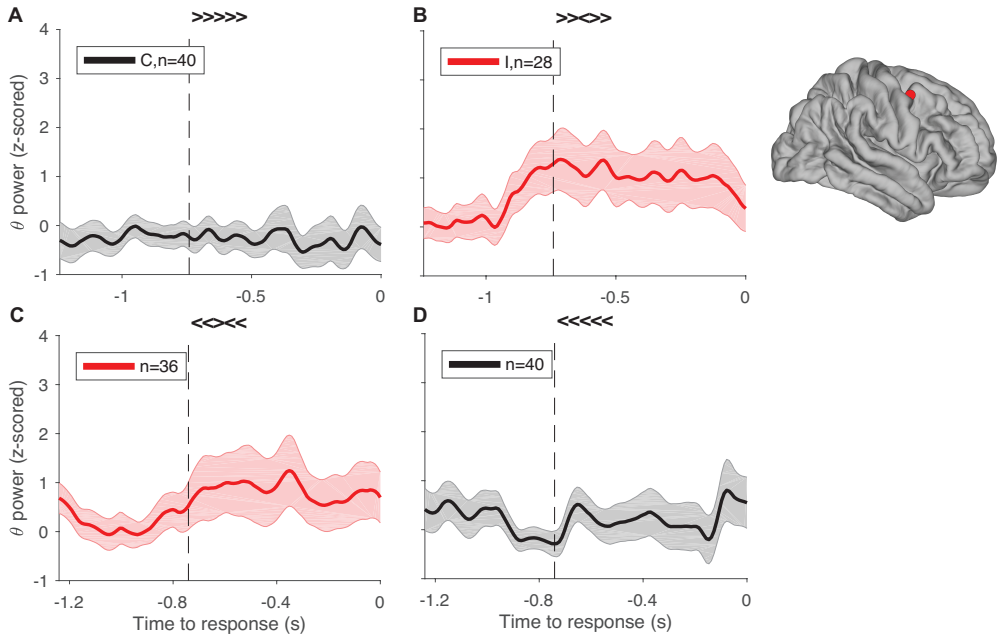

Figure 15. Example Flanker-specific electrode showing within-task invariance in the theta band (right precentral). Z-scored theta power (mean±SEM) aligned to behavioral response time (black for congruent and red for incongruent). Vertical dashed lines denote the average stimulus onsets. Subplot titles indicate specific stimulus types. Conflict modulation occurred in all incongruent target/flanker combinations.

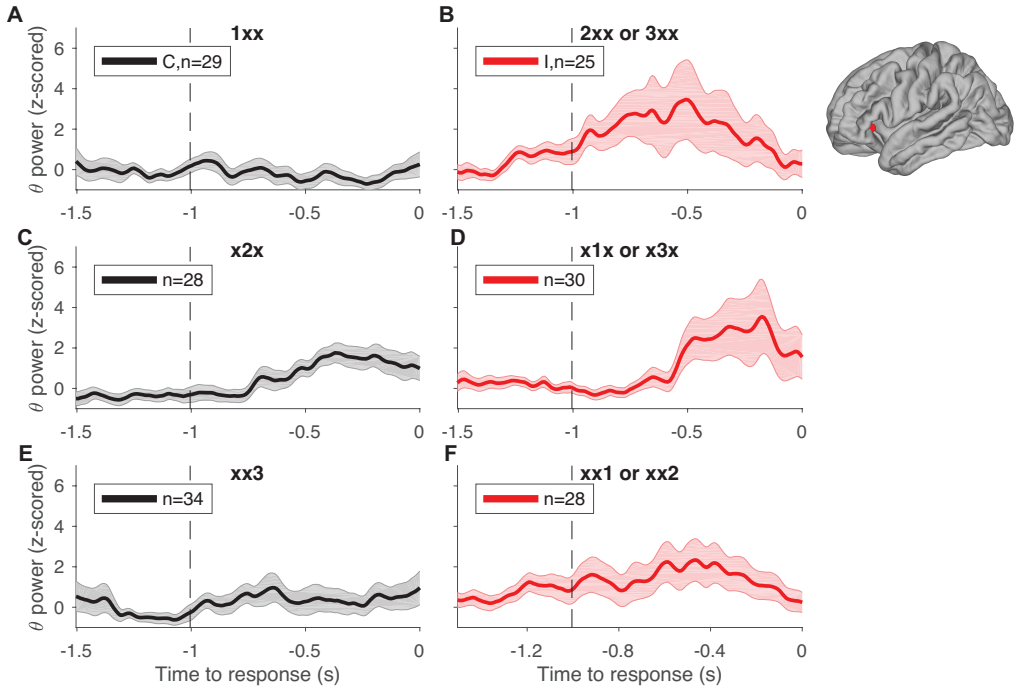

Figure 16. Example Number-specific electrode showing within-task invariance in the theta band (left pars triangularis). Z-scored theta power (mean±SEM) aligned to behavioral response time (black for congruent and red for incongruent). Vertical dashed lines denote the average stimulus onsets. Subplot titles indicate specific stimulus types. Conflict modulation occurred in all incongruent target/distractor combinations.

To characterize the degree of within-task invariance at the electrode ensemble level, we used a machine learning decoding approach to assess whether we could decode the presence of conflict in individual trials (**Figure 13T-U**). In all the decoding analyses, an SVM classifier with a linear kernel was trained after concatenating all the conflict modulated electrodes in each task. We used two neural features: the maximum and the mean band power during each trial (**Methods**), either for the high-gamma band (**Figure 13T**) or the theta band (**Figure 13U**). In all cases, we used cross-validation, separating the data into a training set and an independent test set and we randomly subsampled the data to ensure that the number of congruent trials matched the number of incongruent trials. To evaluate within-task invariance, the classifier was trained using only a subset of the different stimulus combinations and tested on different stimulus combinations. For example, in the first bars in **Figure 13T** and **13U**, the SVM classifier was trained with the neural responses to GREEN/red, GREEN/green, GREEN/blue, BLUE/red, BLUE/green, and BLUE/blue. The classifier's performance was tested using the remaining conditions: RED/red, RED/green, and RED/blue. Even though the classifier was never exposed to the neural responses to any stimulus with the word "RED" during training, the classifier could extrapolate to identify conflict with those novel stimuli in the same task. Similar conclusions were reached for the other possible combinations of training and test stimuli within the Stroop task (**Figure 13T-U**, red bars) and also for the different combinations in the Flanker (yellow bars) and Number (blue bars) tasks. In sum, both at the individual electrode level (**Figures 13A-S**) as well as at the electrode population level (**Figure 13T-U**), and both in the high-gamma (**Figure 13T**) and theta band (**Figure 13U**), the results support the hypothesis that the neural signals modulated by conflict are largely independent of the specific sensory combination of stimuli that give rise to incongruence within each task.

Conflict-modulated electrodes were task-specific

Given the extrapolation across stimuli within a task, we next considered the hypothesis that neural signals representing conflict would also be independent of the specific sensory and motor characteristics of the task. We asked whether electrodes showing conflict modulation were task-specific (i.e., showing activity modulation during incongruent trials in some but not all tasks) or task-invariant (i.e., showing activity modulation during incongruent trials in all three tasks). The examples in **Figures 5, 7-10** illustrate example electrodes with high specificity of conflict modulation. The electrodes in **Figure 5** and **Figure 10** only revealed conflict modulation during the Stroop task (compare left column with middle and right columns). Similarly, the electrode in **Figure 7A** and **Figure 8** showed conflict modulation only during the Flanker task (middle column), and the electrode in **Figure 7B** and **Figure 9** only during the Number task (right column). These types of neural responses were representative of the majority of the data. Out of the total of 134 electrodes that showed conflict modulation in the high-gamma band, 118 electrodes (88%) exhibited modulation in one task but not in the other two tasks. Similarly, out of the total of 109 electrodes that showed conflict modulation in the theta band, 92 electrodes (84%) exhibited modulation in one task but not in the other two tasks.

Although most electrodes demonstrated conflict modulation in one task only, there were 16 electrodes in the high-gamma band (12%) and 17 electrodes (16%) in the theta band that showed conflict modulation in two tasks. Three example electrodes that showed conflict modulation in two tasks are illustrated in **Figures 17-18**. In **Figure 17A**, an electrode in the left inferior parietal cortex exhibited conflict modulation during the Stroop and Flanker tasks, but not during the Number task. Similarly, **Figure 17B** shows an electrode at the right supramarginal gyrus, demonstrating conflict modulation in the Stroop and Number tasks, but not during the Flanker task. **Figure 17C** shows

an electrode in the right insula exhibiting conflict modulation during the Flanker and Number tasks, but not during the Stroop task. These electrodes also showed conflict modulation when the neural signals were aligned to stimulus onset (**Figure 18**). **Table 3** and **Table 4** report the locations and task specificity for all the dual-task modulated electrodes for the high-gamma and theta band respectively.

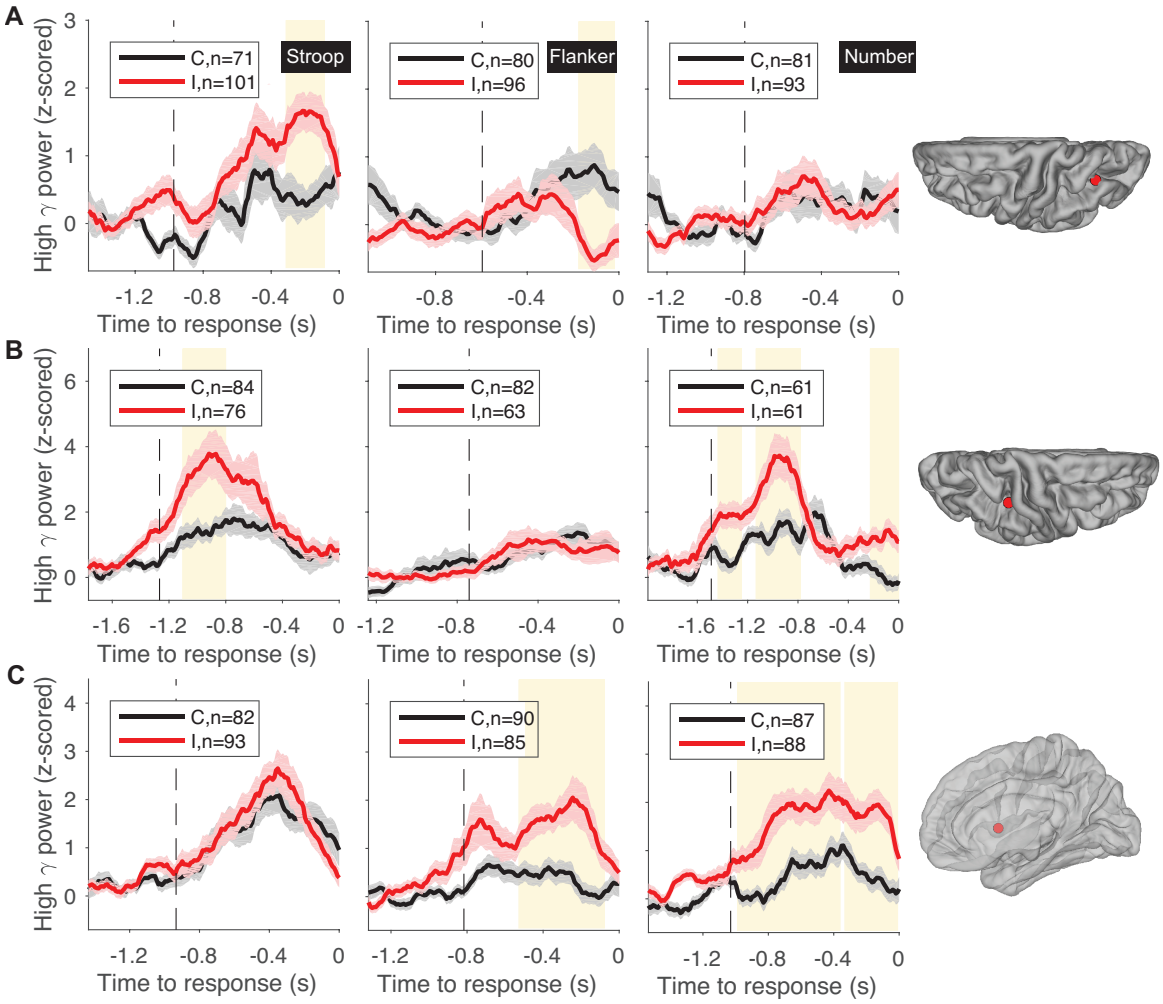

Figure 17. Example electrodes showing conflict modulation in two tasks. **A**. An electrode located in the left inferior parietal lobule (see location on the right) exhibited conflict modulation in the Stroop and Flanker tasks but not in the Number task. Traces show z-scored high gamma power (mean±SEM, black for congruent and red for incongruent) aligned to behavioral response time. Vertical dashed lines denote the average stimulus onsets. Yellow background indicates statistically significant power difference between congruent and incongruent trials (permutation test, 5,000 iterations, α=0.05, Methods). Legend shows the number of congruent (C) and incongruent (I) trials. **B**. An electrode located at the right supramarginal exhibited conflict modulation in the Stroop and Number tasks but not in the Flanker task. **C**. An electrode located in the right insula exhibited conflict modulation in the Flanker and Number tasks but not in the Stroop task. Brain was rendered transparent for better visualization of that electrode in deep structures.

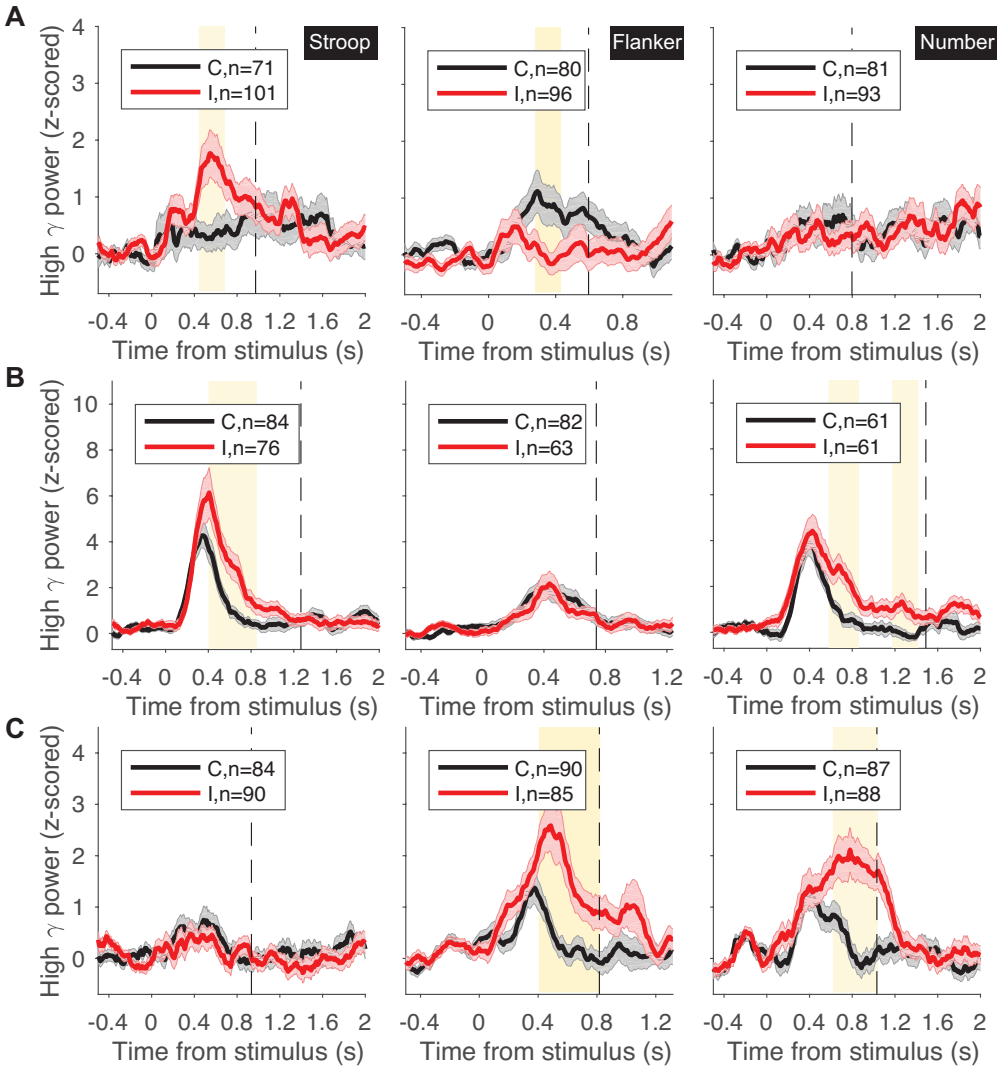

Figure 18. Example dual-task electrodes. Stimulus aligned responses for the electrodes in **Figure 17**.

In sum, most electrodes demonstrated conflict modulation in one task and few electrodes showed conflict modulation in two tasks. Remarkably, we did not find any electrode that was modulated by conflict in all three tasks. Given the complete absence of any task-invariant electrode, we asked whether it is possible that we missed indications of invariance due to our stringent criteria. First, we considered whether it is possible that having elevated activity in the congruent condition could be a prerequisite to observe conflict modulation. The electrode in **Figure 5** showed conflict modulation for the Stroop task but not in the other two tasks. During the Flanker task, this electrode

showed no elevated response whatsoever, and during the Number task, there was a high response with respect to baseline starting about 0.7 seconds before the behavioral response, but this increase was very much the same for congruent and incongruent trials. Thus, there can be activation in the congruent condition without conflict modulation. Similarly, in the example electrode in **Figure 7B**, there was an elevated response in the congruent condition during all three tasks. However, conflict modulation occurred only in the Number task. In total, 302 electrodes showed elevated high-gamma band responses during the congruent condition in at least one task. Among these electrodes, only 80 (26%) also showed conflict modulation. Moreover, the majority of these electrodes (70 out of 80) did not share the same task specificity, i.e., tasks showing conflict modulation did not match tasks displaying elevated responses during the congruent conditions. In sum, multiple electrodes responded during the congruent condition without conflict modulation and multiple electrodes showed conflict modulation only in some task(s) while still showing responses during the congruent condition in other task(s). Thus, an elevated response during the congruent condition is neither necessary nor sufficient to show evidence of conflict modulation. Lack of task invariance cannot be attributed to lack of response during the congruent condition.

Second, the results presented thus far focus on the high-gamma and theta frequency bands. Although different frequency bands of intracranial field potential signals tend to be correlated (Bansal et al., 2012), it is conceivable that some of the electrodes may reveal task invariance in conflict modulation in other frequency bands. To evaluate this possibility, we repeated all the analyses in the following frequency bands **(Methods)**: alpha (8-12 Hz), beta (12-35 Hz), and low gamma (35-70 Hz). We did not consider the delta band because the reaction time, especially of Flanker task, was not long enough to contain a sufficient number of oscillatory cycles to calculate the power. **Table 7** reports the number of electrodes that showed conflict modulation for each task

and for each frequency band. Summarizing **Table 7**, we found conflict modulation in all frequency bands, though the total number of electrodes that showed modulation was the highest in the high-gamma band. Consistent with the results described in the previous sections, the vast majority of electrodes revealed conflict modulation only in one task (alpha: 88%; beta: 94%; low-gamma: 93%). In all frequency bands, we observed a small fraction of electrodes that showed conflict modulation in two tasks. Importantly, we did not find any electrode that showed task-invariance in any of these other frequency bands.

Table 7: Number of conflict-modulated electrodes in each frequency band.

Frequency band	Time-bandwidth product, taper, moving window	One task only			Two tasks			All tasks	Sum
		S	F	N	S+F	S+N	F+N	S+F+N	
high-gamma (70-120 Hz)	5, 7 ,200 ms	37	37	44	1	4	11	0	134
low-gamma (35-70 Hz)	5, 7 ,200 ms	11	30	14	4	0	0	0	59
beta (12-35 Hz)	3, 5, 200 ms	13	26	39	0	4	1	0	83
alpha (8-12 Hz)	2, 3, 500 ms	16	40	23	1	3	7	0	90
theta (4-8 Hz)	2, 3, 500 ms	17	31	44	5	10	2	0	109

In sum, the observations show that most of the electrodes reveal conflict modulation in only one task, and few electrodes show conflict modulation in two tasks. These results lead us to reject our second hypothesis of task invariance in cognitive control at the level of individual electrodes in most part of the brain.

Electrode population level responses revealed task-specific conflict modulation in individual trials

It is conceivable that individual electrodes could show task specificity while an ensemble of multiple electrodes might reflect task invariance. To evaluate this possibility, we investigated whether we could decode the presence of conflict at the electrode population level in individual trials, following the same procedure described in **Figure 13T-U**. Depending on the specific

question about task-specificity, each calculation used different combinations of training and test sets, as described below.

First, we asked whether the population of electrodes modulated by one task could classify the presence of conflict in another task. In **Figure 19**, we trained nine different classifiers using the high-gamma (**Figure 19A**) and theta (**Figure 19C**) band power. The first three bars used Stroop-only electrodes, the middle three bars used Flanker-only electrodes, and the last three bars used Number-only electrodes. The classifiers were trained and tested using cross-validation using only responses from the Stroop task (red), using only responses from the Flanker task (yellow), or using only responses from the Number task (blue). The Stroop-only electrode population yielded significant classification performance when trained and tested on the Stroop task (permutation test, 10,000 iterations, one-sided, $p<0.001$, **Figure 19A** and **19C**, bar 1), and the Flanker-only population yielded significant classification performance when trained and tested on the Flanker task ($p<0.001$, **Figure 19A** and **19C**, bar 5). The population of Number-only electrodes yielded significant classification performance when trained and tested on the Number task ($p<0.001$, **Figure 19A** and **19C**, bar 9), but also when trained and tested on the Stroop task ($p<0.01$, **Figure 19A** and **19C**, bar 7). Although the Number-only electrode population could detect conflict in the Stroop task, the performance on the Number task was still significantly higher than that on the Stroop task ($p<0.001$).

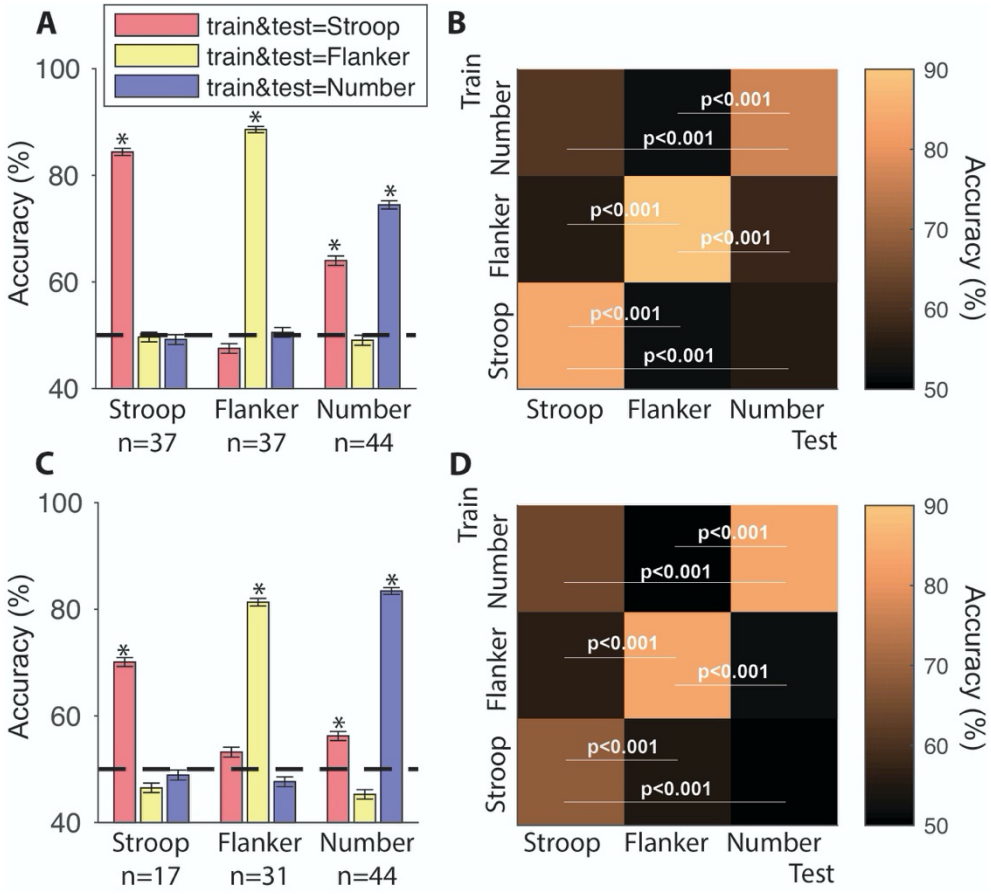

Figure 19. Task-specificity in population-based decoding of conflict in single trials. **A**, **C**. Accuracy of SVM classifier in congruent/incongruent discrimination when using a population of Stroop-specific electrodes (first three bars), Flanker-specific electrodes (next three bars), or Number-specific electrodes (last three bars). The SVM classifier was trained and tested with ten-fold cross-validation × 50 sessions of random sampling of trials using the high-gamma (**A**) and theta (**C**) band power data from the Stroop (red), Flanker (yellow), or Number (blue) task. Asterisks indicate that performances were significantly higher than chance (permutation test, 10,000 iterations, one-sided, $p<0.001$). **B**, **D**. Cross-task training and testing using high-gamma (**B**) and theta (**D**) band power. Here we used the same three populations from part **A** and **C**. The SVM classifier was trained on one task and tested on the other two tasks. The diagonal corresponds to training and testing within the same task and the off-diagonal entries show cross-task extrapolation. P values indicate the comparison between within-task and cross-task testing performances in each electrode population (permutation test with 10,000 iterations, one-sided). Accuracy is reflected by the color of each square (see color map on the right).

We next sought to assess whether a classifier trained only with data from one task could extrapolate to detect conflict in a different task. We trained three classifiers, one using data from the Stroop task only, one using data from the Flanker task only, and one using data from the Number task only using high-gamma (**Figure 19B**) and theta (**Figure 19D**) band power. Then we tested each classifier with data from the Stroop, Flanker, and Number task separately. We

performed this analysis using Stroop-only, Flanker-only, and Number-only electrodes. Note that this was different from the analyses in **Figure 19A** and **19C**, where the training and test data were always from the same task for each classifier, whereas here, the training and test data can be from different tasks. Here, even when the classifier was trained and tested using data from the same task, we still performed cross-validation across trials to avoid overfitting. As expected, the highest classification accuracies were observed for within-task training and testing (diagonal tiles in **Figure 19B** and **19D**). These three conditions not only exhibited better than chance accuracies (high-gamma: 82.6 ± 6.9%; theta: 79.5 ± 8.2%) but also significantly higher performance than all the corresponding cross-task accuracies (high-gamma: 55.3 ±4.3%, **Figure 19B**, $p<0.001$; theta: 53.6 ±5.0%, **Figure 19D**, $p<0.001$). In sum, even at the electrode population level, we observed minimal ability to detect conflict when a decoder was trained and tested in different tasks. However, using exactly the same approach but even fewer trials in each condition, there was high accuracy in distinguishing conflict in individual trials within each task. These results show that at the same electrode population level, within-task invariance is significantly more prominent than cross-task invariance.

Discussion

We studied the neural mechanisms underlying conflict resolution during cognitive control by recording intracranial field potentials from 694 electrodes in 16 subjects who performed three different tasks: Stroop, Flanker, and Number (**Figure 1**). Subjects showed increased reaction times during incongruent trials compared to congruent trials (**Figure 3**), a hallmark of cognitive control (Bush and Shin, 2006; Eriksen and Eriksen, 1974; Mayr et al., 2003; Stroop, 1935). Consistent

with previous studies (Caruana et al., 2014; Gaetz et al., 2013; Koga et al., 2011; Tang et al., 2016), we found robust modulation of neural signals in the high-gamma frequency band when comparing incongruent with congruent trials (**Figures 5, 7-9**). Conflict modulation was also present in the theta band (**Figure 10**) and other frequency bands (**Table 7**). Modulation was evident both when aligning neural signals to the behavioral response and to the stimulus onset, could be appreciated even in single trials, and showed within-task invariance to the different combinations of visual inputs. Surprisingly, despite this robust within-task invariance, most of the electrodes showed task-specificity, with clear incongruent/congruent modulation in only one task but not in the other two. A few electrodes showed task-modulation in two tasks but not the third task.

It is important to note that we focused on the time window within a trial, from the stimulus onset to the behavioral response time, when conflict resolution processes took place. We did not consider pre-stimulus and post-response time periods, which may represent estimation of the conflict element in upcoming trials and post-response feedback signals. Neural responses in these distinct windows may carry different operations and should not be confused with each other, which requires methods with high temporal resolution such as intracranial recordings.

We also considered an electrode ensemble machine learning decoding approach (**Figures 13**, **19**). Population-based decoding is highly sensitive and could, in principle, uncover a task-invariant representation even if we mainly observed specificity in individual electrodes. However, the decoding results also support the conclusion of clear within-task invariance (**Figure 13T-U**) and a largely task-specific representation (**Figure 19**). It is critical to emphasize that we used exactly the same electrode population that covered exactly the same brain regions in the two analyses. These decoding results cannot be ascribed to drifting neural signals or non-stationarities in the data. First, previous work showed that intracranial field potentials tend to be very stable

within a session, and even across recording sessions spanning multiple days (Bansal et al, 2012). Second and most importantly, for the total 18 blocks, each block of one task was always followed by a block of a different task.

Previous studies leveraging a single task precluded the possibility to assess task-invariance and task-specificity in conflict modulation. It is possible to draw inferences about potential invariance by comparing results in different studies; however, precise anatomical comparisons across subjects can be challenging, especially when considering coarse signals that smooth over large numbers of neurons. Inferences across studies do not necessarily imply that the same neural circuits represent conflict in an abstract format. Another potential confound is the distinction between signals aligned to the stimulus or to the behavioral response, which requires a careful comparison of the temporal dynamics of the neural responses. Stimulus-specific neural signals could be misconstrued as conflict modulation if neural responses are aligned to the motor output (e.g., **Figure 6AB**), and motor-specific neural signals could be misinterpreted as conflict modulation if neural responses are aligned to the stimulus onset (e.g., **Figure 6CD**). Thus, either due to using a single task or spatial and temporal averaging, it is difficult to differentiate whether the conflict-associated neural activities in many previous studies reflected task-specific modulation or an abstract conflict signal.

Locations of task-specific responses were widely distributed in the brain (**Figure 11**, **Tables 3-4**). These locations are consistent with many studies documenting responses during cognitive control in the frontal lobe (Bunge et al., 2002; Caruana et al., 2014; Fan et al., 2003; Milham and Banich, 2005; Parris et al., 2019; Robertson et al., 2014; Sheth et al., 2012; Tang et al., 2016), the parietal lobe (Bunge et al., 2002; Bush and Shin, 2006; Coulthard et al., 2008; Fan et al., 2003), the temporal lobe (Bush and Shin, 2006; Fan et al., 2003), the occipital lobe (Egner

and Hirsch, 2005a; Fan et al., 2003; Janssens et al., 2018), and other brain areas such as the insula (Menon and Uddin, 2010). These results suggest that cognitive control processes recruit distributed and largely task-specific networks rather than a single brain region (Dosenbach et al., 2007; Dosenbach et al., 2006; Fan et al., 2003; Marek and Dosenbach, 2018).

It is also important to emphasize that our sampling of brain locations was extensive but certainly not exhaustive (**Table 2**). By combining multiple subjects we achieved a fairly good coverage of most Desikan-Killiany regions (**Figure 4**). However, it is still quite possible that there are other brain regions that represent conflict in a task-invariant fashion that we could not sample here. It is also relevant to note that our study focused on intracranial field potentials; these signals reflect the activities of cell assemblies. It is conceivable that individual neurons might show more or less cross-task invariance than the results reported here. However, studies examining single unit activities in the frontal cortex are also consistent with a lack of task invariance in cognitive control (Ebitz et al., 2020; Fu et al., 2019; Smith et al., 2019). Several studies have shown that frontal cortex neurons demonstrate "mixed selectivity" (Rigotti et al., 2013). Such mixed selectivity is a good summary of the results described here at the level of intracranial field potentials, which seem to reflect a combination of conflict and task-specific demands. Our results are partially echoed by two recent single neuron studies. One study (Fu et al., 2022) examining neurons in the dorsal ACC (dACC) and pre-SMA, regions that were traditionally believed to be primarily involved in domain-general cognitive control processes, found clear evidence of task-specific single neuron activities. This study also showed conflict responses common to two tasks, Stroop and Number, when using the same behavioral response mode (button press). It is possible that the responses in both tasks highlight the common sensorimotor transformation component, especially given that dACC neurons are specialized for representing task-state variables relevant for behavior and are tuned

for a variety of sensory and motor elements (Heilbronner and Hayden, 2016). In contrast, we compared three tasks with either keypress or verbal responses and the fact that a great proportion of dual-task electrodes were selective for the Stroop and Number tasks but not the Flanker task adds weight to this interpretation. Another single-unit study (Ebitz et al., 2020) concluded that firing rates of dACC neurons more likely amplify task-relevant sensorimotor information to facilitate conflict resolution rather than signal conflict abstractly.

We deliberately designed the tasks to be different in terms of the sensory inputs and motor outputs. Conflict emerged upon the discrepancy between color and semantic meaning (Stroop), inconsistent arrow directions (Flanker), and the incongruent position and identity of the target number (Number). Subjects used either verbal responses (Stroop, Number) or keypress responses (Flanker) as output. We conjectured that a general, abstract, signature of cognitive control should be independent of the input and output modalities. However, we found no such task-invariant conflict signals. It is possible that neural signals from electrodes that show conflict modulation in two tasks (e.g., **Figure 17**) correlated with the common aspects of the two tasks. For example, electrodes that showed modulation exclusively during the Stroop and Number tasks (e.g., **Figure 17B**) might be involved in conflict expressed through verbal output, and electrodes that showed modulation in Flanker and Number tasks may be involved in dealing with the Flanker pattern that was present in both tasks. Despite dual-task modulations, the majority of electrodes responded in a task-specific manner, arguably demonstrating engagement in conflict through specific sensory-motor combinations but exhibiting generalized conflict signals within a task. Collectively, our results indicate that cognitive control is orchestrated by distinct and distributed networks and is characterized by within-task invariant and cross-task specific conflict representations.

Materials and Methods

Subjects and recording procedures

Subjects were 16 patients (8 female, ages 12-62, **Table 1**) with pharmacologically-intractable epilepsy treated at Taipei Veterans General Hospital (TVGH), Boston Children's Hospital (BCH), Brigham and Women's Hospital (BWH), and Johns Hopkins Medical Hospital (JHMH). The electrode locations were purely dictated by clinical considerations, precluding any quantitative estimation of sample size at study design. The target number of subjects, 16, was decided during study design based on historical data of electrode distributions from previous studies. This study was approved by the institutional review board in each hospital and was carried out with subjects' informed consent. Subjects were implanted with intracranial depth electrodes (Ad-Tech, Racine, WI, USA). The total number of electrodes was 1,877 (**Table 1**). Neurophysiological data were recorded using XLTEK (Oakville, ON, Canada), Bio-Logic (Knoxville, TN, USA), Nihon Kohden (Tokyo, Japan), and Natus (Pleasanton, CA). The sampling rate was 2048 Hz at BCH and TVGH, 1000 Hz at JHMH, and 512 Hz at BWH. All data were bipolarly referenced. There were no seizure events in any of the sessions. Electrodes in the epileptogenic foci, as well as pathological areas, were removed from analyses.

Task procedures

Each subject completed three tasks in a single recording session: Stroop, Flanker, and Number. A schematic rendering of the tasks is shown in **Figure 1**. Each session contained 18 blocks, with 30 trials of one task (Stroop, Flanker, or Number) per block. The target number of trials was pre-defined based on the results of one of our previous studies (Tang et al, 2016) and

based on a pilot study with 4 healthy volunteers where we confirmed that conflict (i.e., reaction time difference between congruent and incongruent trials) could be robustly detected with this number of trials. Per our IRB protocols, subjects could stop testing at any time; subjects who completed different numbers of blocks are indicated in **Table 1** in bold font. Subjects completed the tasks in normally lighted and quiet rooms. The experiments were written and presented using the Psychtoolbox extension in Matlab_R2016b. Subjects viewed and completed the experiment using a 13-inch Apple Mac laptop. Stimuli subtended approximately 5 degrees of visual angle and were centered on the screen. Before each experiment started, each subject went over a short practice session until the instructions were fully understood. During the actual experiment, no correct/incorrect feedback was provided.

All trials started with 500 ms of fixation, followed by stimulus presentation. The stimulus was presented for 2,000 ms (Stroop, Number), or until the minimum of 2,000 ms and the subject's key response time (Flanker). The stimuli were presented in white (Flanker, Number) or red/green/blue font color (Stroop), on a black background. For those subjects in Taipei, Stroop task stimuli were presented in traditional Chinese characters. Subjects provided a verbal response recorded using a Yeti microphone with an 8,192 Hz sampling rate (Stroop, Number), or a two-alternative keypress response using the left and right keys on the experiment laptop (Flanker).

Electrode localization

We used the iELVis (Groppe et al., 2017) pipeline to localize the depth electrodes. Pre-implant MRI (T1, no contrast) was processed and automatically segmented by Freesurfer (Dale et al., 1999; Reuter et al., 2012), followed by co-registering the post-implant CT to the processed MR images. Electrodes were then identified visually and marked in each subject's co-registered space

using the BioImage Suite (Joshi et al., 2011). Each electrode was assigned an anatomical location (parcellated cortices (Desikan et al., 2006), white matter, subcortical regions, or unknown) using the Freesurfer localization tool. Unknown locations could be due to brain lesions or pathological brain areas. Electrodes in the white matter, ventricles, cerebellum, and unknown locations were excluded from analyses. Out of a total of 1,877 electrodes, we included 694 bipolarly referenced electrodes in the analyses. To show the position of electrodes from different subjects (**Figures 4**, **11**), electrode locations were mapped onto the MNI305 average human brain via an affine transformation (Wu et al., 2018).

Behavioral analyses

The content of verbal responses (Stroop and Number tasks) was transcribed offline. The transcription was blind to the ground truth answers as well as neural responses. The behavioral reaction time for verbal response (Stroop and Number tasks) was determined as the first time the energy of the soundtrack was three standard deviations above the mean energy of the whole trial. Any noise (e.g., door slam, coughing, etc.) before the actual trial response was carefully identified and smoothed to prevent false automatic identification of behavioral response time. The keypress reaction time (Flanker task) was recorded by the Psychtoolbox code.

Preprocessing of intracranial field potential data

A zero-phase digital notch filter (Matlab function "filtfilt") was applied to the broadband signals to remove the AC line frequency at 60 Hz and harmonics. For each electrode and each task, trials with amplitudes (max-min voltage from fixation onset to stimulus off) larger than three standard deviations above the mean amplitude across all trials were considered as containing

artifacts and excluded from analysis (Bansal et al., 2012). The percentage of trials excluded by this criterion was 1.05% (Stroop), 1.25% (Flanker), and 1.29% (Number).

Single electrode analysis of modulation by conflict

We computed the high-gamma band (70-120 Hz), low-gamma band (35-70 Hz), beta band (12-35 Hz), alpha band (8-12 Hz), and theta band (4-8 Hz) power of the intracranial field potential signals by using a multi-taper moving-window spectral estimation method implemented in the Chronux toolbox (Mitra and Bokil, 2008). The time-bandwidth product, number of tapers, and size of moving window used for each frequency band are listed in **Table 7** (Tang et al., 2016). Throughout the paper, we focused on the high-gamma band and theta band signals. The power in the corresponding frequency band was z-scored by subtracting the mean power during the baseline period (500 ms before stimulus onset) and dividing by the standard deviation of the band power during the baseline. Only correct trials were included in the analyses.

First, we examined whether an electrode exhibited any response at all to the stimuli. An electrode was defined as "responsive" if the z-scored high-gamma power during the congruent condition was larger than 1 for at least 150 consecutive milliseconds (15 × 200 ms window shifted by 10 ms), starting from stimulus onset to average behavioral response time. To determine whether an electrode showed conflict modulation, we compared the band power between the congruent and incongruent conditions of each task. For each time bin (200 ms shifted by 10 ms), we compared the band power of incongruent versus congruent trials using a permutation test with 5,000 iterations ($\alpha = 0.05$). An electrode was denoted as showing conflict modulation if the following two criteria were satisfied: (1) The band power of incongruent trials was significantly different from the power in congruent trials for at least 150 consecutive milliseconds (15 × 200 ms window

shifted by 10 ms); (2) Criteria (1) was satisfied in both behavioral response-aligned and stimulus-aligned conditions. When the band power was aligned to behavioral response, selection criteria were applied to the time window starting from the average stimulus onset to the behavioral reaction time. When the band power was aligned to the stimulus, the time window was from stimulus onset to average behavioral reaction time.

An electrode was considered to be visually responsive if the maximum z-scored high-gamma band power was larger than 2 during the 300 milliseconds after stimulus onset. An electrode was considered showing motor responsive if the maximum z-scored high-gamma band power was larger than 2 during 300 milliseconds before behavioral response and the power continually increased during this time window. An electrode was considered to be visually selective for a particular task if it was visually responsive to stimuli only in one task. An electrode was deemed to show motor selectivity if it was motor responsive to verbal output (Stroop and Number) only or keypress (Flanker) only.

To assess the correlation between conflict responses and reaction times (**Figure 12**), a linear regression ("fitlm" function in Matlab) was performed between reaction time and the mean high-gamma or theta band power from stimulus onset to response time of each trial for all conflict-modulated electrodes. An electrode was considered to show a significant correlation if the p value of the linear regression slope was smaller than 0.05.

Classifier analyses

We quantified whether we could distinguish between congruent and incongruent trials in individual trials based on the activity of pseudo-populations formed by multiple electrodes (Liu et al., 2009). We used a linear-kernel support vector machine with ten-fold cross-validation for all

the classifier analyses (**Figures 13**, **19**). Two features were calculated for each trial from each electrode: the mean and maximum band power from average stimulus onset to the behavioral response. These analyses were conducted separately for the high-gamma and theta frequency bands. All data were normalized to zero mean and one standard deviation before each training and testing session. All the classifier performance results reported were based on cross-validated test data. The main text, **Figure 13**, and **Figure 19** describe all the different combinations of training and test data used, which were critical to evaluate within-task and cross-task invariance.

Chapter 2

Neural dynamics of associative working memory

Introduction

Working memory is a cognitive system with limited capacity that stores information temporarily. Items held in working memory can be readily available for mental manipulations like calculation, reordering, and retrieval. An example of working memory task is remembering a verification code sent to one's phone which is immediately forgotten after inputting the numbers. Intracranial recordings, through their superior spatiotemporal resolution, have significantly advanced the understanding of memory processes (Johnson and Knight, 2015; Rutishauser et al., 2021; Sederberg et al., 2007). Results from intracranial EEG are indispensable for formulating the theoretical foundation for the human memory system, guiding and restraining biologically-inspired computational models, and developing neural intervention techniques to treat memory dysfunctions (Ezzyat et al., 2018; Johnson and Knight, 2015; Rutishauser et al., 2021).

Associative working memory is a subtype of working memory, which refers to the ability to associate or link two items that typically lack apparent logical relationship and maintain that association for several seconds or minutes. Associative memory can very well be a form of long-term memory, such as learning to associate a person's name and face and being able to recall such association after many years. A key neural phenomenon under associative memory is that neurons can learn association and gain selectivity to different constituent elements within an association (Ison et al., 2015; Ranganath et al., 2004; Rutishauser et al., 2021; Sakai and Miyashita, 1991; Zhou et al., 2007). Even a single neuron can respond to both entities of a learned association (Ison

et al., 2015; Rutishauser et al., 2021; Sakai and Miyashita, 1991). Researchers have also identified neural responses during associative memory retrieval. These responses may be selective to the content of a cue's associate rather than or in addition to the cue itself and may signal successful associative memory retrieval (Bergmann et al., 2012a; Bunge et al., 2004; Ranganath et al., 2004; Staresina et al., 2019; Zhou et al., 2007).

Although the neural mechanisms of associative working memory have been studied to some extent, several key questions remain to be addressed. First, despite the name "associative" memory, it includes a non-associative component. To illustrate, A and B are two pieces of associated information. To recall B from A, one has to first recognize what A is, which is a recognition memory process but not that of associative memory. Recognition memory refers to the ability to recognize as familiar previously encountered events, objects, people, scenes, and so on. Several studies have reported brain regions or single neurons that are selective for familiarity or novelty (Daselaar et al., 2006; Fried et al., 1997; Knight, 1996; Murray et al., 2014; Park et al., 2014; Rutishauser et al., 2006; Rutishauser et al., 2008; Rutishauser et al., 2015; Viskontas et al., 2006; Yassa and Stark, 2008; Zaehle et al., 2013) as well as those that code for different degrees of familiarity, recency, or memory strength (Montaldi et al., 2006; Park et al., 2014; Rutishauser et al., 2010; Rutishauser et al., 2008; Yassa and Stark, 2008). Since most of these studies used the classic yes/no paradigm, it is unclear how neural activities coordinate recognition memory processes in highly dynamic environments. A piece of information can constantly change its status in memory. When someone is exposed to completely new information, it triggers a fresh and strong mental representation; however, as more information feeds in or memory load increases, that particular information may fade away and its memory trace becomes gradually weaker. When that piece of information emerges and gains attention again, it is back to having a fresh and strong

status in memory. It may appear almost novel if a very long time has passed since its last appearance, or it may be familiar if only a short duration has elapsed. How such dynamics of memory is embodied by neural activities is poorly understood.

Second, several studies have identified the neural correlates of successful associative memory retrieval (Bergmann et al., 2012a; Ranganath et al., 2004; Sakai and Miyashita, 1991; Staresina et al., 2019; Zhou et al., 2007). It is important to know whether these signals imply only successful recall of pair or also contain information about the memory status of pair, for example, its memory strength or level of familiarity. This question relates to the long-lasting debate over whether familiarity and recollection are neurally distinct (Gothe and Oberauer, 2008; Schurgin, 2018; Yonelinas, 2001). Several studies showed that familiarity and recollection are not functionally dissociable, but can be supported by same brain structures (Merkow et al., 2015; Rutishauser et al., 2008; Wais et al., 2010; Wais et al., 2006; Wixted and Squire, 2011). The opposing view, or the dual-process model, claims that recollection and familiarity are independent processes operated by distinct structures (Aggleton and Brown, 1999; Daselaar et al., 2006; Eldridge et al., 2000; Jacoby and Dallas, 1981; Yonelinas, 2001, 2013). Although familiarity and recollection have been commonly discussed under the context of long-term recognition and episodic memory, recollection- and familiarity-like processes may also serve short-term and working memory (Danker et al., 2008; Gothe and Oberauer, 2008; Oberauer, 2005; Yonelinas, 2013); however, the neural constructs of recollection and familiarity pertaining to associative working memory are unclear.

We recorded intracranial field potentials from 20 pharmacologically-intractable epilepsy patients while they were playing the memory matching game (**Figure 20**, **Movie S1**). We focused on the neural activities in the gamma frequency band (30-150 Hz) because a large number of

studies have suggested that the gamma activities are associated with various components of memory processes, including binding object representations and forming perceptual and memory representations (Jensen et al., 2007; Johnson and Knight, 2015; Lundqvist et al., 2016; Mormann et al., 2005; Singer and Gray, 1995; Tallon-Baudry and Bertrand, 1999; Uhlhaas et al., 2011).(van Vugt et al., 2010) Neural activities in the gamma band are thought to reflect synchronous firing of ensembles of neurons that support neuronal communication as well as cortical computation (Jensen et al., 2007; Uhlhaas et al., 2011), have been shown to correlate with neural spiking activities (Buzsaki et al., 2012; Cardin et al., 2009; Lundqvist et al., 2016; Ray and Maunsell, 2011), and demonstrate rich information about both behavior and neurophysiology (Crone et al., 2011; Sederberg et al., 2007). We constructed generalized linear models to characterize how neural responses in the gamma band were modulated and predicted by behavioral parameters that have important implications for memory processes. We showed that gamma activities reflect a wide range of constantly changing working memory status including novelty, recency, and familiarity. Gamma activities are predictive of successful memory retrieval of locational associative information. By comparing match and mismatch, we found that neural responses preceding successful but not unsuccessful associative recall also contain information about the recency or familiarity of the to-be-recalled information.

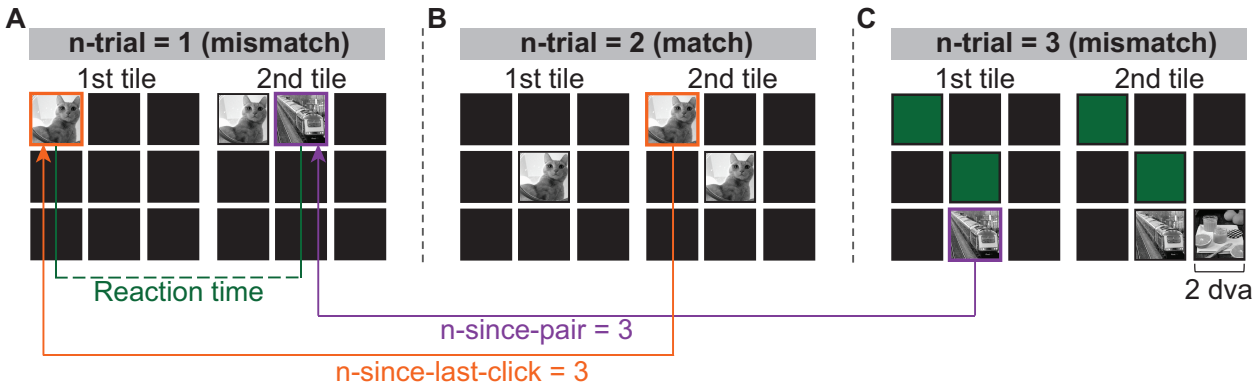

Figure 20. Experimental paradigm. Subjects performed the memory matching game during intracranial neurophysiological recordings with sEEG or EcoG electrodes. Subjects started with a 3×3 board and progressed to more difficult blocks (4×4, 5×5, 6×6, and finally 7×7). If the subject could not complete the current block of board size n×n (n≥4) within a time limit (**Methods**), it would end immediately, and a new block of reduced n started (**Methods**). Subjects would remain at 7×7 boards if completed it within the time limit. A-C illustrate three consecutive trials in a 3×3 board. In each trial, two tiles were flipped sequentially with a self-paced interval. If the two tiles contained different images (**A**, **C**, mismatch), both tiles reset to their original active states after one second. If both tiles contained the same image (**B**, match), they turned green and became fixed after one second. Three behavioral predictors used in the generalized linear models are defined here: reaction time (the time between the first and second tile within a trial), n-since-pair (the number of clicks elapsed since the last time a given tile's matching pair was clicked), and n-since-last-click (the number of clicks elapsed since the same tile was clicked last time).

Results

We recorded intracranial field potentials (IFPs) from 20 participants with pharmacologically-intractable epilepsy implanted with sEEG electrodes (**Table S1**, one patient also had ECoG electrodes). Participants performed the memory matching game (**Figure 20, Movie S1**). Each trial consisted of two self-paced clicks; clicking on a tile revealed an image (**Figure 20A**). Image categories included person, animal, food, vehicle, and indoor scenes. If the two tiles in a trial contained the same image (match, **Figure 20B**), the two tiles turned green and could not be clicked again for the remainder of the block. If the two images were different (mismatch, **Figure 20A** and **20C**), the two tiles turned black and became clickable again. Subjects started in a 3×3 tile board block and progressed to more difficult blocks (4×4, 5×5, 6×6, 7×7 tiles). All tiles had a corresponding match, except for one tile in any board with an odd number of tiles (3×3, 5×5, and

7×7).

Mismatch trials showed longer reaction times and were associated with less recent exposure to matching pairs

The average number of clicks per tile increased with difficulty (board size), as expected (**Figure 21A**). All subjects performed much better than a memoryless model (random clicking, $p<0.001$, permutation test, 5,000 iterations, one-tailed) and worse than a model assuming perfect memory ($p<0.001$, **Figure 21A**). The reaction time (RT) was defined as the interval between the first click and the second click in each trial. The reaction time was longer for mismatch compared with match trials for all board sizes ($p<0.007$, permutation test, 5,000 iterations, **Figure 21B**).

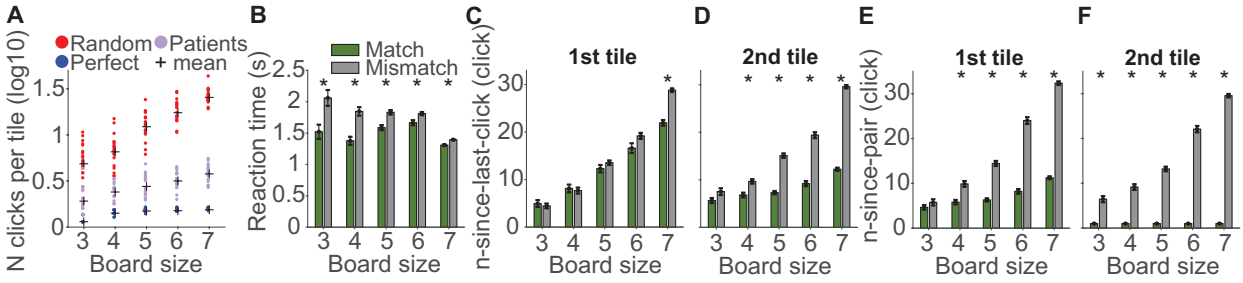

Figure 21. Behaviors of match and mismatch across all subjects. A. Number of clicks per tile of random simulation model (red, n=20), epilepsy patients (purple, n=20), and perfect memory simulation model (blue, n=20) in log scale when playing the memory matching game (**Methods**). Perfect memory simulation models may generate different number of clicks per tile because where to click a new tile was randomized. Plus signs indicate the means of groups. The performance of the epilepsy patients was better than the random model and worse than the perfect model. The number of clicks per tile increased as board size incremented. **B**. Bar plot comparing the reaction time of match and mismatch at different board sizes. Asterisks denote significant differences between match and mismatch trials (permutation test, 5,000 iterations, $\alpha=0.01$). Reaction time of mismatch was longer than match. **C**-**F**. Bar plots comparing n-since-last-click for the first tile (**C**) and the second tile (**D**), and n-since-pair for the first tile (**E**) and second tile (**F**) between match and mismatch at each board size. Asterisks denote significant differences between match and mismatch trials (permutation test, 5,000 iterations, $\alpha=0.01$). For n-since-last-click, we removed trials in which any tile was clicked the first time. For n-since-pair, we removed trials in which any tile's matching pair hadn't been seen. All error bars indicate s.e.m.

For a tile in a given trial, we defined n-since-last-click as the number of clicks since the same tile was clicked last time (**Figure 20**). As expected, n-since-last-click increased with board size ($p<0.001$, linear regression, F-test, **Figure 21C-D**). For the second tile, n-since-last-click was

larger in mismatch compared to match trials for all board sizes except the 3×3 case ($p<0.001$, permutation test, 10,000 iterations, **Figure 21D**). This observation also held for the first tile only for the 7×7 board size ($p<0.001$, **Figure 21C**). Because the two tiles in each trial were flipped sequentially with a self-paced interval (**Figure 20**) rather than simultaneously, the first tile served as a cue, whose matching pair's location was the target information to be recollected. The aim of the second click was to match the cue while that of the first click was not as singular but manifold, for example, exploring new tiles. As a result, where to click the first tile was more random than where to click the second tile, which may explain why the difference in n-since-last-click between match and mismatch was more manifested in the 2nd tile than the 1st tile. Match trials were associated with smaller n-since-last-click, or higher level of familiarity with the cue's pair, compared with mismatch trials (**Figure 21D**).

For a tile in a given trial, we defined n-since-pair as the number of clicks since the last time when its matching pair was seen (**Figure 20**). As expected, n-since-pair increased with the board size given the increased difficulty ($p<0.001$, linear regression, F-test, **Figure 21E-F**). Additionally, the more recent the tile's matching pair was seen, the more likely that the trial was a match. Thus, n-since-pair was larger in mismatch compared to match trials in all cases except the 3×3 board size for the first tile ($p<0.001$, permutation test, 10,000 iterations, **Figure 21E**). For the second tile, n-since-pair for any match trial was always one because the matching pair would have been revealed in the previous click by definition (**Figure 21F**).

Gamma responses detect novelty and track familiarity during non-associative recognition

We recorded intracranial field potentials from 1,750 depth of ECoG electrodes. We included 676 bipolarly referenced electrodes in the gray matter (**Methods**). Electrode locations are shown in **Table 8** and **Figure 22**.

Table 8. Electrode locations.

Region	Left	Right	Total
amygdala	21	24	45
bankssts		7	7
caudalanteriorcingulate	1	3	4
caudalmiddlefrontal	1	5	6
entorhinal	5		5
fusiform	15	10	25
hippocampus	36	31	67
inferiorparietal	11	8	19
inferiortemporal	14	9	23
insula	13	15	28
isthmuscingulate	6		6
lateraloccipital	3	2	5
lateralorbitofrontal	19	32	51
lingual	3	2	5
medialorbitofrontal	4	7	11
middletemporal	53	38	91
paracentral	4		4
parahippocampal	10	13	23
parsopercularis	11	5	16
parsorbitalis	2	3	5
parstriangularis	13	19	32
pericalcarine		1	1
postcentral	8	12	20
posteriorcingulate	4	3	7
precentral	8	21	29
precuneus	3		3
rostralanteriorcingulate	7	7	14
rostralmiddlefrontal	1	4	5
superiorfrontal	13	19	32
superiorparietal	8	12	20
superiortemporal	26	21	47
supramarginal	10	4	14
transversetemporal	6		6
Total	**339**	**337**	**676**

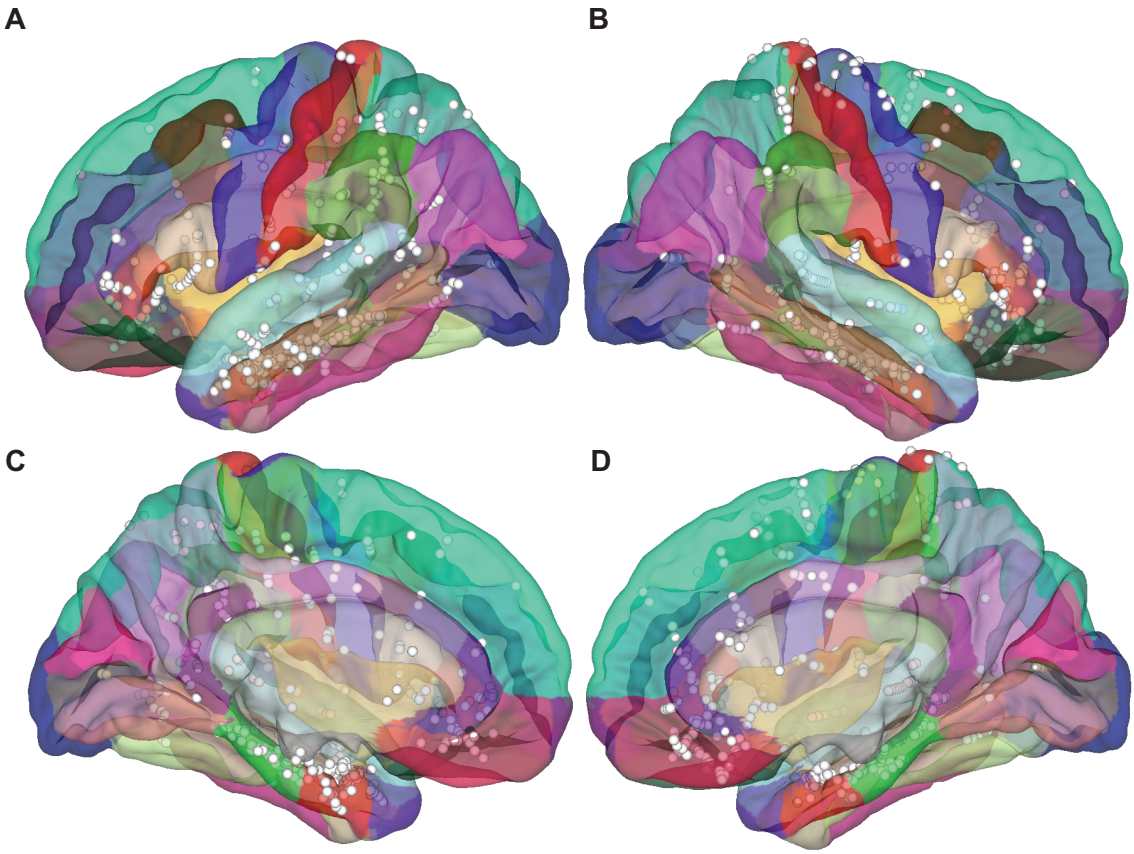

Figure 22. **Locations of electrodes**. Each sphere reflects one of each pair of nearby electrodes that were bipolarly referenced (n=676), overlayed on the Desikan-Killiany atlas with different views: **A**: left lateral; **B**: right lateral; **C**: left medial; **D**: right medial.

We built two generalized linear models (GLM) to characterize how the neural responses depended on the characteristics of each trial. The first model focused on the neural responses to the first tile and the second model on the neural responses to the second tile. In both cases, we focused on the area under the curve (AUC) of the gamma band (30-150 Hz) power in each trial (**Methods**). For the first GLM, the time window started when the first tile was clicked and ended at a time corresponding to the 90th-percentile of the distribution of reaction times. This criterion was a reasonable tradeoff between minimizing overlap with responses after the second tile and maximally capturing information before the second tile. For the second model, the time window started with the second click and ended one second afterward.

We considered 15 predictors for the GLM models, including whether a trial was a match or not, reaction time, n-since-last-click, and n-since-pair. **Table 9** lists all the predictors and their definitions. Several predictors were correlated with each other (**Figure 23C-D**). However, the variance inflation factor (VIF) of each predictor was smaller than 3 for all subjects (**Figure 23A-B**), thus the multi-collinearities between predictors did not harm the performance of our model (**Methods**).

Table 9. Predictors used in the generalized linear models and their definitions. The "indoor scene" category was dropped to avoid the "dummy variable trap" issue in regression models (Methods).

Predictor	Description	Which tile
match	Whether the trial was a match or mismatch	both
n-since-pair*match	How many clicks ago the tile's pair was clicked (match trials only)	first
n-since-last-click	How many clicks ago the same tile was clicked	both
first-click	Whether a tile was clicked the very first time	both
n-times-seen	Number of times the same image (not tile) had been seen (including current)	both
next-match	Whether the next trial was a match or mismatch	both
reaction-time	Time between the first and second tile in second	both
board-size	Total number of tiles in the current block	both
x-position	x position in pixel	both
y-position	y position in pixel	both
distance	Distance between the 2nd tile in the current trial and the 1st tile in the next trial in pixel	both
animal	Image belonged to animal category	both
food	Image belonged to food category	both
person	Image belonged to person category	both
vehicle	Image belonged to vehicle category	both

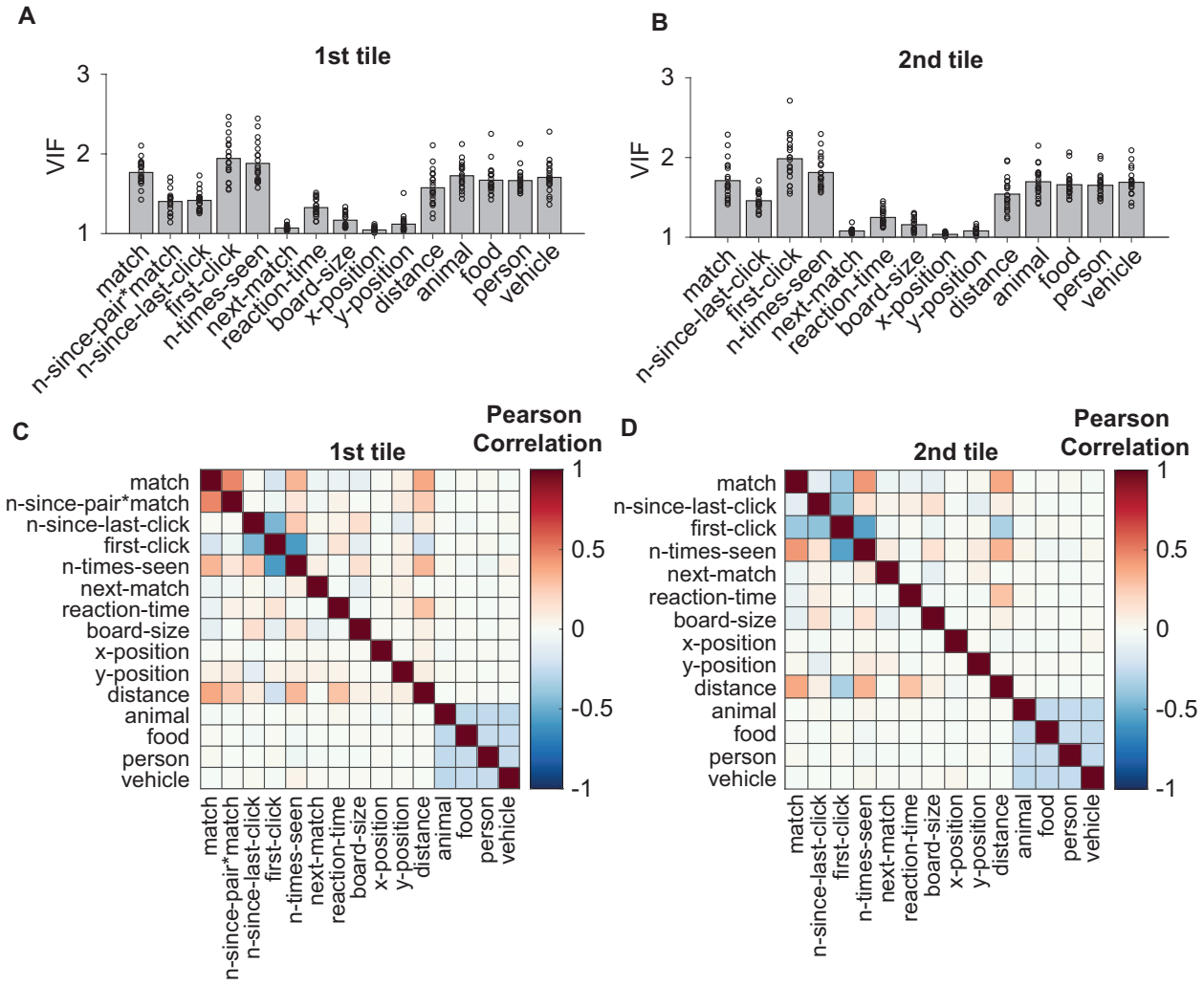

Figure 23. Collinearity and correlation among predictor variables. **A**, **B**. Bar plots indicating the average VIF of each predictor across all subjects. Each dot represents the VIF of that specific predictor for each subject. **C**, **D**. Average Pearson correlation between each pair of predictors for the first tile (**C**) and the second (**D**) tile across all subjects.

When the first tile in a trial was clicked, its status in working memory guided the following actions. If it was a new image, one needed to encode it for future use. Instead, if the tile had been viewed before, it should appear familiar to the subject, and the degree of familiarity may be dependent on how long ago that tile was last seen. We used the "first-click" and the "n-since-last-click" predictors to describe novelty and familiarity. "First-click" refers to whether a tile was seen the very first time. "N-since-last-click" refers to the number of clicks elapsed since last time seeing the same tile. The smaller the n-since-last-click (NSLC), the more recently the same tile was seen

last time, as if it's more familiar. As opposed, the larger the n-since-last-click, the longer ago the same tile was seen last time, as if it's more novel or unfamiliar. According to the generalized linear model, the first-click predictor could predict the gamma power AUC during the 1st tile in 50 electrodes (**Figure 24G**, **Table 10**). The lateral orbitofrontal (LOF) cortex and the pars opercularis (inferior frontal) contained significantly more electrodes than expected from chance ($p<0.01$, bootstrap analysis with 5,000 shuffles, **Methods**). **Figure 24A-F** shows an example electrode in the right LOF gyrus that had first-click as a significant predictor. When a tile was seen the first time, there was a drop of gamma power (**Figure 24B** and **Figure 24D**); in contrast, the gamma power increased for non-first-clicks. The gamma power drop in response to first-clicks was reflected by the negative t-statistic of that predictor in the GLM (**Figure 24A**). Such pattern can be appreciated even at single trial level (**Figure 24C**). **Figure 24D-F** show trial-averaged spectrograms for first-clicks, NSLC = 1 (super familiar), and NSLC >1 (less familiar). **Figure 24G** plots the locations of electrodes where first-click was a significant predictor.

Figure 24. Example first-click and n-since-last-click electrodes. A, **H**: T-statistic of each predictor in the GLM. Asterisks indicate significant predictors for the gamma power AUC. **B**, **I**: Z-scored gamma power of first-click and different n-since-last-click values aligned to the first tile onset (solid lines). Dashed lines indicate the mean reaction times. Multiple dashed lines in (**I**) indicate results after RT equalization (**Methods**). Shaded error bars indicate s.e.m. **C**, **J**: Raster plots showing the z-scored gamma power in individual trials ordered by first-click and then smaller to larger n-since-last-click. Yellow horizontal lines and colored vertical lines indicate n-since-last-click ranges same as in (**B**) and (**I**). **D**-**F**,**K**-**M**. Spectrograms showing the band power aligned to the first tile onset during first-click (**D**, **K**), n-since-last-click=1 (**E**, **L**), and n-since-last-click>1 (**F**, **M**). **G**. Locations of all first-click electrodes during the first tile. Blue: first-click only; red: both first-click and n-since-last-click. **N**. Locations of all n-since-last-click electrodes during the first tile. Orange: n-since-last-click only; red: both n-since-last-click and first-click. All electrodes were shown on one hemisphere for display purpose.

Figure 24 (Continued).

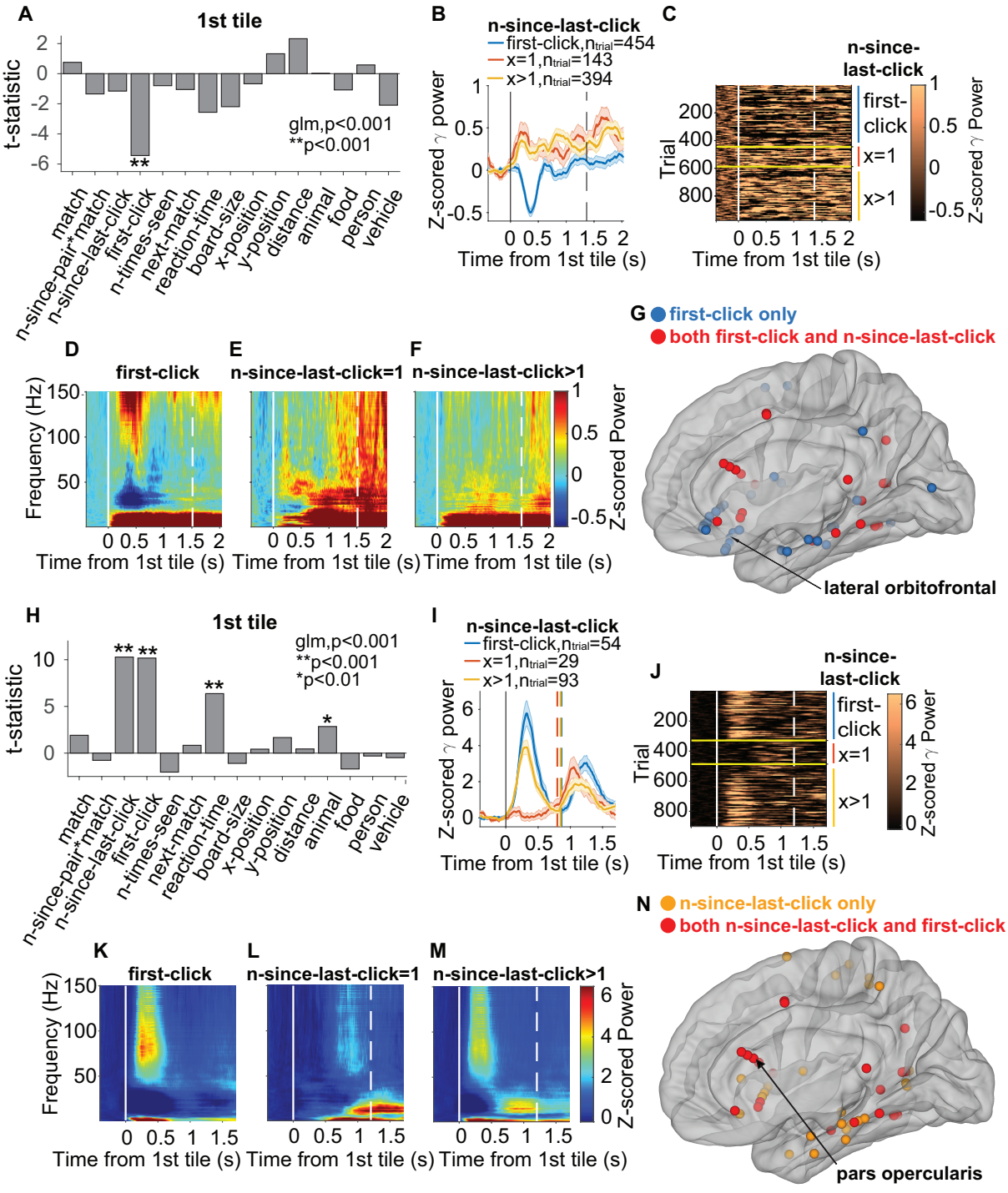

Table 10. Locations of electrodes where first-click was a significant predictor for the gamma power AUC during the 1st tile determined by the GLM. "Number" indicates the number of first-click electrodes in that region.

Location	Number
amygdala	2
bankssts	1
caudalmiddlefrontal	1
fusiform	5
inferiorparietal	3
inferiortemporal	2
insula	1
lateraloccipital	1
lateralorbitofrontal	14
medialorbitofrontal	2
middletemporal	4
parahippocampal	1
parsopercularis	5
parstriangularis	3
precentral	4
rostralmiddlefrontal	1
superiortemporal	1
supramarginal	1
Total	**50**

The n-since-last-click predictor was significant ($p<0.01$, GLM) in 45 (6.7%) electrodes (**Figure 24N**, **Table 11**) during the first tile. **Figure 24H-M** presents an example electrode located in the left pars opercularis where the gamma power levels were predicted by both n-since-last-click and first-click. The positive t-statistic suggests that the larger the n-since-last-click, or the less familiar a tile appeared to be, the higher the gamma power ($p<0.001$, GLM, **Figure 24H**). Tiles that were being viewed the first time were completely novel and induced the highest gamma responses (**Figure 24I**, blue line). Although reaction time was a significant predictor for this electrode, difference in gamma power cannot be entirely due to the difference in reaction time. After performing reaction time equalization (see dashed lines indicating equalized RT in **Figure 24I**), difference in gamma power was still highly obvious. The modulation of the gamma responses by novelty and familiarity was also evident in single trials (**Figure 24J**) as well as in the

spectrograms (**Figure 24K-M**). Such positive correlation between n-since-last-click and gamma power represented the majority of the data (37 out of 45, 82.2%). On the contrary, 8 (17.8%) electrodes showed negative correlation, indicating higher gamma power toward more familiar items. An example electrode in the right pars opercularis was shown in **Figure 25.** For this electrode, the smaller the n-since-last-click, or the more familiar the item, the higher the gamma power. Tiles that were being clicked the first time were completely unfamiliar and were associated with the lowest power (**Figure 25,** blue line). Both the examples in **Figure 24H-M** and **Figure 25** also had the "first-click" as a significant predictor, meaning that they not only tracked along the familiarity gradient but were also able to specifically detect novelty. It is also to be highlighted that successful detection of novelty is a prerequisite for encoding, thus novelty signals may also represent the initial encoding operation. The intersect between first-click and n-since-last-click yielded 20 electrodes (**Figure 24G** and **Figure 24N**, red spheres). Among these 20 electrodes, the signs of the t-statistics of n-since-last-click and first-click predictors were consistent for 18 electrodes (16 positive and 2 negative), meaning that if first-click triggered high response, larger n-since-last-click did so too. Only 2 exhibited the opposite signs. These results indicate that complete novelty largely resembles extremely low familiarity in neural language.

Table 11. Locations of electrodes where n-since-last-click was a significant predictor for the gamma power AUC during the first tile determined by the GLM. "Number" indicates the number of electrodes in that region.

Location	Number
hippocampus	1
bankssts	2
entorhinal	1
fusiform	4
inferiorparietal	4
inferiortemporal	3
insula	4
lateralorbitofrontal	4
middletemporal	4
parahippocampal	3
parsopercularis	4
parstriangularis	1
postcentral	2
precentral	2
rostralmiddlefrontal	1
superiorfrontal	1
superiorparietal	2
superiortemporal	1
supramarginal	1
Total	**45**

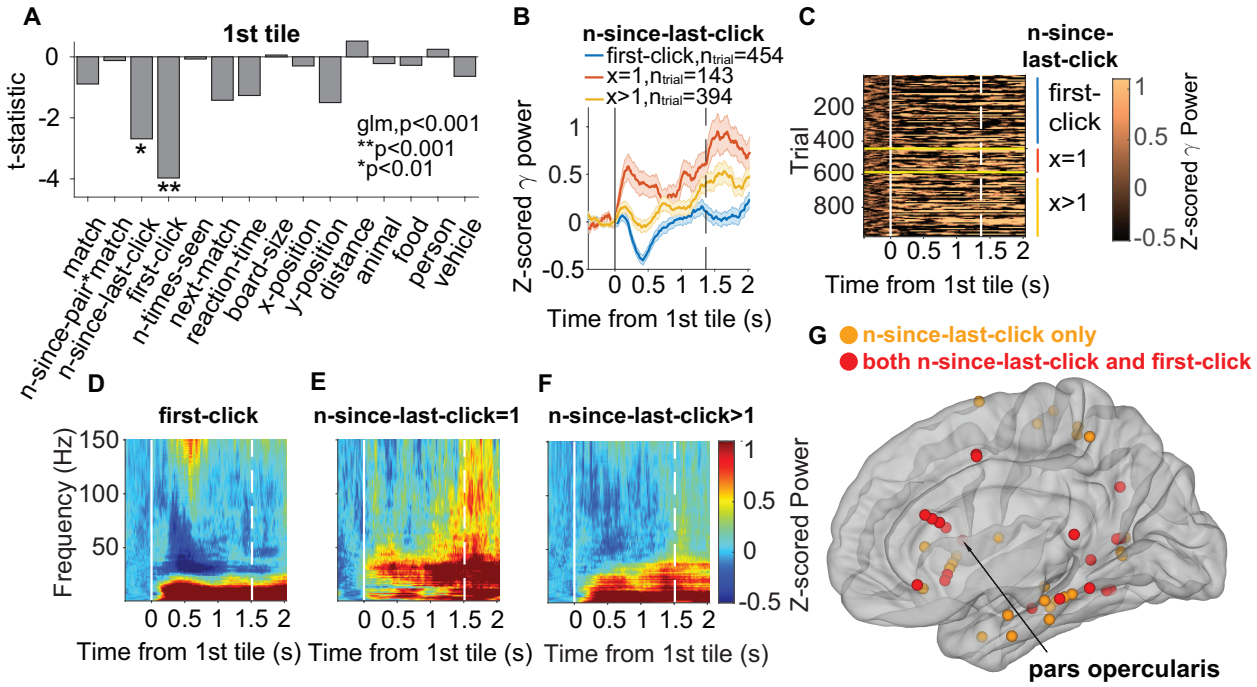

Figure 25. Example electrode located in the right pars opercularis where both first-click and n-since-last-click were significant predictors. A. T-statistic of each predictor in the GLM for the first tile. Asterisks indicate significant predictors for the gamma power AUC. **B**. Z-scored gamma (30-150 Hz) power for first-click (blue), n-since-last-click=1 (red), and n-since-last-click>1 (yellow). Dashed line indicates the mean reaction time. **C**. Raster plot showing the z-scored gamma power in individual trials ordered by first-click and then smaller to larger n-since-last-click. Yellow horizontal lines and colored vertical lines indicate n-since-last-click ranges same as in (**B**). **D**-**F**. Spectrograms showing the band power aligned to the first tile onset during first-click (**D**), n-since-last-click=1 (**E**), and n-since-last-click>1 (**F**). **G**. Locations of all n-since-last-click electrodes during the first tile. Orange: n-since-last-click only; red: both n-since-last-click and first-click. All electrodes were plotted on one hemisphere for display purpose.

First-click and n-since-last-click electrodes could be selective to image category. **Figure 26** shows an example electrode in the left fusiform gyrus where both novelty and vehicle were significant GLM predictors (**Figure 26A**). Novel tiles elicited higher gamma responses than seen tiles (**Figure 26B**). This electrode was selective to images of transport vehicles which led to higher gamma power than images of other categories (**Figure 26C**). **Table 12** shows the locations of all the first-click and n-since-last-click electrodes that were also selective for at least one image category.

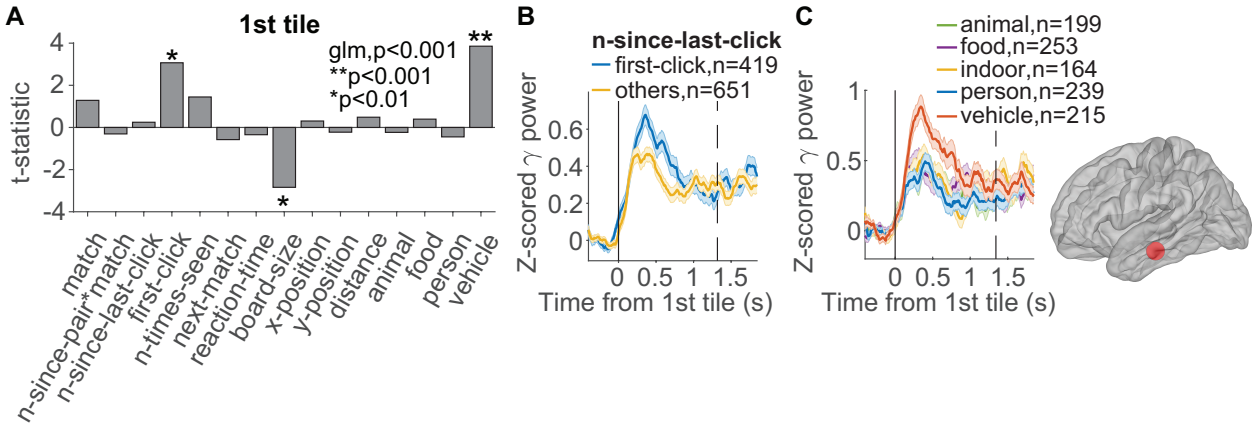

Figure 26. Example first-click electrode in the left fusiform gyrus that was also selective for transport vehicles. **A**. T-statistic of each predictor in the GLM for the first tile. Asterisks indicate significant predictors for the gamma power AUC. **B**. Z-scored gamma (30-150 Hz) power during first-clicks (blue) and non-first-clicks (yellow) aligned to the first tile onset (solid line). Dashed line indicates the mean reaction time. Shaded error bars indicate s.e.m. **C**. Z-scored gamma power during visualization of different categories of images. Electrode location is shown on the right.

Table 12. First-click or n-since-last-click electrodes that were also selective to image category. **: p<0.01 (GLM); ***: p<0.001 (GLM); X: most important predictor. The "indoor scene" category was dropped to follow the common practice of multivariate models to avoid falling into the "dummy variable trap" (**Methods**).

Location	First-click	N-since-last-click	Animal	Food	Person	Vehicle
amygdala	**				***X	
fusiform	**					***X
fusiform	**	**	***X			
fusiform	**	***X				**
inferiorparietal	***	***X			**	
inferiorparietal	***	***X			**	
lateraloccipital	***		**		**	
middletemporal	**				**	
parsopercularis	***	***X	**			
parsopercularis	***X	***			***	
parsopercularis	***X	**	**			**
parstriangularis	**		**X		**	
rostralmiddlefrontal	***X	***			**	
bankssts		**	**		***X	
bankssts		**	**	**	***X	
lateralorbitofrontal		**			***X	
middletemporal		**	***X			
parahippocampal		***				***X
parahippocampal		***				***X

Gamma responses after the first tile and before the second tile signal successful associative memory retrieval of the location of pair

After seeing the first tile, subjects would first attempt to locate the tile's pair. We

hypothesize that upon visualizing the first tile and before the second tile, there would be differential gamma responses forecasting whether the trial was going to be a match or mismatch, indicating successful or unsuccessful associative memory retrieval about the location of pair. Indeed, the match predictor was significant ($p<0.01$, GLM) for 32 electrodes (4.7%, **Figure 27G**, **Table 13**). In the majority of cases (91 %), activity was higher during match trials (successful retrieval) than during mismatch trials (unsuccessful retrieval). The lateral orbitofrontal cortex had significantly more electrodes than expected by chance ($p<0.01$, bootstrap analysis with 5,000 shuffles, **Methods**). Neural responses from an example electrode in the right LOF gyrus are shown in **Figure 27**. The only predictor that significantly contributed to explaining the neural responses was whether the trial was a match or not ($p<0.001$, GLM, **Figure 27A**). Indeed, when aligning the neural responses to the first click, there was a large difference in the neural signals in response to match versus mismatch trials (**Figure 27B**). This difference was also highly evident in individual trials (**Figure 27C** versus **26D**) and power spectrograms (**Figure 27E-F**). The peak in activity happened at approximately 500 ms after the first tile onset (**Figure 27B**).

Table 13. Locations of electrodes where match was a significant predictor for the gamma power AUC during the first tile determined by the GLM. "Number" indicates the number of electrodes in that region.

Location	Number
hippocampus	2
bankssts	1
entorhinal	1
fusiform	1
inferiortemporal	3
insula	4
lateralorbitofrontal	7
middletemporal	5
parahippocampal	1
parsopercularis	3
precentral	1
precuneus	1
rostralanteriorcingulate	1
supramarginal	1
Total	**32**

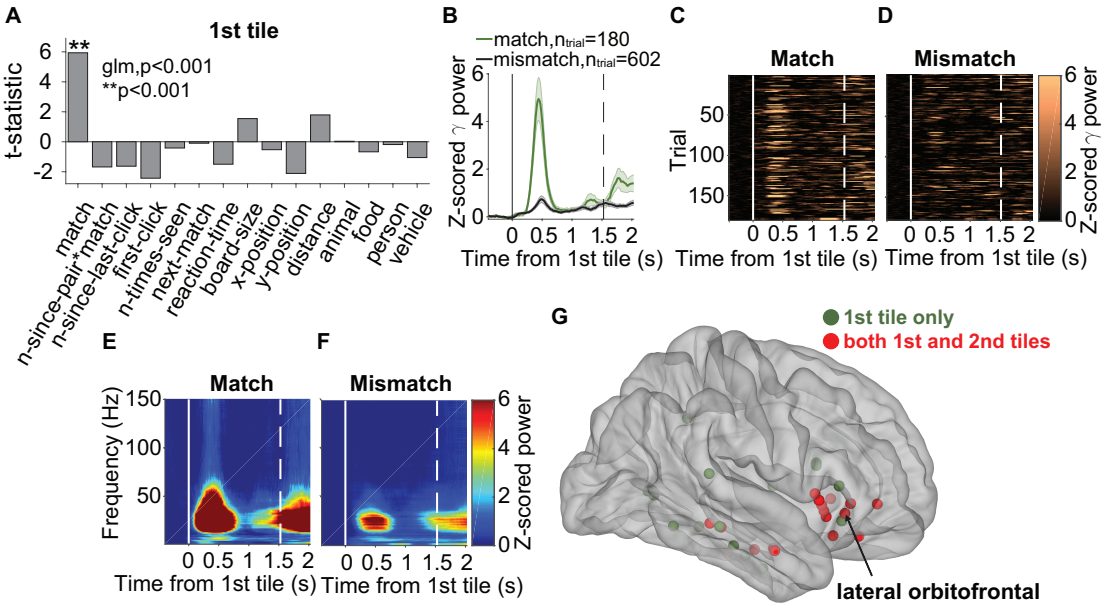

Figure 27. An example electrode in the right lateral orbitofrontal gyrus where "match" was a significant predictor for gamma activities during the 1st tile. A. T-statistic of each predictor in the GLM for the first tile. Asterisks indicate significant predictors for the gamma power AUC. **B**. Z-scored gamma (30-150 Hz) power during match (green) and mismatch (black) trials aligned to the first tile onset (solid line). Dashed line indicates the mean reaction time. Shaded error bars indicate s.e.m. **C-D**. Raster plots showing the z-scored gamma power in individual trials. For display purpose, trial number of match and mismatch was equalized (**Methods**). **E-F**. Spectrograms showing the band power during match and mismatch trials aligned to the first tile onset. **G**. Locations of all electrodes where "match" was a significant predictor during the first tile only (green) and during both tiles (red) plotted on one hemisphere.

Figure 28 shows another example match-electrode located in the left middle temporal gyrus where the match was a significant predictor for the gamma band activities after the 1st tile and before the 2nd tile. Similar to the LOF electrode in **Figure 27**, the gamma power during match trials was higher than during mismatch trials. However, the pattern of this modulation was different in several ways. First, the power increase was sustained rather than transient (compare **Figure 28B** versus **Figure 27B**). Second, the frequency that was modulated by match was higher than the LOF electrode in **Figure 27** (compare **Figure 28C-D** versus **Figure 27E-F**). The sustained change in gamma power was also evident in individual trials (**Figure 28E-F**). These results suggest that the middle temporal gyrus and the LOF cortex might be functionally distinct during associative memory retrieval although the neural responses in both regions were predictive of whether there was successful recall of the pair's location.

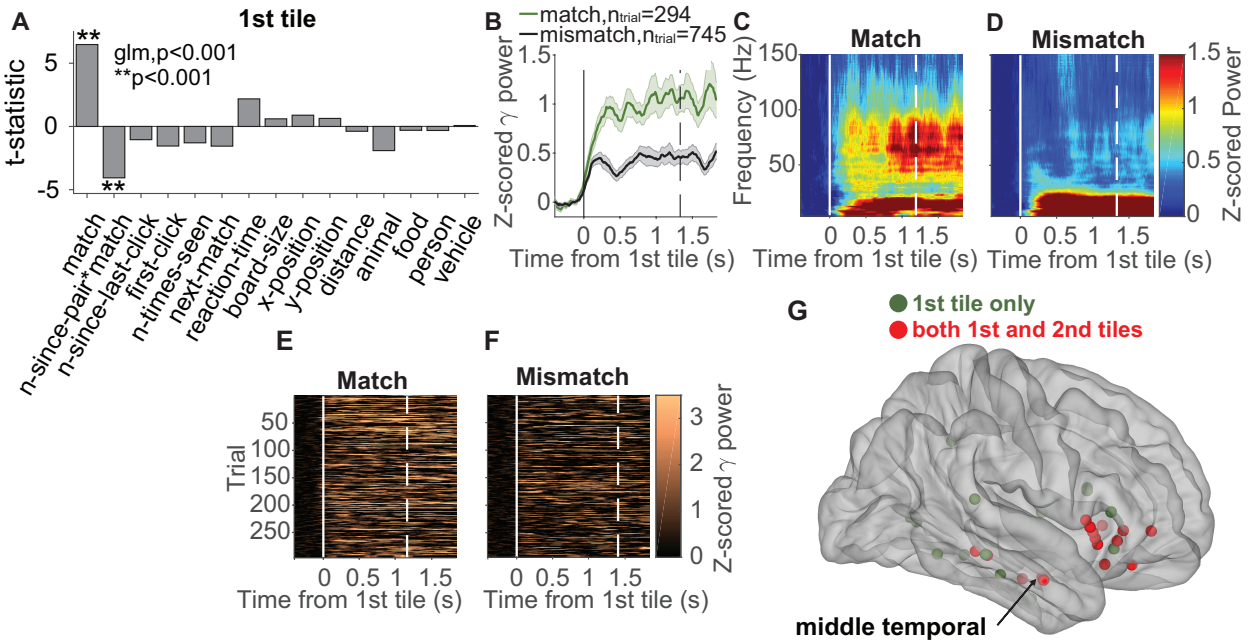

Figure 28. Example electrode located in the left middle temporal gyrus where "match" was a significant predictor for gamma activities during the 1st tile. A. T-statistic of each predictor in the GLM for the first tile. Asterisks indicate significant predictors for the gamma power AUC. **B**. Z-scored gamma (30-150 Hz) power during match (green) and mismatch (black) trials aligned to the first tile onset (solid line). Dashed line indicates the mean reaction time. **C-D**. Spectrograms showing the band power during match and mismatch trials aligned to the first tile onset. **E-F**. Raster plots showing the z-scored gamma power in individual match (**E**) and mismatch (**F**) trials. For display purpose, trial number of match and mismatch was equalized (**Methods**). **G**. Locations of all electrodes where "match" was a significant predictor during the first tile only (green) and during both tiles (red) plotted on one hemisphere.

Gamma responses capture the status of pair despite its absence during successful recall

We have demonstrated that differential neural responses could predict match or mismatch before trial outcome. Next, we asked whether gamma activities during this time partakes in capturing the status of pair in working memory. The n-since-pair predictor is related to the memory strength or familiarity with a tile's matching pair. The smaller the n-since-pair, the more recently the tile's pair had been seen, the stronger its memory representation. We considered only match trials for this predictor (n-since-pair*match) because there was no successful recall in mismatch. N-since-pair*match was a significant predictor ($p<0.01$, GLM) for the gamma power AUC during

the first tile in 15 (2.2%) electrodes (**Figure 29I**, **Table 14**). **Figure 29** shows an n-since-pair*match electrode in the left middle temporal gyrus. The gamma power progressively decreased as the currently seeing cue's matching pair became more distant in memory (**Figure 29B**) and thus exhibited a negative t-statistic (**Figure 29A**). We hypothesize that n-since-pair*match electrodes with negative t-statistics are more sensitive to recently viewed pairs and make them readily available for retrieval. We then asked whether there was any electrode showing the reverse phenomenon, that is, higher gamma activity toward a tile whose pair was seen a long time ago, which would have a positive t-statistic. Indeed, three electrodes (20%) had this property. We hypothesize that this type of electrodes are responsible for keeping distant items in working memory and make them available when needed.

Table 14. Locations of electrodes where n-since-pair was a significant predictor for the gamma power AUC during the first tile determined by the GLM. "Number" indicates the number of electrodes in that region.

Location	Number
insula	3
lateralorbitofrontal	3
medialorbitofrontal	1
middletemporal	3
parahippocampal	1
parsopercularis	1
precentral	1
precuneus	1
supramarginal	1
Total	**15**

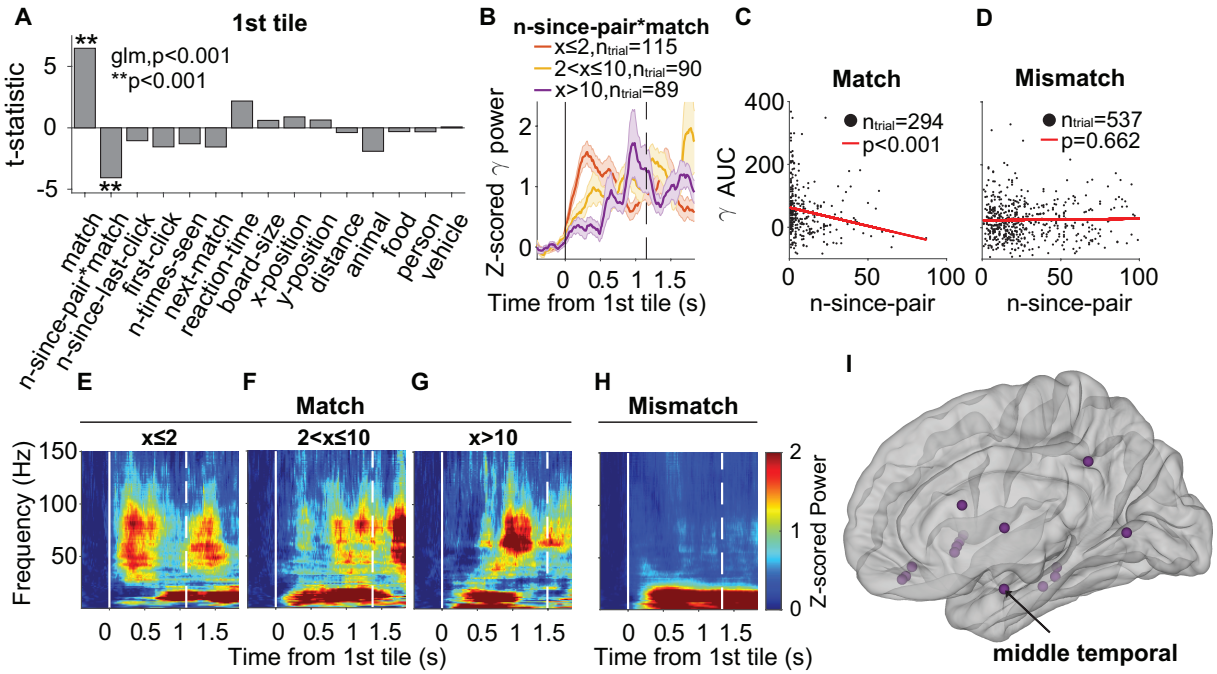

Figure 29. An example n-since-pair*match electrode located in the left middle temporal gyrus. A. T-statistic of each predictor in the GLM for the first tile. Asterisks indicate significant predictors for the gamma power AUC. **B**. Z-scored gamma power during match trials with different n-since-pair values. Signals were aligned to the first tile onset (solid line). Dashed line indicates the mean reaction time. Shaded error bars represent s.e.m. **C**-**D**. Scatter plots of gamma power AUC vs. n-since-pair for match (**C**) and mismatch (**D**). Each dot represents data from one trial. Red lines represent linear fits of the data. **E**-**G**. Spectrograms showing the band power aligned to the first tile onset for match trials with different n-since-pair ranges. **H**. Spectrogram showing the band power aligned to the first tile for mismatch trials. **I**. Locations of all n-since-pair*match electrodes plotted on one hemisphere.

To address the question of whether recollection and familiarity are dissociated during associative memory retrieval, we supplemented a linear regression analysis using n-since-pair as the independent variable and the gamma power AUC as the dependent variable for match and mismatch trials separately (**Figure 29C-D**). Nearly all n-since-pair*match electrodes could significantly predict the gamma power of match trials from n-since-pair values (see p values in **Table 15**). While none of these electrodes could predict gamma power AUC from n-since-pair using mismatch trials ($p>0.06$, linear regression) although where were much more mismatch than match trials. Therefore, the neural code of the pair's recency or grade of familiarity existed only during successful, but not unsuccessful, associative memory retrieval. It is also important to

emphasize that these neural responses were in the absence of the pair but only in the presence of the cue. Moreover, 11 of the n-since-pair*match electrodes (73.3%) also had "match" as a significant predictor (**Figure 29** is an example), meaning that they encoded both successful retrieval and the memory strength or familiarity with the tile's pair. In all, these results suggest that recollection and familiarity can be orchestrated by the same brain region and do not support the dual-process model in the context of associative working memory.

Table 15. Linear regression of gamma power AUC versus n-since-pair in match and mismatch trials for all n-since-pair*match electrodes. P values indicate significance of linear fits.

Region	p (match)	p (mismatch)
insula	0.005	0.978
insula	0.002	0.343
insula	0.001	0.856
lateralorbitofrontal	0.006	0.848
lateralorbitofrontal	0.003	0.178
lateralorbitofrontal	0.001	0.408
medialorbitofrontal	0.012	0.352
middletemporal	0.001	0.701
middletemporal	0.005	0.738
middletemporal	0.007	0.919
parahippocampal	0.070	0.068
parsopercularis	0.029	0.231
precentral	0.000	0.698
precuneus	0.010	0.484
supramarginal	0.006	0.331

Gamma responses after the second tile

Next we examined neural activities after the second tile, when the trial result (match or mismatch) became clearly apparent to the subjects. We built a separate GLM using the same predictors except for n-since-pair*match because it was less relevant for mental operations after the second tile by definition. Similar to the first tile, we found electrodes where the gamma power could be dictated by first-click and/or n-since-last-click. First-click was a significant ($p<0.01$, GLM) predictor in 24 electrodes (**Figure 30H**, **Table 16**) and n-since-last-click a significant

predictor in 9 electrodes (**Figure 30G**, **Table 17**). First-click electrodes were distributed along the ventral visual stream and the prefrontal cortex (**Figure 30H**) while n-since-last-click mostly in the frontal lobe (**Figure 30G**). Most of the n-since-last-click electrodes were also selective for novelty (**Figure 30G**, red spheres). Similar to the first tile, first-click and n-since-last-click may represent the degree of novelty or familiarity with the second tile. The drastic reduction in the number of both types of electrodes compared to the 1st tile may be due to that there were two images (same if match, different if mismatch) on the screen at the same time, which decreased the "neural resolution" of familiarity. The lateral orbitofrontal cortex contained 4 NSLC electrodes for the second tile, which were significantly more than chance ($p<0.01$, bootstrap analysis, 5,000 shuffles). An example electrode in the LOF cortex is shown in **Figure 30**. This electrode exhibited the highest gamma responses toward novel and also less familiar tiles (**Figure 30B**, blue and yellow lines) but lower responses to familiar ones (red line). Such distinction can also be clearly observed in single trials (**Figure 30C**) and power spectrograms (**Figure 30D-F**).

Table 16. Locations of electrodes where first-click was a significant predictor for the gamma power AUC during the second tile determined by the GLM. "Number" indicates the number of electrodes in that region.

Location	Number
amygdala	1
hippocampus	1
bankssts	1
fusiform	2
insula	2
lateraloccipital	2
lateralorbitofrontal	4
lingual	1
parahippocampal	1
parsopercularis	2
postcentral	2
precentral	2
precuneus	1
rostralmiddlefrontal	1
superiortemporal	1
Total	**24**

Table 17. Locations of electrodes where n-since-last-click was a significant predictor for the gamma power AUC during the second tile determined by the GLM. "Number" indicates the number of electrodes in that region.

Location	Number
fusiform	1
lateralorbitofrontal	4
parsopercularis	1
postcentral	1
precentral	1
rostralmiddlefrontal	1
Total	**9**

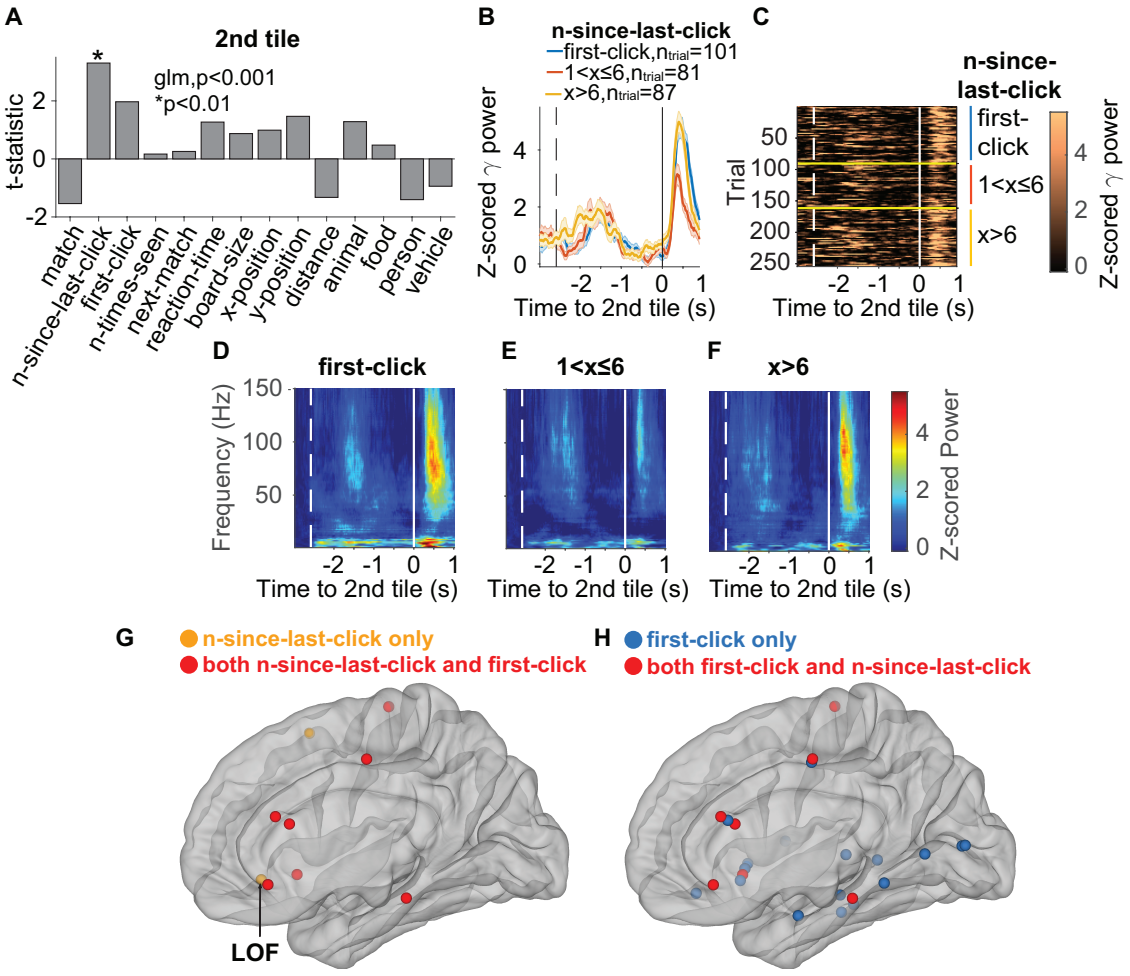

Figure 30. Example n-since-last-click electrode during the 2nd tile. A. T-statistic of each predictor in the GLM for the 2nd tile. Asterisks indicate significant predictors for the gamma power AUC. **B**. Z-scored gamma power of first-click and different n-since-last-click values aligned to the 2nd tile onset (solid line). Dashed line indicates the mean onset of the 1st tile. Shaded error bars indicate s.e.m. **C**. Raster plot showing z-scored gamma power in individual trials ordered by first-click and then smaller to larger n-since-last-click. Yellow horizontal lines and colored vertical lines indicate n-since-last-click ranges same as in (**B**). **D-F**. Trial-averaged spectrograms showing the band power aligned to the 2nd tile onset. **G**. Locations of all n-since-last-click electrodes during the 2nd tile. Orange: n-since-last-click only; red: both n-since-last-click and first-click. **H**. Locations of all first-click electrodes during the 2nd tile. Blue: first-click only; red: both first-click and n-since-last-click. All electrodes were plotted on one hemisphere for display purpose.

Next we examined whether neural activities after the 2nd tile could be dictated by match or mismatch. For gamma power AUC during 1 second after the 2nd tile, the match predictor was statistically significant ($p<0.01$, GLM) for 112 electrodes (**Figure 31G**, **Table 18**). There were 17 electrodes where the match predictor was significant during both the 1st and 2nd tiles (**Figure 31G**, red spheres). The lateral orbitofrontal cortex and the insula had significantly more electrodes than chance ($p<0.01$, bootstrap analysis with 5,000 shuffles, **Methods**). **Figure 31** shows an example electrode located in the right insula where the match was a significant predictor for the gamma band activities after the 2nd tile ($p<0.001$, GLM, **Figure 31A**). The neural signals during match trials were stronger than during mismatch trials (**Figure 31B**, **E-F**). Such difference also reliably manifested in individual trials (**Figure 31C-D**). For this electrode, there was no differential responses between match and mismatch before the 2nd tile ($p=0.2$, GLM for the 1st tile). Thus these activities were attributed to the visualization of the 2nd tile but not the 1st tile, suggesting that increased gamma activities might be exclusive of the positive feedback (match) after the 2nd tile was revealed. Same as this insular electrode, the majority (78.2%) of match-electrodes after the 2nd tile could not predict gamma responses before the 2nd tile from match or mismatch. These electrodes were arguably not involved in associative memory retrieval but in feedback processes.

Table 18. Locations of electrodes where match was a significant predictor for the gamma power AUC during the 2nd tile determined by the GLM. "Number" indicates the number of electrodes in that region.

Location	Number
amygdala	2
hippocampus	6
bankssts	3
caudalmiddlefrontal	2
entorhinal	1
inferiorparietal	4
inferiortemporal	5
insula	12
lateraloccipital	1
lateralorbitofrontal	21
lingual	1
medialorbitofrontal	3
middletemporal	14
parahippocampal	2
parsopercularis	6
parstriangularis	8
postcentral	2
posteriorcingulate	1
precentral	7
rostralanteriorcingulate	1
rostralmiddlefrontal	1
superiorfrontal	2
superiorparietal	3
supramarginal	3
Total	**112**

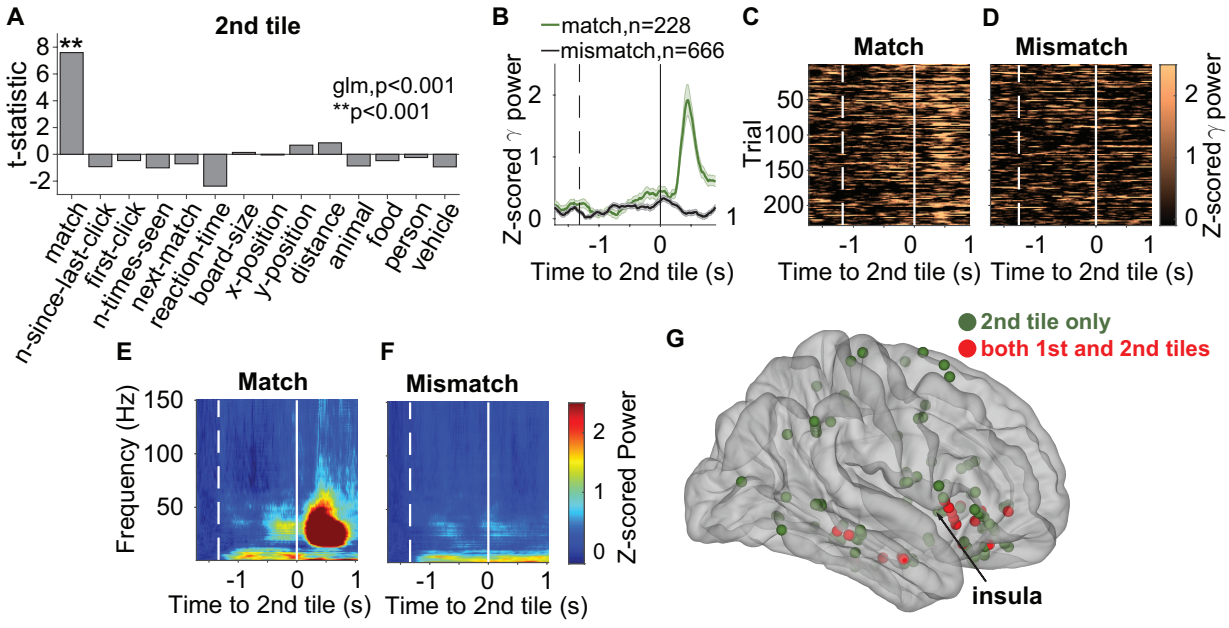

Figure 31. Example electrode in the right insula where match was a significant predictor for gamma responses during the 2nd tile. A. T-statistic of each predictor in the GLM for the 2nd tile. Asterisks indicate significant predictors for the gamma power AUC. **B**. Z-scored gamma power during match and mismatch trials aligned to the 2nd tile onset (solid line). Dashed line indicates the mean onset of the 1st tile. **C**-**D**. Raster plots showing the z-scored gamma power in individual trials. **E**-**F**. Spectrograms showing the band power during match and mismatch. **G**. Locations of all electrodes where "match" was a significant predictor during the 2nd tile only (green) and during both tiles (red). All electrodes were plotted on one hemisphere for display purpose.

Discussion

We studied the neural mechanisms of working memory by recording intracranial field potentials from 676 bipolarly-referenced electrodes in the gray matter (**Figure 22, Table 8**) from 20 pharmacologically-intractable epilepsy patients implanted with sEEG electrodes (one patient also had ECoG electrodes). During intracranial EEG monitoring, patients played the memory matching game (**Figure 20, Movie S1**), for which they were instructed to find pairs of images from image matrices of different sizes. We leveraged generalized linear models to assess the relative contribution from different behavioral predictors to the gamma activities.

Subjects performed significantly better than a memoryless model and worse than a perfect

memory model (**Figure 21A**). We observed distinct behaviors of match and mismatch. The reaction time for mismatch was significantly longer than match (**Figure 21B**), which may suggest a distinction in confidence level toward match versus mismatch. We used the variable n-since-last-click to estimate the degree of familiarity with a tile that was being viewed and n-since-pair the degree of familiarity with a currently seeing tile's matching pair. In addition, n-since-last-click also reflected the relative working memory load from last time seeing the same tile until the current visualization, and n-since-pair reflected the relative working memory load from the most recent view of the tile's pair until the current viewing of the tile. Since memory load was higher in more difficulty blocks that contained more tiles, both n-since-last-click and n-since-pair (NSP) increased as board size incremented (**Figure 21C-F**). The level of familiarity with the to-be-clicked second tile in a trial, either indexed by the recency since last seeing itself (NSLC), or by the recency since last seeing its pair (NSP), more strongly influenced the trial result (match or mismatch) than that with the first tile (**Figure 21C-E**).

We used generalized linear models to assess how the gamma band power could be predicted by behavioral parameters that have important implications for memory processes (**Table 9**). First of all, we asked whether we could differentiate from the neural responses if the currently viewing tile was novel or familiar. This process was necessary to: (1) support the initial encoding and maintenance of the tile's content and location in working memory if it was novel (Knight, 1996); (2) rapidly recognize it, if seen before, in order to proceed with retrieving the location of its associated pair. We found electrodes that were selective to novel information and those that could record recency or encode the degree of familiarity. Several electrodes carried both properties and for these electrodes, novel and very unfamiliar items triggered similar neural responses (**Figure 24I** and **Figure 30B**). Responses to novelty or familiarity can be bimodal, that is, they

increase to familiar information and decrease to novel information (**Figure 25**, pars opercularis), or vice versa. The lateral orbitofrontal cortex and the pars opercularis were significantly rich in novelty-selective electrodes (bootstrap analysis, **Methods**). Though didn't reach statistical significance, the medial temporal lobe (MTL) also contained a good number of electrodes that were selective for novelty, gradient of familiarity, or both (**Figure 24G**, **N**). These results are consistent with previous works that reported the role of the prefrontal cortex (Duzel et al., 2004; Friedman et al., 2001; Kishiyama et al., 2009; Knight, 1984; Montaldi et al., 2006; Tulving et al., 1996) and the medial temporal lobe (Knight, 1996; Montaldi et al., 2006; Rutishauser et al., 2006; Rutishauser et al., 2021; Rutishauser et al., 2008; Tulving et al., 1996; Viskontas et al., 2006; Xiang and Brown, 1998) in processing novelty and familiarity. Neural responses that were modulated by novelty of familiarity could also be selective to image category (**Figure 26**). Despite some extent of category selectivity, the fact that familiarity gradient can be observed at the level of local field potential suggests that the representation of familiarity gradient may not be item-specific, that is, different items or images may elicit the same neural response as long as their degree of familiarity is the same. This speculation is consistent with the notion that single memory cells are rarely sharply tuned to particular sensory features (Funahashi et al., 1989; Fuster and Jervey, 1982; Rutishauser et al., 2021; Zhou et al., 2007). Different from the traditional and widely used yes or no long-term memory paradigm, we administered to subjects a working memory game during which a same item could constantly and rapidly change its memory status. We demonstrated that neural activities in the gamma band can be quickly updated to capture highly dynamic working memory processes.

We demonstrated that differential gamma responses were predictive of successful associative memory retrieval of the pair's location. Such difference can be observed before the

trial outcome or in the absence of the pair. These differential responses can be transient (**Figure 27B**) or sustained (**Figure 28B**), and may vary in time and frequency (compare **Figure 27C-D** with **Figure 28C-D**), which are suggestive of plausibly different operations during associative working memory retrieval. For example, transient activities may signal fast events like suddenly knowing the trial outcome (match or mismatch) and sustained responses may correspond to enduring processes such as maintenance and retrieval. It is possible that sustained activities can be due to averaging across temporally distinct neural activities (Lundqvist et al., 2016); however, it has been reported that single neurons in the hippocampus show sustained firing rate increase during associative retrieval (Staresina et al., 2019). Bootstrap analysis revealed that the lateral orbitofrontal cortex is a critical region for associative working memory retrieval, providing an important update to the existing literature that mostly emphasizes the role of the hippocampus and its surrounding MTL structures in long-term associative memory (Bergmann et al., 2012a; Duncan et al., 2009; Mayes et al., 2007; Ranganath et al., 2004; Rutishauser et al., 2021; Rutishauser et al., 2008; Sakai and Miyashita, 1991; Staresina et al., 2019). It is also worth mentioning that the mechanisms of associative retrieval might be fundamentally different between long-term and working memory (Bergmann et al., 2012a, b), an important domain to be explored by future studies.

Neural responses that were predictive of successful associative retrieval may also encode the degree of familiarity with the currently viewing cue's pair despite the absence of the pair. We used n-since-pair to index the degree of familiarity with a cue's pair when only the cue was shown on the screen. N-since-pair*match electrodes signaled both successful associative memory retrieval and the grade of familiarity because modulation of the gamma activities by n-since-pair was only observed in match but not mismatch (**Table 15**, **Figure 29C-D**). Moreover, 73% of n-

since-pair*match electrodes also had "match" as a significant predictor. Therefore, these results are in contradictory with the dual-process model but support the idea that recollection and familiarity can be operated by the same brain structure, consistent with results from several studies that addressed this topic using long-term memory paradigms (Merkow et al., 2015; Rutishauser et al., 2008; Wais et al., 2010; Wais et al., 2006; Wixted and Squire, 2011).

Neural activities after the second tile can also be modulated by different behavioral parameters. Similar to the first tile, neural responses after the 2nd tile were modulated by novelty (first-click) or familiarity (n-since-last-click) with respect to the 2nd tile. The number of such electrodes was much fewer than 1st tile, likely due to the simultaneous presence of two images on the screen. Gamma power after the 2nd tile could be predicted by match or mismatch in 112 widely distributed electrodes. Most of these electrodes did not show differential gamma responses before the 2nd tile, suggesting that their roles might be less related to associative memory processes but more to feedbacks, like signaling reward or failure. Neural activities after the 2nd tile may also be involved in strategic decision making, for example, erasing already-matched tiles from the working memory cache to offer space for new information. Such execution is important for proper working memory functioning because working memory is a system with limited capacity. Regions to be highlighted here are the lateral orbitofrontal cortex and the insula (bootstrap analysis, **Methods**).

Taken together, we took advantage of the high spatiotemporal resolution of human intracranial EEG to study the mechanisms underlying working memory. We administered a memory matching game where each item rapidly changed its status in working memory, allowing for examining how neural activities can track highly dynamic working memory processes. We also investigated both non-associative and associative aspects of working memory by carefully

dissociating neural responses toward the cue and those toward the cue's pair. Gamma activities were selective to novelty and represent different degrees of familiarity. Differential gamma responses preceded successful associative retrieval of the pair's location. The lateral orbitofrontal cortex and the pars opercularis, among several others, were highly involved in working memory operations. In sum, neural activities in the gamma band capture a wide array of working memory status and orchestrate both non-associative and associative working memory processes.

Materials and Methods

Task paradigm

Participants performed our implementation of the classical memory matching game (**Figure 20, Movie S1**). The game involved remembering the location and content of a set of tiles to find all the matching pairs. Each block contained a square board of n×n tiles. In the beginning, all tiles were shown in black. In each trial, participants chose one tile, then a second tile, by clicking on them in a self-paced fashion. At the end of each trial, either the two tiles revealed the same image (match) or not (mismatch). If the tiles matched, then those tiles turned green 1,000 ms after the second click and they could not be clicked again for the remainder of the block. If the tiles did not match, then they turned black (revealable) again 1,000 ms after the second click. When all tiles had turned green, i.e., all matches had been found, the block ended and another block began. During each block, the map between positions and objects was fixed. The game always started with a block of size 3×3 and progressed to more difficult blocks (4×4, 5×5, 6×6, and finally 7×7). Blocks that held an odd number of tiles (3×3, 5×5, and 7×7) contained one distractor image (a human face) which had no corresponding pair. For each block except the 3×3 board, there was a

limit for the total time elapsed (2 minutes for 4×4, 3.3 min for 5×5, 4.8 min for 6×6, 8.2 min for 7×7). If a participant did not complete a block within the time limit, that block ended and a new, easier, block started. This was achieved by reducing n by 1, except when n=7, where it was reduced by 2. Conversely, when participants successfully completed a block with a board of size n within the time limit, they moved on to a more difficult block by increasing n by 1. When participants completed a 7×7 block, they would do further n=7 blocks. There was no image repetition across the entire session.

All the object images were from the Microsoft COCO 2017 validation dataset (Lin, 2015) and were rendered in grayscale and square shape. We included a balanced numbers of images from 5 categories: person, animal, food, vehicle, and indoor scenes. All the images were rendered on a 13-inch Apple MacBook Pro laptop. The size of each tile was 0.75×0.75 inches (2×2 degrees of visual angle, dva) and the separation between two adjacent tiles was 0.125 inch (0.33 dva) for board size n=7 and 0.25 inch (0.67 dva) for the others. The game implementation was written and presented using the Psychtoolbox extension (Brainard, 1997; Pelli, 1997) in Matlab_2016b (Mathworks, Natick, MA).

Epilepsy subjects and recording procedures

We recorded intracranial field potentials from 20 patients with pharmacologically intractable epilepsy (12-52 years old, 9 female, **Table S1**) undergoing monitoring at Boston Children's Hospital (Boston, US), Brigham and Women's Hospital (Boston, US), and Xuanwu Hospital (Beijing, China). All recording sessions were seizure-free. All patients had normal or corrected-to-normal vision. The study protocol was approved by each hospital's institutional review board. Experiments were run under patients' or their legal guardians' informed consent.

One patient at Brigham and Women's Hospital (BWH) was implanted with both stereo encephalography (sEEG) and electrocorticography (ECoG) electrodes while all other patients had only sEEG electrodes (Ad-tech, USA; ALCIS, France). Intracranial field potentials were recorded with Natus (Pleasanton, CA) and Micromed (Italy). The sampling rate was 2048 Hz at Boston Children's Hospital, 512 Hz or 1024 Hz at BWH, and 512 Hz at Xuanwu Hospital. Electrode trajectories were determined based on clinical purposes for precisely localizing suspected epileptogenic foci and surgically treating epilepsy (Fried et al., 2014).

Behavioral analyses

We created two computational models to simulate behavior assuming perfect memory and void memory (chance performance, **Figure 21A**). The perfect memory model remembered all revealed tiles without forgetting. The random model simulated random clicking. We calculated the reaction time (RT, time between two clicks in a trial), n-since-last-click (the number of clicks since the same tile was clicked), and n-since-pair (number of clicks since the last time when the matching pair was seen). We compared these variables for match and mismatch trials at each board size (**Figure 21B-F**, permutation test, 10,000 iterations, α=0.01). For n-since-last-click, we removed trials in which any tile was clicked the first time. For n-since-pair, we removed trials in which any tile's matching pair hadn't been seen. We defined random match as a match trial in which the second tile had never been seen before; such trials were excluded from both the behavioral and neurophysiological analyses. We used the F-test for linear regression models to assess whether RT, n-since-last-click, and n-since-pair significantly covaried with board size; this was done separately for match and mismatch, and for 1st and 2nd tiles.

Electrode localization

Electrodes were localized using the iELVis (Groppe et al., 2017) toolbox. We used Freesurfer (Dale et al., 1999) to segment the preimplant magnetic resonance (MR) images, upon which post-implant CT was then rigidly registered to. Electrodes were marked in the CT aligned to preimplant MRI using Bioimage Suite (Joshi et al., 2011). Each electrode was assigned to an anatomical location using the Desikan-Killiany (Desikan et al., 2006) atlas for subdural grids and strips or FreeSurfer's volumetric brain segmentation for depth electrodes. Out of 1,750 electrodes in total, we included 676 bipolarly referenced electrodes in the gray matter (**Table 8**). Discarded electrodes were due to bipolar referencing (one removed from each shank), locations in white matter or pathological sites, or containing large artifacts. Electrode locations were mapped onto the MNI305 average brain via affine transformation (Wu et al., 2018) for display purposes.

Preprocessing of intracranial field potential data

Bipolar subtraction was applied to each pair of neighboring electrodes on each shank of depth electrodes or surface grid/strip electrodes (Wang et al., 2021). A zero-phase digital notch filter (Matlab function "filtfilt") was applied to the bipolarly subtracted broadband signals to remove the line frequency at 60 Hz (BCH, BWH) or 50 Hz (Xuanwu) and their harmonics. For each electrode, trials whose amplitudes (voltage range) were larger than 5 standard deviations from the mean amplitude across all trials were considered containing potential artifacts and discarded from further analyses (Bansal et al., 2012). For the first tile, the time window for artifact rejection was from 400 ms before click until 1 second after the average RT. For the second tile, the time window was [400 ms + average reaction time] before the second click until 1 second after the second click. Across all electrodes, we rejected 1.75% of all trials for the 1st tile and 1.73% for the

2nd tile.

Time-frequency decomposition

The gamma band (30-150 Hz) power was computed using the Chronux toolbox (Mitra and Bokil, 2008). We used a time-bandwidth product of 5 and 7 leading tapers, a moving window size of 200 ms, and a step size of 10 ms (Xiao et al., 2022). For each trial, the power was normalized by subtracting the mean gamma band power during the baseline (400 ms before 1st tile) and dividing by the standard deviation of the gamma power during baseline. Spectrograms were rendered with the Fieldtrip toolbox (Oostenveld et al., 2011). For all the subjects, there were more mismatch than match trials. In the raster plots, we subsampled the mismatch trials, keeping those trials whose reaction times were closest to the mean reaction time of match trials. All random matches were excluded from analyses.

Generalized linear model

We used generalized linear models (GLM) (Friston et al., 1995; Worsley and Friston, 1995) to analyze the relationship between the gamma band power and behavioral parameters. We used two different GLMs, one using neural responses between the 1st and 2nd tile, and the other using neural responses after the 2nd tile. For the first GLM, the time window started when the first tile was clicked and ended at a time corresponding to the 90th-percentile of the distribution of reaction times. This criterion showed to be a reasonable tradeoff between minimum overlap with responses after the 2nd tile and maximum amount of information captured. For the second GLM, we used 1 second after the 2nd tile as the window of analysis. The response variable to be fit by the GLM was the area under the curve (AUC) of the gamma band power during the defined time windows. **Table**

9 describes the behavioral parameters that were considered as predictors in the GLM. From the five image categories, we dropped the "indoor" category to avoid falling into the "dummy variable trap", which is a common practice of using multivariate models.

We performed multicollinearity analysis to discard the presence of highly correlated predictors that could impair the performance of the model (Dormann et al., 2013; Welsch and Kuh, 1977). We calculated the variance inflation factor (VIF) for each predictor to detect the presence of multicollinearities. A VIF of 1 indicated that there was no correlation with other predictors. The larger the VIF, the higher the correlation, and a VIF greater than 5 indicated very high correlation that could significantly harm the performance of the model. For all subjects in our analysis, the VIFs of all predictors were smaller than 3 (**Figure 23**).

For each predictor, we calculated the parameter estimate (beta coefficient) from the least mean squares fit of the model to the data, the t-statistic (beta coefficient divided by its standard error), and the p-value to test the effect of each predictor on the gamma power AUC. A beta coefficient or t-statistic of zero indicates that the predictor has no effect at all on the response. A predictor was considered significant if the GLM model differed from a constant model ($p<0.01$) and if the p-value for that predictor was smaller than 0.01.

Bootstrap analysis

To determine if any brain region contained significantly more electrodes than expected by chance considering any GLM predictor, we randomly sampled the same number (n) of electrodes as those where that predictor was significant, from the total electrode population for 10,000 iterations. Taking the "match" predictor as an example, from the total 676 locations, we randomly sampled 32 locations (the number of match-significant electrodes) for 10,000 times, and calculated

the p value of any location, for example the hippocampus, as the number of times when n(hippocampus, real) was smaller than n(hippocampus, sampled). If $p<0.01$, we considered that brain region as containing significantly more electrodes than purely expected by chance.

Conclusion

In this work, we used intracranial EEG to investigate the neural mechanisms of human cognitive control and working memory, which are important cognitive functions that are critical to everyday life. We observed distinct neural activities in response to the presence versus absence of conflicting information. Three cognitive control tasks were administered to epilepsy patients implanted with depth electrodes, in order to investigate the potentially common and distinctive aspects of conflict resolution processes. The tasks were designed to implicate different processing systems, response modalities, and task demands to evaluate the domain-general notion of cognitive control processes. We demonstrated that conflict is represented abstractly within each task, irrespective of different sensory combinations. On the contrary, conflict is largely represented differentially in different tasks. Robust neural conflict effect in one task can be completely absent in another task. Taken together, our results suggest a within-task invariant, but cross-task specific mechanism of conflict resolution during cognitive control.

Also using intracranial EEG with human epilepsy patients, we investigated core neural mechanisms underlying working memory. Subjects performed a memory matching game which allowed us to examine both associative and non-associative working memory processes. Neural responses in the gamma band (30-150 Hz) are able to distinguish novel from familiar information, and "grade" familiarity. Such capacity of recording the degree of familiarity not only applies to image in sight but also to associated information needed to be recollected, which is not yet visible. This property of neural responses emerges only during successful, but not unsuccessful, associative retrieval. Additionally, we identified gamma activities that forecast successful associative working memory retrieval.

In summary, this study uncovered key neural mechanisms underlying human cognitive control and working memory, providing important updates to the current theories and guidance for developing treatments for cognitive control and memory dysfunctions.

Appendix

Movie S1. Task paradigm. See online video.

Table S1. Subject information.

Subject no.	Hospital	Gender	Age
1	Xuanwu	female	22
2	Xuanwu	male	32
3	Xuanwu	male	19
4	Xuanwu	male	33
5	Xuanwu	male	35
6	Xuanwu	male	21
7	Xuanwu	male	26
8	Xuanwu	male	23
9	Xuanwu	female	47
10	Xuanwu	female	21
11	Xuanwu	female	26
12	BWH	female	32
13	BWH	female	31
14	BWH	female	26
15	BWH	male	22
16	BWH	female	52
17	BWH	male	44
18	BCH	male	18
19	BCH	female	12
20	BCH	male	15

References

Aggleton, J.P., and Brown, M.W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. Behav Brain Sci *22*, 425-444; discussion 444-489.

Assef, E.C., Capovilla, A.G., and Capovilla, F.C. (2007). Computerized stroop test to assess selective attention in children with attention deficit hyperactivity disorder. Span J Psychol *10*, 33-40.

Baddeley, A. (1992). Working memory. Science *255*, 556-559.

Baddeley, A.D. (1986). Working memory (Oxford
New York: Clarendon Press ;
Oxford University Press).

Banich, M.T. (2019). The Stroop Effect Occurs at Multiple Points Along a Cascade of Control: Evidence From Cognitive Neuroscience Approaches. Front Psychol *10*, 2164.

Bansal, A.K., Singer, J.M., Anderson, W.S., Golby, A., Madsen, J.R., and Kreiman, G. (2012). Temporal stability of visually selective responses in intracranial field potentials recorded from human occipital and temporal lobes. J Neurophysiol *108*, 3073-3086.

Barbey, A.K., Koenigs, M., and Grafman, J. (2011). Orbitofrontal contributions to human working memory. Cereb Cortex *21*, 789-795.

Barch, D.M., Braver, T.S., Akbudak, E., Conturo, T., Ollinger, J., and Snyder, A. (2001). Anterior cingulate cortex and response conflict: effects of response modality and processing domain. Cereb Cortex *11*, 837-848.

Baudena, P., Halgren, E., Heit, G., and Clarke, J.M. (1995). Intracerebral potentials to rare target and distractor auditory and visual stimuli. III. Frontal cortex. Electroencephalogr Clin Neurophysiol *94*, 251-264.

Bawden, D., and Robinson, L. (2009). Az informacio arnyoldalai: az informacios tulterheles, az informacio okozta szorongas es mas ellentmondasok, patologias jelensegek. =. Library Review / Konyvtari Figyelo *19*, 485-489.

Bergmann, H.C., Rijpkema, M., Fernandez, G., and Kessels, R.P. (2012a). Distinct neural correlates of associative working memory and long-term memory encoding in the medial temporal lobe. Neuroimage *63*, 989-997.

Bergmann, H.C., Rijpkema, M., Fernandez, G., and Kessels, R.P. (2012b). The effects of valence and arousal on associative working memory and long-term memory. PLoS One *7*, e52616.

Blackman, R.K., Crowe, D.A., DeNicola, A.L., Sakellaridi, S., MacDonald, A.W., 3rd, and Chafee, M.V. (2016). Monkey Prefrontal Neurons Reflect Logical Operations for Cognitive Control in a Variant of the AX Continuous Performance Task (AX-CPT). J Neurosci *36*, 4067-4079.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., and Cohen, J.D. (2001). Conflict monitoring and cognitive control. Psychol Rev *108*, 624-652.

Botvinick, M.M., Cohen, J.D., and Carter, C.S. (2004). Conflict monitoring and anterior cingulate cortex: an update. Trends Cogn Sci *8*, 539-546.

Bowles, B., Crupi, C., Mirsattari, S.M., Pigott, S.E., Parrent, A.G., Pruessner, J.C., Yonelinas, A.P., and Kohler, S. (2007). Impaired familiarity with preserved recollection after anterior temporal-lobe resection that spares the hippocampus. Proc Natl Acad Sci U S A *104*, 16382-16387.

Brainard, D.H. (1997). The Psychophysics Toolbox. Spat Vis *10*, 433-436.

Bunge, S.A., Burrows, B., and Wagner, A.D. (2004). Prefrontal and hippocampal contributions to visual associative recognition: interactions between cognitive control and episodic retrieval. Brain Cogn *56*, 141-152.

Bunge, S.A., Hazeltine, E., Scanlon, M.D., Rosen, A.C., and Gabrieli, J.D. (2002). Dissociable contributions of prefrontal and parietal cortices to response selection. Neuroimage *17*, 1562-1571.

Bush, G., and Shin, L.M. (2006). The Multi-Source Interference Task: an fMRI task that reliably activates the cingulo-frontal-parietal cognitive/attention network. Nature protocols *1*, 308-313.

Buzsaki, G., Anastassiou, C.A., and Koch, C. (2012). The origin of extracellular fields and currents--EEG, ECoG, LFP and spikes. Nat Rev Neurosci *13*, 407-420.

Cardin, J.A., Carlen, M., Meletis, K., Knoblich, U., Zhang, F., Deisseroth, K., Tsai, L.H., and Moore, C.I. (2009). Driving fast-spiking cells induces gamma rhythm and controls sensory responses. Nature *459*, 663-667.

Carter, C.S., Braver, T.S., Barch, D.M., Botvinick, M.M., Noll, D., and Cohen, J.D. (1998). Anterior cingulate cortex, error detection, and the online monitoring of performance. Science *280*, 747-749.

Carter, C.S., and van Veen, V. (2007). Anterior cingulate cortex and conflict detection: an update of theory and data. Cogn Affect Behav Neurosci *7*, 367-379.

Caruana, F., Uithol, S., Cantalupo, G., Sartori, I., Lo Russo, G., and Avanzini, P. (2014). How action selection can be embodied: intracranial gamma band recording shows response competition during the Eriksen flankers test. Front Hum Neurosci *8*, 668.

Cavanagh, J.F., and Frank, M.J. (2014). Frontal theta as a mechanism for cognitive control. Trends Cogn Sci *18*, 414-421.

Cave, C.B., and Squire, L.R. (1992). Intact verbal and nonverbal short-term memory following damage to the human hippocampus. Hippocampus *2*, 151-163.

Cole, M.W., Yeung, N., Freiwald, W.A., and Botvinick, M. (2009). Cingulate cortex: diverging data from humans and monkeys. Trends Neurosci *32*, 566-574.

Coulthard, E.J., Nachev, P., and Husain, M. (2008). Control over conflict during movement preparation: role of posterior parietal cortex. Neuron *58*, 144-157.

Courchesne, E., Hillyard, S.A., and Galambos, R. (1975). Stimulus novelty, task relevance and the visual evoked potential in man. Electroencephalogr Clin Neurophysiol *39*, 131-143.

Courtney, S.M., Petit, L., Maisog, J.M., Ungerleider, L.G., and Haxby, J.V. (1998). An area specialized for spatial working memory in human frontal cortex. Science *279*, 1347-1351.

Crane, J., and Milner, B. (2005). What went where? Impaired object-location learning in patients with right hippocampal lesions. Hippocampus *15*, 216-231.

Crone, N.E., Korzeniewska, A., and Franaszczuk, P.J. (2011). Cortical gamma responses: searching high and low. Int J Psychophysiol *79*, 9-15.

Crone, N.E., Miglioretti, D.L., Gordon, B., and Lesser, R.P. (1998). Functional mapping of human sensorimotor cortex with electrocorticographic spectral analysis. II. Event-related synchronization in the gamma band. Brain : a journal of neurology *121 (Pt 12)*, 2301-2315.

Dale, A.M., Fischl, B., and Sereno, M.I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. Neuroimage *9*, 179-194.

Danker, J.F., Hwang, G.M., Gauthier, L., Geller, A., Kahana, M.J., and Sekuler, R. (2008). Characterizing the ERP Old-New effect in a short-term memory task. Psychophysiology *45*, 784-793.

Daselaar, S.M., Fleck, M.S., and Cabeza, R. (2006). Triple dissociation in the medial temporal lobes: recollection, familiarity, and novelty. J Neurophysiol *96*, 1902-1911.

Davelaar, E.J., and Stevens, J. (2009). Sequential dependencies in the Eriksen flanker task: a direct comparison of two competing accounts. Psychon Bull Rev *16*, 121-126.

de Chastelaine, M., Mattson, J.T., Wang, T.H., Donley, B.E., and Rugg, M.D. (2017). Independent contributions of fMRI familiarity and novelty effects to recognition memory and their stability across the adult lifespan. Neuroimage *156*, 340-351.

Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T.*, et al.* (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage *31*, 968-980.

Diamond, A. (2013). Executive functions. Annual review of psychology *64*, 135-168.

Diana, R.A., Yonelinas, A.P., and Ranganath, C. (2007). Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends Cogn Sci *11*, 379-386.

Dormann, C.F., Elith, J., Bacher, S., Buchmann, C., Carl, G., Carré, G., Marquéz, J.R.G., Gruber, B., Lafourcade, B., Leitão, P.J.*, et al.* (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. Ecography *36*, 27-46.

Dosenbach, N.U., Fair, D.A., Miezin, F.M., Cohen, A.L., Wenger, K.K., Dosenbach, R.A., Fox, M.D., Snyder, A.Z., Vincent, J.L., Raichle, M.E.*, et al.* (2007). Distinct brain networks for adaptive and stable task control in humans. Proc Natl Acad Sci U S A *104*, 11073-11078.

Dosenbach, N.U., Visscher, K.M., Palmer, E.D., Miezin, F.M., Wenger, K.K., Kang, H.C., Burgund, E.D., Grimes, A.L., Schlaggar, B.L., and Petersen, S.E. (2006). A core system for the implementation of task sets. Neuron *50*, 799-812.

Dubey, A., and Ray, S. (2019). Cortical Electrocorticogram (ECoG) Is a Local Signal. J Neurosci *39*, 4299-4311.

Duncan, K., Curtis, C., and Davachi, L. (2009). Distinct memory signatures in the hippocampus: intentional States distinguish match and mismatch enhancement signals. J Neurosci *29*, 131-139.

Duzel, E., Habib, R., Guderian, S., and Heinze, H.J. (2004). Four types of novelty-familiarity responses in associative recognition memory of humans. Eur J Neurosci *19*, 1408-1416.

Ebitz, R.B., Smith, E.H., Horga, G., Schevon, C.A., Yates, M.J., McKhann, G.M., Botvinick, M.M., Sheth, S.A., and Hayden, B.Y. (2020). Human dorsal anterior cingulate neurons signal conflict by amplifying task-relevant information. bioRxiv, 2020.2003.2014.991745.

Egner, T., and Hirsch, J. (2005a). Cognitive control mechanisms resolve conflict through cortical amplification of task-relevant information. Nat Neurosci *8*, 1784-1790.

Egner, T., and Hirsch, J. (2005b). The neural correlates and functional integration of cognitive control in a Stroop task. Neuroimage *24*, 539-547.

Eldridge, L.L., Knowlton, B.J., Furmanski, C.S., Bookheimer, S.Y., and Engel, S.A. (2000). Remembering episodes: a selective role for the hippocampus during retrieval. Nat Neurosci *3*, 1149-1152.

Eriksen, B.A., and Eriksen, C.W. (1974). Effects of noise letters upon the identification of a target letter. Perception & Psychophysics *16*, 143-149.

Ezzyat, Y., Wanda, P.A., Levy, D.F., Kadel, A., Aka, A., Pedisich, I., Sperling, M.R., Sharan, A.D., Lega, B.C., Burks, A.*, et al.* (2018). Closed-loop stimulation of temporal cortex rescues functional networks and improves memory. Nat Commun *9*, 365.

Fan, J., Flombaum, J.I., McCandliss, B.D., Thomas, K.M., and Posner, M.I. (2003). Cognitive and brain consequences of conflict. Neuroimage *18*, 42-57.

Fried, I., MacDonald, K.A., and Wilson, C.L. (1997). Single neuron activity in human hippocampus and amygdala during recognition of faces and objects. Neuron *18*, 753-765.

Fried, I., Rutishauser, U., Cerf, M., and Kreiman, G. (2014). Single neuron studies of the human brain : probing cognition (Cambridge, Massachusetts: The MIT Press).

Friedman, D., Cycowicz, Y.M., and Gaeta, H. (2001). The novelty P3: an event-related brain potential (ERP) sign of the brain's evaluation of novelty. Neurosci Biobehav Rev *25*, 355-373.

Friston, K.J., Holmes, A.P., Poline, J., Grasby, P., Williams, S., Frackowiak, R.S., and Turner, R. (1995). Analysis of fMRI time-series revisited. Neuroimage *2*, 45-53.

Fu, Z., Beam, D., Chung, J.M., Reed, C.M., Mamelak, A.N., Adolphs, R., and Rutishauser, U. (2022). The geometry of domain-general performance monitoring in the human medial frontal cortex. Science *376*, eabm9922.

Fu, Z., Wu, D.J., Ross, I., Chung, J.M., Mamelak, A.N., Adolphs, R., and Rutishauser, U. (2019). Single-Neuron Correlates of Error Monitoring and Post-Error Adjustments in Human Medial Frontal Cortex. Neuron *101*, 165-177 e165.

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

Fuster, J.M., and Alexander, G.E. (1971). Neuron activity related to short-term memory. Science *173*, 652-654.

Fuster, J.M., and Jervey, J.P. (1982). Neuronal firing in the inferotemporal cortex of the monkey in a visual memory task. J Neurosci *2*, 361-375.

Gaetz, W., Liu, C., Zhu, H., Bloy, L., and Roberts, T.P. (2013). Evidence for a motor gamma-band network governing response interference. Neuroimage *74*, 245-253.

Giovanello, K.S., Verfaellie, M., and Keane, M.M. (2003). Disproportionate deficit in associative recognition relative to item recognition in global amnesia. Cogn Affect Behav Neurosci *3*, 186-194.

Goghari, V.M., and MacDonald, A.W., 3rd (2009). The neural basis of cognitive control: response selection and inhibition. Brain Cogn *71*, 72-83.

Goschke, T. (2014). Dysfunctions of decision-making and cognitive control as transdiagnostic mechanisms of mental disorders: advances, gaps, and needs in current research. Int J Methods Psychiatr Res *23 Suppl 1*, 41-57.

Gothe, K., and Oberauer, K. (2008). The integration of familiarity and recollection information in short-term recognition: modeling speed-accuracy trade-off functions. Psychol Res *72*, 289-303.

Gratton, G., Coles, M.G., and Donchin, E. (1992). Optimizing the use of information: strategic control of activation of responses. J Exp Psychol Gen *121*, 480-506.

Gratton, G., Cooper, P., Fabiani, M., Carter, C.S., and Karayanidis, F. (2018). Dynamics of cognitive control: Theoretical bases, paradigms, and a view for the future. Psychophysiology *55*.

Groman, S.M., and Jentsch, J.D. (2012). Cognitive control and the dopamine D(2)-like receptor: a dimensional understanding of addiction. Depress Anxiety *29*, 295-306.

Groppe, D.M., Bickel, S., Dykstra, A.R., Wang, X., Megevand, P., Mercier, M.R., Lado, F.A., Mehta, A.D., and Honey, C.J. (2017). iELVis: An open source MATLAB toolbox for localizing and visualizing human intracranial electrode data. J Neurosci Methods *281*, 40-48.

Hanslmayr, S., Pastotter, B., Bauml, K.H., Gruber, S., Wimber, M., and Klimesch, W. (2008). The electrophysiological dynamics of interference during the Stroop task. J Cogn Neurosci *20*, 215-225.

Heilbronner, S.R., and Hayden, B.Y. (2016). Dorsal Anterior Cingulate Cortex: A Bottom-Up View. Annu Rev Neurosci *39*, 149-170.

Helfrich, R.F., and Knight, R.T. (2016). Oscillatory Dynamics of Prefrontal Cognitive Control. Trends Cogn Sci *20*, 916-930.

Hyman, S.E. (2007). The neurobiology of addiction: implications for voluntary control of behavior. Am J Bioeth *7*, 8-11.

Ison, M.J., Quian Quiroga, R., and Fried, I. (2015). Rapid Encoding of New Memories by Individual Neurons in the Human Brain. Neuron *87*, 220-230.

Jacoby, L.L., and Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. J Exp Psychol Gen *110*, 306-340.

Janssens, C., De Loof, E., Boehler, C.N., Pourtois, G., and Verguts, T. (2018). Occipital alpha power reveals fast attentional inhibition of incongruent distractors. Psychophysiology *55*.

Jensen, O., Kaiser, J., and Lachaux, J.P. (2007). Human gamma-frequency oscillations associated with attention and memory. Trends Neurosci *30*, 317-324.

Jiang, Y., Haxby, J.V., Martin, A., Ungerleider, L.G., and Parasuraman, R. (2000). Complementary neural mechanisms for tracking items in human working memory. Science *287*, 643-646.

Johnson, E.L., and Knight, R.T. (2015). Intracranial recordings and human memory. Curr Opin Neurobiol *31*, 18-25.

Joshi, A., Scheinost, D., Okuda, H., Belhachemi, D., Murphy, I., Staib, L.H., and Papademetris, X. (2011). Unified framework for development, deployment and robust testing of neuroimaging algorithms. Neuroinformatics *9*, 69-84.

Kishiyama, M.M., Yonelinas, A.P., and Knight, R.T. (2009). Novelty enhancements in memory are dependent on lateral prefrontal cortex. J Neurosci *29*, 8114-8118.

Knight, R. (1996). Contribution of human hippocampal region to novelty detection. Nature *383*, 256-259.

Knight, R.T. (1984). Decreased response to novel stimuli after prefrontal lesions in man. Electroencephalogr Clin Neurophysiol *59*, 9-20.

Koga, S., Rothermel, R., Juhasz, C., Nagasawa, T., Sood, S., and Asano, E. (2011). Electrocorticographic correlates of cognitive control in a Stroop task-intracranial recording in epileptic patients. Hum Brain Mapp *32*, 1580-1591.

Kohler, S., Danckert, S., Gati, J.S., and Menon, R.S. (2005). Novelty responses to relational and non-relational information in the hippocampus and the parahippocampal region: a comparison based on event-related fMRI. Hippocampus *15*, 763-774.

Kucewicz, M.T., Cimbalnik, J., Matsumoto, J.Y., Brinkmann, B.H., Bower, M.R., Vasoli, V., Sulc, V., Meyer, F., Marsh, W.R., Stead, S.M.*, et al.* (2014). High frequency oscillations are associated with cognitive processing in human recognition memory. Brain *137*, 2231-2244.

Kuusinen, V., Cesnaite, E., Perakyla, J., Ogawa, K.H., and Hartikainen, K.M. (2018). Orbitofrontal Lesion Alters Brain Dynamics of Emotion-Attention and Emotion-Cognitive Control Interaction in Humans. Front Hum Neurosci *12*, 437.

Lara, A.H., and Wallis, J.D. (2015). The Role of Prefrontal Cortex in Working Memory: A Mini Review. Front Syst Neurosci *9*, 173.

Lesh, T.A., Niendam, T.A., Minzenberg, M.J., and Carter, C.S. (2011). Cognitive control deficits in schizophrenia: mechanisms and meaning. Neuropsychopharmacology *36*, 316-338.

Lhermitte, F. (1986). Human autonomy and the frontal lobes. Part II: Patient behavior in complex and social situations: the "environmental dependency syndrome". Ann Neurol *19*, 335-343.

Li, Y.S., Nassar, M.R., Kable, J.W., and Gold, J.I. (2019). Individual Neurons in the Cingulate Cortex Encode Action Monitoring, Not Selection, during Adaptive Decision-Making. J Neurosci *39*, 6668-6683.

Lin, T.M., M; Belongie, S; Bourdev, L; Girshick, R; Hays, J; Perona, P; Ramanan, D; Zitnick, C. L; Dollar, P; (2015). Microsoft COCO: Common Objects in Context. arXiv.

Liston, C., Matalon, S., Hare, T.A., Davidson, M.C., and Casey, B.J. (2006). Anterior cingulate and posterior parietal cortices are sensitive to dissociable forms of conflict in a task-switching paradigm. Neuron *50*, 643-653.

Liu, H., Agam, Y., Madsen, J.R., and Kreiman, G. (2009). Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. Neuron *62*, 281-290.

Lundqvist, M., Rose, J., Herman, P., Brincat, S.L., Buschman, T.J., and Miller, E.K. (2016). Gamma and Beta Bursts Underlie Working Memory. Neuron *90*, 152-164.

MacLeod, C.M. (1991). Half a century of research on the Stroop effect: an integrative review. Psychol Bull *109*, 163-203.

Marek, S., and Dosenbach, N.U.F. (2018). The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. Dialogues Clin Neurosci *20*, 133-140.

Mayes, A., Montaldi, D., and Migo, E. (2007). Associative memory and the medial temporal lobes. Trends Cogn Sci *11*, 126-135.

Mayr, U., Awh, E., and Laurey, P. (2003). Conflict adaptation effects in the absence of executive control. Nat Neurosci *6*, 450-452.

Menon, V., and D'Esposito, M. (2022). The role of PFC networks in cognitive control and executive function. Neuropsychopharmacology *47*, 90-103.

Menon, V., and Uddin, L.Q. (2010). Saliency, switching, attention and control: a network model of insula function. Brain Struct Funct *214*, 655-667.

Merkow, M.B., Burke, J.F., and Kahana, M.J. (2015). The human hippocampus contributes to both the recollection and familiarity components of recognition memory. Proc Natl Acad Sci U S A *112*, 14378-14383.

Milham, M.P., and Banich, M.T. (2005). Anterior cingulate cortex: an fMRI analysis of conflict specificity and functional differentiation. Hum Brain Mapp *25*, 328-335.

Milham, M.P., Banich, M.T., Webb, A., Barad, V., Cohen, N.J., Wszalek, T., and Kramer, A.F. (2001). The relative involvement of anterior cingulate and prefrontal cortex in attentional control depends on nature of conflict. Brain Res Cogn Brain Res *12*, 467-473.

Miller, E.K. (2000). The prefrontal cortex and cognitive control. Nat Rev Neurosci *1*, 59-65.

Miller, E.K., and Cohen, J.D. (2001). An integrative theory of prefrontal cortex function. Annu Rev Neurosci *24*, 167-202.

Miller, E.K., Erickson, C.A., and Desimone, R. (1996). Neural mechanisms of visual working memory in prefrontal cortex of the macaque. J Neurosci *16*, 5154-5167.

Mitra, P., and Bokil, H. (2008). Observed brain dynamics (New York ; Oxford: Oxford University Press).

Montaldi, D., Spencer, T.J., Roberts, N., and Mayes, A.R. (2006). The neural system that mediates familiarity memory. Hippocampus *16*, 504-520.

Mormann, F., Fell, J., Axmacher, N., Weber, B., Lehnertz, K., Elger, C.E., and Fernandez, G. (2005). Phase/amplitude reset and theta-gamma interaction in the human medial temporal lobe during a continuous word recognition memory task. Hippocampus *15*, 890-900.

Mukamel, R., and Fried, I. (2012). Human intracranial recordings and cognitive neuroscience. Annu Rev Psychol *63*, 511-537.

Muller, N.G., Machado, L., and Knight, R.T. (2002). Contributions of subregions of the prefrontal cortex to working memory: evidence from brain lesions in humans. J Cogn Neurosci *14*, 673-686.

Murray, R.J., Brosch, T., and Sander, D. (2014). The functional profile of the human amygdala in affective processing: insights from intracranial recordings. Cortex *60*, 10-33.

Nakamura, K., Roesch, M.R., and Olson, C.R. (2005). Neuronal activity in macaque SEF and ACC during performance of tasks involving conflict. J Neurophysiol *93*, 884-908.

Norman, Y., Yeagle, E.M., Khuvis, S., Harel, M., Mehta, A.D., and Malach, R. (2019). Hippocampal sharp-wave ripples linked to visual episodic recollection in humans. Science *365*.

Oberauer, K. (2005). Binding and inhibition in working memory: individual and age differences in short-term recognition. J Exp Psychol Gen *134*, 368-387.

Oehrn, C.R., Hanslmayr, S., Fell, J., Deuker, L., Kremers, N.A., Do Lam, A.T., Elger, C.E., and Axmacher, N. (2014). Neural communication patterns underlying conflict detection, resolution, and adaptation. J Neurosci *34*, 10438-10452.

Oostenveld, R., Fries, P., Maris, E., and Schoffelen, J.M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. Comput Intell Neurosci *2011*, 156869.

Opitz, B., Mecklinger, A., Friederici, A.D., and von Cramon, D.Y. (1999). The functional neuroanatomy of novelty processing: integrating ERP and fMRI results. Cereb Cortex *9*, 379-391.

Owen, A.M., Downes, J.J., Sahakian, B.J., Polkey, C.E., and Robbins, T.W. (1990). Planning and spatial working memory following frontal lobe lesions in man. Neuropsychologia *28*, 1021-1034.

Paller, K.A., and McCarthy, G. (2002). Field potentials in the human hippocampus during the encoding and recognition of visual stimuli. Hippocampus *12*, 415-420.

Park, J., Lee, H., Kim, T., Park, G.Y., Lee, E.M., Baek, S., Ku, J., Kim, I.Y., Kim, S.I., Jang, D.P.*, et al.* (2014). Role of low- and high-frequency oscillations in the human hippocampus for encoding environmental novelty during a spatial navigation task. Hippocampus *24*, 1341-1352.

Parris, B.A., Wadsley, M.G., Hasshim, N., Benattayallah, A., Augustinova, M., and Ferrand, L. (2019). An fMRI Study of Response and Semantic Conflict in the Stroop Task. Front Psychol *10*, 2426.

Paus, T., Petrides, M., Evans, A.C., and Meyer, E. (1993). Role of the human anterior cingulate cortex in the control of oculomotor, manual, and speech responses: a positron emission tomography study. J Neurophysiol *70*, 453-469.

Pelli, D.G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. Spat Vis *10*, 437-442.

Ranganath, C., Cohen, M.X., Dam, C., and D'Esposito, M. (2004). Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. J Neurosci *24*, 3917-3925.

Ray, S., and Maunsell, J.H. (2011). Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. PLoS Biol *9*, e1000610.

Reuter, M., Schmansky, N.J., Rosas, H.D., and Fischl, B. (2012). Within-subject template estimation for unbiased longitudinal image analysis. Neuroimage *61*, 1402-1418.

Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. Science *306*, 443-447.

Rigotti, M., Barak, O., Warden, M.R., Wang, X.J., Daw, N.D., Miller, E.K., and Fusi, S. (2013). The importance of mixed selectivity in complex cognitive tasks. Nature *497*, 585-590.

Robertson, J.A., Thomas, A.W., Prato, F.S., Johansson, M., and Nittby, H. (2014). Simultaneous fMRI and EEG during the multi-source interference task. PLoS One *9*, e114599.

Rutishauser, U., Mamelak, A.N., and Schuman, E.M. (2006). Single-trial learning of novel stimuli by individual neurons of the human hippocampus-amygdala complex. Neuron *49*, 805-813.

Rutishauser, U., Reddy, L., Mormann, F., and Sarnthein, J. (2021). The Architecture of Human Memory: Insights from Human Single-Neuron Recordings. J Neurosci *41*, 883-890.

Rutishauser, U., Ross, I.B., Mamelak, A.N., and Schuman, E.M. (2010). Human memory strength is predicted by theta-frequency phase-locking of single neurons. Nature *464*, 903-907.

Rutishauser, U., Schuman, E.M., and Mamelak, A.N. (2008). Activity of human hippocampal and amygdala neurons during retrieval of declarative memories. Proc Natl Acad Sci U S A *105*, 329-334.

Rutishauser, U., Ye, S., Koroma, M., Tudusciuc, O., Ross, I.B., Chung, J.M., and Mamelak, A.N. (2015). Representation of retrieval confidence by single neurons in the human medial temporal lobe. Nat Neurosci *18*, 1041-1050.

Sakai, K., and Miyashita, Y. (1991). Neural organization for the long-term memory of paired associates. Nature *354*, 152-155.

Sani, I., Stemmann, H., Caron, B., Bullock, D., Stemmler, T., Fahle, M., Pestilli, F., and Freiwald, W.A. (2021). The human endogenous attentional control network includes a ventro-temporal cortical node. Nat Commun *12*, 360.

Schurgin, M.W. (2018). Visual memory, the long and the short of it: A review of visual working memory and long-term memory. Atten Percept Psychophys *80*, 1035-1056.

Scoville, W.B., and Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. J Neurol Neurosurg Psychiatry *20*, 11-21.

Sederberg, P.B., Schulze-Bonhage, A., Madsen, J.R., Bromfield, E.B., Litt, B., Brandt, A., and Kahana, M.J. (2007). Gamma oscillations distinguish true from false memories. Psychol Sci *18*, 927-932.

Shenhav, A., Botvinick, M.M., and Cohen, J.D. (2013). The expected value of control: an integrative theory of anterior cingulate cortex function. Neuron *79*, 217-240.

Sheth, S.A., Mian, M.K., Patel, S.R., Asaad, W.F., Williams, Z.M., Dougherty, D.D., Bush, G., and Eskandar, E.N. (2012). Human dorsal anterior cingulate cortex neurons mediate ongoing behavioural adaptation. Nature *488*, 218-221.

Simon, J.R., and Berbaum, K. (1990). Effect of conflicting cues on information processing: the 'Stroop effect' vs. the 'Simon effect'. Acta Psychol (Amst) *73*, 159-170.

Singer, W., and Gray, C.M. (1995). Visual feature integration and the temporal correlation hypothesis. Annu Rev Neurosci *18*, 555-586.

Smith, E.E., Jonides, J., Koeppe, R.A., Awh, E., Schumacher, E.H., and Minoshima, S. (1995). Spatial versus Object Working Memory: PET Investigations. J Cogn Neurosci *7*, 337-356.

Smith, E.H., Horga, G., Yates, M.J., Mikell, C.B., Banks, G.P., Pathak, Y.J., Schevon, C.A., McKhann, G.M., 2nd, Hayden, B.Y., Botvinick, M.M*., et al.* (2019). Widespread temporal coding of cognitive control in the human prefrontal cortex. Nat Neurosci *22*, 1883-1891.

Squire, L.R., and Zola-Morgan, S. (1991). The medial temporal lobe memory system. Science *253*, 1380-1386.

Staresina, B.P., Reber, T.P., Niediek, J., Bostrom, J., Elger, C.E., and Mormann, F. (2019). Recollection in the human hippocampal-entorhinal cell circuitry. Nat Commun *10*, 1503.

Stroop, J.R. (1935). Studies of interference in serial verbal reactions (Nashville, Tenn.: George Peabody College for Teachers), pp. 19 p.

Tallon-Baudry, C., and Bertrand, O. (1999). Oscillatory gamma activity in humans and its role in object representation. Trends Cogn Sci *3*, 151-162.

Tang, H., Yu, H.Y., Chou, C.C., Crone, N.E., Madsen, J.R., Anderson, W.S., and Kreiman, G. (2016). Cascade of neural processing orchestrates cognitive control in human frontal cortex. Elife *5*.

Tully, L.M., Lincoln, S.H., Liyanage-Don, N., and Hooker, C.I. (2014). Impaired cognitive control mediates the relationship between cortical thickness of the superior frontal gyrus and role functioning in schizophrenia. Schizophr Res *152*, 358-364.

Tulving, E., Markowitsch, H.J., Craik, F.E., Habib, R., and Houle, S. (1996). Novelty and familiarity activations in PET studies of memory encoding and retrieval. Cereb Cortex *6*, 71-79.

Turken, A.U., and Swick, D. (1999). Response selection in the human anterior cingulate cortex. Nat Neurosci *2*, 920-924.

Uhlhaas, P.J., Pipa, G., Neuenschwander, S., Wibral, M., and Singer, W. (2011). A new look at gamma? High- (>60 Hz) gamma-band activity in cortical networks: function, mechanisms and impairment. Prog Biophys Mol Biol *105*, 14-28.

van Veen, V., and Carter, C.S. (2005). Separating semantic conflict and response conflict in the Stroop task: a functional MRI study. Neuroimage *27*, 497-504.

van Veen, V., Cohen, J.D., Botvinick, M.M., Stenger, V.A., and Carter, C.S. (2001). Anterior cingulate cortex, conflict monitoring, and levels of processing. Neuroimage *14*, 1302-1308.

van Vugt, M.K., Schulze-Bonhage, A., Litt, B., Brandt, A., and Kahana, M.J. (2010). Hippocampal gamma oscillations increase with memory load. J Neurosci *30*, 2694-2699.

Viskontas, I.V., Knowlton, B.J., Steinmetz, P.N., and Fried, I. (2006). Differences in mnemonic processing by neurons in the human hippocampus and parahippocampal regions. J Cogn Neurosci *18*, 1654-1662.

Wais, P.E., Squire, L.R., and Wixted, J.T. (2010). In search of recollection and familiarity signals in the hippocampus. J Cogn Neurosci *22*, 109-123.

Wais, P.E., Wixted, J.T., Hopkins, R.O., and Squire, L.R. (2006). The hippocampus supports both the recollection and the familiarity components of recognition memory. Neuron *49*, 459-466.

Wang, C., Ulbert, I., Schomer, D.L., Marinkovic, K., and Halgren, E. (2005). Responses of human anterior cingulate cortex microdomains to error detection, conflict monitoring, stimulus-response mapping, familiarity, and orienting. J Neurosci *25*, 604-613.

Wang, J., Tao, A., Anderson, W.S., Madsen, J.R., and Kreiman, G. (2021). Mesoscopic physiological interactions in the human brain reveal small-world properties. Cell Rep *36*, 109585.

Warren, J.M., Akert, K., and Pennsylvania State University. (1964). The frontal granular cortex and behavior (New York,: McGraw-Hill).

Welsch, R.E., and Kuh, E. (1977). Linear regression diagnostics (National Bureau of Economic Research).

Widge, A.S., Heilbronner, S.R., and Hayden, B.Y. (2019). Prefrontal cortex and cognitive control: new insights from human electrophysiology. F1000Res *8*.

Wixted, J.T., and Squire, L.R. (2011). The medial temporal lobe and the attributes of memory. Trends Cogn Sci *15*, 210-217.

Worsley, K.J., and Friston, K.J. (1995). Analysis of fMRI time-series revisited—again. Neuroimage *2*, 173-181.

Wu, J., Ngo, G.H., Greve, D., Li, J., He, T., Fischl, B., Eickhoff, S.B., and Yeo, B.T.T. (2018). Accurate nonlinear mapping between MNI volumetric and FreeSurfer surface coordinate systems. Hum Brain Mapp *39*, 3793-3808.

Xiang, J.Z., and Brown, M.W. (1998). Differential neuronal encoding of novelty, familiarity and recency in regions of the anterior temporal lobe. Neuropharmacology *37*, 657-676.

Xiao, Y., Chou, C.-C., Cosgrove, G.R., Crone, N.E., Stone, S., Madsen, J.R., Reucroft, I., Shih, Y.-C., Weisholtz, D., Yu, H.-Y.*, et al.* (2022). Task-specific neural processes underlying conflict resolution during cognitive control. bioRxiv, 2022.2001.2016.476535.

Yassa, M.A., and Stark, C.E. (2008). Multiple signals of recognition memory in the medial temporal lobe. Hippocampus *18*, 945-954.

Yonelinas, A.P. (2001). Components of episodic memory: the contribution of recollection and familiarity. Philos Trans R Soc Lond B Biol Sci *356*, 1363-1374.

Yonelinas, A.P. (2013). The hippocampus supports high-resolution binding in the service of perception, working memory and long-term memory. Behav Brain Res *254*, 34-44.

Zaehle, T., Bauch, E.M., Hinrichs, H., Schmitt, F.C., Voges, J., Heinze, H.J., and Bunzeck, N. (2013). Nucleus accumbens activity dissociates different forms of salience: evidence from human intracranial recordings. J Neurosci *33*, 8764-8771.

Zhou, Y.D., Ardestani, A., and Fuster, J.M. (2007). Distributed and associative working memory. Cereb Cortex *17 Suppl 1*, i77-87.

Zilverstand, A., Huang, A.S., Alia-Klein, N., and Goldstein, R.Z. (2018). Neuroimaging Impaired Response Inhibition and Salience Attribution in Human Drug Addiction: A Systematic Review. Neuron *98*, 886-903.