Title
Spatiotemporal Dynamics of Working Memory in Humans

Author
Johnson, Elizabeth

Publication Date
2016

Peer reviewed|Thesis/dissertation

Spatiotemporal Dynamics of Working Memory in Humans

by

Elizabeth Loss Johnson

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Psychology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Robert Knight, Chair
Professor Jonathan Wallis
Professor Matthew Walker
Professor Ming Hsu

Fall 2016

Abstract

Spatiotemporal Dynamics of Working Memory in Humans

by

Elizabeth Loss Johnson

Doctor of Philosophy in Psychology

University of California, Berkeley

Professor Robert Knight, Chair

Working memory (WM) is the ability to hold information for online processing. As the basis of long-term memory formation and a fundamental construct of thinking, it is paramount that we understand how WM works. Distributed network models posit that the prefrontal cortex (PFC) supports WM by coordinating top-down control over other regions involved in sensory representation and long-term memory. We utilized an episodic memory paradigm that probes WM for identity, spatial, and temporal information to examine the PFC dependent model of WM. In two studies, multimodal electrophysiology data reveal that PFC control over WM is fundamentally dynamic in nature, and that WM is dependent on activity distributed across anterior and posterior cortical regions. Results challenge the simple PFC model of WM.

Ch. 1 presents evidence from intracranial recordings that frontal and medial temporal lobe (MTL) theta rhythms carry WM-related activity, and uncovers two WM systems. The PFC-MTL system exhibits bidirectional interaction that shifts with msec precision in response to task demands. In contrast, MTL rhythms direct activity in the orbitofrontal cortex via theta rhythms that do not vary with task demands. These findings support a bidirectional PFC-MTL system in humans – in which theta rhythms subserve executive control during episodic memory formation. Ch. 2 presents evidence from patients with unilateral PFC damage, which shows that a posteriorly-sourced alpha-beta network provides adequate resources for well above-chance WM accuracy. However, when the PFC is intact, PFC low theta activity increases commensurate with executive demand, and PFC-sourced slow rhythms and posteriorly-sourced alpha-beta rhythms travel in opposite directions to support optimal WM.

Ch. 3 reviews 15 years of intracranial research on human memory, and considers the potential of intracranial electrophysiology as a technique to address unresolved questions in the neuroscience of human memory. Ch. 4 presents key themes from this work for younger readers; specifically, it introduces the concepts of cross-frequency coupling between theta rhythms and fast activities in the MTL, and inter-regional PFC-MTL synchrony for memory formation. In a second public outreach piece, appendix 1 introduces the logic of neuropsychology to younger readers to understand why memories of music are resilient to the deleterious effects of amnesia and dementia. Appendix 2 shows that WM develops in children along with increases in sustained attention, and appendix 3 reviews evidence that executive control develops commensurate with

PFC connectivity across distributed neural networks. Finally, appendices 4-5 present applications of research on WM and control, together delineating behavioral and neural underpinnings of optimal relational reasoning in neurologically healthy adults.

Table of Contents

Preface

Memory allows us to hold information in mind and continuously update our autobiographies as we amass experience. When this intricate system breaks down, as in dementia, stroke, or traumatic brain injury, individuals lose the sense of self in time. Given high dementia rates (Van Cauwenberghe, Van Broeckhoven & Sleegers, 2016) and rapidly advancing technologies targeted at memory enhancement (Reardon, 2015), it is essential to define the mechanisms that govern this intricate system.

Mounting evidence points to a balancing act of timing and coordination (Johnson & Knight, 2015), irrespective of whether the brain is holding information in working memory (WM) or remembering it in episodes over the long term (Ranganath, Cohen & Brozinsky, 2005; Oztekin, Davachi & McElree, 2010; cf. Bergmann, Rijpkema, Fernández & Kessels, 2012). This contrasts with modular views of WM as a prefrontal cortex (PFC) function (Goldman-Rakic, 1995) and long-term memory as a hippocampal function (Annese et al., 2014). Instead, episodic memories may be formed via PFC control, sensory information representations (Sreenivasan, Curtis & D'Esposito, 2014; Lara & Wallis, 2015; Eriksson, Vogel, Lansner, Bergstöm & Nyberg, 2015), and hippocampal mechanisms for combining information into episodes (Hanslmayr, Staresina & Bowman, 2016). This distributed network is coordinated through precise temporal synchrony.

However, scrutiny of the literature suggests that wherever we look, we find correlates of memory function. Parietal regions may play conjunctive roles in information binding (Shimamura, 2011) and in frontoparietal networks for control (Duncan, 2013), while thalamic function may subserve PFC-hippocampal communication (Sweeney-Reed et al., 2014). Furthermore, evidence on which regions play causal versus peripheral roles (Goldman-Rakic, 1995; Mackey, Devinsky, Doyle, Meager & Curtis, 2016; Ranganath & D'Esposito, 2005) and on how activity unfolds over time (Hanslmayr et al., 2016; Hanslmayr & Staudigl, 2014) is rife with contradiction. A more dynamic perspective suggests that, instead of mnemonic demands initiating a particular cascade of synchronizations through a distributed network, available systems quickly adapt to support the successful formation of episodic memories.

Mammalian neural circuitries support ensemble or multi-unit activity, such that any one neuron becomes increasingly trivial as an ensemble increase in size (Yuste, 2015). In the PFC, populations of neurons display diverse and evolving response profiles (Fusi, Miller & Rigotti, 2016; Stokes et al., 2013; Warden & Miller, 2010; Barak, Tsodyks & Romo, 2010). Importantly, subsequent memory accuracy is commensurate with such multi-unit adaptability (Rigotti et al., 2013; Balaguer-Ballester, Lapish, Seamans & Durstewitz, 2011), which may explain apparently contradictory behaviors following frontal lesions (Duncan, 2013). Similar dynamic coding has also been observed in hippocampal populations (Fusi et al., 2016; Eichenbaum, 2016), which respond non-randomly to distinct episodic features (McKenzie et al., 2014). Taken together, these data suggest that frontal and hippocampal regions support episodic memory formation by interacting dynamically, as a function of task demands.

Ch. 1 and 2 present converging evidence from multimodal electrophysiology and neuropsychology for dynamic, network-level substrates of executive control for human memory. Subjects encoded, maintained, and subsequently selected the identity, spatial relation, or

temporal relation episodic features that denote a real-world memory. Intracranial recordings of the PFC, orbitofrontal cortex (OFC; including the OFC and medial frontal polar areas), and medial temporal lobe (MTL; including the hippocampus and surrounding areas) revealed a bidirectional PFC-MTL system that juggled these episodic features per task demands (Johnson et al., in prep.). In contrast, MTL slow rhythms influenced OFC activities in a parallel system that was invariant to task demands. These findings demonstrate initial evidence of bidirectional PFC-MTL communication in humans – and furthermore, that this system shifts dynamically to subserve executive control during WM.

In a complementary project, electrophysiology data recorded noninvasively in individuals with unilateral lesions to the PFC revealed a PFC source for the slow rhythmic substrate of executive control (Johnson et al., in prep.). When the PFC was not compromised, slow rhythmic activity increased in the PFC per task demands, concurrent with directional connectivity from the PFC to widespread parieto-occipital regions. These effects were absent in patients, which indicates that the entire network was affected when part of the PFC was compromised. In contrast, rhythms originating in more posterior regions precessed toward the frontal lobe at higher frequencies (Hillebrand et al., 2016), irrespective of PFC integrity and task demands. Performance was consistently optimal when both networks were active, with bidirectional, multiplexed rhythms crossing paths between frontal and parieto-occipital regions to support executive control. When the PFC was compromised, the central-posterior network provided adequate, though not optimal, resources for WM.

These findings are incompatible with modular views of the PFC in WM (Goldman-Rakic, 1995; see Sreenivasan et al., 2014; Lara & Wallis, 2015; Postle, 2016). Instead, we observed that the PFC interacts dynamically and rhythmically with multiple other regions to support executive control during episodic memory formation. Taken together, results highlight an emergent shift away from single-neuron and region-specific approaches to cognitive neuroscience – and build on millennia of inquiry into human memory. At the same time, they raise more questions than they answer. Ch. 3 considers the potential of human intracranial electrophysiology as a technique to address some of these questions, with emphasis on the specifics of PFC involvement in memory (Johnson & Knight, 2015). Ch. 4 focuses on core concepts in the neuroscience of human memory for younger audiences, providing the key findings from intracranial research on WM (Johnson & Helfrich, 2016). This public outreach piece introduces such questions to our most open minds, who may address them going forward.

Acknowledgements

Supposedly, a doctoral degree signals that the holder of said degree has come out on the other end of an existential crisis. Not so.

Here is where I get to extend the biggest thanks ever to the people who always put up with me, most notably Margaret Loss, Clark Johnson, and Susanna Segaloff; my advisors, Robert Knight and David Gallo; and the car that can fly, Keith Johnson.

Three cheers to something good in 2016!

Ch. 1: Dynamic Frontotemporal Systems for Episodic Working Memory

Citation: Johnson, E. L., Adams, J. N., Solbakk, A-K., Endestad, T., Larsson, P. G., Lin, J. J. & Knight, R. T. Dynamic frontotemporal systems for episodic working memory (in prep.).

Working memory (WM) is a core neural function necessary to hold information in an online state. We employed intracranial electrophysiology to delineate how human frontotemporal regions support the encoding, maintenance, and processing of episodic features in WM. We observed that theta rhythms carried WM-related activity and uncovered two parallel theta-modulated frontotemporal systems. The medial temporal lobe (MTL) and lateral prefrontal cortex (PFC) system exhibited concurrent bidirectional interaction, which shifted dynamically with executive demand. In contrast, there was a unidirectional MTL theta influence on orbitofrontal (OFC) activity that was invariant to executive demand. Locally, we observed double-dissociations such that MTL was involved in dynamic executive control at encoding and maintenance, with PFC and OFC at processing. Together, these results are incompatible with simple models of PFC-guided network function for WM. Instead, we reveal a rapid, parallel frontotemporal network wherein theta rhythms subserve executive control during episodic memory formation.

1.1 Main Text

The ability to simultaneously hold multiples pieces of information in working memory (WM) allows us to learn, guides behavior, and enables our present neural states to adaptively evolve over time. Contrary to long-held modular views of the selective role of prefrontal cortex (PFC) in WM (Goldman-Rakic, 1995), mounting evidence suggests that WM is linked to distributed networks governed by PFC control over sensory and long-term information representations (Sreenivasan et al., 2014). Further, research with invasive stereotactic recordings – the only technique that enables direct access to the human hippocampus with high spatiotemporal resolution – implicates hippocampal theta rhythms in modulating local high frequency band activity during WM (Axmacher et al., 2010; Leszczyński et al., 2015; Johnson & Knight, 2015).

The hippocampus and surrounding medial temporal lobe structures (together, MTL) are critical for long-term memory (Annese et al., 2014), especially for 'episodes' with distinct spatiotemporal context (Eichenbaum, 2014). Indeed, human hippocampal recordings have shown that long-term memory formation is associated with spiking activity locked to theta phase (Rutishauser et al., 2010) and propose theta-phase multiplexing as a mechanism for episodic memory (Rey et al., 2014; Staudigl & Hanslmayr, 2013; Hanslmayr et al., 2016). Research in rodents (Place et al., 2016; Hallock et al., 2016) and macaques (Brincat & Miller, 2015) reveals that slow rhythms route bidirectional information flow between the PFC and MTL, co-temporaneous with distinct local spiking activities, at different points during the encoding and retrieval of episodic-like memories. These findings raise the possibility that PFC and MTL theta rhythms subserve executive control during episodic memory formation in humans.

We employed the unique spatial and spectrotemporal resolution of invasive electrophysiology to investigate how the encoding, maintenance, and subsequent selection of episodic information unfolds in real time across frontotemporal regions during WM. Ten human subjects (37 ± 13

years of age, 8 males) completed a task designed to operationalize episodic WM (Fig. 1.1A, Materials and Methods). On each trial, subjects encoded two colored shapes in specific spatiotemporal positions in preparation for a subsequent test on either the identity of each shape in the pair, or on the spatial or temporal relationship between the shapes in the pair. Episodic focus was designated at the start of each trial by a cue word – either IDENTITY or RELATION, for two conditions at encoding-maintenance – and then again mid-delay by a test prompt – either SAME (identity), or TOP/BOTTOM (spatial relation) or FIRST/SECOND (temporal relation), for three conditions at processing. Subjects were proficient at the task (accuracy range 0.75-1, chance 0.5), and mixed-effects models confirmed that any effects of episodic focus in the electrophysiology data could not be attributed to difficulty (accuracy $p > 0.75$, correct-trial response time $p > 0.12$; identity 0.90 ± 0.04, 1659 ± 701 msec; spatial relation 0.92 ± 0.09, 1335 ± 574 msec; temporal relation 0.89 ± 0.09, 1475 ± 734 msec).

Signals were recorded directly from MTL (36 electrodes), lateral PFC (118 electrodes), and orbitofrontal (OFC, 102 electrodes) regions, for a total of 256 region-of-interest (ROI) electrodes (Materials and Methods; Fig. 1.1B). For each electrode, we computed the Hilbert transformation in 24 logarithmically-spaced, partially overlapping frequency bands between 1-200 Hz, and then selected time windows for the 200-msec pretrial baseline, 1500-msec encoding-maintenance period, and 900-msec processing period (Fig. 1.1A). Then, we took a data-driven approach to first quantify the mechanisms underlying successful WM, and then test whether and how focusing on one versus another episodic feature within a memory would shift those mechanisms.

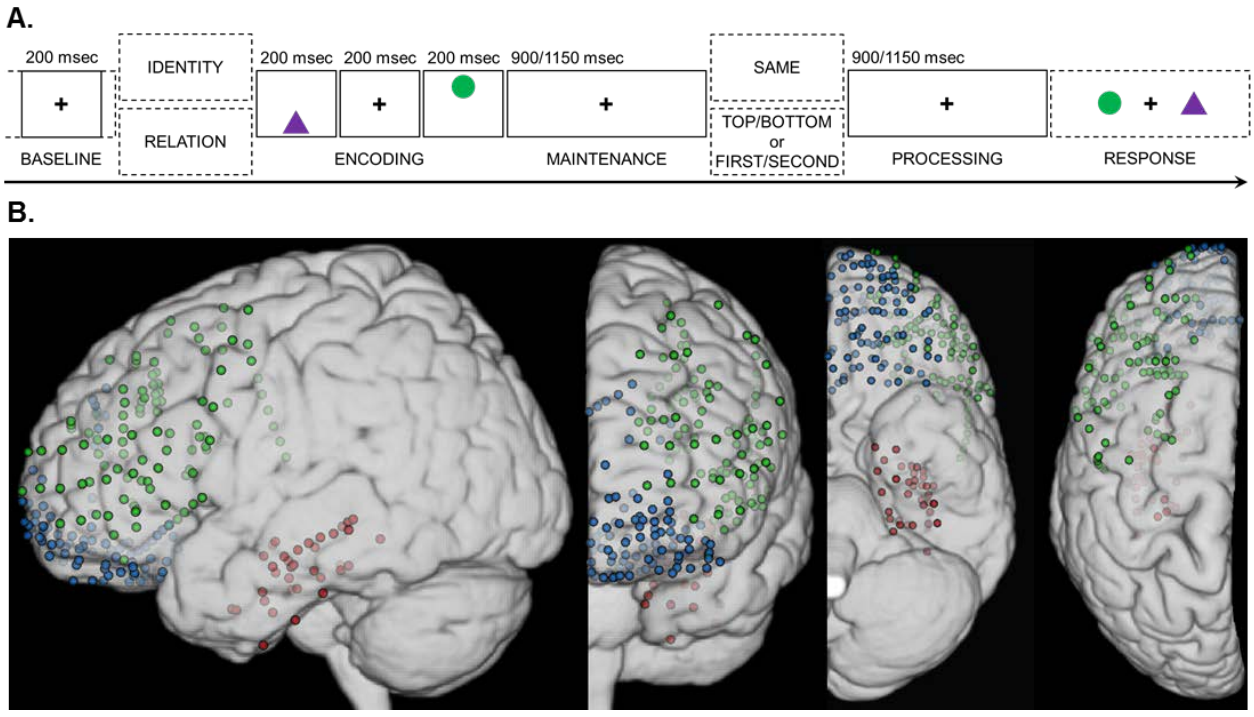

Fig. 1.1. Single-trial WM design (A) and electrode coverage for all 10 subjects, normalized to the left hemisphere and color-coded by ROI (B). After a 1-sec pretrial fixation, subjects were cued to focus on either IDENTITY or RELATION information at encoding (A). Then, two shapes were presented rapidly in one of two spatial positions and one of two temporal positions.

After a 900- or 1150-msec maintenance fixation, the test prompt appeared, followed by a processing fixation of the same length. WM was tested in a two-alternative forced choice test. In the identity test, subjects indicated whether the pair was the SAME pair they just studied. In the spatiotemporal relation test, subjects indicated which shape fit the TOP/BOTTOM spatial or FIRST/SECOND temporal relation prompt. Electrode coverage is shown in four views of the left hemisphere (B), from left to right: sagittal, coronal, ventral, and dorsal. Red, MTL; green, PFC; blue, OFC.

Single-trial analyses of task-evoked spectrotemporal power revealed sustained theta increases and variable, faster activity bursts (centered between 16-152.5 Hz) throughout the encoding-maintenance and processing periods in each ROI (Materials and Methods). We also observed alpha (9.5-13.5 Hz) increases in the OFC parallel to variable, alpha-range decreases in the PFC (Hanslmayr et al., 2016). The spectrotemporal characteristics of these activity bursts varied by trial, but were significant when averaged across all correct trials ($|z| > 2.58$, $p < 0.01$; Fig. 1.2A-C). The higher-frequency activities were partially suggestive of distinct, population-level spiking (Scheffer-Teixeira et al., 2013; Buzsáki & Wang, 2012; Sherman et al., 2016; Lundqvist et al., 2016). Per-electrode cluster-based statistics (Maris & Oostenvald, 2007) indicated that there were no effects of episodic condition in any ROI. Trial-by-trial fluctuations dominated the episodic WM encoding and delay periods above any overall condition-related patterns in 249/256 electrodes (cf. Lundqvist et al., 2016; Stokes & Spaak, 2016), suggesting that sustained theta and bursts of faster activity index successful WM in frontotemporal regions.

For each subject, we seeded the MTL electrodes and extracted the maximal task-evoked theta frequency, which spanned the canonical range of 3.5-8-Hz centroids, for use in subsequent analyses (Materials and Methods). Then, we quantified single-trial cross-electrode correlations at peak theta to test whether individual task-evoked slow rhythmic activity linked frontotemporal regions in a simultaneously active network. Analysis of trial-by-trial fluctuations in cross-electrode correlations revealed frontotemporal network theta co-activation in each subject ($z > 1.96$, $p < 0.05$; Fig. 1.2D). Per-subject tests did not detect any effects of episodic condition during encoding-maintenance or processing (all Kruskal-Wallis $p > 0.11$), suggesting that frontotemporal network theta co-activation is common to successful WM across domains.

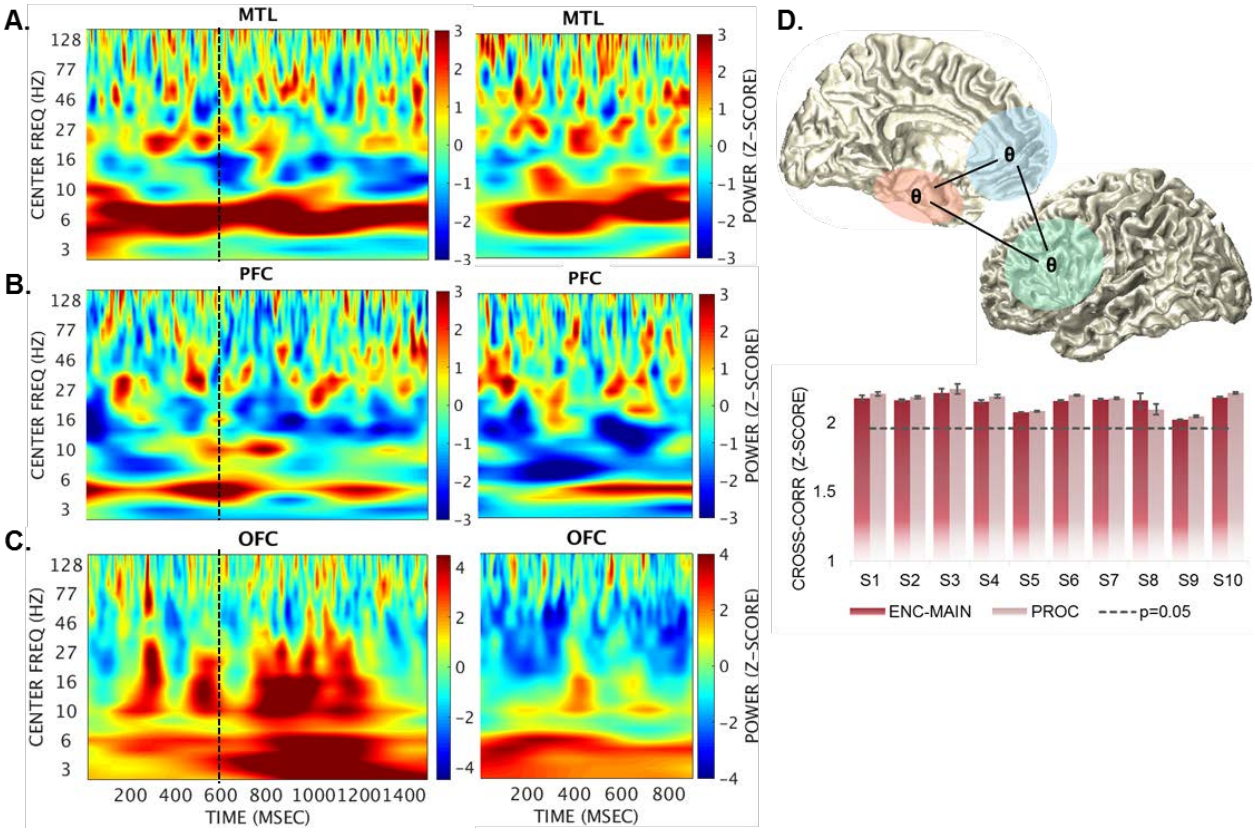

Fig. 1.2. Mean task-evoked spectrotemporal power over encoding-maintenance (left) and processing (right) for single electrodes in the MTL (A), PFC (B), and OFC (C) regions. Single-trial analyses revealed sustained theta power within ROIs (A-C), and cross-electrode correlations at individual peak theta frequency within and between ROIs (D). Vertical dotted line = maintenance fixation start (Fig. 1.1A). IDEN, identity; SPA, spatial relation; TEMP, temporal relation; θ, theta; CROSS-CORR, cross-electrode correlation; ENC-MAIN, encoding-maintenance; PROC, processing.

Because WM was correlated with sustained theta and bursts of faster activity, we hypothesized that effects of episodic condition would emerge as dynamic, theta-phase modulations of higher-frequency activities. If theta multiplexing shifted as a function of episodic condition, then results would reveal a frontal and/or temporal theta-modulated system for executive control. To capture modulation of burst-patterned activities, we computed time-frequency representations of theta phase-amplitude coupling (PAC; Voytek et al., 2013) by condition for each electrode, and across each MTL-PFC and MTL-OFC electrode pair (Materials and Methods). Recent data suggest that PAC fluctuations predict population-level spiking patterns (Murta et al., 2016), making PAC an ideal metric for tracking dynamic shifts in activity bursts. Single-subject PAC encoding-maintenance and processing data were *z*-scored against the pretrial baseline to delineate task-evoked PAC in each electrode and electrode pair, and then submitted to whole-ROI group linear mixed effects models to test for system-level dynamics. Inter-regional PAC pairs were modeled as indices of effective connectivity by first subtracting each frontal ROI phase-MTL amplitude coupling *z*-score from its reverse, yielding positive values if MTL → PFC/OFC and negative values if PFC/OFC → MTL.

Because the baseline-standardization method is critically affected by differences in trial counts, we replicated all models using PAC data *z*-scored on time-resolved surrogate distributions (Materials and Methods). This procedure controls for differences in the number of trials as well as task-related burst amplitude characteristics with msec precision, and offers an alternative method for computing true PAC (Voytek et al., 2013; Aru et al., 2015). All reported results were significant at the single-electrode level, and at the Bonferroni-adjusted alpha threshold for whole-ROI analyses (described below), irrespective of whether input PAC data had been standardized on the pretrial baseline or on surrogate distributions. Because the baseline-standardization method yielded more conservative estimates of true PAC, we chose to report data and statistics obtained using that method. We do not report outcomes that were significant using one or the other, but not both, standardization techniques.

Single-electrode theta PAC reached task-evoked significance in bursts spread across time and amplitude frequency ($z > 1.96$, $p < 0.05$; Fig. 1.S1-S2), and differed by episodic condition in all ROIs ($|z| > 1.96$, $p < 0.05$). We also observed bursting patterns in PAC across inter-regional pairs of electrodes ($z > 1.96$, $p < 0.05$; Fig. 1.3), and by condition ($|z| > 1.96$, $p < 0.05$), suggesting that frontotemporal theta flexibly coordinates activity both locally and across networks (cf. van der Meij et al., 2012; Friese et al., 2013; Voytek et al., 2015). Furthermore, pair-wise comparisons of inter-regional PAC by directionality revealed bursts of theta PAC in each direction at different spectrotemporal points ($|z| > 1.96$, $p < 0.05$, Fig. 1.3, right), demonstrating concurrent, bidirectional frontotemporal interaction during successful episodic WM (cf. Hallock et al., 2016; Brincat & Miller, 2015).

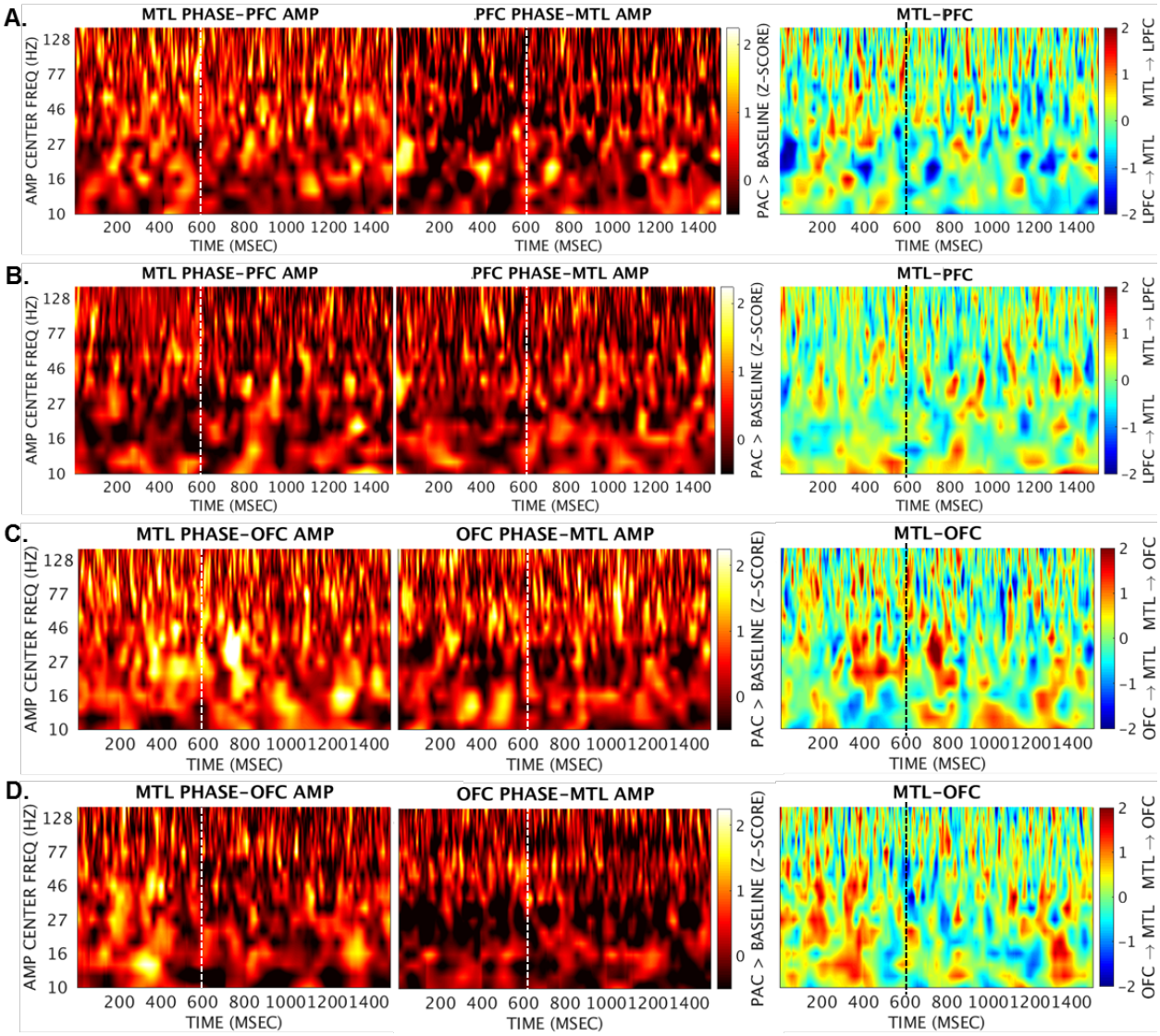

Fig. 1.3. Inter-regional MTL-PFC (A-B) and MTL-OFC (C-D) PAC by episodic condition and directionality in single inter-regional electrode pairs. Inter-regional PAC reached significance in bursts, irrespective of whether the MTL (left) or PFC/OFC (middle) provided theta phase, shown here during encoding-maintenance. Tests of directionality revealed effective MTL-PFC PAC in each direction (right) at different spectrotemporal points during the encoding-maintenance of identity (A) and relation (B) information, $|z| > 1.96$, $p < 0.05$. In contrast, MTL-OFC PAC was largely unidirectional (right), with MTL theta carrying OFC bursts during the encoding-maintenance of identity (C) and relation (D) information. Vertical dotted line = maintenance fixation start (Fig. 1.1A). AMP, amplitude.

We submitted all PAC data to linear mixed effects models by ROI to test whether the bursts differed in dynamic patterns by episodic condition (Materials and Methods). By modeling inter-regional PAC as indices of directionality, we tested whether bidirectional communication was reliably weighted toward MTL or PFC/OFC theta phase by spectrotemporal characteristics and condition. A significant time-frequency interaction would indicate that the spectrotemporal

bursts of PAC reflected a true bursting pattern and not random variation. Under the hypothesis that theta phase would selectively coordinate executive demand, we operationalized results as shifts (or not) in the dynamics of a frontotemporal system. A system was dynamic if episodic condition was sufficient to change local or effective PAC in both time and frequency – that is, observation of a three-way interaction with condition. If an amplitude-providing ROI was dynamic, we can deduce that theta in the phase-providing ROI modulated the bursts of activity revealed in the task-evoked power analysis (Fig. 1.2A-C), supporting our hypothesis and supporting a theta mechanism for executive control.

The models revealed spectrotemporal bursting patterns that selectively shifted by ROI, period during the trial, and episodic condition (Fig. 1.4A). At encoding-maintenance, local PAC occurred in bursts in the MTL (cf. Axmacher et al., 2010; Leszczyński et al., 2015) and effective PAC occurred in bursts throughout the frontotemporal network ($z > 4.5$, $p < 5\times10^{-6}$). The three-way interactions mirrored PAC bursting patterns ($z > 4.1$, $p < 3\times10^{-5}$), suggesting MTL-hubbed theta-modulated dynamics. At processing, bursting patterns shifted to frontal ROIs and dynamics again mirrored PAC bursting patterns. Local PFC and OFC PAC became dynamic ($z > 3.3$, $p < 9\times10^{-4}$), while MTL and effective MTL-OFC PAC dynamics were attenuated ($p_{corr} > 0.05$, $z_{diff} > 2.0$). Effective MTL-PFC PAC remained dynamic ($z > 4.0$, $p < 5\times10^{-5}$), suggesting a transfer of an MTL-hubbed information encoding role (cf. McKenzie et al., 2014; Eichenbaum, 2016) to an PFC-hubbed selection role (cf. Sreenivasan et al., 2014; Fusi et al., 2016). This pattern supports an executive control system subserved by bidirectional MTL-PFC theta interactions. Overall, the correspondence between time-frequency and three-way interactions reveal dynamic theta-based modulation for executive control.

Turning to omnibus patterns, effects of frequency and two-way interactions with episodic condition delineate complementary MTL-PFC and MTL-OFC systems for WM. Frontotemporal network interaction peaked at select amplitude frequencies, revealing a dynamic, net bidirectional relationship between the MTL and PFC that depended on executive demand. At encoding-maintenance, a frequency by condition interaction indicated that MTL-PFC interactions depended on amplitude frequency as a function of episodic condition (Fig. 1.4B-C; $p < 5\times10^{-7}$). We observed a double-dissociation such that PFC theta influenced MTL amplitude in the beta range as a function of center frequency and episodic condition; activity related to identity information was maximally modulated at 16 Hz while relation activity was most pronounced at 26.5 Hz. At the same time, twin dissociations showed that MTL theta influenced PFC amplitudes maximally for identity activity at 12 Hz, and relation over identity activity at gamma (38.5 and 64 Hz) and high-frequency broadband (HFB; 91-107.5) centroids. Come processing, both the main effect of frequency and the interaction were significant, and an overall pattern of increased PFC theta influence on MTL amplitudes emerged ($p < 0.002$). Still, MTL-PFC interactions remained bidirectional, and amplitude frequency by condition effects exhibited variability throughout the frequency spectrum, suggestive of flexible, inter-regional executive control.

The MTL-PFC interactions demonstrated a theta-phase multiplexed system for bidirectional communication during episodic WM (cf. Brincat & Miller, 2015; Hallock et al., 2016). In contrast, MTL theta directed MTL-OFC interactions, which were invariant to amplitude center frequency and executive demand (Fig. 1.4B, 4D). Throughout encoding-maintenance and

processing, main effects of frequency indicated that MTL theta phase directed OFC amplitudes at alpha and beta, broadly defined (9.5-32 Hz) and HFB (76.5-107.5 Hz; $p < 3\times10^{-7}$). Taken together, MTL-PFC and MTL-OFC interactions exhibited divergent profiles during episodic memory formation, with MTL-PFC interactions selectively reflecting executive control.

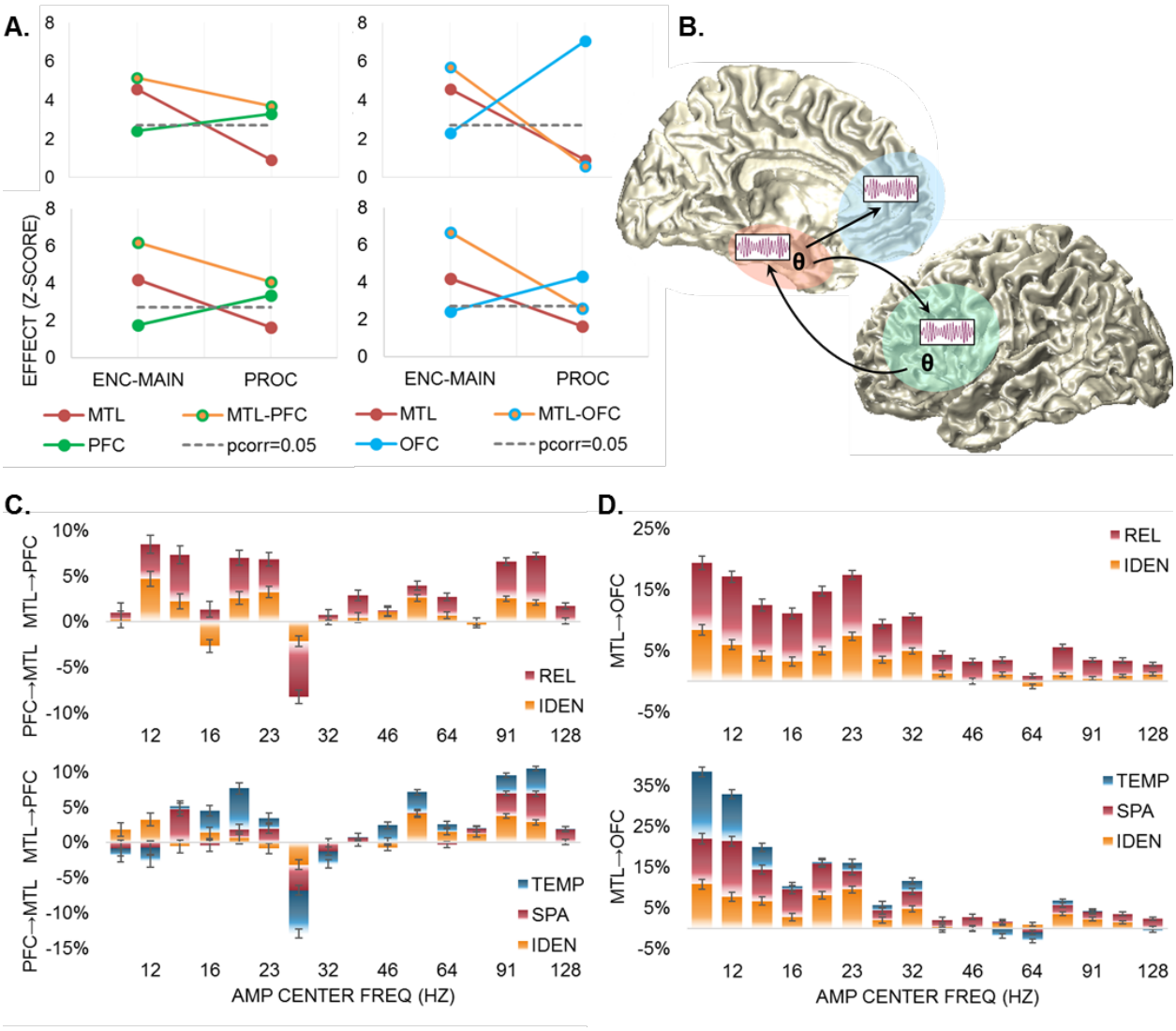

Fig. 1.4. Characteristics of local and effective PAC in time and frequency, and their ROI-specific dynamics. PAC occurred in spectrotemporal bursts (A) at encoding-maintenance in local MTL, and effective MTL-PFC and MTL-OFC; and at processing in local PFC and OFC, and effective MTL-PFC ($p < 0.002$, top). Bursting patterns were mirrored by three-way interactions with condition (bottom). Condition was sufficient to shift the time-frequency characteristics of effective MTL-OFC and MTL-PFC PAC at encoding-maintenance, $p < 7\times10^{-10}$, but only MTL-PFC PAC remained dynamic through processing, $p < 5\times10^{-5}$. Local PAC exhibited role reversals, showing dynamic organization in MTL at encoding-maintenance, $p < 3\times10^{-5}$, then in PFC and OFC at processing, $p < 9\times10^{-4}$. Collapsing the data over time revealed that MTL theta modulated both PFC and OFC amplitudes, while only PFC theta showed the reverse (B). We observed bidirectional PAC multiplexing between MTL and PFC (C), which shifted by condition

at encoding-maintenance (top), $p < 5\times10^{-7}$, and processing (bottom), $p < 0.002$. In contrast, the MTL-OFC system was unidirectional (D), with maximal MTL theta modulation at lower OFC amplitudes, $p < 3\times10^{-7}$. θ, theta; AMP, amplitude; IDEN, identity; REL, relation; SPA, spatial relation; TEMP, temporal relation; ENC-MAIN, encoding-maintenance; PROC, processing.

In sum, our results uncover two complementary frontotemporal systems with dissociable electrophysiological profiles based on dynamic, theta-multiplexed activity during successful WM. First, our results support a distributed network account of WM (Sreenivasan et al., 2014). Our data are incompatible with models of PFC-guided network function; instead, we reveal that executive control is governed by bidirectional MTL-PFC interactions during WM. Because the MTL differentially interacted with the PFC and OFC, co-temporaneous with relative shifts in local network hubs, we show that MTL-PFC and MTL-OFC long-range systems are functionally dissociable. Second, theta rhythms linked regions in the frontotemporal network and modulated faster bursts of activity at different spatiotemporal points during WM. Our results extend recent findings of trial-by-trial fluctuations in macaque ensemble activity (Lundqvist et al., 2016; Stokes & Spaak, 2016) to human cognition, and reveal that they extend beyond the frontal lobe to long-range MTL interactions. Likewise, results demonstrate that theta rhythms carry per-trial activity, substantiating theta multiplexing as an underlying mechanism for episodic memory formation (Rey et al., 2014; Staudigl & Hanslmayr, 2013; Hanslmayr et al., 2016; Place et al., 2016; Axmacher et al., 2010).

We provide the first demonstration of parallel bidirectional communication between the MTL and PFC in humans – and demonstrate that it unfolds with millisecond precision as a function of executive demand during WM. In contrast, OFC activities were on the receiving end of MTL theta modulations, and showed dynamic local PAC specific to information selection at processing. Given proposals that OFC neurons encode goals (Tsujimoto et al., 2011) and that the orbitofrontal cortex maintains a cognitive map of task-related variables, analogous to that of the anatomically connected hippocampus (Wikenheiser & Schoenbaum, 2016), we suggest that the OFC plays a role in distributed memory systems that may be guided by retrieval objectives. Overall, the MTL, PFC, and OFC exhibit unique, interactive roles in dynamic systems for episodic memory formation, each showing evidence of theta-modulated activities, engaged on a network level to coordinate executive demands.

1.2. References

1. Annese, J., Schenker-Ahmed, N. M., Bartsch, H., Maechler, P., Sheh, C., Thomas, N., … Corkin, S. (2014). Postmortem examination of patient H.M.'s brain based on histological sectioning and digital 3D reconstruction. *Nature Communications*, 5, 3122.
2. Aru, J., Aru, J., Priesemann, V., Wibral, M., Lana, L., Pipa, G., … Vicente, R. (2015). Untangling cross-frequency coupling in neuroscience. *Current Opinion in Neurobiology*, 31, 51–61.
3. Axmacher, N., Henseler, M. M., Jensen, O., Weinreich, I., Elger, C. E., & Fell, J. (2010). Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proceedings of the National Academy of Sciences of the United States of America*, 107(7), 3228–3233.

4. Berens, P. (2009). CircStat: A MATLAB toolbox for circular statistics. *Journal of Statistical Software*, 31(10), 1–21.
5. Brincat, S. L., & Miller, E. K. (2015). Frequency-specific hippocampal-prefrontal interactions during associative learning. *Nature Neuroscience*, 18(4), 1–10.
6. Buzsáki, G., & Wang, X.-J. (2012). Mechanisms of Gamma Oscillations. *Annual Review of Neuroscience*, 35, 203–225.
7. Cramer, A. O. J., van Ravenzwaaij, D., Matzke, D., Steingroever, H., Wetzels, R., Grasman, R. P. P. P., … Wagenmakers, E.-J. (2015). Hidden Multiplicity in Multiway ANOVA: Prevalence, Consequences, and Remedies. *Psychological Bulletin Review*. http://doi.org/10.3758/s13423-015-0913-5
8. Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods*, 134(1), 9–21.
9. Eichenbaum, H. (2014). Time cells in the hippocampus: a new dimension for mapping memories. *Nature Reviews Neuroscience*, 15(11), 732–44.
10. Eichenbaum, H. (2016). Still searching for the engram. *Learning & Behavior*. http://doi.org/10.3758/s13420-016-0218-1
11. Flinker, A., Korzeniewska, A., Shestyuk, A. Y., Franaszczuk, P. J., Dronkers, N. F., Knight, R. T., & Crone, N. E. (2015). Redefining the role of Broca's area in speech. *PNAS*, 112(9), 2871–2875.
12. Friese, U., Köster, M., Hassler, U., Martens, U., Trujillo-Barreto, N., & Gruber, T. (2013). Successful memory encoding is associated with increased cross-frequency coupling between frontal theta and posterior gamma oscillations in human scalp-recorded EEG. *NeuroImage*, 66, 642–647.
13. Fusi, S., Miller, E. K., & Rigotti, M. (2016). Why neurons mix: High dimensionality for higher cognition. *Current Opinion in Neurobiology*, 37, 66–74.
14. Goldman-Rakic, P. (1995). Cellular basis of working memory. *Neuron*, 14(3), 477–485.
15. Haegens, S., Cousijn, H., Wallis, G., Harrison, P. J., & Nobre, A. C. (2014). Inter- and intra-individual variability in alpha peak frequency. *NeuroImage*, 92, 46–55.
16. Hallock, H. L., Wang, A., & Griffin, A. L. (2016). Ventral Midline Thalamus Is Critical for Hippocampal-Prefrontal Synchrony and Spatial Working Memory. *Journal of Neuroscience*, 36(32), 8372–8389.
17. Hanslmayr, S., Staresina, B. P., & Bowman, H. (2016). Oscillations and Episodic Memory: Addressing the Synchronization/Desynchronization Conundrum. *Trends in Neurosciences*, 39(1), 16–25.
18. Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, 59(4), 434–446.
19. Johnson, E. L., & Knight, R. T. (2015). Intracranial recordings and human memory. *Current Opinion in Neurobiology*, 31, 18–25.
20. Leszczyński, M., Fell, J., & Axmacher, N. (2015). Rhythmic Working Memory Activation in the Human Hippocampus. *Cell Reports*, 13(6), 1272–1282.
21. Lundqvist, M., Rose, J., Herman, P., Brincat, S. L., Buschman, T. J., & Miller, E. K. (2016). Gamma and beta bursts underlie working memory. *Neuron*. http://doi.org/10.1016/j.neuron.2016.02.028
22. Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, 164(1), 177–190.

23. McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D., & Eichenbaum, H. (2014). Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron*, 83(1), 202–215.
24. Murta, T., Chaudhary, U., Tierney, T., Dias, A., Leite, M., Carmichael, D., … Lemieux, L. (2016). Phase-amplitude coupling and the BOLD signal: A simultaneous intracranial EEG (icEEG) - fMRI study in humans performing a finger-tapping task. *NeuroImage*. http://doi.org/10.1016/j.neuroimage.2016.08.036
25. Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational Intelligence and Neuroscience*, 2011. http://doi.org/10.1155/2011/156869
26. Place, R., Farovik, A., Brockmann, M., & Eichenbaum, H. (2016). Bidirectional prefrontal-hippocampal interactions support context-guided memory. *Nature Neuroscience*, 19(8). http://doi.org/10.1038/nn.4327
27. Rey, H. G., Fried, I., & Quian Quiroga, R. (2014). Timing of single-neuron and local field potential responses in the human medial temporal lobe. *Current Biology*, 24(3), 299–304.
28. Rutishauser, U., Ross, I. B., Mamelak, A. N., & Schuman, E. M. (2010). Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature*, 464(7290), 903–907.
29. Scheffer-Teixeira, R., Belchior, H., Leão, R. N., Ribeiro, S., & Tort, A. B. L. (2013). On high-frequency field oscillations (>100 Hz) and the spectral leakage of spiking activity. *Journal of Neuroscience*, 33(4), 1535–1539.
30. Sherman, M. A., Lee, S., Law, R., Haegens, S., Thorn, C. A., & Hämäläinen, M. S. (2016). Neural mechanisms of transient neocortical beta rhythms: Converging evidence from humans, computational modeling, monkeys, and mice. *PNAS*. http://doi.org/10.1073/pnas.1604135113
31. Shirhatti, V., Borthakur, A., & Ray, S. (2016). Effect of Reference Scheme on Power and Phase of the Local Field Potential. *Neural Computation*, 882–913.
32. Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, 18(2), 82–89.
33. Staudigl, T., & Hanslmayr, S. (2013). Theta oscillations at encoding mediate the context-dependent nature of human episodic memory. *Current Biology*, 23(12), 1101–1106.
34. Stokes, M., & Spaak, E. (2016). The Importance of Single-Trial Analyses in Cognitive Neuroscience. *Trends in Cognitive Sciences*. http://doi.org/10.1016/j.tics.2016.05.008
35. Trongnetrpunya, A., Nandi, B., Kang, D., Kocsis, B., Schroeder, C. E., & Ding, M. (2016). Assessing Granger Causality in Electrophysiological Data: Removing the Adverse Effects of Common Signals via Bipolar Derivations. *Frontiers in Systems Neuroscience*, 9.
36. Tsujimoto, S., Genovesio, A., & Wise, S. P. (2011). Frontal pole cortex: Encoding ends at the end of the endbrain. *Trends in Cognitive Sciences*, 15(4), 169–176.
37. van der Meij, R., Kahana, M., & Maris, E. (2012). Phase-Amplitude Coupling in Human Electrocorticography Is Spatially Distributed and Phase Diverse. *Journal of Neuroscience*, 32(1), 111–123.
38. Voytek, B., D'Esposito, M., Crone, N., & Knight, R. T. (2013). A method for event-related phase/amplitude coupling. *NeuroImage*, 64(1), 416–424.

39. Voytek, B., Kayser, A. S., Badre, D., Fegen, D., Chang, E. F., Crone, N. E., … D'Esposito, M. (2015). Oscillatory dynamics coordinating human frontal networks in support of goal maintenance. *Nature Neuroscience*. http://doi.org/10.1038/nn.4071
40. Wikenheiser, A. M., & Schoenbaum, G. (2016). Over the river, through the woods: cognitive maps in the hippocampus and orbitofrontal cortex. *Nature Reviews Neuroscience*, 17(8), 513–523.

Acknowledgements: We thank Dr. Josef Parvizi for assisting in the collection of data at Stanford Medical Center. We also thank S. M. Griffin for creating Fig. 1.1B, as well as S. M. Szczepanski, J. W. Y. Kam, R. F. Helfrich, F. Foo, and Y. M. Fonken for helpful discussions, and C. D. Dewar and J. Lubell for technical assistance. This work was supported by grants from the National Institutes of Health (2R37NS21135 to R.T.K., NINDS NS060993 K23 to J.J.L.), Nielsen Corporation (to R.T.K.), UC Irvine School of Medicine Bridge Fund (to J.J.L.), Research Council of Norway (240389/F20 to A.K.S. and T.E.), and University of Oslo Internal Fund (to A.K.S. and T.E.).

1.3. Materials and Methods

1.3.1. Experimental Design

We tested episodic working memory (WM) in a single-trial visuospatial task paradigm (Fig. 1.1A). After a 1-sec pretrial fixation, a starting screen indicated whether the following pair of stimuli would be tested for IDENTITY or spatiotemporal RELATION information. Two colored shapes were then presented for 200 msec each in one of two vertical spatial positions and one of two temporal positions. After a 900- or 1150-msec maintenance period, a test prompt appeared indicating what type of information would be tested. After a processing period of the same length, two shapes were presented on the horizontal axis and subjects responded in a two-alternative forced choice test, yielding a 0.5 chance rate. In the identity test, subjects indicated whether the pair was the SAME pair they just studied; half of the pairs show two old shapes ("yes") and half the pairs show one old shape and one new shape ("no"). In the spatial relation test, subjects indicated which shape had been on the TOP or BOTTOM, and in the temporal relation test, which shape had been presented FIRST or SECOND.

The length of maintenance and processing delay periods was randomly jittered at 900- or 1150-msec to preclude anticipatory or in-phase mechanisms. The task was fully counterbalanced with 120 trials split evenly between identity, spatial, and temporal conditions, chosen randomly from a pool of 150 trials. No stimuli were repeated across trials. An experimenter went through the task instructions and a set of six practice trials with each subject, who was permitted to repeat the practice trials by request.

We submitted group accuracy and correct-trial response time data to logit and linear mixed-effects models (fitglme.m, fitlme.m), respectively, with condition as the fixed effect and subject as the random effect, to confirm that the three conditions did not differ in difficulty (Jaeger, 2008). We analyzed electrophysiological data for correct trials.

1.3.2. Data Acquisition

We report data from ten human subjects (37 ± 13 years of age, 8 males) who were undergoing intracranial monitoring to localize epileptic foci in preparation for surgical resection. Electrodes were implanted based solely on the clinical needs of each patient, and we selected datasets for inclusion via off-site review of individual neuroanatomical coverage. These datasets were collected at one of three hospitals: University of California, Irvine (UCI; seven subjects with stereotactic and/or subdural implants), Oslo University Hospital (OUH; two subjects with stereotactic implants), or Stanford Medical Center (Stanford; one subject with stereotactic and subdural implants). All patients gave informed written consent in accordance with the respective institutional review board.

UCI and OUH data were acquired using Nihon Kohden recording systems, and Stanford data were acquired using a Tucker Davis Technologies recording system. Data were sampled at 10 kHz (UCI), 5 kHz (UCI), 1.526 kHz (Stanford), or 512 Hz (OUH). We resampled UCI and Stanford data offline to 1 kHz (resample.m), and preprocessed all datasets through a standard pipeline.

1.3.3. Data Preprocessing

Each patient's raw electrophysiological traces were manually inspected by a neurologist (R.T.K.) in order to exclude any pathological channels and mark epochs containing ictal activity. Remaining channels were high-pass filtered above 1 Hz and low-pass filtered below 200 Hz (165 Hz for OUH data) using zero-phase finite impulse response filters (eegfilt.m), and then filtered for 60-Hz (50-Hz for OUH data) line noise harmonics using a discrete Fourier transform and demeaned (ft_preprocessing.m). We manually re-inspected the data to mark any channels and epochs containing residual noise or abnormal signal for exclusion.

We determined individual channel locations in a group review of each patient's structural and post-implant scans under at least one neurologist. Channels were included by anatomical placement in three regions of interest (ROIs): medial temporal lobe (MTL: hippocampus, or parahippocampal, perirhinal, or entorhinal area), lateral prefrontal cortex (PFC: inferior, middle, or superior frontal, or premotor area), and orbitofrontal cortex (OFC: orbitofrontal area or medial frontal pole). We referenced every clean channel within an ROI to its next adjacent channel, spaced at 0.5 cm within that ROI, using bipolar montages to create a new set of virtual electrodes with minimized volume conduction (Shirhatti et al., 2016; Trongnetrpunya et al., 2016). The final dataset contained 256 traces across ten subjects, comprised of 3.6 ± 2.0 MTL, 11.8 ± 8.4 PFC, and 10.2 ± 14.4 OFC virtual electrodes per subject.

We then split each subject's data into trials with a 1-sec buffer on either side, and excluded any trials overlapping with epochs that had been marked for ictal or other abnormal activity. We split all remaining trials into: (1) a 200-msec pretrial baseline window extending from 250-50 msec before the start screen; (2) a 1500-msec encoding-maintenance window extending from the onset of the first stimulus, irrespective of delay jitter; and (3) a 900-msec processing window extending from the start of the processing period, irrespective of delay jitter. Each of the three within-trial windows was padded out to 7.5 sec, for a minimum of three cycles' buffer at 1 Hz on

either side, to minimize filtering-induced edge artifacts. We quantified all metrics per electrode, separately for pretrial, encoding-maintenance, and processing data. Then, we standardized encoding-maintenance and processing data outputs against a common pretrial baseline, and applied inferential statistics.

1.3.4. Event-Related Potential (ERP) Analysis

We computed and analyzed single-trial, event-related potentials (ERPs) separately by electrode for each subject. We passed the data through a low-pass, zero-phase finite impulse response filter at 30 Hz, and absolute baseline-corrected the encoding-maintenance and processing outputs on the all-trial, temporal mean of the pretrial data window. Then, we separately analyzed ERP data for condition differences between identity and relation trials at encoding-maintenance, and between identity, spatial, and temporal trials at processing. Statistical testing employed a Monte Carlo method with cluster-based maximum correction for multiple comparisons on the time dimension within each electrode (Maris & Oostenvald, 2007). An independent-samples *t*-test (or *F*-test for three conditions) was used to identify clusters of contiguous time points showing a difference between conditions, thresholded at 0.05, two-tailed, and then the *t*-statistics (or *F*-statistics) were summed over all data points per cluster to calculate cluster size. Then, the condition labels were randomly shuffled and the same clustering procedure was applied; we repeated this procedure 1000 times to create a null distribution. Observed clusters were considered significant if fewer than 10% of randomizations yielded a larger effect (ft_timelockstatistics.m). No condition differences were detected.

1.3.5. Task-Evoked Spectrotemporal Power Analyses

We computed and analyzed single-trial, task-evoked power separately by electrode for each subject. We passed the data through 24 logarithmically-spaced bandpass, zero-phase finite impulse response filters between 1-192 Hz (1-162 Hz and 23 filters for OUH data), with partial overlap. We applied the Hilbert transformation to each of the 24 filtered time series, yielding a series of complex numbers, from which we extracted and squared the absolute values to produce real power values. We baseline-standardized the encoding-maintenance and processing outputs against the pretrial data window using a statistical bootstrapping procedure (cf. Flinker et al., 2015). First, we pooled the pretrial outputs of all trials into a single time series for each electrode and frequency. Then, we randomly selected and averaged *t* data points (*t* = number of trials in that subject's dataset) from the pooled time series; we repeated this procedure 1000 times to create a normal distribution of pretrial baseline power values. Encoding-maintenance and processing data for each trial and time point were *z*-scored on the baseline distribution to assess the significance of task-evoked changes.

Statistical testing of condition differences in task-evoked power employed the equivalent procedure as for ERPs, but with clustering on the time and frequency dimensions (ft_freqstatistics.m). Across all 256 electrodes, seven (1 MTL + 3 PFC + 3 OFC in five subjects) showed condition differences in any time-frequency cluster during encoding-maintenance or processing. We excluded these seven electrodes from further analysis.

1.3.6. Individual Theta Band and Cross-Correlation Analyses

The task-evoked power analysis revealed significantly increased, sustained activity centered in the theta range throughout MTL, PFC, and OFC electrodes (Fig. 1.2A-C). We derived the peak frequency ranges for theta-based analyses on an individual basis from each subject's resultant MTL activity at encoding-maintenance. We selected task-evoked power in seven extended theta-range frequencies centered between 2.75 Hz and 9.5 Hz (cf. Haegens et al., 2014), then averaged over the trial and time dimensions to reveal maximal frequency. Individual peaks spanned the theta center frequencies: $n = 3$ at 3.5 Hz, $n = 1$ at 4.5 Hz, $n = 2$ at 5.5 Hz, $n = 3$ at 6.5 Hz, and $n = 1$ at 8 Hz. We visually inspected the task-evoked power data without averaging to confirm that these peak frequencies exhibited comparable responses across subjects.

Using the MTL electrodes as seeds, we computed cross-electrode correlations on single-trial, task-evoked power at the individual peak theta band for each subject. Concurrent activation would indicate that individual peak task-evoked theta extends to frontal regions and links frontotemporal regions in a simultaneously active network, while a null result would show that MTL peak theta does not extend to frontal regions – irrespective of whether frontotemporal regions would appear linked at another frequency. We applied the Pearson correlation to all electrode combinations over the time dimension to produce per-trial correlation coefficients and associated *p*-values (corrcoef.m) for encoding-maintenance and processing. This analysis revealed frontotemporal network co-activation at individual peak theta frequency on a per-trial basis. For statistical testing of condition differences, we normalized the correlation coefficients using Fisher's *z*-transformation and then submitted them to Kruskal-Wallis tests. No condition differences were detected.

1.3.7. Theta Phase-Amplitude Coupling (PAC) Analyses

We quantified and analyzed cross-frequency phase-amplitude coupling (PAC) separately by electrode for each subject. First, we split the data into conditions and subtracted the trial-averaged time series from each single-trial time series to minimize any contamination from simultaneous voltage changes on phase consistency or PAC (Brincat & Miller, 2015). Then, spectral decomposition employed the equivalent procedure as for computation of spectrotemporal power. We extracted the angle from each individual's theta band complex time series to produce the phase time series for each electrode and trial. We then separately extracted the absolute values from each complex time series to produce a matrix of amplitude time series by frequency for each electrode and trial. We computed event-related PAC between theta phase and the full range of amplitudes, by frequency, within each electrode and also between pairs of electrodes, yielding local and inter-regional PAC.

To compute local PAC, we applied the Pearson correlation between theta phase and each frequency amplitude via sliding window on the trial dimension at every instantaneous time point (circ_corrcl.m). This method produces time-frequency representations of PAC correlation coefficients for each electrode (Voytek et al., 2013). We standardized the encoding-maintenance and processing outputs against the temporal mean of the pretrial data window using Fisher's *z*-tests. This transformed the PAC correlation coefficients to *z*-scores to assess the significance of task-evoked changes in true PAC. Note that this transformation method is more conservative

than traditional surrogate testing, as both true and spurious PAC comprise pretrial data, while only spurious PAC should survive in a surrogate distribution of correlations between the real phase (or amplitude) and a shuffled amplitude (or phase) series of any data (Aru et al., 2015).

However, because Fisher's *z*-tests are critically affected by differences in trial counts, we replicated all reported PAC data standardized on time-resolved surrogate distributions. At each time-frequency point, we randomly shuffled the amplitudes across trials and re-calculated PAC; we repeated this procedure 100 times to create a surrogate distribution of PAC correlation coefficients derived from identical amplitude and trial parameters to the original but with spurious timing relative to real theta phase. Then, raw PAC was *z*-scored on the surrogate distribution to assess the significance of true PAC with msec precision (Voytek et al., 2013). Single-electrode PAC standardized on the pretrial baseline indeed returned more conservative estimates of true PAC – e.g., *z*-scores of 2.5 versus 3.5. For this reason, and for consistency with the single-electrode condition and directionality comparisons (described below), we report data obtained using the baseline-standardized PAC.

Then, we separately analyzed PAC data for condition differences between identity and relation trials at encoding-maintenance, and between identity and spatial relation trials, and identity and temporal relation trials, at processing. Statistical testing employed Fisher's *z*-tests, outputting *p*-values with positive *z*-scores if identity > relation PAC and negative *z*-scores if identity < relation PAC, equivalent to a two-tailed test. Similar to the baseline transformation, this testing method reveals condition-related changes in true PAC.

We computed inter-regional PAC between pairs of electrodes for all MTL-PFC and MTL-OFC electrode pairs within left and right hemispheres, respectively. For each pair, we applied the equivalent correlation procedure between the theta phase of the MTL (PFC/OFC) electrode and each frequency amplitude of the PFC/OFC (MTL) electrode. Then, we applied the equivalent baseline transformation to assess the significance of task-evoked changes in MTL (PFC/OFC) phase-PFC/OFC (MTL) amplitude coupling. We separately analyzed inter-regional PAC outputs for condition differences, as described above, and also for directionality by condition at each electrode pair (cf. Voytek et al., 2015). Statistical testing of directionality again employed Fisher's *z*-tests, outputting *p*-values with positive *z*-scores if MTL phase → PFC/OFC amplitude and negative *z*-scores if PFC/OFC phase → MTL amplitude. This testing method reveals effective connectivity as direction-related changes in true inter-regional PAC.

1.3.8. Region-of-Interest (ROI) PAC Analyses

We tested the spectrotemporal patterns of PAC and corresponding condition effects in non-overlapping, linear mixed effects models (LMEMs) of PAC data by region of interest (ROI). We selected baseline-standardized PAC by condition at encoding-maintenance and processing, respectively, down-sampled the data to 10-msec resolution, and pooled it across subjects. We cut data at amplitudes with center frequencies outside of the 9.5-128-Hz range common to all subjects, yielding sixteen frequencies for the LMEMs. We modeled the inter-regional theta PAC by directionality by subtracting PFC/OFC phase-MTL amplitude coupling *z*-scores from MTL phase-PFC/OFC amplitude coupling *z*-scores at each electrode pair. This procedure transformed

inter-regional PAC into indices of effective connectivity, with positive values if MTL → PFC/OFC and negative values if PFC/OFC → MTL.

For local PAC LMEMs, we considered all 35 MTL, 115 PFC, and 99 OFC electrodes as samples of the respective ROI and fit them as random effects. For inter-regional PAC LMEMs, all within-subject, within-hemisphere electrode pairs were considered samples of the respective ROI interaction; we fit MTL-PFC LMEMs with 111 electrode random effects and MTL-OFC LMEMs with 107 electrode random effects. In addition to electrodes, we fit the 10 subjects to each LMEM as random effects, and fit the 2 (3) conditions at encoding-maintenance (processing), 150 (90) time points at encoding-maintenance (processing), and 16 frequencies as fixed effects (fitlme.m). Results were Bonferroni-corrected for multiple comparisons (3 main + 4 interaction fixed effects; Cramer et al., 2015), yielding an adjusted alpha threshold: $p_{corr} = 0.05/7 \approx 0.0071$.

Finally, we replicated the LMEMs using input PAC data standardized on time-resolved surrogate distributions. All reported LMEM results were significant irrespective of whether raw PAC was standardized on the temporal mean of pretrial baseline window or on time-resolved surrogate distributions. We do not report outcomes that were supra-threshold using one input data type but not both. All reported whole-ROI outcomes were obtained using baseline-standardized PAC because it returned more conservative estimates of single-electrode PAC (Fig. 1.4).

1.3.9. Code

We analyzed all data in MATLAB (The MathWorks, Inc., Natick, MA) using custom scripts. Some electrophysiological data analyses adopted functions available in open-source MATLAB toolboxes: FieldTrip (Oostenvald et al., 2011), CircStat (Berens, 2009), and EEGLAB (Delorme & Makeig, 2004). The behavioral task was programmed in E-Prime Professional 2.0 (Psychology Software Tools, Pittsburgh, PA).

1.4. Supplementary Figures

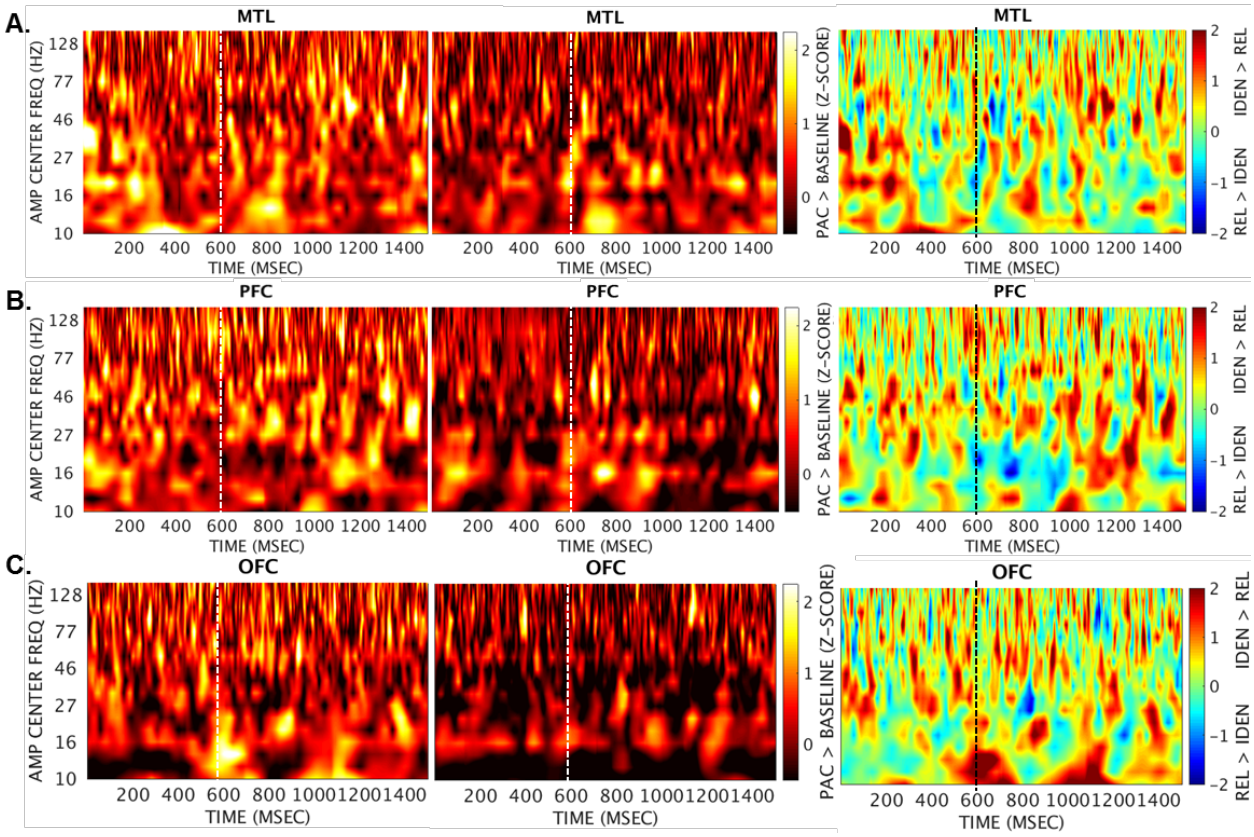

Fig. 1.S1. Local MTL (A), PFC (B), and OFC (C) PAC by episodic condition in single-ROI electrodes. Local PAC reached significance in bursts during the encoding-maintenance of identity (left) and relation (middle) information, $z > 1.96$, $p < 0.05$. While condition comparisons (right) might suggest more coupling for identity than relation information, this effect was not replicated using PAC data standardized on time-resolved surrogate distributions (Materials and Methods). Vertical dotted line = maintenance fixation start (Fig. 1.1A).

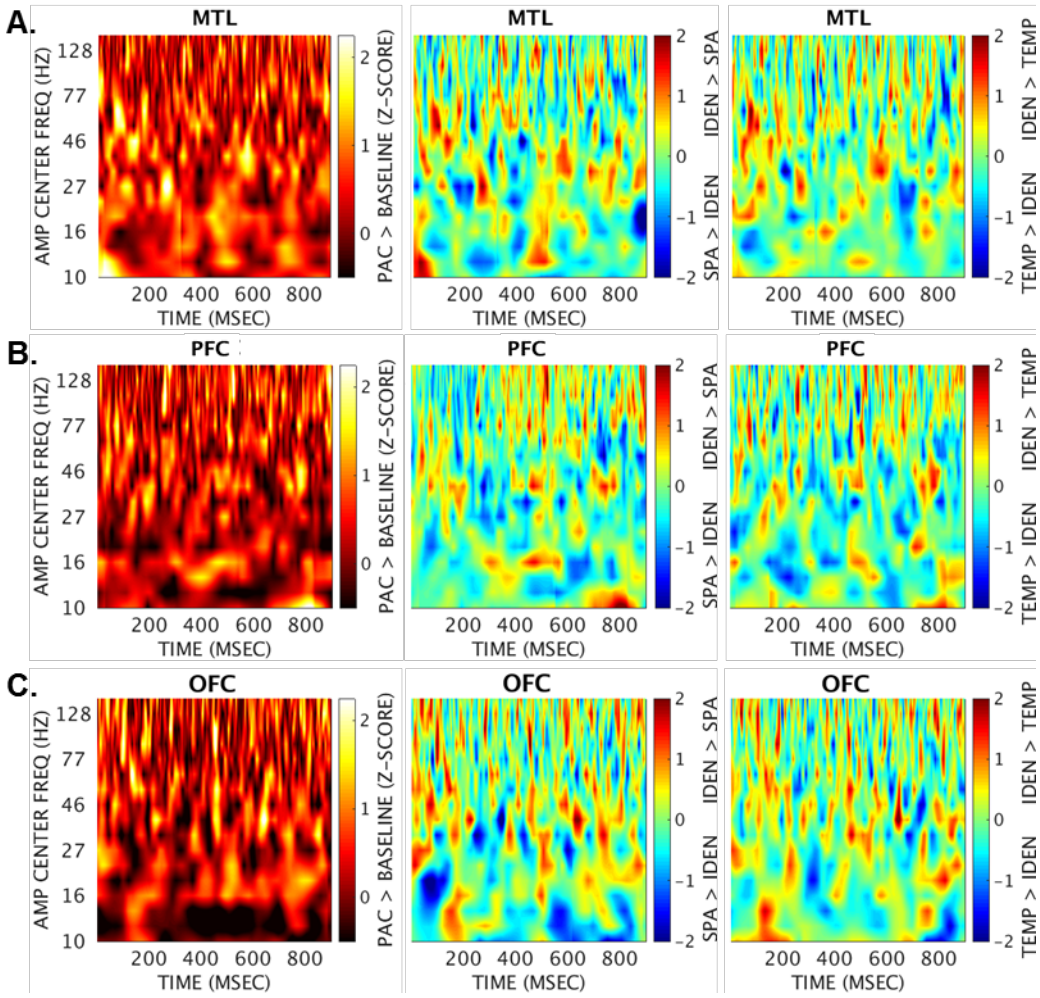

Fig. 1.S2. Local MTL (A), PFC (B), and OFC (C) PAC by episodic condition in single-ROI electrodes. The bursting pattern continued through processing, as shown for identity information (left), and in comparison to spatial relation (middle) and temporal relation (right) information. Condition tests revealed significant coupling for each condition at different spectrotemporal points, $|z| > 1.96$, $p < 0.05$. IDEN, identity; SPA, spatial relation; TEMP, temporal relation.

Ch. 2: Causal Evidence That Bidirectional Neural Rhythms Support Optimal Working Memory

Citation: Johnson, E. L., Dewar, C. D., Solbakk, A-K., Endestad, T., Meling, T. R. & Knight, R. T. Causal evidence that bidirectional neural rhythms support optimal working memory (in prep.).

We investigated the effects of unilateral prefrontal cortex (PFC) damage on the neural networks supporting the encoding, maintenance, and active processing of information in working memory (WM). In healthy controls, low theta activity in PFC increased with processing demands, concurrent with delta-theta directional connectivity from PFC to widespread parieto-occipital regions, as measured using the Phase Slope Index. These effects were compromised in patients. Low theta activity was attenuated over the lesion and the frontal-to-posterior directional rhythms were absent bilaterally. In contrast, alpha-beta rhythms originating in posterior regions preceded frontal rhythms, independent of PFC integrity and task demands. These findings reveal independent but complimentary systems, with one originating in PFC and the other in posterior regions. Performance was consistently optimal when bidirectional, multiplexed rhythms between frontal and parieto-occipital regions were recruited to support executive control. When PFC was compromised, the posterior network provided adequate cognitive resources for well above-chance WM performance.

2.1. Introduction

The ability to maintain and select pieces of information in working memory (WM) allows us to plan, solve problems, and adapt our underlying neural states over time. Building on traditional views that attribute WM function to the prefrontal cortex (PFC; Goldman-Rakic, 1995), distributed network accounts posit that WM is supported by long-range interactions coordinated through PFC-driven, top-down control (Miller & Cohen, 2001; Sreenivasan et al., 2014; Lara & Wallis, 2015; Postle, 2016; Eriksson et al., 2015; Lorenc et al., 2015). However, examination of individuals with PFC damage – which provides sufficient evidence to draw causal links between neuroanatomy and behavior – indicates that PFC plays an important, yet selective role in WM (Szczepanski & Knight, 2014). A meta-analysis of 166 PFC patients revealed that performance on strictly mnemonic tasks (e.g., maintaining information in WM) was unaffected by PFC lesions (D'Esposito & Postle, 1999). Instead, performance depended critically on whether a task imposed additional executive demands (e.g., selecting or manipulating information in WM; also Barbey et al., 2013; Petrides & Milner, 1982). Once executive demands were imposed, performance decrements were commensurate with the extent of frontal damage (Müller & Knight, 2006). Taken together, these findings suggest that PFC exerts control over information representations maintained across distributed regions, but only as necessitated by task demands.

Neural oscillations provide a plausible mechanism for PFC-guided network control over information in WM as task demands unfold in real time (Helfrich & Knight, 2016). Slow rhythmic synchrony between PFC and distal regions – including the hippocampus (Anderson et al., 2010; Brincat & Miller, 2015; Place et al., 2016), thalamus (Sweeney-Reed et al., 2014), and temporo-parietal regions (Watrous et al., 2013; Burke et al., 2013) – is consistently correlated with successful encoding and retrieval in memory (Johnson & Knight, 2015). Within PFC, focusing attentional resources within the context of multi-item WM was reflected in delta-theta

(1-7 Hz) and alpha band (8-16 Hz) activity in non-human primates (Lara & Wallis, 2014). Specific item representations, in contrast, have been found in posterior sensory regions (Postle, 2016; Ku et al., 2015; Galeano Weber et al., 2016; Bettencourt & Xu, 2016; Fallon et al., 2016), although recent studies suggest that information may be represented in higher cortical areas as well (Ester et al., 2016; Serences, 2016). Notably, cross-frequency coupling between frontal theta (~5 Hz) oscillations and parieto-occipital gamma band (>30 Hz) activity is enhanced during the encoding of visual information that is subsequently remembered (Friese et al., 2013).

We recorded the 64-channel electroencephalogram (EEG) in patients with PFC lesions to investigate the causal influence of PFC integrity on local and long-range rhythmic activity during the encoding, maintenance, and subsequent processing of information in WM. Fourteen unilateral PFC patients (Fig. 2.1A, 2.S1; 46 ± 16 years of age, 15 ± 3 years of education, 6 males) and 20 age- and education-matched, healthy controls completed a lateralized WM task that targeted the lesioned or intact hemisphere (Fig. 2.1B, Methods). On each trial, subjects encoded two shapes in specific spatiotemporal positions in preparation for a subsequent test on either the identity of each shape in the pair, or on the spatial or temporal relationship between the shapes in the pair. They were instructed to maintain central fixation because the shapes would be presented rapidly (200 msec each) on the left or right side of the screen (Duarte et al., 2005; Voytek et al., 2010a, 2010b). We confirmed the lateralized visual hemifield manipulation with eyegaze position data (Methods). Condition was designated at the start of each trial by a cue word, either IDENTITY or RELATION, and then again mid-delay by a test prompt – either SAME (identity), or TOP/BOTTOM (spatial relation) or FIRST/SECOND (temporal relation). The test prompt was presented mid-delay to impose executive demands at processing. This critical manipulation allowed us to look at how WM unfolded over time, first at encoding and maintenance, and then as the task demanded subjects to actively process information for a memory test.

We tested the hypothesis that theta rhythms would flexibility coordinate PFC influence over parieto-occipital regions per executive demands in the service of WM. However, behavioral outcomes demonstrated that, while patients were impaired relative to controls, they were able to do the task well above chance (Fig. 2.1C). This crucial result reveals that PFC does not play a unitary role in WM. Because all patients exhibited above-chance accuracy, we hypothesized that WM would be dually supported by a separate network that was not under top-down PFC control. To assess the impact of unilateral PFC lesions on whole-brain activity, we took a data-driven approach to first quantify the mechanisms underlying successful WM in controls and then investigate how these mechanisms were implemented in patients. All between-groups statistical testing employed a Monte Carlo method with 95% cluster-based maximum correction for multiple comparisons (Maris & Oostenveld, 2007).

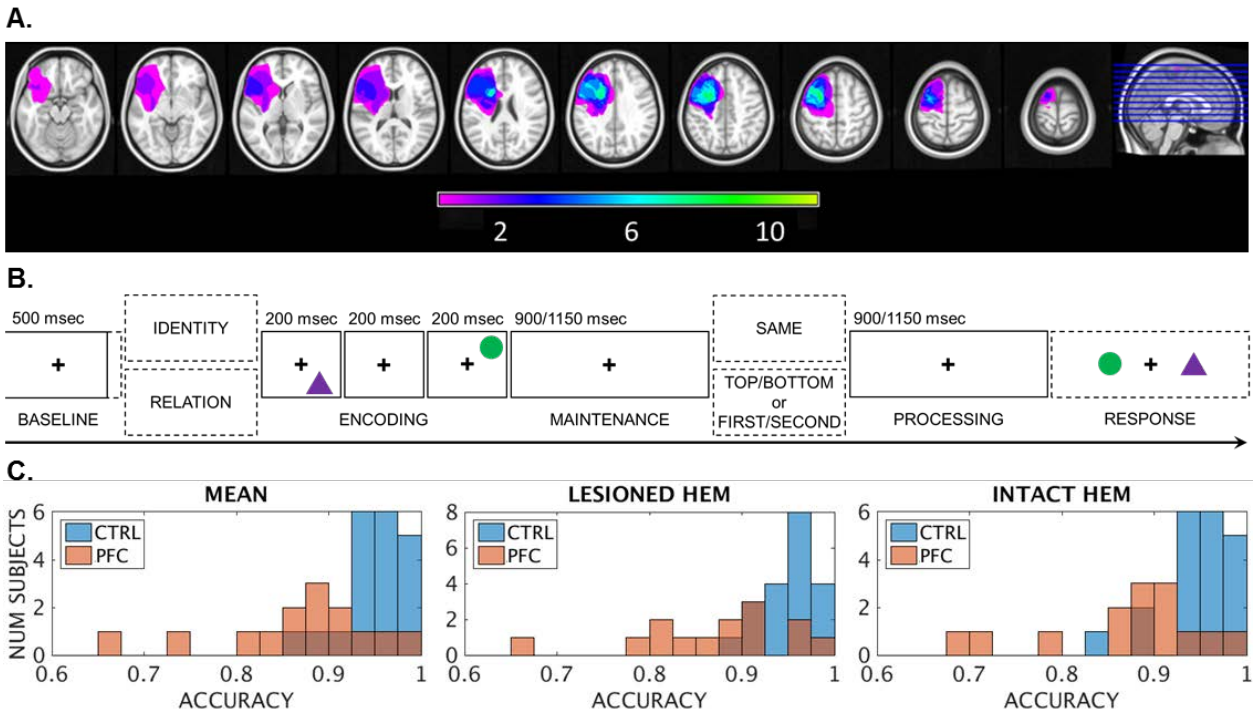

Fig. 2.1. Reconstruction of the extent of PFC lesion overlap for all 14 patients normalized to the left hemisphere (A), single-trial lateralized WM design (B), and distribution of performance by group (C). Color scale (A) = number of patients with lesions at the specified site. After a 2-sec pretrial fixation, subjects were cued to focus on either IDENTITY or RELATION information at encoding (B). Then, two shapes were presented rapidly to the left or right visual hemifield in one of two spatial positions and one of two temporal positions. After a 900- or 1150-msec maintenance fixation, the test prompt appeared, followed by a processing fixation of the same length. WM was tested in a two-alternative forced choice test, yielding a 0.5 chance rate. In the identity test, subjects indicated whether the pair was the SAME pair they just studied. In the spatiotemporal relation test, subjects indicated which shape fit the TOP/BOTTOM spatial or FIRST/SECOND temporal relation prompt. Patients exhibited impaired accuracy across conditions relative to controls (C), $p < 10\times10^{-5}$, irrespective of whether the stimuli were encoded in the lesioned or intact hemisphere, $p > 0.08$ (note: $p_{corr} \approx 0.0071$). HEM, hemisphere; CTRL, controls; PFC, PFC patients.

2.2. Results

2.2.1. Behavior

Accuracy was tested in logit mixed-effects models with group, visual hemifield, and condition as fixed effects, and subjects as random effects (Jaeger, 2008). Because accuracy did not differ between patients as a function of which hemisphere was lesioned ($p > 0.34$), we swapped right-hemisphere lesioned patient data across the midline so that all lesions were normalized to the left hemisphere (right visual hemifield) to comprise a single patient group (Methods). Then, we randomly swapped the data for half of the controls ($n = 20/2 = 10$) to preclude any inter-hemispheric variation from confounding lesion-related outcomes. Results revealed that patients

were impaired at the task relative to controls (Fig. 2.1C; 0.87 ± 0.08 versus 0.95 ± 0.03, $p < 10\times10^{-5}$). There were no differences between conditions or visual hemifields (condition $p > 0.31$, group*field $p > 0.08$; $p_{corr} \approx 0.0071$; Cramer et al., 2015), permitting analyses of data pooled across conditions and further suggesting that unilateral PFC lesions have a bilateral influence on the neural networks supporting WM. We analyzed EEG data for all correct trials by visual field and group, collapsed across identity, spatial relation, and temporal relation conditions (Methods).

2.2.2. Event-Related Potentials

Analysis of event-related potentials (ERPs) confirmed the bilateral and network-scale influence of unilateral PFC lesions. We bandpass-filtered the EEG data between 1-30 Hz, divided the traces into time windows for a 500-msec pretrial baseline, 1500-msec encoding-maintenance period, and 900-msec processing period, and then corrected encoding-maintenance and processing data on the pretrial baseline (Fig. 2.1B, Methods). Patients showed attenuated positive-polarity ERPs in parieto-occipital channels when stimuli were presented to the visual hemifield contralateral to PFC damage. Effects were bilateral (Fig. 2.2), and reached significance in the intact hemisphere ~200-300 msec after the offset of the second stimulus (cluster $p = 0.03$; cf. Barceló et al., 2000; Yago et al., 2004). Importantly, the effect did not persist when stimuli were presented to the visual hemifield contralateral to the intact hemisphere (cluster $p > 0.33$), revealing that unilateral PFC lesions impacted long-range networks on a trial-wise basis, depending on which hemisphere was taxed.

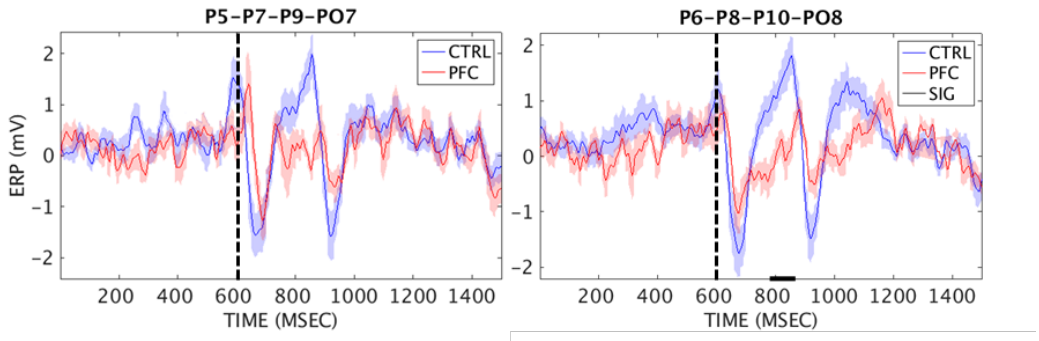

Fig. 2.2. Mean ERPs over encoding-maintenance in select parieto-occipital channels by hemisphere and group. Patients exhibited diminished positive-polarity ERPs relative to controls early in the maintenance period, bilaterally. The effect reached significance in a cluster of channels (P6-P8-P10-PO8) in the intact hemisphere (cluster $p = 0.03$, marked in black on the x-axis [right]). Exact data reflect outcomes when stimuli were encoded in the lesioned hemisphere. Vertical dotted line = maintenance fixation start (Fig. 2.1B). CTRL, controls; PFC, PFC patients; SIG, significant.

2.2.3. Spectrotemporal Power

Spectral decomposition was computed using a Hanning taper in evenly spaced frequency bins between 1-40 Hz, and then encoding-maintenance and processing data for all correct trials were standardized on the pretrial baseline (Methods). Single-subject analyses of task-evoked spectrotemporal power revealed that successful WM delay activity was linked to two causally

dissociable, frequency-dependent mechanisms. Visual processing was marked by increased parieto-occipital beta-gamma (12-35 Hz) activity with narrowband alpha desynchronization during encoding, followed by widespread alpha-beta (9-24 Hz) decreases throughout maintenance and processing (Fig. 2.3; $|z| > 1.96$, $p < 0.05$). These patterns did not differ by group, indicating that sustained alpha-beta suppression for WM was independent of PFC, consistent with sources in more posterior regions (Hillebrand et al., 2016; Xie et al., 2016). Spectrotemporal power did not differ by visual hemifield presentation.

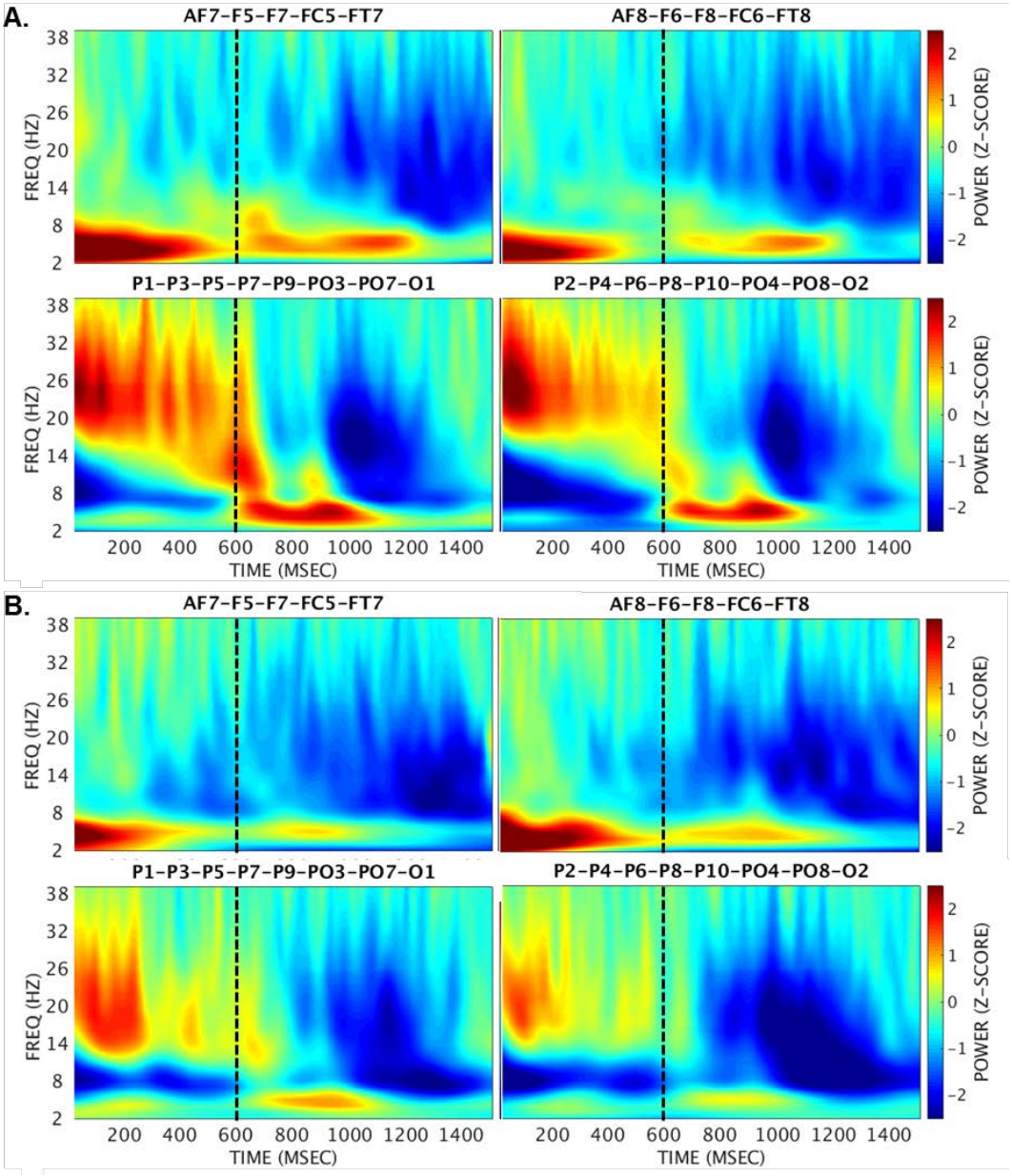

Fig. 2.3. Mean task-evoked spectrotemporal power over encoding-maintenance for controls (A) and PFC patients (B) in lateral frontal (top) and parieto-occipital (bottom) regions. Single-subject analyses revealed anterior and posterior theta increases, and posterior beta-gamma (12-35 Hz) increases at encoding followed by a decrease at maintenance. A narrowband alpha decrease was

observed in posterior channels throughout the encoding-maintenance period. While parieto-occipital power appears diminished in PFC patients relative to controls, contrasts did not survive statistical testing ($z_{diff} < 1.96$, $p > 0.05$; cluster $p = 1$). Exact data reflect outcomes when stimuli were encoded in the lesioned hemisphere. Vertical dotted line = maintenance fixation start (Fig. 2.1B).

In contrast, encoding-maintenance (Fig. 2.1B) was marked by anterior delta-theta (2-7 Hz) activity, which increased and then remained elevated through processing (Fig. 4A-B; $z > 3.29$, $p < 0.001$). Patients did not show the same increase after executive demands were imposed. The hemispheric asymmetry test revealed a cluster centered in low theta (3-4 Hz) that spanned ~0-700 msec of the processing period (Fig. 4C; cluster $p \leq 0.04$). Patients exhibited less of an increase in channels proximal to the lesion site than in the contralateral channels, highlighting a PFC origin for the slow rhythmic substrate of executive control. The effect was significant irrespective of whether stimuli were encoded in the lesioned or intact hemisphere, revealing bilateral low theta for executive control in PFC.

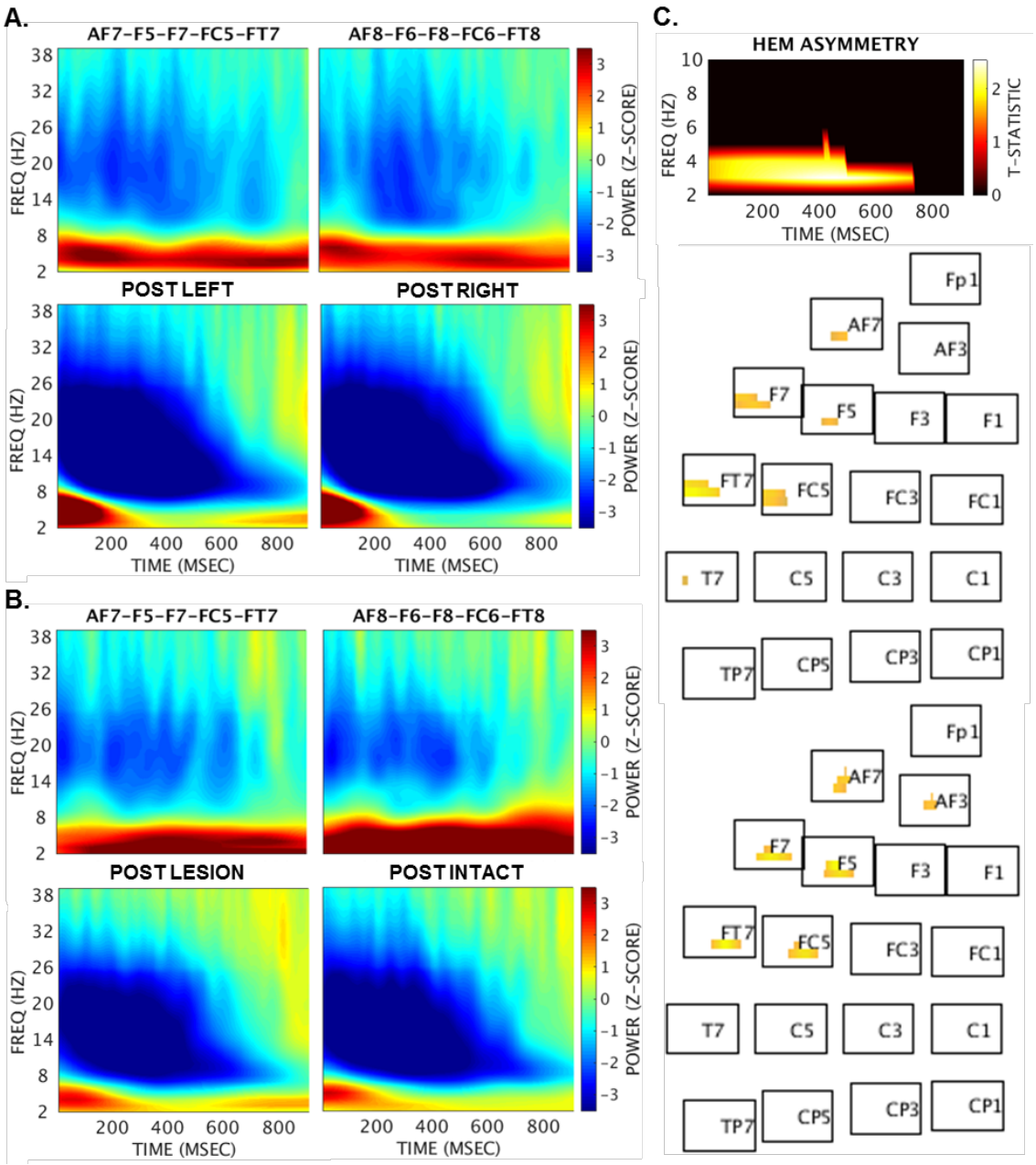

Fig. 2.4. Mean task-evoked spectrotemporal power over processing for controls (A) and PFC patients (B) in lateral frontal (top) and parieto-occipital (bottom) regions, and the anterior hemispheric asymmetry effect (C). Single-subject analyses revealed sustained delta-theta (2-7 Hz) in anterior regions and widespread alpha-beta (9-24 Hz) power suppression. Exact data reflect outcomes when stimuli were encoded in the lesioned hemisphere (A-B). Patients showed less of an increase in low theta (3-4 Hz) activity proximal to the lesion site than in contralateral channels (C), cluster $p \leq 0.04$ (masked), uncovering a PFC origin for low theta rhythms underlying executive control in WM. The hemispheric asymmetry effect (top) reflects the mean *t*-statistics collapsed across visual field presentations for the cluster of channels (AF7-F5-F7-FC5-FT7) with overlapping outcomes when stimuli were presented to the right (middle) versus

left visual field (bottom). POST LEFT/LESION, channels P1-P3-P5-P7-P9-PO3-PO7-O1; POST RIGHT/INTACT, channels P2-P4-P6-P8-P10-PO4-PO8-O2; HEM, hemisphere.

2.2.4. Directional Connectivity

Next, we investigated the temporal dynamics of directional connectivity between regions at encoding, maintenance, and processing using the Phase Slope Index (PSI; Nolte et al., 2008). The PSI tracks whether the slope of the phase lag between A-B channel pairs exhibits consistent polarity across several adjacent frequency bins; positive PSI indicates that channel A → channel B, while negative PSI indicates the reverse, and zero PSI indicates either zero or an evenly balanced lead/lag relationship between A and B. PSI was quantified separately for the delta-theta (2-7 Hz) and alpha-beta (9-24 Hz) ranges based on the results of the spectrotemporal power analysis (Methods). If PFC signals directed visual information per executive demands, then the diminished PFC low theta activity observed at processing should also cause diminished PFC → parieto-occipital PSI in patients relative to controls. Note that alpha-beta PSI was purported to be elevated in patients reflecting neuroplasticity to compensate for attenuated executive control (cf. Voytek et al., 2010a).

Task-evoked delta-theta rhythms precessed from parieto-occipital to frontal regions at the offset of the second stimulus in both groups (Fig. 2.5A; $z \leq -10$, $p < 2\times10^{-23}$), suggesting the end of bottom-up information transfer at encoding in that frequency range. Then, time-resolved connectivity patterns revealed a shift in directionality so that frontal regions led parieto-occipital regions mid-maintenance, but only in controls ($z > 1.96$, $p < 0.05$). Additional group differences emerged at various points early in encoding (Fig. 2.5B) and over the course of the maintenance period (Fig. 2.5C), so that even when control PSI was not supra-threshold in the PFC-led direction, controls still displayed greater PFC leads than patients (all cluster $p < 0.05$). Moreover, graph theoretical network analysis (Methods) revealed an absence of PFC-led delta-theta PSI in patients throughout the task. We observed this network effect bilaterally and regardless of whether stimuli had been presented to the lesioned or intact hemisphere, demonstrating that unilateral PFC lesions compromised the entire PFC-sourced network supporting executive control. Group differences were most pronounced mid-maintenance when stimuli were presented to the intact hemisphere (Fig. 2.5C, bottom; all cluster $p < 0.05$), confirming that the entire delta-theta network was attenuated by unilateral PFC lesions.

Delta-theta PSI continued to increase in controls with task demands such that the PFC lead peaked early in processing (Fig. 2.5D, $z \geq 10$, $p < 2\times10^{-23}$), while patient PSI remained at zero. Cluster-based statistics confirmed the between-groups differences across channels and in network-wide patterns (Fig. 2.5E, all cluster $p < 0.05$), which were strongest at 100-200 msec and persisted over ~0-700 msec of the processing period. These differences occurred irrespective of which hemisphere had previously encoded the stimuli so that the hemispheric differences which emerged during the maintenance period (Fig. 2.5C) decreased by processing. Network-wide outcomes were apparent in bilateral central-posterior regions, revealing causal evidence that PFC slow rhythms influenced widespread poster sensory and nearby regions commensurate with executive demand. In contrast, alpha-beta rhythms from parieto-occipital regions led PFC throughout encoding, maintenance, and processing, and were invariant to executive demand (Fig. 2.6). Furthermore, PFC lesions did not affect alpha-beta PSI at any point during the encoding or

delay periods, corroborating an independent, more posterior source for alpha-beta rhythms during WM.

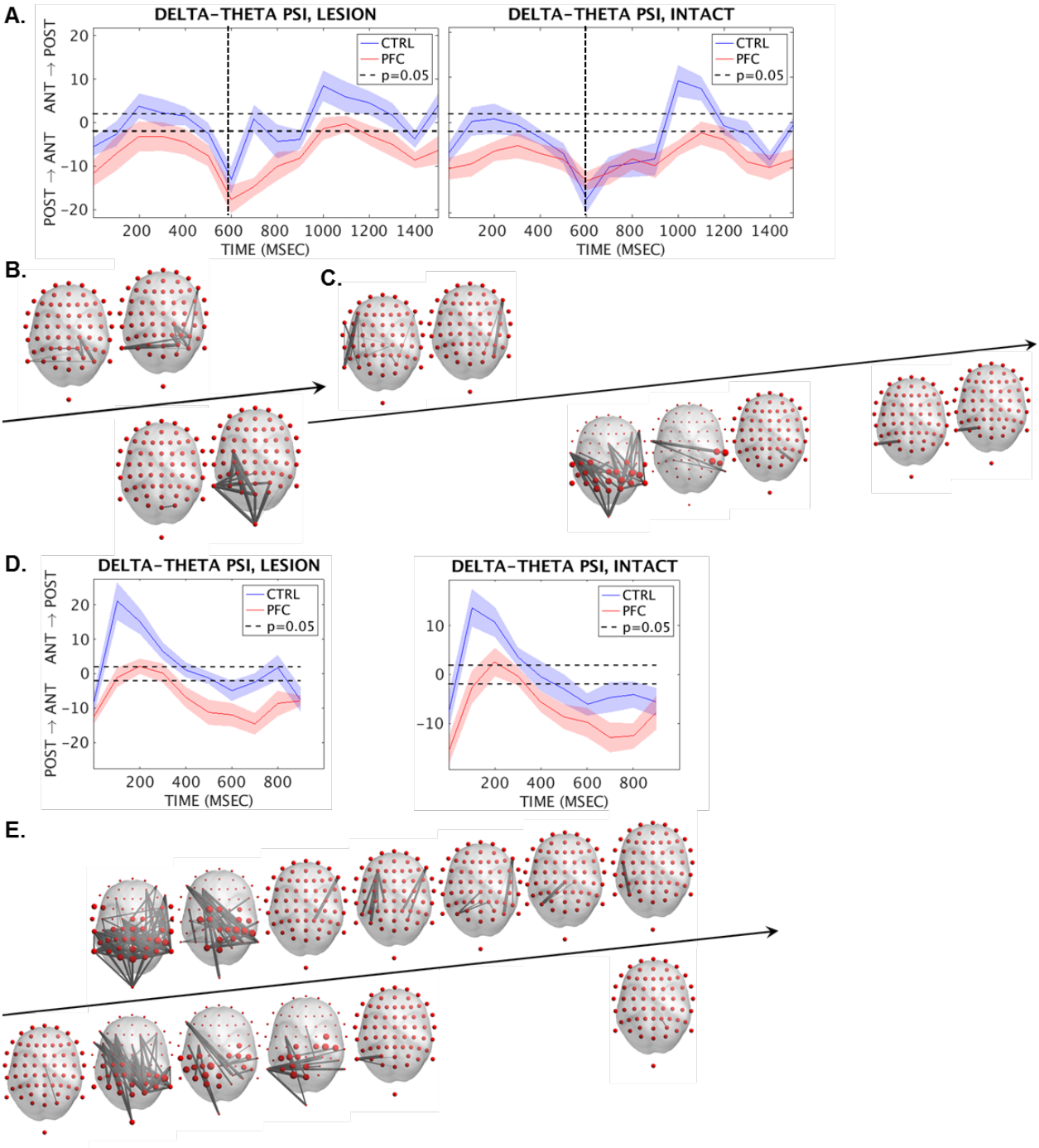

Fig. 2.5. Mean task-evoked delta-theta (2-7 Hz) PSI at encoding-maintenance (A) and processing (D) as a function of visual hemifield, and the results of between-groups statistical testing. Parieto-occipital regions led frontal regions at the end of encoding in both groups ($z \leq -10$, $p < 2\times10^{-23}$), followed by a shift in directionality at mid-maintenance specific to controls (A), irrespective of visual hemifield ($z > 1.96$, $p < 0.05$). Channel-wise comparisons revealed less

delta-theta lead in patients 100-200 msec into encoding (B) when stimuli were presented to the lesioned hemisphere (top) and 200-300 msec into encoding when stimuli were presented to the intact hemisphere (bottom). Similar effects were observed at maintenance (C), with group differences appearing earlier when stimuli were presented to the lesioned hemisphere (at 700-800 msec [top] versus 1000-1200 and 1400-1500 msec [bottom]). At processing, the PFC lead increased in controls ($z \geq 10$, $p < 2\times10^{-23}$), while patients continued to exhibit no PFC → parieto-occipital PSI (D). Channel-wise comparisons revealed less delta-theta lead in patients throughout the first 700 msec of processing, irrespective of hemifield of presentation at encoding, with the most pronounced effects in anterior-to-posterior PSI at 100-200 msec (E). Statistics topographies (B-C, E) = PSI (between-channel lines) and network (red circles) *t*-statistics scaled proportionally to the outcome magnitude per 100-msec time point. ANT LESION, channels AF7-F5-F7-FC5-FT7; ANT INTACT, channels AF8-F6-F8-FC6-FT8; POST LESION, channels P1-P3-P5-P7-P9-PO3-PO7-O1; POST INTACT, channels P2-P4-P6-P8-P10-PO4-PO8-O2; CTRL, controls.

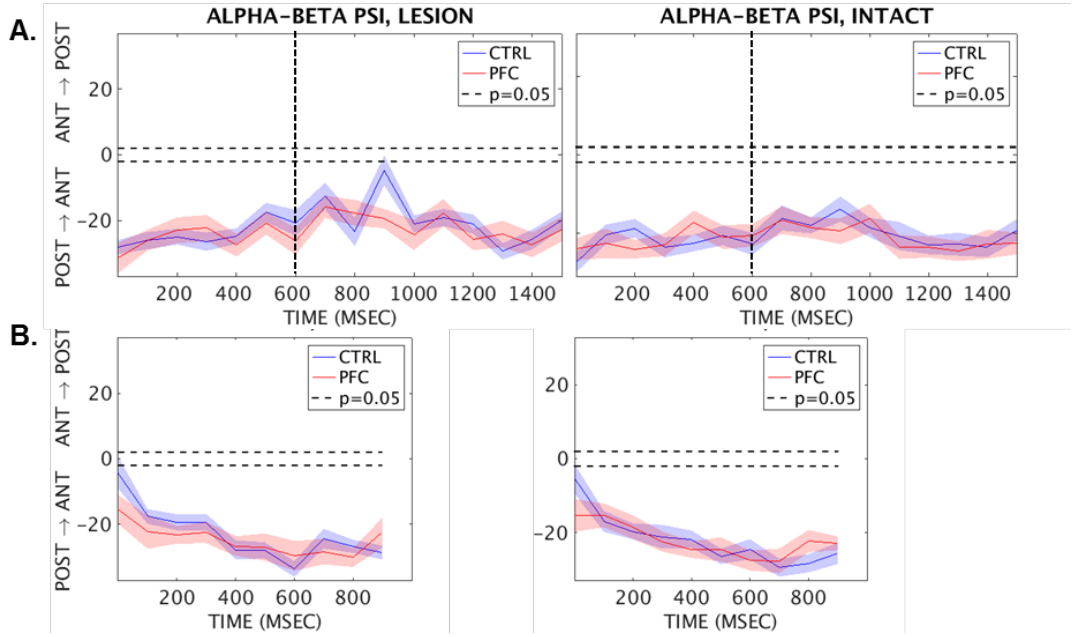

Fig. 2.6. Mean task-evoked alpha-beta (9-24 Hz) PSI at encoding-maintenance (A) and processing (B) as a function of visual hemifield. Parieto-occipital regions led frontal regions irrespective of executive demands ($z < –1.96$, $p < 0.05$). No differences were observed as a function of visual hemifield or group. Vertical dotted line = maintenance fixation start (Fig. 2.1B). ANT LESION, channels AF7-F5-F7-FC5-FT7; ANT INTACT, channels AF8-F6-F8-FC6-FT8; POST LESION, channels P1-P3-P5-P7-P9-PO3-PO7-O1; POST INTACT, channels P2-P4-P6-P8-P10-PO4-PO8-O2; CTRL, controls.

2.3. Discussion

We examined the causal influence of PFC integrity on visuospatial WM ability and the oscillatory substrates in real time as a function of executive demand. Results uncover two long-range neural networks with dissociable neuroanatomical and electrophysiological profiles that

together govern optimal WM performance. Low theta activity was attenuated at the lesion site following presentation of the test prompt, concurrent with diminished frontal-to-posterior directional connectivity in the delta-theta range in patients. These findings extend correlative evidence using electrophysiological recordings (Raghavachari et al., 2013; Güntekin & Başar, 2015), and provide the first causal demonstration that the slow rhythmic substrate of executive control is dependent on PFC – and these oscillations dynamically influence parieto-occipital regions depending on whether information was being maintained or actively processed in WM. While these results are consistent with models of PFC-guided network function during WM (Sreenivasan et al., 2014; Lara & Wallis, 2015; Postle, 2016; Eriksson et al., 2015), our behavioral data reveal that PFC-guided network function is not necessary for WM unless additional executive demands are imposed for active processing.

Instead, our results provide causal evidence that alpha-beta rhythms originating in central-posterior regions supported above-chance WM performance. The widespread decreases in power observed over the maintenance and processing periods may reflect frontoparietal control and/or recruitment of the dorsal attention network (Sadaghiani & Kleinschmidt, 2016; Xie et al., 2016). The notion that such domain-general physiological signatures underpin WM function is not new (Postle, 2006), but their mechanistic interplay with PFC function remains controversial. We show that alpha-beta rhythms precessed from parieto-occipital to frontal regions (Hillebrand et al., 2016). Critically, we observed that this network was unaffected by task demands or PFC lesions, and showed no signs of compensatory neuroplasticity, inconsistent with prior accounts of PFC lesion-related, within-region outcomes (e.g., Voytek et al., 2010a). The observed lack of interplay with PFC is in accord with proposals that parieto-occipital alpha-beta oscillatory activity is a substrate for the purely mnemonic component of WM (cf. D'Esposito & Postle, 1999; Postle et al., 1999; Galeano Weber et al. 2016).

Taken together, the results support our hypothesis that dual-networks govern WM – and that one network works independently of PFC. The findings demonstrate that delta-theta and alpha-beta rhythms comprise divergent substrates of WM function. While the central-posterior alpha-beta network is adequate for well above-chance performance, optimal WM depends on the additional recruitment of a dynamic slow rhythm driving executive control from PFC.

2.4. References

1. Anderson, K. L., Rajagovindan, R., Ghacibeh, G. A., Meador, K. J. & Ding, M. Theta oscillations mediate interaction between prefrontal cortex and medial temporal lobe in human memory. *Cereb. Cortex* 20, 1604–1612 (2010).
2. Bettencourt, K. C. & Xu, Y. Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nat. Neurosci.* 19, 150–157 (2015).
3. Barbey, A.K., Koenigs, M. & Grafman, J. Dorsolateral prefrontal contributions to human working memory. *Cortex* 49, 1195–1205 (2013).
4. Barceló, F., Suwazono, S. & Knight, R. T. Prefrontal modulation of visual processing in humans. *Nat. Neurosci.* 3, 399–403 (2000).
5. Brincat, S. L. & Miller, E. K. Frequency-specific hippocampal-prefrontal interactions during associative learning. *Nat. Neurosci.* 18, 1–10 (2015).

6. Burke, J. F. et al. Synchronous and asynchronous theta and gamma activity during episodic memory formation. *J. Neurosci.* 33, 292–304 (2013).
7. Cohen, M. X. *Analyzing Neural Time Series Data: Theory and Practice*. MIT Press (2014).
8. Cohen, M. X. Comparison of different spatial transformations applied to EEG data: A case study of error processing. *Int. J. Psychophysiol.* 97, 245–257 (2015).
9. Cramer, A. O. J. et al. Hidden Multiplicity in Multiway ANOVA: Prevalence, Consequences, and Remedies. *Psychol. Bull. Rev.* (2015). doi:10.3758/s13423-015-0913-5
10. D'Esposito, M. & Postle, B. R. The dependence of span and delayed-response performance on prefrontal cortex. *Neuropsychologia* 37, 1303–1315 (1999).
11. Delorme, A. & Makeig, S. EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21 (2004).
12. Duarte, A., Ranganath, C. & Knight, R. T. Effects of Unilateral Prefrontal Lesions on Familiarity, Recollection, and Source Memory. *J. Neurosci.* 25, 8333–8337 (2005).
13. Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F. & Nyberg, L. Neurocognitive Architecture of Working Memory. *Neuron* 88, 33–46 (2015).
14. Ester, E. F., Sutterer, D. W., Serences, J. T. & Awh, E. Feature-Selective Attentional Modulations in Human Frontoparietal Cortex. *J. Neurosci.* 36, 8188–8199 (2016).
15. Fallon, S. J., Zokaei, N. & Husain, M. Causes and consequences of limitations in visual working memory. *Ann. N. Y. Acad. Sci.* 1369, 40–54 (2016).
16. Flinker, A. et al. Redefining the role of Broca's area in speech. *PNAS* 112, 2871–5 (2015).
17. Friese, U. et al. Successful memory encoding is associated with increased cross-frequency coupling between frontal theta and posterior gamma oscillations in human scalp-recorded EEG. *NeuroImage* 66, 642–647 (2013).
18. Galeano Weber, E. M., Hahn, T., Hilger, K. & Fiebach, C. J. Distributed patterns of occipito-parietal functional connectivity predict the precision of visual working memory. *NeuroImage* (2016). doi:10.1016/j.neuroimage.2016.10.006
19. Goldman-Rakic, P. S. Cellular basis of working memory. *Neuron* 14, 477–485 (1995).
20. Gómez-Herrero, G. Automatic artifact removal (AAR) toolbox v1. 3 (Release 09.12. 2007) for MATLAB. Tampere University of Technology (2007).
21. Güntekin, B. & Başar, E. Review of evoked and event-related delta responses in the human brain. *Int. J. Psychophysiol.* (2015). doi:10.1016/j.ijpsycho.2015.02.001
22. Helfrich, R. F. & Knight, R. T. Oscillatory Dynamics of Prefrontal Cognitive Control. *Trends Cogn. Sci.* (2016). doi:10.1016/j.tics.2016.09.007
23. Hillebrand, A. et al. Direction of information flow in large-scale resting-state networks is frequency-dependent. *PNAS* 113, 3867–3872 (2016).
24. Hipp, J. F. & Siegel, M. Dissociating neuronal gamma-band activity from cranial and ocular muscle activity in EEG. *Front. Hum. Neurosci.* 7, 338 (2013).
25. Islam, M. K., Rastegarnia, A. & Yang, Z. Methods for artifact detection and removal from scalp EEG: A review. *Clin. Neurophysiol.* (2016). doi:10.1016/j.neucli.2016.07.002
26. Jaeger, T. F. Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *J. Mem. Lang.* 59, 434–446 (2008).

27. Johnson, E. L. & Knight, R. T. Intracranial recordings and human memory. *Curr. Opin. Neurobiol.* 31, 18–25 (2015).
28. Kayser, J. & Tenke, C. E. Issues and considerations for using the scalp surface Laplacian in EEG/ERP research: A tutorial review. *Int. J. Psychophysiol.* 97, 189–209 (2015).
29. Ku, Y., Bodner, M. & Zhou, Y. Di. Prefrontal cortex and sensory cortices during working memory: quantity and quality. *Neurosci. Bull.* 31, 175–182 (2015).
30. Lara, A. H. & Wallis, J. D. Executive control processes underlying multi-item working memory. *Nat. Neurosci.* 17, 876–883 (2014).
31. Lara, A. H. & Wallis, J. D. The role of prefrontal cortex in working memory: a mini review. *Front. Syst. Neurosci.* 9 (2015).
32. Lorenc, E. S., Lee, T. G., Chen, A. J.-W. & D'Esposito, M. The effect of disruption of prefrontal cortical function with transcranial magnetic stimulation on visual working memory. *Front. Syst. Neurosci.* 9 (2015).
33. Maris, E. & Oostenveld, R. Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods* 164, 177–190 (2007).
34. Miller, E. K. & Cohen, J. D. An Integrative Theory of Prefrontal Cortex Function. *Annu. Rev. Neurosci.* 24, 167–202 (2001).
35. Müller, N. G. & Knight, R. T. The functional neuroanatomy of working memory: Contributions of human brain lesion studies. *Neuroscience* 139, 51–58 (2006).
36. Nolte, G. et al. Robustly estimating the flow direction of information in complex physical systems. *Phys. Rev. Lett.* 100, 1–4 (2008).
37. Oostenveld, R., Fries, P., Maris, E. & Schoffelen, J. M. FieldTrip: Open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Comput. Intell. Neurosci.* 2011, (2011).
38. Perrin, F., Pernier, J., Bertrand, O. & Echallier, J. F. Spherical splines for scalp potential and current density mapping. *Electroencephalogr. Clin. Neurophysiol.* 72, 184–187 (1989).
39. Petrides, M. & Milner, B. Deficits on subject-ordered tasks after frontal- and temporal-lobe lesions in man. *Neuropsychologia* 20, 249–262 (1982).
40. Place, R., Farovik, A., Brockmann, M. & Eichenbaum, H. Bidirectional prefrontal-hippocampal interactions support context-guided memory. *Nat. Neurosci.* 19, (2016).
41. Postle, B. R. How does the brain keep information 'in mind'? *Curr. Dir. Psychol. Sci.* (2016). doi:10.1177/0963721416643063
42. Postle, B. R. Working memory as an emergent property of the mind and brain. *Neuroscience* 139, 23–38 (2006).
43. Postle, B. R., Berger, J. S. & D'Esposito, M. Functional neuroanatomical double dissociation of mnemonic and executive control processes contributing to working memory performance. *PNAS* 96, 12959–12964 (1999).
44. Raghavachari, S. et al. Theta Oscillations in Human Cortex During a Working-Memory Task: Evidence for Local Generators. *J. Neurophysiol.* 95, 1630–1638 (2013).
45. Sadaghiani, S. & Kleinschmidt, A. Brain Networks and alpha-Oscillations: Structural and Functional Foundations of Cognitive Control. *Trends Cogn. Sci.* (2016). doi:10.1016/j.tics.2016.09.004
46. Serences, J. T. Neural mechanisms of information storage in visual short-term memory. *Vision Res.* 128, 53–67 (2016).

47. Sreenivasan, K. K., Curtis, C. E. & D'Esposito, M. Revisiting the role of persistent neural activity during working memory. *Trends Cogn. Sci.* 18, 82–89 (2014).
48. Sweeney-Reed, C. M. et al. Corticothalamic phase synchrony and cross-frequency coupling predict human memory formation. *eLife* 3, e05352 (2014).
49. Szczepanski, S. M. & Knight, R. T. Insights into Human Behavior from Lesions to the Prefrontal Cortex. *Neuron* 83, 1002–1018 (2014).
50. Voytek, B. et al. Dynamic Neuroplasticity after Human Prefrontal Cortex Damage. *Neuron* 68, 401–408 (2010a).
51. Voytek, B. & Knight, R. T. Prefrontal cortex and basal ganglia contributions to visual working memory. *PNAS* 107, 18167–18172 (2010b).
52. Watrous, A. J., Tandon, N., Conner, C. R., Pieters, T. & Ekstrom, A. D. Frequency-specific network connectivity increases underlie accurate spatiotemporal memory retrieval. *Nat. Neurosci.* 16, 349–356 (2013).
53. Xia, M., Wang, J. & He, Y. BrainNet Viewer: A Network Visualization Tool for Human Brain Connectomics. *PLoS One* 8 (2013).
54. Xie, Y., Feng, Z., Xu, Y., Bian, C. & Li, M. The different oscillation patterns of alpha band in the early and later stages of working memory maintenance. *Neurosci. Lett.* 633, 220–226 (2016).
55. Yago, E., Duarte, A., Wong, T., Barceló, F. & Knight, R. T. Temporal kinetics of prefrontal modulation of the extrastriate cortex during visual attention. *Cogn. Affect. Behav. Neurosci.* 4, 609–617 (2004).

Acknowledgements: We thank D. Scabini for coordinating all patient testing efforts in Berkeley. We also thank J. Lubell for assistance testing patients in Oslo, A. Jafarpour for assistance writing code to process the Eyelink 1000 eyetracker data, and V. Piai, F. Foo, R. F. Helfrich, and S. M. Szczepanski for helpful discussions. This work was supported by grants from the National Institutes of Health (2R37NS21135 to R.T.K.), Nielsen Corporation (to R.T.K.), Research Council of Norway (240389/F20 to A.K.S., T.E., and T.R.M.), and University of Oslo Internal Fund (to A.K.S. and T.E.).

2.5. Methods

2.5.1. Subjects

We report data from 14 patients with unilateral prefrontal cortex (PFC) lesions (Fig. 1A and S1; 46 ± 16 years of age, 15 ± 3 years of education, 43% male) and 20 age- and education-matched, healthy controls (44 ± 19 years of age, 16 ± 3 years of education, 55% male). Patients had normal vision, estimated IQ in the normal range, and no other neurological or psychiatric diagnoses. Each patient was examined by a neurologist (R.T.K.) or neurosurgeon (T.R.M.) prior to testing, and final eligibility was determined through review of each patient's structural magnetic resonance imaging (MRI) scans. Patients were included by lesion focus in unilateral inferior, middle, and/or superior frontal gyrus. Half presented with left-hemisphere lesions and half presented with right-hemisphere lesions (n = 7 each). Injury was in the chronic phase (8 ± 6 years since incident or surgery). Independent-samples t-tests assuming unequal variance confirmed that the control group matched on demographics (age p = 0.72, education p = 0.24). Subjects were tested at one of two sites: University of California, Berkeley (UCB; five patients

with lesions due to stroke and all controls), or Oslo University Hospital (OUH; nine patients with lesions due to tumor resection). All subjects gave informed written consent in accordance with the respective institutional review board.

2.5.2. Experimental Design

We tested working memory (WM) in a single-trial, lateralized, visuospatial task paradigm (Fig. 2.1B). After a 2-sec pretrial fixation, a starting screen indicated whether the following pair of stimuli would be tested for IDENTITY or spatiotemporal RELATION information. Then, following a 100-msec central fixation, two colored shapes were presented for 200 msec each in one of two vertical spatial positions and one of two temporal positions. Each shape was presented to the left or right of a central fixation cross to target the contralateral hemisphere (Duarte et al., 2005; Voytek et al., 2010a, 2010b). After a 900- or 1150-msec maintenance period, a test prompt appeared indicating what type of information would be tested. The test prompt was presented mid-delay to elicit executive control mechanisms during a processing period of the same length. Then, two shapes were presented full field on the horizontal axis and subjects responded in a two-alternative forced choice test, yielding a 0.5 chance rate. In the identity test, subjects indicated whether the pair was the SAME pair they just studied; half of the pairs show two old shapes ("yes") and half the pairs show one old shape and one new shape ("no"). In the spatial relation test, subjects indicated which shape had been on the TOP or BOTTOM, and in the temporal relation test, which shape had been presented FIRST or SECOND.

The length of maintenance and processing delay periods was randomly jittered at 900- or 1150-msec to preclude anticipatory or in-phase mechanisms. The task was fully counterbalanced with 240 trials split evenly between the left and right visual hemifields, and then sub-split evenly into identity, spatial, and temporal conditions, chosen randomly from a pool of 250 trials. No stimuli were repeated across trials. An experimenter went through the task instructions and a set of six practice trials with each subject, who was permitted to repeat the practice trials by request.

2.5.3. Data Acquisition

Subjects were tested in a sound-attenuated recording room. The electroencephalogram (EEG) was collected using a 64 + 8 channel BioSemi ActiveTwo amplifier (BioSemi, Amsterdam, Netherlands) sampled at 1024 Hz. The horizontal electrooculogram (EOG) was recorded at both external canthi, and the vertical EOG was monitored with a right inferior eye electrode and a superior eye/frontopolar electrode. Electrode impedances were kept below 20 kΩ. All EEG data were re-referenced offline to the mean potential of two earlobe electrodes (pop_biosig.m), resampled to 256 Hz (pop_resample.m), and then preprocessed through a standard pipeline.

Continuous eyegaze positions were recorded to exclude any trials *post-hoc* in which stimuli had been encoded in the ipsilateral hemifield. UCB eyetracking data were collected using an Eyelink 1000 optical tracker (SR Research, Ontario, Canada) sampled at 1 kHz and OUH eyetracking data were collected using an iView X optical tracker (SMI, Teltow, Germany) sampled at 60 Hz. Subjects' head movements were restrained using a custom wooden chin rest to minimize contamination of anterior-channel EEG traces.

2.5.4. Behavioral Data Analysis

First, we established that patients with left and right PFC damage did not differ in task accuracy. We excluded any trials marked based on eyegaze position, and then calculated the mean accuracy per patient by visual hemifield (right or left) and condition (identity, spatial relation, or temporal relation). Accuracy data were submitted to a logit mixed-effects model (fitglme.m) with lesion hemisphere, visual hemifield, and condition as the fixed effects, and subject as the random effect (Jaeger, 2008). No significant effects were found.

Then, visual hemifield presentations were swapped for the right-lesioned patients ($n = 7$) to create a single patient group with lesions normalized to the right visual field. We randomly selected half of the controls ($n = 20/2 = 10$) for the identical swapping procedure to preclude any confounding influence from inter-hemispheric differences. This step ensured that all results would be a function of lesioned versus intact and not left versus right hemisphere. We calculated the mean accuracy for each subject by visual hemifield (lesion/right or intact/left) and condition, and submitted the accuracy data to a mixed logit model with group (patient or control), visual hemifield, and condition as the fixed effects, and subject as the random effect. The main effect of group and the three-way interaction passed an alpha threshold of 0.05. However, neither result survived the Bonferroni-correction for multiple comparisons (3 main + 4 interaction fixed effects; Cramer et al., 2015), which yielded an adjusted alpha threshold: $p_{corr} = 0.05/7 \approx 0.0071$. Because the visual hemifield and condition effects did not approach significance ($p > 0.15$), we submitted the same data to a model with group as the only fixed effect to confirm that patients were impaired at the task.

Because several of the patients exhibited general slowed or impaired motor function following injury, we did not compare response time between patients and controls.

2.5.5. Data Preprocessing

Eyetracking data were analyzed for both 200-msec stimulus presentation epochs relative to the within-trial temporal mean position over the 100-msec central fixation preceding the first stimulus. Any trial in which gaze position drifted from the center to include the ipsilateral visual hemifield during stimulus presentation was marked for exclusion from all behavioral and EEG analyses. This ensured that analyses would be isolated to the lesioned or intact hemisphere at encoding.

Each subject's EEG traces were high-pass filtered above 1 Hz and low-pass filtered below 70 Hz using zero-phase finite impulse response filters (eegfilt.m), and demeaned. Electromyography (EMG) artifacts were removed automatically using the AAR external plug-in with the default 30-sec sliding window (pop_autobssemg.m; Gomez-Herrero, 2007; Islam et al., 2016). Traces were then filtered for 60-Hz (50-Hz for OUH data) line noise harmonics using a discrete Fourier transform (ft_preprocessing.m). We then split each subject's data into trials with a 1-sec buffer on either side, excluded any that had been marked for exclusion based on eyegaze position, and manually inspected the data to mark any channels containing abnormal signal. Next, we used independent components analysis to remove artifactual signal components from all remaining channels (ft_componentanalysis.m). These artifacts constituted EOG and microsaccadic

movements, auricular components, heartbeat, and residual cranial muscle activity (Hipp et al., 2013). Any channels that had been marked as abnormal were then replaced via interpolation of the mean of the nearest neighbors (7.6 neighbors per channel on average; ft_channelrepair.m). Finally, the data were manually re-inspected to identify and discard any trials containing epochs with residual noise.

We applied the surface spline Laplacian (ft_scalpcurrentdensity.m) to all clean EEG data to enhance spatial resolution and minimize volume conduction (Perrin et al., 1989; Kayser & Tenke, 2015; Cohen, 2015, 2014). Then, channels in patients with right-hemisphere lesions (n = 7) and 10 controls were swapped across the midline (cf. Voytek et al., 2010a, 2010b).

We split the trials into: (1) a 500-msec pretrial baseline window extending from 550-50 msec before the start screen; (2) a 1500-msec encoding-maintenance window extending from the onset of the first stimulus, irrespective of delay jitter; and (3) a 900-msec processing window extending from the start of the processing period, irrespective of delay jitter. Each of the three within-trial windows was padded out to 7.5 sec, for a minimum of three cycles' buffer at 1 Hz on either side, to minimize filtering-induced edge artifacts.

2.5.6. Electroencephalogram (EEG) Analyses and Statistical Testing

Except for the baseline Fourier analysis (described below), we analyzed EEG data for all correct trials as a function of visual hemifield. We analyzed activity between 1-40 Hz, and quantified all metrics per electrode and trial separately for pretrial, encoding-maintenance, and processing data. Then, we standardized encoding-maintenance and processing data outputs against a common pretrial baseline, computed the per-subject mean, and applied inferential statistics between groups. This way, inter-subject differences in trial counts might affect subject-wise estimates but not between-groups statistics. Within-channel EEG metrics were also tested for group-by-hemisphere interactions – termed hemispheric asymmetry. First, we calculated inter-hemisphere contrasts by subtracting the data in the 27 intact-hemisphere channels from the data in the homologous lesioned-hemisphere channels (e.g., F5–F6), excluding the midline. Then, we tested the contrast data for between-groups effects.

All statistical testing employed a Monte Carlo method with cluster-based maximum correction for multiple comparisons (Maris & Oostenveld, 2007). An independent-samples t-test was used to identify clusters of contiguous data points showing a difference between groups, thresholded at 0.05, two-tailed, and then the t-statistics were summed over all data points per cluster to calculate cluster size. Then, the group labels were randomly shuffled and the same clustering procedure was applied; we repeated this procedure 1000 times to create a null distribution. Observed clusters were considered significant if fewer than 5% of randomizations yielded a larger effect.

2.5.7. Event-Related Potential (ERP) Analysis

We passed the data through a low-pass, zero-phase finite impulse response filter at 30 Hz, and absolute baseline-corrected the encoding-maintenance and processing outputs on the all-correct-trial, temporal mean of the pretrial data window. Then, we visually inspected the data for event-

related potentials (ERPs), and submitted the encoding-maintenance and processing data, respectively, to group and hemispheric asymmetry tests with clustering on the time and space dimensions (neighboring channels ≥ 2; ft_timelockstatistics.m).

2.5.8. Baseline Fourier Analysis

We quantified the raw power spectra over the 500-msec pretrial baseline period for all trials. Spectral decomposition was performed using a Hanning taper over a frequency-dependent sliding window of three cycles, with 2-Hz bandwidth in steps of 1 Hz, yielding a series of complex numbers from which absolute values were extracted and squared to produce real power values (ft_freqanalysis.m). We tested the data for group and hemispheric asymmetry effects with clustering on the frequency and space dimensions (ft_freqstatistics.m). The hemispheric asymmetry test revealed a cluster between 6-18 Hz in anterior channels (Fig. 2.S2; cluster p = 0.004), confirming that pretrial baseline differences must be reconciled prior to between-groups tests of task-evoked activity.

2.5.9. Task-Evoked Spectrotemporal Power Analyses

Spectrotemporal decomposition employed the equivalent procedure as for computation of baseline spectral power. We baseline-standardized the encoding-maintenance and processing outputs against the pretrial data window using a statistical bootstrapping procedure (cf. Flinker et al., 2015). First, the pretrial outputs of all trials were pooled into a single time series for each channel and frequency. Then, we randomly selected and averaged t data points (t = number of trials in that subject's dataset) from the pooled time series; we repeated this 1000 times to create a normal distribution of pretrial baseline power values. Encoding-maintenance and processing data for each trial and time point were z-scored on the baseline distribution to assess the significance of task-evoked changes. Then, we submitted the task-evoked power data to group and hemispheric asymmetry tests with clustering on the time, frequency, and space dimensions (ft_freqstatistics.m).

2.5.10. Directional Connectivity Analyses

The task-evoked power analysis revealed significantly increased, sustained activity centered in delta-theta (2-7 Hz) in anterior channels; and increased alpha-beta (9-24 Hz) activity in posterior channels during stimulus presentation followed by widespread suppression during delay in central and posterior channels. We quantified directional connectivity using the Phase Slope Index (PSI; Nolte et al., 2008) separately for delta-theta and alpha-beta (9-24 Hz) frequency ranges. First, we subtracted the trial-averaged time series from each single-trial time series to minimize any contamination from simultaneous voltage changes on phase consistency (Brincat & Miller, 2015). Then, spectral decomposition employed the equivalent procedure as for computation of spectrotemporal power, sampled at 100-msec resolution. However, instead of extracting the absolute values, we computed single-trial cross-spectral density directly from the complex numbers.

The PSI was quantified for each trial-by-time point, and then averaged over trials and submitted to surrogate testing (code adapted from data2psiX.m; Cohen, 2014). We randomly shuffled the

frequency bins at each time-frequency point 1000 times to create a normal distribution of surrogate data, and *z*-scored all outputs on the surrogate distributions to assess true PSI. Finally, encoding-maintenance and processing PSI data were absolute baseline-corrected on the temporal mean of the pretrial PSI data to isolate true, task-evoked directionality. We performed between-groups tests on task-evoked PSI data at encoding-maintenance and processing, with clustering on the time, frequency, and space dimensions. Then, we averaged all significant PSI results across frequencies to obtain delta-theta and alpha-beta connectivity time courses, which we visualized topographically.

2.5.11. Network Analysis

We applied graph theoretical metrics to assess node degree (ft_networkanalysis.m) – i.e., the degree to which each channel shares a significant (PSI $|z| > 1.96$, $p < 0.05$) directional connection to all other channels. This computation was performed on the surrogate-standardized PSI data, prior to applying the absolute baseline correction. Then, the node data were absolute baseline-corrected on the pretrial node data, employing the equivalent method as for the PSI data to isolate true, task-evoked network shifts in directionality. Together with PSI outputs, this analysis produced a per-subject connectome for every time point in each of the delta-theta and alpha-beta networks. We performed *post-hoc* between-groups tests on task-evoked node degree data if significant effects were observed in the PSI data. This step permitted us to assess whether between-groups differences in specific inter-regional connections would have a net effect on region- or network-wide lead/lag characteristics. Any significant network effects were then averaged across frequencies to obtain time courses, and visualized topographically with the PSI effects.

2.5.12. Code

We analyzed all data in MATLAB (The MathWorks, Inc., Natick, MA) using custom scripts. Electrophysiological data analyses adopted functions available in open-source MATLAB toolboxes – FieldTrip (Oostenveld et al., 2011) and EEGLAB (Delorme & Makeig, 2004) – or code corresponding to Cohen (2014). Group differences in directional connectivity were plotted using the open-source toolbox BrainNet Viewer (Xia et al., 2013). The behavioral task was programmed in E-Prime Professional 2.0 (Psychology Software Tools, Pittsburgh, PA).

2.6. Supplementary Figures

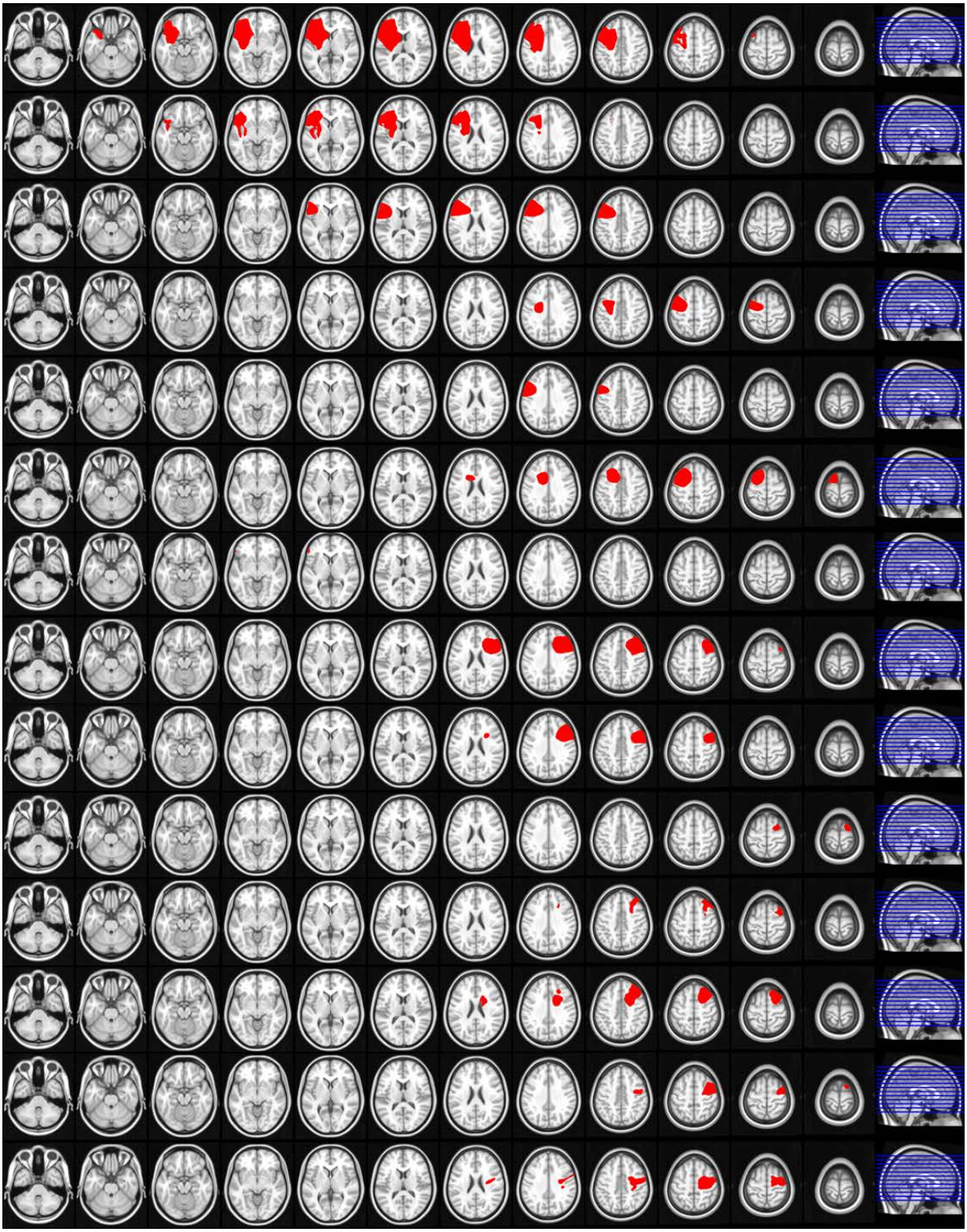

Fig. 2.S1. Individual computerized reconstructions of structural scans for all 14 PFC patients. Red = lesion site.

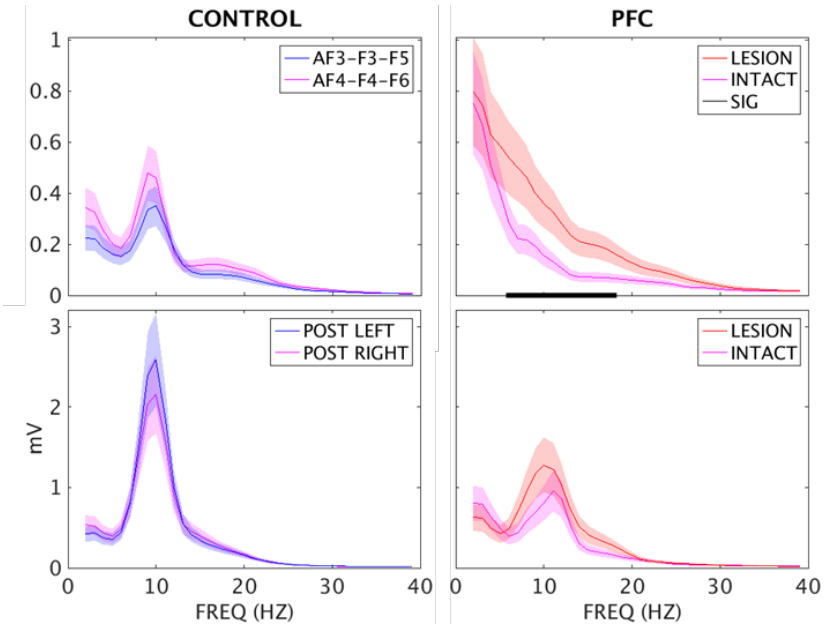

Fig. 2.S2. Mean raw energy per frequency over the 500-msec pretrial baseline period in the lesioned/left versus intact/right hemisphere as a function of group. Controls (left) exhibited clear peaks at alpha (~10 Hz) in frontal (top) and parieto-occipital (bottom) regions that did not differ between hemispheres. In contrast, PFC patients (right) did not show an alpha peak in anterior channels in either hemisphere, but the mean energy was elevated between 6-18 Hz in a cluster of channels proximal to the lesion site (AF3-F3-F5), relative to the contralateral channels (hemispheric asymmetry cluster $p = 0.004$, marked in black on the x-axis [top right]). While the posterior alpha peak appears diminished in patients relative to controls, contrasts did not survive statistical testing (between-groups $p > 0.05$, uncorrected). POST LEFT/LESION, channels P1-P3-P5-P7-P9-PO3-PO7-O1; POST RIGHT/INTACT, channels P2-P4-P6-P8-P10-PO4-PO8-O2; SIG, significant.

Ch. 3: Intracranial Recordings and Human Memory

Citation: Johnson, E. L. & Knight, R. T. Intracranial recordings and human memory. *Curr. Opin. Neurobiol.* 31 (2015).

Recent work involving intracranial recording during human memory performance provides superb spatiotemporal resolution on mnemonic processes. These data demonstrate that the cortical regions identified in neuroimaging studies of memory fall into temporally distinct networks and the hippocampal theta activity reported in animal memory literature also plays a central role in human memory. Memory is linked to activity at multiple interacting frequencies, ranging from 1 to 500 Hz. High-frequency responses and coupling between different frequencies suggest that frontal cortex activity is critical to human memory processes, as well as a potential key role for the thalamus in neocortical oscillations. Future research will inform unresolved questions in the neuroscience of human memory and guide creation of stimulation protocols to facilitate function in the damaged brain.

3.1. Introduction

Our ability to act based on personal experience drawn from memory is central to everyday life, and defines our individual identity. Human memory function is susceptible to a wide range of neurological insults. For instance, dementia and associated memory dysfunction are reaching epidemic levels as our population ages [1]. We must understand the precise neural mechanisms governing memory to make inroads into the understanding of normal as well as disordered memory. Subdural and depth recordings — termed electrocorticography (ECoG) or intracranial electroencephalography (iEEG) — offer superb temporal and spatial resolution that is unparalleled in the study of human cognition. The present review focuses on the contributions of subdural and depth recordings obtained during the successful encoding and immediate or delayed retrieval of memories in humans. We argue that ECoG/iEEG informs unresolved questions in the study of human memory and is yielding insights necessary for the development of novel interventions to facilitate memory function in the damaged brain. We will use the term ECoG to subsume both subdural (epicortical) and depth (subcortical) recordings.

It is well-known that the hippocampus and surrounding medial temporal lobe (MTL) structures are necessary for episodic memory (long-term memory for personal events) [1,2], and that lateral prefrontal cortex (PFC) is necessary for working memory (active storage and processing in memory) [3,4]. However, working and episodic memory, although often approached as separate topics in psychology, both depend on MTL–PFC interactions. What is the nature of these and other inter-regional interactions, and might they be fractionated depending on the type of memory in question? Furthermore, how do neural networks differentially support encoding and retrieval operations, even within a given type of memory? Might the PFC play a domain-general role in memory — that is, a global role that is not specific to stimulus modality or encoding or retrieval operation — that is comparable and/ or complementary to the role of the MTL?

Lesion studies and functional magnetic resonance imaging (fMRI) reveal the where of memory function. Scalp EEG and magnetoencephalography (MEG) reveal the when and, for low-frequency spectral activity and event-related potentials (ERPs), the how of memory. In contrast,

ECoG has superb spatiotemporal resolution and can measure an expanded frequency range of activity, including high-frequency responses and, in rare instances, single-unit activity (SUA). Thus, ECoG can reveal the how of human memory across an extended scope of the neurophysiology of memory in humans. For instance, Burke and colleagues [5] reported the results of a large-scale ECoG study of subsequent memory (SM) — that is, measures of neural activity correlated with or predictive of later remembering (see [6]). Their data reveal that regions previously identified using fMRI fall into two networks that exhibit spatiotemporally distinct patterns of 64–95 Hz gamma band power activity within the first 1.5 s following encoding of each later recalled word — first in the ventral visual pathway and MTL, and then across association regions including left-lateralized inferior frontal, posterior parietal, and ventrolateral temporal cortices. The authors suggest that these networks reflect higher-order visual processes followed by top-down control mechanisms [5]. Jacobs et al. [7] used MTL depth recordings to reveal that neuronal firing is phase-locked to oscillatory activity in the delta, theta, and gamma frequency bands in humans. Comparable invasive recording has been traditionally restricted to animals; thus, human ECoG recording provides a powerful bridge to the animal literature on memory processing.

3.2. Event-Related Potentials and Medial Temporal Lobe Function

Intracranial recordings provide the spatial resolution needed to explore temporal dynamics of memory within subregions of the MTL. Axmacher and colleagues [8] examined SM using ECoG and fMRI, and found that, unlike words that were successfully recalled, words that were later forgotten deactivate the hippocampus at encoding — manifested by a positive direct current response in ECoG and negative blood-oxygen-level dependent (BOLD) response in fMRI. This effect was apparent for words presented both early and late in a list, suggesting a continuum of hippocampal involvement over long and short retention durations. Other studies of encoding as a function of subsequent recall demonstrate that the successful encoding of words is associated with an early (300–400 ms) negative ERP in the rhinal cortex, followed by a late (500 ms or later) positive hippocampal ERP [9,10]. Viewing encoding as a function of subsequent recognition, as opposed to recall, several studies suggest that SM is linked to negative ERPs in the hippocampus. In preparation for immediate recognition, a negative ERP is observed 300–500 ms after stimulus presentation and again upon presentation of the probe at retrieval [11]. In preparation for delayed recognition, a late hippocampal negative ERP is observed following stimulus presentation [12–14]. The differential ERP effects observed in hippocampal activity depending on whether SM is measured by recall or recognition suggest that the hippocampus plays a selective role in recollection (i.e., specific memory, in contrast to strength-based familiarity; see [15]). However, additional data indicate that there is more to the human MTL story than ERPs. For instance, negative-polarity ERPs have been shown to correlate with high-gamma activity [14] (see section on high-frequency responses, below).

Intracranial recordings have also informed the long-standing debate over the difference between recollection and familiarity [15]. Staresina et al. [16] demonstrated that the perirhinal cortex and hippocampus are qualitatively dissociable at retrieval, revealing that the magnitude of the ERP in each MTL subregion region differs between the successful recognition of an item versus a source detail (e.g., background color), versus correct rejection of an item. The hippocampus shows enhanced ERP activity during the retrieval of source information as compared to item retrieval or

correct rejection [16]. Rutishauser et al. [17] used microelectrodes to record activity of single neurons (SUA) in the hippocampus and amygdala. Encoding is associated with sustained local SUA — with the highest spike rate observed during encoding of items later recollected (here, retrieved with correct source information), followed by items considered familiar (retrieved without source information), and the lowest SUA rate observed for items not recognized — supporting a continuous strength model of retrieval. Together, the high spatiotemporal resolution of ECoG provides evidence supporting both a continuous strength model of hippocampal function [16,17] and a dual-process model of retrieval by MTL subregion [16].

3.3. Low-Frequency Responses and Memory

The animal literature has provided robust evidence of oscillations in the theta frequency band (3–8 Hz) at encoding in MTL structures (e.g., [18]). In humans, direct intra-MTL depth recordings and/or cortical surface recordings have shown that presentation of a subsequently remembered stimulus resets theta activity or alters theta power in the hippocampus, rhinal cortex, and/ or amygdala [10,12–14,19,20], as well as temporal, frontal, and/or parietal–occipital association cortices [21–23,24]. Shifts in other frequency bands tend to co-occur with these shifts in theta. Specifically, theta and alpha phase resets in temporal, parietal, and occipital cortices [21]; theta, alpha, and beta phase resets [12] or power decreases [20] in the MTL; and right cortical theta power increases in the midst of widespread gamma increases [25] have been linked to SM. Employing SUA measurements, Rutishauser et al. [19] reported that theta phase resets are tightly coupled with local spiking activity — that is, theta phase-SUA coupling (see [26]). Critically, hippocampal theta phase-SUA coupling predicts subsequent long-term recognition as well as participants' confidence in their responses at retrieval. Suthana and colleagues [27] found that stimulation of the entorhinal cortex during encoding resets the theta phase in the hippocampus and enhances spatial memory, suggesting a causal role for hippocampal theta activity in SM.

Patterns of theta and successive alpha band power increases in the hippocampus and rhinal cortex just before stimulus presentation have also been shown to predict subsequent recognition [28] (Fig. 3.1). Fell et al. [28] proposed that this pattern reflects the coupling of activated contextual information (theta) and top-down control processes (alpha). This intracranial finding demonstrates the importance of preparatory membrane excitability in successful encoding.

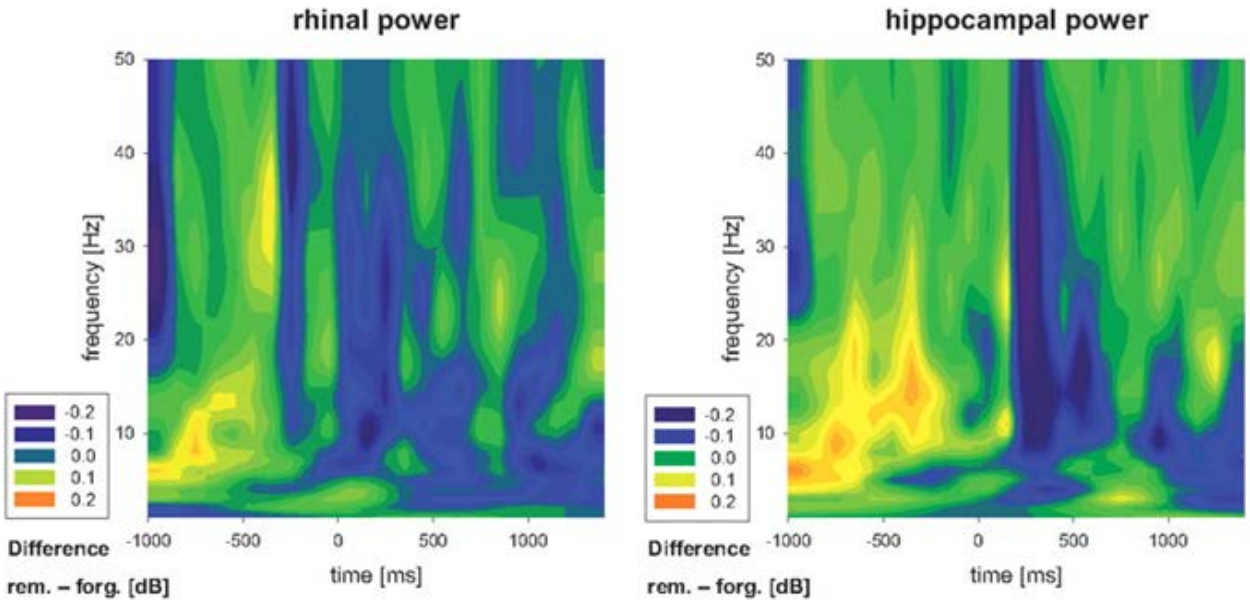

Fig. 3.1. Normalized power difference plots for the contrast of subsequently remembered > forgotten words in the rhinal cortex (left) and hippocampus (right) reported in [28]. Power increases are shown primarily in the theta and alpha bands before stimulus presentation at encoding. Adapted from [28] with permission.

There is activity in multiple frequencies occurring simultaneously during memory encoding and retrieval. Successful encoding in humans has been linked to two distinct hippocampal oscillations at the edges of the theta band. Lega et al. [29] found that the ‘slow-theta’ exhibits higher power at stimulus presentation and is selectively coupled with power in the gamma frequency band. (See Box 3.1 for a description of interactions between different frequency bands, i.e., cross-frequency coupling.) This pattern is also observed just prior to recall; taken together, Lega et al. [29] argue that this 3 Hz activity may be the human analogue of the theta oscillations observed in animals. At retrieval, Watrous et al. [30] showed that coherence between MTL, PFC, and parietal cortex is increased for correctly retrieved source information. Spatial memory is linked to 1-4 Hz oscillations while temporal memory is linked to 7-10 Hz oscillations in these regions, supporting a multiplexing mechanism wherein different frequency bands support distinct memory operations (Fig. 3.2). Although multiplexing is not specific to memory, these results suggest that the phenomenon of frequency multiplexing within and across regions may be central to human memory capacity [31].

Box 3.1. Phase-Amplitude Coupling in Memory

Cross-frequency coupling refers to interactions between different frequency bands; it can refer to the coupling of different phases or amplitudes, or the phase of one frequency and the amplitude of another (that is, phase-amplitude coupling). Coupling between higher-frequency amplitude and lower-frequency phase has been linked with a variety of cognitive and motor functions [62–64], and proposed as a mechanism for the short-range coding and inter-regional communication and integration of information [51,52]. Axmacher and colleagues [44] reported that successful maintenance of multiple items in working memory is associated with enhanced coupling between theta phase and beta and low-gamma (14–50 Hz, peak at 28 Hz) amplitudes in the hippocampus (also [46]). Further- more, memory load modulates theta phase to incorporate additional envelopes of higher-frequency activity, with theta peaking at 7 Hz. At retrieval, Foster et al. [33] reported that recognition of remote autobiographical memories correlates with enhanced coupling between theta phase and high-frequency amplitude. The magnitude of phase-locking between the hippocampus and retrosplenial cortex in the theta band peaks 300–400 ms before high-frequency (70–180 Hz) peak amplitude.

Phase-amplitude coupling also supports communication pathways between the thalamus and frontal cortex during successful memory encoding and retrieval. Staudigl et al. [49] revealed that successful retrieval is linked to thalamus-frontal synchrony and enhanced coupling between inter-regional beta phase and 55-80 Hz gamma amplitude. Furthermore, they found that the beta activity modulated patterns of gamma power.

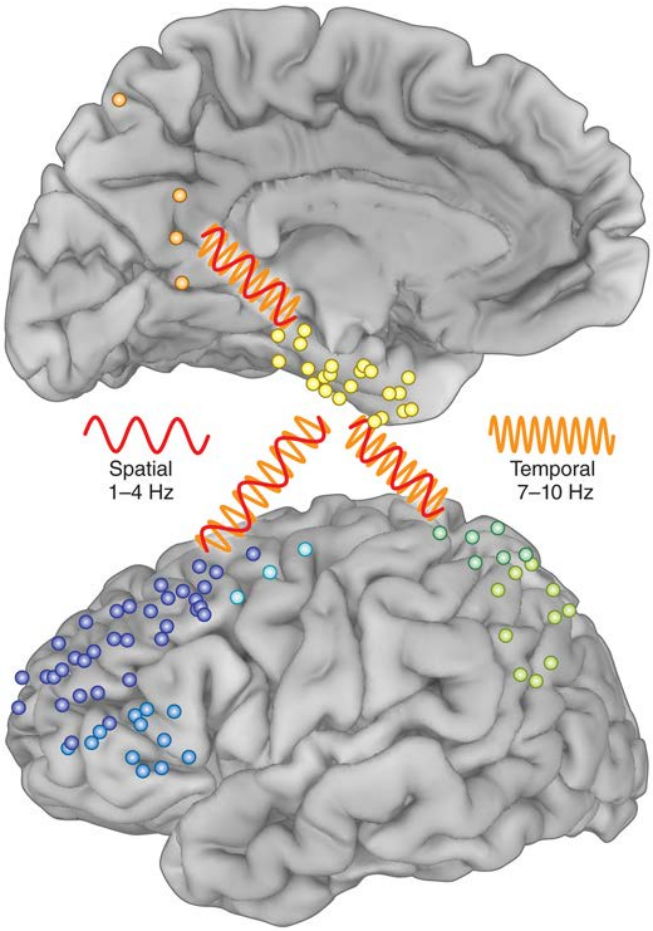

Fig. 3.2. Individual subdural recording sites from the patients studied in [30]; blue, prefrontal; green, parietal; orange, precuneus; yellow, parahippocampal. The red oscillation (1–4 Hz) represents coherence between brain regions during spatial memory and the orange oscillation (7–10 Hz) represents coherence between these regions during temporal memory. Adapted from [30,31] with permission.

Evidence from multiple studies indicates that theta is involved in timing coordinated activity within and across regions during successful encoding and retrieval. Within the hippocampus, theta power first increases and later decreases following presentation of stimuli that are subsequently recognized [14]. Fell et al. [13] showed that synchrony between the hippocampus and rhinal cortex varies as a function of frequency, with increased synchrony in delta and theta bands, but increased followed by decreased synchrony in the low-gamma band (28-46 Hz); phase-locking within each region also varies over time as a function of frequency. Using both subdural and depth recordings to examine network activity, Burke et al. [24] showed that theta and gamma power increase in a posterior-to-anterior direction with time over widespread cortical as well as hippocampal regions. Importantly, while some spectral modulations co-occur with local inter-regional synchrony and others with local or global asynchrony, synchronous activity for verbal SM is hubbed in the left PFC. An outstanding question in the intracranial recording of memory concerns whether the PFC may serve a domain-general, causal role as part of an MTL–

PFC network, or whether the role of neocortical regions is dictated by domain-specific representations at encoding and/or cognitive operations at retrieval.

Anderson and colleagues [32] showed that increased theta power in the MTL precedes increased theta power in the PFC during successful recall, and serves to synchronize the two regions. Two studies reported that retrieval of remote autobiographical memories (that is, memories encoded before entering the laboratory) is associated with theta band coherence between the MTL and other regions — such as phase-locking between the MTL and retrosplenial cortex [33]. Steinvorth et al. [34] showed that layers of the entorhinal cortex that project outward to the cortex exhibit theta activity during retrieval that is synchronized with theta activity in the frontal and temporal cortices. Layers that project inward to the hippocampus, however, show increased gamma activity. Taken together, these findings suggest that theta is important in long-range communication during successful retrieval.

3.4. High-Frequency Responses and Memory

Activity in gamma and higher-frequency bands (50-500 Hz) is important for representing information within neural regions. Notably, activity in frequency ranges above 70–80 Hz correlates with local spiking activity [35,36], suggesting that these high-frequency responses reflect SUA. High-frequency power increases [37] and oscillatory activity [38] have been shown to represent individual stimuli in the neocortex, and gamma band activity is sensitive to differences between stimuli [39]. High-frequency signals are less than a microvolt in amplitude at the scalp, placing them in the noise range of scalp EEG recording. Thus, reliable data in high-frequencies is generally limited to subdural or depth recording (cf. [22]). Importantly, ECoG studies demonstrate that low-frequency and high-frequency activity often share an inverse relationship [5,20,24,25,38-40]. Likewise, while measures of high-frequency amplitude show overlap with the BOLD measures of fMRI, low-frequency activity is anatomically dissociated with BOLD measures [41].

Intracranial recordings reveal an important role for high-frequency responses in successful episodic memory encoding and retrieval. Sederberg and colleagues [42] demonstrated that SM is linked to power increases in the 28–100 Hz gamma range in subdural and depth electrodes in the hippocampus, temporal cortex, and PFC at encoding (also [5,20,22]), and that this pattern is reinstated just before correctly recalling words. Kucewicz et al. [38] reported that encoding images induces oscillations from 50 to 500 Hz within the primary visual cortex as well as limbic and higher cortical regions, consistent with the visual processing stream (also [5]), and successful recall is linked with increased 50–500 Hz oscillatory activity in widespread higher cortical regions. Within the hippocampus, Park et al. [40] revealed a role for high-gamma (51–100 Hz) but not low-gamma, delta, or theta, in successful encoding during navigation. Axmacher et al. [14] found that 70–90 Hz high-gamma power is selectively increased during processing of unexpected items at multiple points during encoding in preparation for a recognition test. In addition to subserving encoding in conjunction with synchronization and desynchronization in the theta band, this high-gamma activity is also correlated with the hippocampal N500 ERP. Foster et al. [43] showed that 70–180 Hz power peaks in the hippocampus-connected posteromedial cortex after 400 ms during retrieval of autobiographical memories.

Gamma and higher-frequency band power responses have also been used to study working memory, suggesting a complex picture of working memory function that involves oscillations and sequenced spiking activity. Axmacher et al. [44] demonstrated that maintenance of multiple items in working memory is linked to the coupling of neural assemblies in the 25–100 Hz gamma range with theta phase in the hippocampus. Furthermore, increasing the number of items (i.e., memory load), is associated with modulating this cross-frequency coupling by increasing the length of theta cycles to incorporate additional pockets of higher-frequency power increases (Box 1). Indeed, Roux and Uhlhaas [45] argued that 'theta-gamma' (gamma range: 30–200 Hz) coupling subserves the organization of items maintained in sequence (also [46]), while alpha-gamma coupling may also support the active inhibition of task-irrelevant processes integral to complex working memory.

3.5. Inter-Regional Coherence and Memory

Memory is supported by both the local connectivity of MTL subregions, and the distributed connectivity of the MTL to other cortical and subcortical regions [1]. Intracranial recordings can provide insight into the dynamics of both short-range and long-range communication. Fell and colleagues reported that successful encoding is associated with hippocampus–rhinal cortex coherence in the delta and theta bands [13,47] and, in the gamma band, with early hippocampus–rhinal cortex synchronization and later desynchronization [13,48] (also [49]). Comparable patterns of within-theta and within-gamma bands also occur over left cortical regions [24]. Synchronization within theta and gamma bands is both multiplexed and dissociable throughout the MTL and cortex in SM [24]. At retrieval, multiplexing occurs as a function of the source being retrieved — with spatial and temporal information traveling along different frequency bands but within comparable MTL–PFC-parietal networks (Fig. 3.2) [30,31]. Anderson et al. [32] suggested that one mechanism of coherence between the MTL and PFC is supported by synchronous activity in the theta band during successful recall.

Intracranial recordings of human cognition also suggest how communication might occur between synchronized regions. Staudigl and colleagues [50] recorded activity in a patient with intrathalamic depth electrodes as well as frontal cortex coverage with scalp EEG. They found that successful retrieval is associated with increased synchrony between the thalamus and PFC in the beta band as well as coupling between beta oscillations and gamma power. Indeed, it has been has proposed that coupling between the phase of lower frequencies and the amplitude of higher frequencies enhances local cortical processing, facilitating transmission of information across synchronized brain regions (Box 1) [51,52].

3.6. Intracranial Recordings and Reinstatement

Episodic retrieval often involves the reinstatement of neural activity patterns elicited during encoding [53]. This phenomenon is frequently studied using fMRI, limiting precise timing of these spatially localized patterns. Several ECoG studies report data from both the encoding and retrieval phases of long-term memory paradigms. Kucewicz et al. [38] recorded 50–500 Hz activity and reported a dissociation such that encoding of pictures induces more oscillatory activity in the occipital and parahippocampal cortices than retrieval, consistent with a model of bottom-up visual processing. Retrieval, in contrast, induces more high-frequency activity in the

temporal and frontal cortices than encoding, consistent with top-down cognitive processing. Ekstrom and colleagues [54] recorded SUA and local field potentials in the hippocampus and entorhinal cortex during a navigation task and at subsequent recognition. Their data reveal a dissociation between SUA and power increases in broadband potentials, theta and gamma (30–100 Hz) bands, with increased SUA at encoding and increased local field potential activity at retrieval. Reinstatement may be specific to a subset of neurons in these MTL regions.

Context reinstatement — that is overlap of perceptual, conceptual, and/or categorical details based on similarity between encoding and retrieval — is evident in association cortices. Manning and others identified semantic components of neural activity during the encoding of words and reported that the resultant power spectra are reactivated in temporal and frontal cortices according to semantic clustering on subsequent free recall [55-57]. Overlap in neural activity between encoding and recall is not only similar, but also predictive of recall of items similar in context [55,56]; this pattern is more evident in the cortex than in the hippocampus [57]. Morton et al. [56] further reported that reactivation occurs at all frequencies studied (2–100 Hz) in the temporal cortex but does not occur at any frequency in the occipital cortex. Finally, these studies as well as [29] and [42] reported that reactivation of neural activity observed at encoding occurs just before retrieval, revealing that context reinstatement may be part of preparing to retrieve information and not part of retrieval itself.

3.7. Open Questions and Directions in Intracranial Recordings and Memory

Intracranial recordings offer unparalleled spatiotemporal resolution in the study of human memory and capture high-frequency band responses, effectively bridging the study of memory across human and animal species, and raising the possibility of answering important, unresolved questions in the neuroscience of memory. ECoG studies of working and episodic memory in humans demonstrate the relationships between neuronal spiking and high-frequency activity with oscillations of different frequency bands, which regions interact to support memory and how these regions interact at different stages of memory processing, and the precise source of ERPs generated in memory performance. Recent ECoG data suggest that the PFC is a key hub for successful encoding in humans [24], providing evidence that the frontal cortex plays a domain-general, causal mechanism in memory networks, Furthermore, emerging subcortical depth recordings in both animals and humans suggest that the driving source of neocortical oscillations may be thalamic [50,58].

Subdural and depth recordings also shed light on multiple questions posed in the psychology of memory. For instance, ECoG provides support on the neural level for multiple models of recollection-based versus familiarity-based retrieval [15–17], and Hanslmayr and Staudigl [59] argued that encoding and retrieval data support Endel Tulving's principle of encoding specificity [60]. ECoG data also indicate that the relationship between successful encoding and retrieval operations – that is, reinstatement – is both spatially and temporally complex.

Intracranial recordings will also provide guidance on how to create stimulation protocols to facilitate function in the damaged brain (see [27,61]). For instance, what is the precise nature of MTL–PFC interactions, and which mechanisms of inter-regional interaction play a causal role in successful memory formation and/or retrieval? Much of the extant neocortical data support a role

of the PFC in the representation of encoded information, but suggest that the PFC may be a hub for multiplexing and successful encoding of stimuli (see [24]). Might the PFC support memory processes in a domain-general way, irrespective of the type of encoding modality or retrieval operation? Finally, because of its fine spatiotemporal resolution, ECoG offers the means for determining the oscillatory and phase parameters of potential therapeutic stimulation, as well as the precise location and timing of application to best facilitate function (also [61]). If there is a causal and domain-general frontal mechanism governing memory function, and if the source of that mechanism is indeed thalamic, these regions may present alternative stimulation sites — allowing possibility of memory facilitation for patients with MTL damage.

3.8. References and Recommended Reading

1. Dickerson BC, Eichenbaum H: The episodic memory system: neurocircuitry and disorders. *Neuropsychopharmacology* 2010, 35:86-104.
2. Scoville WB, Milner B: Loss of recent memory after bilateral hippocampal lesions. *J Neurol Neurosurg Psychiatry* 1957, 20:11-21.
3. Goldman-Rakic P: Cellular basis of working memory. *Neuron* 1995, 14:477-485.
4. Müller NG, Knight RT: The functional neuroanatomy of working memory: contributions of human brain lesion studies. *Neuroscience* 2006, 139:51-58.
5. Burke JF, Long NM, Zaghloul KA, Sharan AD, Sperling MR, Kahana MJ: Human intracranial high-frequency activity maps episodic memory formation in space and time. *NeuroImage* 2014, 85:834-843.

   Using subdural and depth recordings in a cohort of 98 patients, the authors report that the neural regions identified in functional magnetic resonance imaging (fMRI) as associated with successful encoding fall into two spatiotemporally distinct neural networks — likely reflecting bottom-up perceptual processing and top-down control, respectively. This large-scale study demonstrates how intracranial recordings expand on traditional high-spatial resolution imaging methods available to study human memory.

6. Paller KA, Wagner AD: Observing the transformation of experience into memory. *Trends Cogn Sci* 2002, 6:93-102.
7. Jacobs J, Kahana MJ, Ekstrom AD, Fried I: Brain oscillations control timing of single-neuron activity in humans. *J Neurosci* 2007, 27:3839-3844.
8. Axmacher N, Elger CE, Fell J: Working memory-related hippocampal deactivation interferes with long-term memory formation. *J Neurosci* 2009, 29:1052-1060.
9. Fernández G, Effern A, Grunwald T, Pezer N, Lehnertz K, Dümpelmann M, Van Roost D, Elger CE: Real-time tracking of memory formation in the human rhinal cortex and hippocampus. *Science* 1999, 285:1582-1585.
10. Mormann F, Fernández G, Klaver P, Weber B, Elger CE, Fell J: Declarative memory formation in hippocampal sclerosis: an intracranial event-related potentials study. *NeuroReport* 2007, 18:317-321.
11. Paller KA, McCarthy G: Field potentials in the human hippocampus during the encoding and recognition of visual stimuli. *Hippocampus* 2002, 12:415-420.

12. Mormann F, Fell J, Axmacher N, Weber B, Lehnertz K, Elger CE, Fernández G: Phase/amplitude reset and theta-gamma interaction in the human medial temporal lobe during a continuous word recognition memory task. *Hippocampus* 2005, 15:890-900.
13. Fell J, Ludowig E, Rosburg T, Axmacher N, Elger CE: Phase-locking within human mediotemporal lobe predicts memory formation. *NeuroImage* 2008, 43:410-419.
14. Axmacher N, Cohen MX, Fell J, Haupt S, Dümpelmann M, Elger CE, Schlaepfer TE, Lenartz D, Sturm V, Ranganath C: Intracranial EEG correlates of expectancy and memory formation in the human hippocampus and nucleus accumbens. *Neuron* 2010, 65:541-549.
15. Yonelinas AP: The nature of recollection and familiarity: a review of 30 years of research. *J Mem Lang* 2002, 46:441-517.
16. Staresina BP, Fell J, Dunn JC, Axmacher N, Henson RN: Using state-trace analysis to dissociate the functions of the human hippocampus and perirhinal cortex in recognition memory. *PNAS* 2013, 110:3119-3124.

    The authors use medial temporal lobe depth recordings to reveal that the hippocampus and perirhinal cortex support qualitatively dissociable processes in successful item and source recognition, and that the hippocampus is sensitive to retrieval demands. These results show support on the neural level (with event-related potentials) for multiple models of memory retrieval and address a long-standing debate in the psychology and neuroscience of human memory.

17. Rutishauser U, Schuman EM, Mamelak A: Activity of human hippocampal and amygdala neurons during retrieval of declarative memories. *PNAS* 2008, 105:329-334.
18. Düzel E, Penny WD, Burgess N: Brain oscillations and memory. *Curr Opin Neurobiol* 2010, 20:143-149.
19. Rutishauser U, Ross IB, Mamelak AN, Schuman EM: Human memory strength is predicted by theta-frequency phase-locking of single neurons. *Nature* 2010, 464:903-907.
20. Matsumoto JY, Stead M, Kucewicz MT, Matsumoto AJ, Peters PA, Brinkmann BH, Danstrom JC, Goerss SJ, Marsh WR, Meyer FB, Worrell GA: Network oscillations modulate interictal epileptiform spike rate during human memory. *Brain* 2013, 136:2444-2456.
21. Rizzuto DS, Madsen JR, Bromfield EB, Schulze-Bonhage A, Seelig D, Aschenbrenner-Scheibe R, Kahana MJ: Reset of human neocortical oscillations during a working memory task. *PNAS* 2003, 100:7931-7936.
22. Long NM, Burke JF, Kahana MJ: Subsequent memory effect in intracranial and scalp EEG. *NeuroImage* 2014, 84:488-494.
23. Sederberg PB, Kahana MJ, Howard MW, Donner EJ, Madsen JR: Theta and gamma oscillations during encoding predict subsequent recall. *J Neurosci* 2003, 23:10809-10814.
24. Burke JF, Zaghloul KA, Jacobs J, Williams RB, Sperling MR, Sharan AD, Kahana MJ: Synchronous and asynchronous theta and gamma activity during episodic memory formation. *J Neurosci* 2013, 33:292-304.

    On the basis of subdural and depth recordings in a cohort of 68 patients, the authors propose that successful encoding is linked to two major types of widespread-network spectral modulations: those reflecting local synchrony, and those that reflect either local

synchrony concurrent with global asynchrony across regions or asynchronous modulations of neural activity. This study builds on previous electrophysiological work to show the important role of asynchronous activity in memory formation, and suggests a key role for the frontal cortex in human memory networks.

25. Sederberg PB, Schulze-Bonhage A, Madsen JR, Bromfield EB, McCarthy DC, Brandt A, Tully MS, Kahana MJ: Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cereb Cortex* 2007, 17:1190-1196.
26. Buzsáki G, Anastassiou CA, Koch C: The origin of extracellular fields and currents—EEG, ECoG, LFP and spikes. *Nat Rev Neurosci* 2012, 13:407-420.
27. Suthana N, Haneef Z, Stern J, Mukamel R, Behnke E, Knowlton B, Fried I: Memory enhancement and deep-brain stimulation of the entorhinal area. *N Engl J Med* 2012, 366:502-510.
28. Fell J, Ludowig E, Staresina BP, Wagner T, Kranz T, Elger CE, Axmacher N: Medial temporal theta/alpha power enhancement precedes successful memory encoding: evidence based on intracranial EEG. *J Neurosci* 2011, 31:5392-5397.
29. Lega BC, Jacobs J, Kahana MJ: Human hippocampal theta oscillations and the formation of episodic memories. *Hippocampus* 2012, 22:748-761.
30. Watrous AJ, Tandon N, Conner CR, Pieters T, Ekstrom AD: Frequency-specific network connectivity increases underlie accurate spatiotemporal memory retrieval. *Nat Neurosci* 2013, 16:349-356.

    Using subdural recordings across widespread medial temporal, frontal, and parietal regions, the authors show that retrieval of spatial and temporal context information is associated with network activity across these regions, hubbed in the medial temporal lobe. Importantly, spatial memory retrieval is linked to 1–4 Hz oscillations while temporal memory retrieval is linked to 7–10 Hz oscillations, showing that frequency multiplexing occurs in neural networks during successful memory retrieval.

31. Knight RT, Eichenbaum H: Multiplexed memories: a view from human cortex. *Nat Neurosci* 2013, 16:257-258.
32. Anderson KL, Rajagovindan R, Ghacibeh GA, Meador KJ, Ding M: Theta oscillations mediate interaction between prefrontal cortex and medial temporal lobe in human memory. *Cereb Cortex* 2010, 20:1604-1612.
33. Foster BL, Kaveh A, Dastjerdi M, Miller KJ, Parvizi J: Human retrosplenial cortex displays transient theta phase locking with medial temporal cortex prior to activation during autobiographical memory retrieval. *J Neurosci* 2013, 33:10439-10446.
34. Steinvorth S, Wang C, Ulbert I, Schomer D, Halgren E: Human entorhinal gamma and theta oscillations selective for remote autobiographical memory. *Hippocampus* 2010, 20:166-173.
35. Ray S, Maunsell JHR: Different origins of gamma rhythm and high-gamma activity in macaque visual cortex. *PLoS Biol* 2011, 9:e1000610.
36. Cardin JA, Carlén M, Meletis K, Knoblich U, Zhang F, Deisseroth K, Tsai L-H, Moore CI: Driving fast-spiking cells induces gamma rhythm and controls sensory responses. *Nature* 2009, 459:663-667.

37. Jacobs J, Kahana MJ: Neural representations of individual stimuli in humans revealed by gamma-band electrocorticographic activity. *J Neurosci* 2009, 29:10203-10214.
38. Kucewicz MT, Cimbalnik J, Matsumoto JY, Brinkmann BH, Bower MR, Vasoli V, Sulc V, Meyer F, Marsh WR, Stead SM, Worrell GA: High-frequency oscillations are associated with cognitive processing in human recognition memory. *Brain* 2014, 137:2231-2244.

    The authors test the role of oscillations in multiple high-frequency bands, ranging from 50–500 Hz, in subdural and depth recordings of successful encoding and retrieval. The data reveal that encoding pictures is linked to activity along the bottom-up visual processing stream — with distinct patterns detected in limbic and neocortical regions — and retrieval is linked to association cortices, likely reflecting cognitive control. This study demonstrates the role of fast activity and network synchronization in memory.

39. van Vugt MK, Schulze-Bonhage A, Sekuler R, Litt B, Brandt A, Baltuch G, Kahana MJ: Intracranial electroencephalography reveals two distinct similarity effects during item recognition. *Brain Res* 2009, 1299:33-44.
40. Park J, Lee H, Kim T, Park G, Lee E, Baek S, Ku J, Kim IY, Kim SI, Jang DP, Kang JK: Role of low- and high-frequency oscillations in the human hippocampus for encoding environmental novelty during a spatial navigation task. *Hippocampus* 2014 http://dx.doi.org/10.1002/hipo.22315.

    On the basis of hippocampal depth recordings taken during multiple sessions of virtual navigation, the authors show a double-dissociation between lower-frequency and higher-frequency oscillations such that delta and theta activity increases incrementally as environments become less novel, while low-gamma activity decreases. High-gamma (51-100 Hz) activity decreases only in sessions of high recall success, revealing a specific role for high-gamma oscillations in successful encoding.

41. Khursheed F, Tandon N, Tertel K, Pieters TA, Disano MA, Ellmore TM: Frequency-specific electrocorticographic correlates of working memory delay period fMRI activity. *NeuroImage* 2011, 56:1773-1782.
42. Sederberg PB, Schulze-Bonhage A, Madsen JR, Bromfield EB, Litt B, Brandt A, Kahana MJ: Gamma oscillations distinguish true from false memories. *Psychol Sci* 2007, 18:927-932.
43. Foster BL, Dastjerdi M, Parvizi J: Neural populations in human posteromedial cortex display opposing responses during memory and numerical processing. *PNAS* 2012, 109:15514-15519.
44. Axmacher N, Henseler MM, Jensen O, Weinreich I, Elger CE, Fell J: Cross-frequency coupling supports multi-item working memory in the human hippocampus. *PNAS* 2010, 107:3228-3233.
45. Roux F, Uhlhaas PJ: Working memory and neural oscillations: alpha-gamma versus theta-gamma codes for distinct WM information? *Trends Cogn Sci* 2014, 18:16-25.
46. Lisman JE, Jensen O: The θ–γ neural code. *Neuron* 2013, 77:1002-1016.

47. Fell J, Klaver P, Elfadil H, Schaller C, Elger CE, Fernandez G: Rhinal-hippocampal theta coherence during declarative memory formation: interaction with gamma synchronization? *Eur J Neurosci* 2003, 17:1082-1088.
48. Fell J, Klaver P, Lehnertz K, Grunwald T, Schaller C, Elger CE, Fernández G: Human memory formation is accompanied by rhinal–hippocampal coupling and decoupling. *Nat Neurosci* 2001, 4:1259-1264.
49. Fell J, Axmacher N: The role of phase synchronization in memory processes. *Nat Rev Neurosci* 2011, 12:105-118.
50. Staudigl T, Zaehle T, Voges J, Hanslmayr S, Esslinger C, Hinrichs H, Schmitt FC, Heinze H-J, Richardson-Klavehn A: Memory signals from the thalamus: early thalamocortical phase synchronization entrains gamma oscillations during long-term memory retrieval. *Neuropsychologia* 2012, 50:3519-3527.
51. Canolty RT, Knight RT: The functional role of cross-frequency coupling. *Trends Cogn Sci* 2010, 14:506-515.
52. Sadeh B, Szczepanski SM, Knight RT: Oscillations and behavior: the role of phase–amplitude coupling in cognition. In *Cognitive electrophysiology of attention: signals of the mind*. Edited by Mangun G. Elsevier; 2014:268-282.
53. Rissman J, Wagner AD: Distributed representations in memory: insights from functional brain imaging. *Annu Rev Psychol* 2012, 63:101-128.
54. Ekstrom AD, Viskontas I, Kahana MJ, Jacobs J, Upchurch K, Bookheimer S, Fried I: Contrasting roles of neural firing rate and local field potentials in human memory. *Hippocampus* 2007, 17:606-617.
55. Manning JR, Polyn SM, Baltuch GH, Litt B, Kahana MJ: Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *PNAS* 2011, 108:12893-12897.
56. Morton NW, Kahana MJ, Rosenberg EA, Baltuch GH, Litt B, Sharan AD, Sperling MR, Polyn SM: Category-specific neural oscillations predict recall organization during memory search. *Cereb Cortex* 2013, 23:2407-2422.

    The authors show that patterns of category-specific oscillatory activity identified via depth recordings and scalp electroencephalography (EEG) during the encoding of words predict whether words will be successfully recalled and, furthermore, the clustering of words at recall. This study provides high-resolution spatiotemporal data in support of retrieved-context models of memory, demonstrating how category information may be used effectively to organize search in memory.

57. Manning JR, Sperling MR, Sharan A, Rosenberg EA, Kahana MJ: Spontaneously reactivated patterns in frontal and temporal lobe predict semantic clustering during memory search. *J Neurosci* 2012, 32:8871-8878.
58. Saalmann YB, Pinsk MA, Wang L, Li X, Kastner S: The pulvinar regulates information transmission between cortical areas based on attention demands. *Science* 2012, 337:753-756.
59. Hanslmayr S, Staudigl T: How brain oscillations form memories – a processing based perspective on oscillatory subsequent memory effects. *NeuroImage* 2014, 85:648-655.
60. Tulving E, Thomson D: Encoding specificity and retrieval processes in episodic memory. *Psychol Rev* 1973, 80:352-373.

61. Lee H, Fell J, Axmacher N: Electrical engram: how deep brain stimulation affects memory. *Trends Cogn Sci* 2013, 17:574-584.
62. Durschmid S, Zaehle T, Kopitzki K, Schmitt FC, Voges J, Henze H, Knight RT, Hinrichs H: Phase-amplitude cross-frequency coupling in the human nucleus accumbens tracks action monitoring during cognitive control. *Front Hum Neurosci* 2013, 7:634.
63. Durschmid S, Quandt F, Kramer U, Hinrichs H, Schulz R, Pannek H, Chang EF, Knight RT: Oscillatory dynamics track motor learning in human cortex. *PLoS ONE* 2014, 9:e89576.
64. Szczepanski SM, Crone NE, Kuperman RA, Auguste KI, Parvizi J, Knight RT: Dynamic changes in phase-amplitude coupling facilitate spatial attention control in fronto-parietal cortex. *PLoS Biol* 2014, 12(8):e1001936.

Acknowledgements: We would like to thank Dr. Noa Ofen for valuable contributions in the planning of this piece. This research was supported by National Institute of Neurological Disorders and Stroke (NINDS) Grant 2R37NS21135 and the Nielsen Corporation.

Ch. 4: Public Outreach: How Brain Cells Make Memories

Citation: Johnson, E. L. & Helfrich, R. F. How brain cells make memories. *Front. Young Minds* 4 (2016).

Remembering a lot of things at the same time is difficult. As an experiment, read these numbers: 07041776. Then, close your eyes and try to say them aloud, in order. How did you do? We would guess that you remembered around half of the numbers. Now, try again but think of the same numbers as a date: 07-04-1776. Did you remember more of the numbers this time? You just demonstrated something called working memory. Working memory ("WM" for short) is the ability to hold onto and process pieces of information. WM activates when you experience and remember events in your life, learn new facts, talk to people, read, and do math. WM is a basic human behavior. As shown in the numbers experiment, WM has limited capacity. How does the brain support WM? And, what is happening in the brain that limits our capacity to store multiple memories at the same time?

4.1. How does the brain build memories?

Remembering is a complicated process. We will tell you about a few tricks that the human brain uses to remember a lot of things at the same time; but, first it is important to understand how memory works in the brain. The memory system is built into different parts of the brain. One key player is the part of your brain directly behind your forehead. This part is called the frontal lobe. When you think (or think about thinking!), you use your frontal lobe. Another key player in the memory system is buried deep inside your brain. This part is called the hippocampus, and it is very important for long-term memories, for instance, what you remember about how the brain builds memories tomorrow or another time in the future. For a picture of these brain structures and more information about the hippocampus, have a look at Ref. [1]. While looking at the brain structures tells us where memories are built, it does not tell us how memories are built. This article explains how brain cells build memories. We will explain why it is difficult to remember many things at the same time and then show you a few ways to improve your own memory.

To investigate the working memory (WM) system, we record electrical signals from people's brains while they hold onto and process pieces of information. We ask people to remember things, such as numbers, words, or pictures. Then, our electrical recordings show us what brain cells called "neurons" do when people remember things after a short time (usually between 1 s and 1 min) [2]. When neurons are active, they deliver very small electric currents (much smaller than currents from wall sockets). These WM experiments show that the electric currents change depending on how much information you remember.

Working memory (WM) is the ability to hold onto and process pieces of information.

Normally, you have to remember a lot of things at once. For example, to understand this article, you have to remember what you just read while you are still reading. Maybe you're also thinking about what is for dinner, where you're having dinner, and when you have to be there. Remembering all of these different things depends on an electric current that cycle three to eight

times per second [2–5]. This means that some neurons in the human brain fire together over and over again between three and eight times in 1 s [3].

We use computer tools to analyze different components of the brain signal. In Fig. 4.1, we show you a picture of what the brain signal looks like and what the component for WM looks like. We also show you the component that is active when you process sensory information, such as the things you see or hear. You can see that sensory processing is much faster (between 30 and 100 cycles per second) than the WM component. Our brains have a trick to make memories out of this very fast sensory activity. The brain uses the slower, three to eight cycle waves from the WM system to group the faster sensory activity together [4]. Fig. 4.1 shows that you can have, for example, seven faster cycles of sensory information within one WM cycle. This is a very effective trick for organizing seven pieces of information in WM. This trick also explains why it is hard to remember more than seven things at the same time. The speed of the WM system seems to be limited to three to eight cycles per second; that is, the same seven things cycle over and over again three to eight times per second in WM. This limitation seems to restrict the number of items that we can remember at the same time.

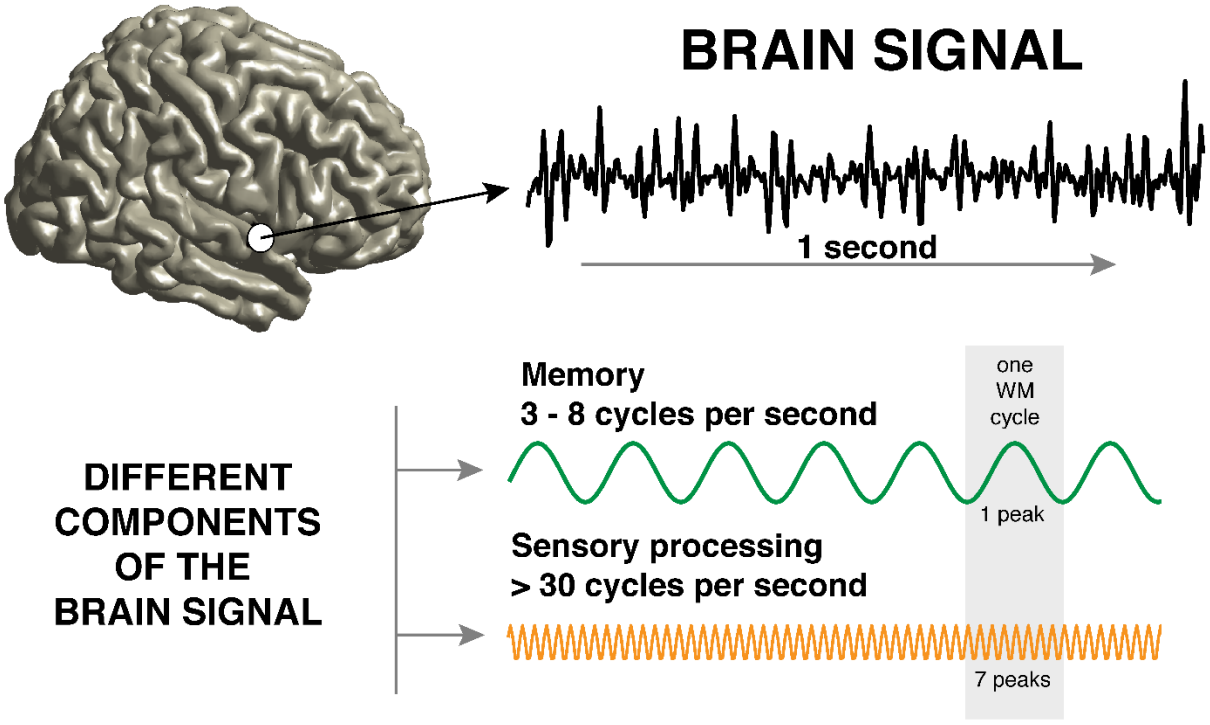

Fig. 4.1. Recording from a human brain: the brain signal (top) looks noisy and random. We can break down the signal into different components, which have different functions. For example, electric currents that cycle three to eight times per second are involved in memory function. Much faster currents (30 or more cycles per second) represent sensory information, such as the things you see or hear. So, the brain signal shown at the top seems to be a combination of different currents, allowing our brains to do a lot of things simultaneously. As shown in the gray-shaded area, one cycle of the slow WM activity can include, for example, seven cycles of the sensory information.

It should come as no surprise that the brain's WM function, which helps you think, read, and do math, is complicated. One reason for WM is so complicated is that the memory system is built into different parts of the brain, and these different parts of the brain need to talk to each other. Different parts of the brain talk to each other when neurons fire at the same time. To understand what this means, think about how several different people play instruments together in a band. The tune sounds good when the rhythms are coordinated, which means they go together in time. In the brain, we can see the memory rhythm cycling in different parts of the brain at the same time [2].

Another reason for WM is complicated is that the electrical signals in any one part of the brain are already very complex. As we showed you in Fig. 4.1, researchers have to break down the brain signal in order to make sense of it. The brain signal is so complicated, because the brain does a lot of things at the same time. Our brains do these different things at different speeds, a phenomenon called "multiplexing." Think about a song being played with different instruments together. By breaking down these sounds, we are able to differentiate the guitar's rhythm from rhythms made by the saxophone or drums. It is similar with the brain signal, where we use computer analysis to separate the memory rhythm (the slow component) from other rhythms, such as the sensory processing rhythm (the faster component).

4.2. Why is our ability to remember limited?

Think back to the WM experiment where we asked you to memorize the numbers 07041776 in order, and then to try again with the date 07-04-1776. Notice that it was easier to remember the numbers as a date, or set of three pieces of information, than it was to remember all eight numbers in order. And, have you noticed that 07-04-1776 is another way to write July 4th, 1776, the day the United States declared independence? You will probably have no trouble remembering this one piece of information. We show you more examples in Box 1.

Box 4.1. What is one piece of information?

Remembering a lot of different things at the same time is difficult, especially if something does not have a meaning to you. We asked you to remember the following eight numbers in order: 07041776. Then, we asked you to think about the numbers as a set of three pieces of information in a date: 07-04-1776. Finally, we told you that was another way of writing the single piece of information: July 4th, 1776, the day the United States declared independence. All of a sudden, eight meaningless numbers became one meaningful piece of information – and that was easier to remember. If you did not remember the eight numbers in order, but you could remember Independence Day, then you successfully "chunked," or grouped together, the information in your memory. Here are two more examples:

1. Let us say you meet five new friends at school and you have to remember their names: Daniel, Emily, Colin, Anna, and Bob. Try to recall their names in 1 min. How did you do? Maybe you forgot one or two! Well, what if we tell you that your five new friends' names start with the first five letters of the alphabet? A ... B ... C ...
2. Try to remember random colors; let us say: red, white, blue, black, red, and yellow. Again, that is difficult, but if you know that the first three colors are in the American flag and the last three colors are in the German flag, it might be easier. We show you how your brain might fit this information into a WM cycle in Fig. 4.2.

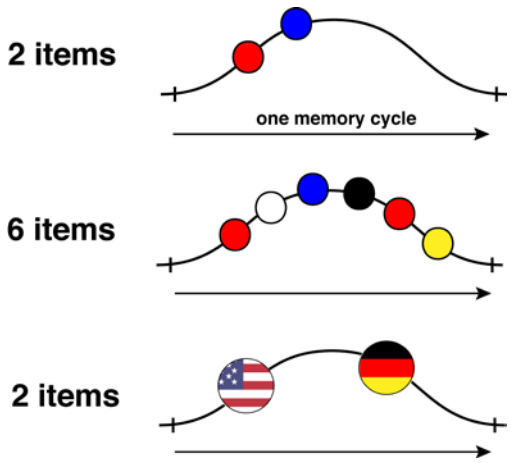

Fig. 4.2. Chunking information in one WM cycle: let us say you have to remember a set of different colors in the right order. Holding onto two colors in your memory is simple. It gets more complicated if you have to remember six colors, especially if one color (here, red) appears twice. As you can see, the order becomes more important with more information to hold in memory. It probably helps if we tell you that all the colors appear in the American and German flags, in that order. That is, it probably helps if we show you a trick to put these six colors into two "chunks," making the colors easier to remember. The brain uses another trick to organize multiple pieces of information in three to eight cycles of electrical activity per second. First, different pieces of information are put into different time slots within one memory cycle. Then, our brains can automatically use timing to group different pieces of information together into chunks. When we chunk a lot of information into fewer, more meaningful pieces of information (such as two-colored flags versus six colors), we are helping our brains use their own organizational tricks!

We have long understood that WM capacity is limited to approximately five to seven pieces of information [4]. As shown in Box 1, a piece of information can refer to something relatively meaningless like a single number or a color, or something more meaningful like a date or a flag.

As seen in Fig. 4.1 (gray-shaded area), approximately seven cycles of sensory information fit into one WM cycle. To understand what this means, consider how we represent pieces of information in the brain. Think about Box 4.1 (Fig. 4.2), where you used the sensory process of vision to study the colors and the colored flags. To memorize a list of numbers, you would use vision to read the numbers; alternatively, you would use hearing if someone were to read the numbers to you. Electrical activity that cycles 30 or more times per second is involved in representing brain activity during sensory processing [4]. So, if you try to remember more than seven pieces of information at the same time, your brain might process all of the sensory information, but you might not actually remember all of the information later. This may be because more than seven items exceeds the capacity of the slower memory component – that is, the WM cycle [4].

Now, let us do the math. If faster electrical currents cycle an average of 30 times in a second and slower electrical currents cycle an average of 5 times in a second, how many faster electrical cycles fit into each slower electrical cycle (30 ÷ 5)? The answer is 6, which is consistent with our limited WM capacity of five to seven pieces of information. The number of times faster electrical currents fit into the slower WM electrical cycles might actually determine our WM capacity limits [4]! Have another look at Fig. 4.1 for a picture of how the number of pieces of information (represented in electrical activity that cycles 30 or more times per second) can fit into a memory cycle (three to eight cycles per second). This relationship between fast- and slow-cycling electrical activities is important to how neurons make memories. Depending on how much information needs to be processed, the brain can speed up or slow down the slower, WM wave within the range of three to eight repeating cycle per second. Therefore, this slow rhythm can adapt, which helps the brain group the fast rhythm into meaningful pieces of information.

Thinking again to Box 4.1 (Fig. 4.2), it may be easier to remember two-colored flags than to remember six colors in order, but it is more complicated to remember two flags of three colors each than it is to remember two single colors. Scientists are currently doing experiments to figure out how the brain supports WM for more complicated pieces of information – or, “chunks” of multiple pieces of information.

As mentioned above, the slower-cycling WM electrical activity is adaptive. This means that the WM cycle might slow down, from eight to three cycles per second, to incorporate more pieces of sensory information in one WM cycle [4]. Another way that the brain supports WM for chunked information is that the WM cycle organizes the sensory information in order based on timing [5]. In Fig. 4.2, the first red-colored item occurs before the blue item, and the two red items occur at different times, separated in order by three other items. When we only have to remember two items, red and blue, the order of red-then-blue is simple. But, when we have to remember six items, the timing becomes more complicated – and important. This means that as we hold onto and process more and more pieces of information, the order in which different pieces of information enter the WM cycle becomes more and more important to WM function. The fast-

cycling electrical activity, which represents pieces of sensory information and reflects the firing of neurons [2], actually occurs in ordered time slots within the slower WM cycle [5].

Putting it all together, electrical signals recorded from the human brain show us that we hold onto and process pieces of information in coordinated patterns of activity [2–5]. Neurons make memories by firing together in specific parts of the brain. That might be one mechanism for remembering multiple pieces of information at the same time. This complicated WM system allows us to make memories, and it may also be the reason why remembering a lot of things at the same time is so hard!

4.3. Improve Your Memory with Science

Think back to Fig. 4.2 again. In order to remember all six-colored items, we asked you to think about them as two-colored flags. As we described, this is called "chunking," in which you combine a lot of information into fewer, more manageable chunks. Chunking is a very effective strategy for remembering multiple things at the same time. As mentioned above, the brain uses the timing of the faster electrical waves to incorporate more and more information into each slower WM cycle. By chunking a lot of information into a single item or event, we are allowing our brains to handle more pieces of information. Usually, the brain automatically breaks incoming information into manageable pieces, making the information easier to process. You can also actively use chunking to improve your memory, for example, when you study to learn information.

You can associate and combine different pieces of information, however, you would like in order to make chunks. Your brain can actually use the timing of different cycles of electrical activity to make sense of relationships based on time, space, emotions, or anything else that holds meaning for you [6]. For example, think about two events that happened yesterday, such as talking to a friend and eating dinner. Which one happened first? Did they happen in different places? Did one make you laugh? Each of these questions adds meaning to the events, allowing you to chunk the events together in WM. So, make stories!

You may have guessed from our example WM experiments that the first author of this review is from the United States. Barring extensive brain damage, she will never forget the image of the American flag or the date July 4th, 1776, because it has particular meaning for her. What country do you think the second author is from? We showed you the American flag first and the German flag second (... yes, Germany!). It also helps to visualize your stories.

Finally, because we hold onto pieces of information through cycles of repetitive electrical activity, memory improves with repetition. Make sure to tell your stories to help your brain hold onto your memories.

4.4. References

1. Davachi, L., and Shohamy, D. 2014. Thanks for the memories.... *Front. Young Minds* 2:23. doi:10.3389/frym.2014.00023

2. Johnson, E. L., and Knight, R. T. 2015. Intracranial recordings and human memory. *Curr. Opin. Neurobiol.* 31:18–25. doi:10.1016/j.conb.2014.07.021
3. 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* 464:903–7. doi:10.1038/nature08860
4. Lisman, J. E., and Jensen, O. 2013. The theta-gamma neural code. *Neuron* 77:1002–16. doi:10.1016/j.neuron.2013.03.007
5. Axmacher, N., Henseler, M. M., Jensen, O., Weinreich, I., Elger, C. E., and Fell, J. 2010. Cross-frequency coupling supports multi-item working memory in the human hippocampus. *Proc. Natl. Acad. Sci. U.S.A.* 107:3228–33. doi:10.1073/ pnas.0911531107
6. Craik, F. I., and Lockhart, R. S. 1972. Levels of processing: a framework for memory research. *J. Verbal Learn. Verbal Behav.* 11(6):671–84. doi:10.1016/S0022-5371(72)80001-X

Acknowledgements: We extend huge thanks to Dr. Robert Knight. This work was supported by NIH Grant R37NS21135 (E.L.J.) and the Alexander von Humboldt Foundation (R.F.H.).

Conclusion and Future Directions

By applying advanced neuroscientific techniques to human memory research, the current projects uncovered the dynamic and rhythmic nature of frontal and posterior control over information in working memory (WM). Intracranial data revealed parallel theta-modulated frontal-medial temporal (MTL) systems for WM (Johnson et al., in prep.; see Johnson & Knight, 2015), including a bidirectional prefrontal (PFC)-MTL system and a unidirectional MTL-to-orbitofrontal cortex (OFC) system. Importantly, PFC-MTL interactions depended on which episodic feature (identity, spatial relation, or temporal relation) subjects focused on in WM. This is the first demonstration of bidirectional PFC-MTL interactions in the human brain (cf. Place et al., 2016; Brincat & Miller, 2015; Hallock, Wang & Griffin, 2016) – and critically shows that executive demands drive the direction of interaction. Local MTL, PFC, and OFC activities also shifted as a function of executive demand, such that focusing on one episodic feature or another during maintenance recruited the MTL, while subsequent selection recruited frontal regions. The concurrent persistence of dynamic, bilateral PFC-MTL interactions pinpointed a distributed system for executive control in the context of WM.

At the same time, the PFC ostensibly does not govern WM unless additional executive demands are imposed for active processing (cf. D'Esposito & Postle, 1999). Contrary to traditional views of WM as tantamount to PFC integrity (Goldman-Rakic, 1995; see Postle, 2016; Sreenivasan et al., 2014; Lara & Wallis, 2015; Eriksson et al., 2015), lesion data revealed that alpha-beta rhythms originating in posterior regions provided adequate resources for proficient, albeit not consistent, WM function. However, when the PFC was fully intact, dissociable PFC-driven slow rhythms traveled in the opposite direction, so that multiplexed rhythms precessing in different directions between frontal and parieto-occipital regions together subserved optimal WM function (Johnson et al., in prep.). PFC-sourced low theta activity increased commensurate with executive demand, while the directionality of PFC-parieto-occipital slow rhythmic interactions shifted as a function of whether information was being encoded or maintained in WM. Taken together, results of the current projects delineate the fundamentally dynamic nature of PFC control in WM, and show that oscillations in the theta range coordinate distributed frontal interactions per available neural resources.

As the cornerstone of memory formation and thinking, it is paramount that we understand how WM works. These results illuminate phenomena in the human brain that corroborate and critically extend non-human animal data (cf. Fusi et al., 2016; McKenzie et al., 2014; Stokes et al., 2013; Warden & Miller, 2010; Barak et al., 2010; see Johnson & Knight, 2015), and bolster millennia of inquiry into the nature of human memory. Moreover, they show that dynamic association cortex activity also manifests at the network level, with theta oscillations influencing long-range interactions as dual functions of executive demand and individual structural integrity.

An accurate understanding of memory is especially pertinent as the technology sector aims to enhance human memory function (Reardon, 2015), and given the high rates of dementia in our aging population (Van Cauwenberghe et al., 2016). By bringing the field closer to such an understanding, the current results propel basic science – but raise more questions than they answer. A dynamic perspective of WM suggests that, instead of mnemonic demands initiating a particular cascade of synchronizations through a distributed network, available systems quickly

adapt to match current goals with an available trajectory through state space (Stokes, 2015; also Rose et al., 2016). The current results demonstrate that the network synchrony and dynamic state accounts are not mutually exclusive. Future research may consider an updated proposal: what makes memory function so intricate is the interaction of careful synchrony and dynamic state space.

Results of the current projects illuminate the potential of analytic approaches which track information that exists outside the scope of overtly measurable activity. For instance, bidirectional interactions between the PFC and distal regions were contingent on frequency, highlighting multiplexed rhythms for successful WM (cf. Watrous, Tandon, Conner, Pieters & Ekstrom, 2013) in the absence of variations in region-specific activity. Additional analyses may adopt multivariate approaches designed to pinpoint patterns in the spatial underpinnings of WM (Stokes et al., 2013; Wolff, Ding, Myers & Stokes, 2015; Rose et al., 2016) – and crucially explore outcomes across WM, long-term memory, and other cognitive domains. Finally, state-space analysis is ideally suited to detect patterns of dynamic potential during the pretrial intervals that precede successful versus unsuccessful encoding; that is, such novel analytic approaches may uncover the neural substrates that predict subsequent memory accuracy. Joint application of multiple analytic techniques to single datasets would not only address the interaction proposal directly, but also address the prospect of effectively recreating neural states for memory function.

References

1. Annese, J. et al. Postmortem examination of patient H.M.'s brain based on histological sectioning and digital 3D reconstruction. *Nat. Commun.* 5 (2014).
2. Balaguer-Ballester, E., Lapish, C. C., Seamans, J. K. & Durstewitz, D. Attracting dynamics of frontal cortex ensembles during memory-guided decision-making. *PLoS Comp. Biol.* 7, e1002057 (2011).
3. Barak, O., Tsodyks, M. & Romo, R. Neuronal population coding of parametric working memory. *J. Neurosci.* 30, 9424–9430 (2010).
4. Bergmann, H. C., Rijpkema, M., Fernández, G. & Kessels, R. P. C. Distinct neural correlates of associative working memory and long-term memory encoding in the medial temporal lobe. *NeuroImage* 63, 989–997 (2012).
5. D'Esposito, M. & Postle, B. R. The dependence of span and delayed-response performance on prefrontal cortex. *Neuropsychologia* 37, 1303–1315 (1999).
6. Duncan, J. The structure of cognition: Attentional episodes in mind and brain. *Neuron* 80, 35–50 (2013).
7. Eichenbaum, H. Still searching for the engram. *Learning & Behavior* (2016). doi:10.3758/s13420-016-0218-1
8. Eriksson, J., Vogel, E. K., Lansner, A., Bergstöm, F. & Nyberg, L. Neurocognitive architecture of working memory. *Neuron* 88, 33–46 (2015).
9. Foo, F. & Johnson, E. L. The resilience of our memory for music. *Front. Young Minds* (submitted).
10. Fusi, S., Miller, E. K. & Rigotti, M. Why neurons mix: High dimensionality for higher cognition. *Curr. Opin. Neurobiol.* 37, 66–74 (2016).
11. Goldman-Rakic, P. S. Cellular basis of working memory. *Neuron* 14, 477–485 (1995).
12. Hallock, H. L., Wang, A. & Griffin, A. L. Ventral midline thalamus is critical for hippocampal-prefrontal synchrony and spatial working memory. *J. Neurosci.* 36, 8372–8389 (2016).
13. Hanslmayr, S., Staresina, B. P. & Bowman, H. Oscillations and episodic memory: Addressing the synchronization/desynchronization conundrum. *Trends Neurosci.* 39, 16–25 (2016).
14. Hanslmayr, S. & Staudigl, T. How brain oscillations form memories - A processing based perspective on oscillatory subsequent memory effects. *NeuroImage* 85, 648–655 (2014).
15. Hillebrand, A. et al. Direction of information flow in large-scale resting-state networks is frequency-dependent. *PNAS* 113, 3867–3872 (2016).
16. Johnson, E. L. et al. Dynamic frontotemporal systems for episodic working memory (in prep.).
17. Johnson, E. L., Dewar, C. D., Solbakk, A-K., Endestad, T., Meling, T. R. & Knight, R. T. Causal evidence that bilateral rhythms support optimal working memory (in prep.).
18. Johnson, E. L. & Helfrich, R. F. How brain cells make memories. *Front. Young Minds* 4 (2016).
19. Johnson, E. L. & Knight, R. T. Intracranial recordings and human memory. *Curr. Opin. Neurobiol.* 31, 18–25 (2015).
20. Lara, A. H. & Wallis, J. D. The role of prefrontal cortex in working memory: A mini review. *Front. Syst. Neurosci.* 9 (2015).

21. Mackey, W. E., Devinsky, O., Doyle, W. K., Meager, M. R. & Curtis, C. E. Human dorsolateral prefrontal cortex is not necessary for spatial working memory. *J. Neurosci.* 36, 2847–2856 (2016).
22. McKenzie, S., Frank, A. J., Kinsky, N. R., Porter, B., Rivière, P. D. & Eichenbaum, H. Hippocampal representation of related and opposing memories develop within distinct, hierarchically organized neural schemas. *Neuron* 83, 202–215 (2014).
23. Oztekin, I., Davachi, L. & McElree, B. Are representations in working memory distinct from representations in long-term memory? Neural evidence in support of a single store. *Psych. Sci.* 21, 1123–1133 (2010).
24. Postle, B. R. How does the brain keep information 'in mind'? *Curr. Dir. Psychol. Sci.* 25, 151–156 (2016).
25. Ranganath, C., Cohen, M. X. & Brozinsky, C. J. Working memory maintenance contributes to long-term memory formation: neural and behavioral evidence. *J. Cogn. Neurosci.* 17, 994–1010 (2005).
26. Ranganath, C. & D'Esposito, M. Directing the mind's eye: Prefrontal, inferior and medial temporal mechanisms for visual working memory. *Curr. Opin. Neurobiol.* 15, 175–182 (2005).
27. Reardon, S. Memory-enhancement trials move into humans. *Nature* 527, 15–16 (2015).
28. Rigotti, M. et al. The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 1–6 (2013).
29. Rose, N. S. et al. Reactivation of latent working memories with transcranial magnetic stimulation. *Science* 354, 1136–1139 (2016).
30. Shimamura, A. P. Episodic retrieval and the cortical binding of relational activity. *Cogn. Affect. Behav. Neurosci.* 11, 277–291 (2011).
31. Sreenivasan, K. K., Curtis, C. E. & D'Esposito, M. Revisiting the role of persistent neural activity during working memory. *Trends Cogn. Sci.* 18, 82–89 (2014).
32. Stokes, M. G. 'Activity-silent' working memory in prefrontal cortex: A dynamic coding framework. *Trends Cogn. Sci.* 19, 394–405 (2015).
33. Stokes, M. G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D. & Duncan, J. Dynamic coding for cognitive control in prefrontal cortex. *Neuron* 78, 364–375 (2013).
34. Sweeney-Reed, C. M. et al. Corticothalamic phase synchrony and cross-frequency coupling predict human memory formation. *eLife* 3, e05352 (2014).
35. Van Cauwenberghe, C., Van Broeckhoven, C. & Sleegers, K. The genetic landscape of Alzheimer disease: Clinical implications and perspectives. *Genet. Med.* 18, 421–430 (2016).
36. Warden, M. R. & Miller, E. K. Task-Dependent Changes in Short-Term Memory in the Prefrontal Cortex. *J. Neurosci.* 30, 15801–15810 (2010).
37. Watrous, A. J., Tandon, N., Conner, C. R., Pieters, T. & Ekstrom, A. D. Frequency-specific network connectivity increases underlie accurate spatiotemporal memory retrieval. *Nat. Neurosci.* 16, 349–356 (2013).
38. Wolff, M. J., Ding, J., Myers, N. E. & Stokes, M. G. Revealing hidden states in visual working memory using electroencephalography. *Front. Syst. Neurosci.* 9 (2015).
39. Yuste, R. From the neuron doctrine to neural networks. *Nat. Rev. Neurosci.*, 16, 487–497 (2015).

Appendix 1: Public Outreach: The Resilience of Our Memory for Music

Citation: Foo, F.* & Johnson, E. L.* The resilience of our memory for music. *Front. Young Minds* (submitted). *equal contribution

Have you ever wondered what happens in your brain when you think about your favorite songs? Recent research has revealed an area of the brain that is active when we listen to music that we know. This musical memory area is separate from the parts of your brain you use to remember things you have learned in school, and details about events in your life. In this article, we will show you where the musical memory area is, and why our memory for music is often resilient to brain diseases that cause memory loss.

A1.1. Our Memory for Music

Try this simple exercise: go to your music library, pick a song, and play the first 3 seconds of it. Give yourself 1 point if you managed to sing or hum at least the next 5 seconds of that song. Do this for 20 songs. How many points did you score? We would not be surprised if it is more than 15. Now, think about what you just did. You effortlessly recalled the pitch, rhythm, and maybe even lyrics of more than 15 songs in a short amount of time. That is a lot of data you managed to summon from your brain, just like that.

Our brains possess a remarkable ability to form and retrieve memories of music, even when we are not conscious of doing so. For example, if you hear a catchy song, you would most likely be able to remember parts of it a few days later. After hearing it several times, you might know it by heart. Think about how much more effort it takes to learn information from a textbook, or remember the details of day-to-day events in your life. Even more intriguing, musical memories seem to be exceptionally preserved in people who suffer from amnesia – that's the clinical term for memory loss.

In his popular book *Musicophilia*, Dr. Oliver Sacks recounted the story of musician and musicologist Clive Wearing who, after suffering a devastating brain infection called herpes encephalitis, was unable to "retain an impression of anything for more than a blink" [1, Ch. 15]. He was also unable to remember almost his entire past, but he could play pieces of music on the piano from memory and mouth the melodies while conducting a choir. The case of Mr. Wearing is not unique – non-musicians with severe amnesia can also show lasting traces of musical memory [1, Ch. 29]. What, then, is so special about their ability to remember music, even when they cannot remember just about anything else? In this article, we will show you where musical memories may be imprinted in the brain, and how these traces can survive while other memories are lost.

A1.2. How do musical memories differ from other long-term memories?

The formation and retrieval of long-term memory traces (i.e., pieces of personal experiences and knowledge) involve the coordinated participation of multiple brain regions. For instance, when you remember what something looked like, you are using your occipital lobe, which is involved in vision. When you remember what you were thinking about, or wonder how something may

have happened differently, you are using your frontal lobe (which is important for thinking). When you think about other moments in time, such as your own past (or future!), you engage a network of brain regions including both the temporal and frontal lobes [2-3]. All of these representations are put together in a specific region called the hippocampus, located within the temporal lobes, to form a memory.

Mr. Wearing's brain infection destroyed his hippocampus and nearby regions, causing amnesia. The famous patient Henry Molaison, known by his initials as "H.M.," also suffered severe memory loss after doctors surgically removed his hippocampus and the tips of both temporal lobes (see Fig. A1.1). By studying the cases of Mr. Wearing, H.M., and other individuals with amnesia, we can conclude that the hippocampus and neighboring regions of the temporal lobe are crucial for long-term memory function.

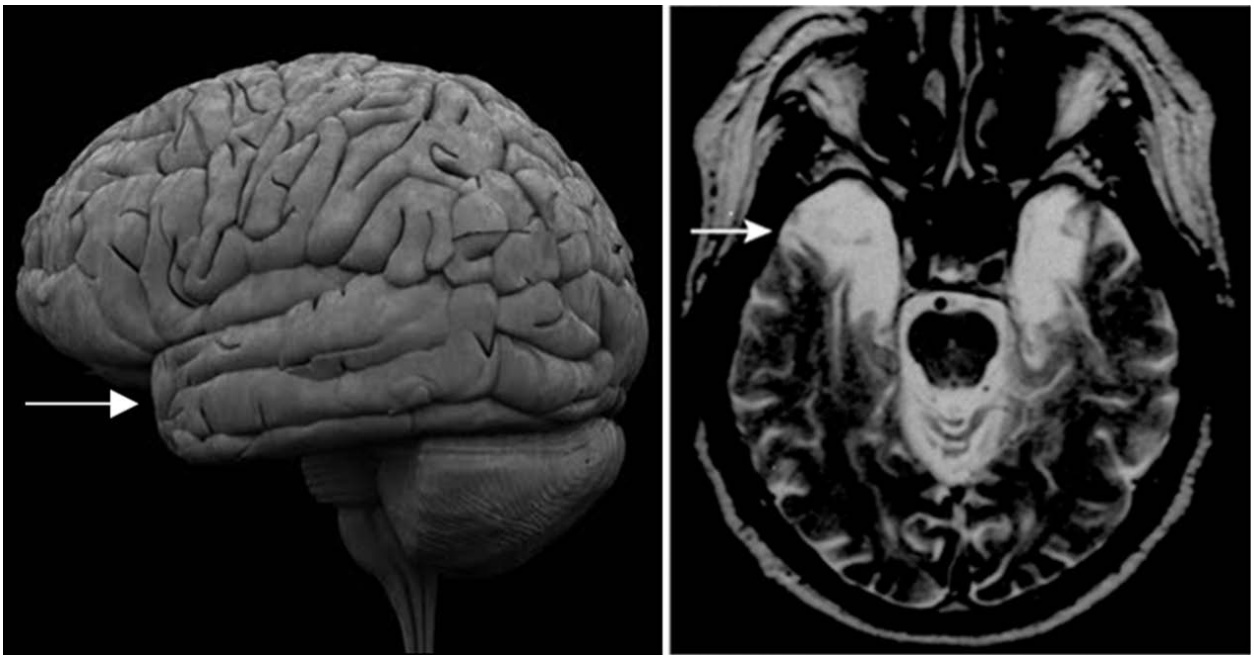

Fig. A1.1. Patient H.M.'s brain after surgery on the temporal lobes. The picture on the left shows a side view of the outside of the brain. The arrow is pointing at the tip of the temporal lobe. The picture on the right shows a cross-section of H.M.'s brain, from the viewpoint of someone looking up from his feet. The arrow is pointing to the bright portions of his temporal lobes, which are filled with fluid after removal of the brain tissue. Without these portions of the temporal lobes, H.M. could not form new memories or retrieve memories of events that happened in the years leading up to surgery. Adapted from [2] with permission.

Another condition that is associated with damage to the hippocampus and nearby regions of the temporal lobe is dementia – a neurological illness that affects one's memory, thinking, and social abilities. Individuals with dementia show progressive cognitive decline, which usually begins with mild amnesia and then gradually worsens over many years to the point where they can no longer care for themselves [3]. As the disease progresses, their brains show more and more damage throughout the network of temporal, frontal, and other regions involved in thinking about themselves in time [2-4]. Yet, despite profound memory loss and a warped sense of self,

these individuals often show striking memory for music [1, Ch. 29]. Their musical memory traces somehow survive widespread brain damage when other long-term memory traces do not. Might musical memories be processed somewhere else in the brain, in an area separate from the network of regions involved in long-term memory?

To answer this question, researchers recorded the brain responses of 32 healthy young adults as they listened to carefully pre-selected snippets of well-known, recently known, and completely unknown pieces of music [5]. The well-known pieces were selected from Top 10 songs between 1977-2007, nursery rhymes, and oldies, while the unknown songs were individually selected by looking at people's listening habits and choices on Amazon and Pandora. An hour before the participants' brains were scanned, they heard half of the songs in the unknown group twice, so that these songs would make up the 'recently known' group. The researchers observed that two specific regions, called the ventral pre-supplementary motor area and the caudal anterior cingulate gyrus, were significantly more active when participants heard well-known songs compared to recently known or unknown songs. These regions are collectively marked in red in the top row of Fig. A1.2. Additionally, a computer could accurately predict whether a particular song was well-known, recently known, or unknown just by analyzing the patterns of brain activity in these regions. Taken together, the results of this study reveal a "musical memory area" (MMA) that enables us to remember our favorite songs. Importantly, this MMA is separate from the hippocampus and the temporal lobe that we know are necessary for long-term memory function.

A1.3. Why might musical memory be preserved in patients with Alzheimer's disease?

Now that we have identified the MMA, let's take a closer look at what happens in the brains of people who have Alzheimer's disease (AD). AD is the most common form of dementia, and it most frequently afflicts people close to your grandparents' age. People who suffer from AD start to lose their memories of who they are, where they have been, and what they have done. In later stages of the disease, they begin to lose their ability to speak, do simple everyday tasks, plan, solve problems, and interact well with other people. In other words, AD progresses beyond amnesia, impacting their sense of self in time and, eventually, every aspect of their lives [2-3].

In order to diagnose AD, doctors look for multiple symptoms in the brain. These include [4]:

I) Cortical or grey matter atrophy (shrinkage of the brain because brain cells are dying);
II) An increased presence of β-amyloid plaques (a sticky buildup that causes brain cells to die); and
III) A reduced amount of glucose uptake, otherwise known as cerebral glucose hypometabolism (the brain is not consuming enough sugar to function properly).

To visualize structural damage from cortical atrophy (I), doctors use magnetic resonance imaging (MRI). To visualize buildup of β-amyloid plaques (II) and glucose hypometabolism (III), doctors use positron emission tomography (PET). If you would like to learn more about AD and its diagnosis, check out ref. [3].

This brings us to the second, third, and fourth rows of Fig. A1.2. Following identification of the MMA in healthy young adults (top row), the researchers took MRI and PET scans of 20

individuals diagnosed with AD who did not have musical training [5]. They found that the MMA showed some of the *lowest* amounts of cortical atrophy and glucose hypometabolism out of the entire brain (second and third rows, Fig. A1.2). In contrast, cortical atrophy was observed in the same portions of the temporal lobe that, once removed, caused amnesia in patient H.M. (Fig. A1.1). The other regions showing high amounts of cortical atrophy or glucose hypometabolism are part of the widespread network of regions involved in thinking about oneself in time [2]. If AD damages those regions and spares the MMA, it makes sense that musical memories survive while other long-term memories, as well as one's sense of self in time, are lost.

The researchers also observed a similar amount of β-amyloid plaque buildup in the MMA compared to other brain regions (fourth row, Fig. A1.2) [5]. As β-amyloid plaques generally appear first in the development of AD, sometimes even before noticeable memory loss [4], they argued that the MMA remains in the early stages of AD longer than the temporal lobe and other regions. AD progression typically follows from β-amyloid plaque buildup → glucose hypometabolism → cortical atrophy, making the MMA among the last brain regions to degenerate [4]. To understand what this means, think about how a sticky buildup causing brain cells to die would appear first before brain cells actually die; this progression happens earlier in the temporal lobe than in the MMA. As such, the MMA is well-preserved over the course of AD, even when it has ravaged most other parts of the brain.

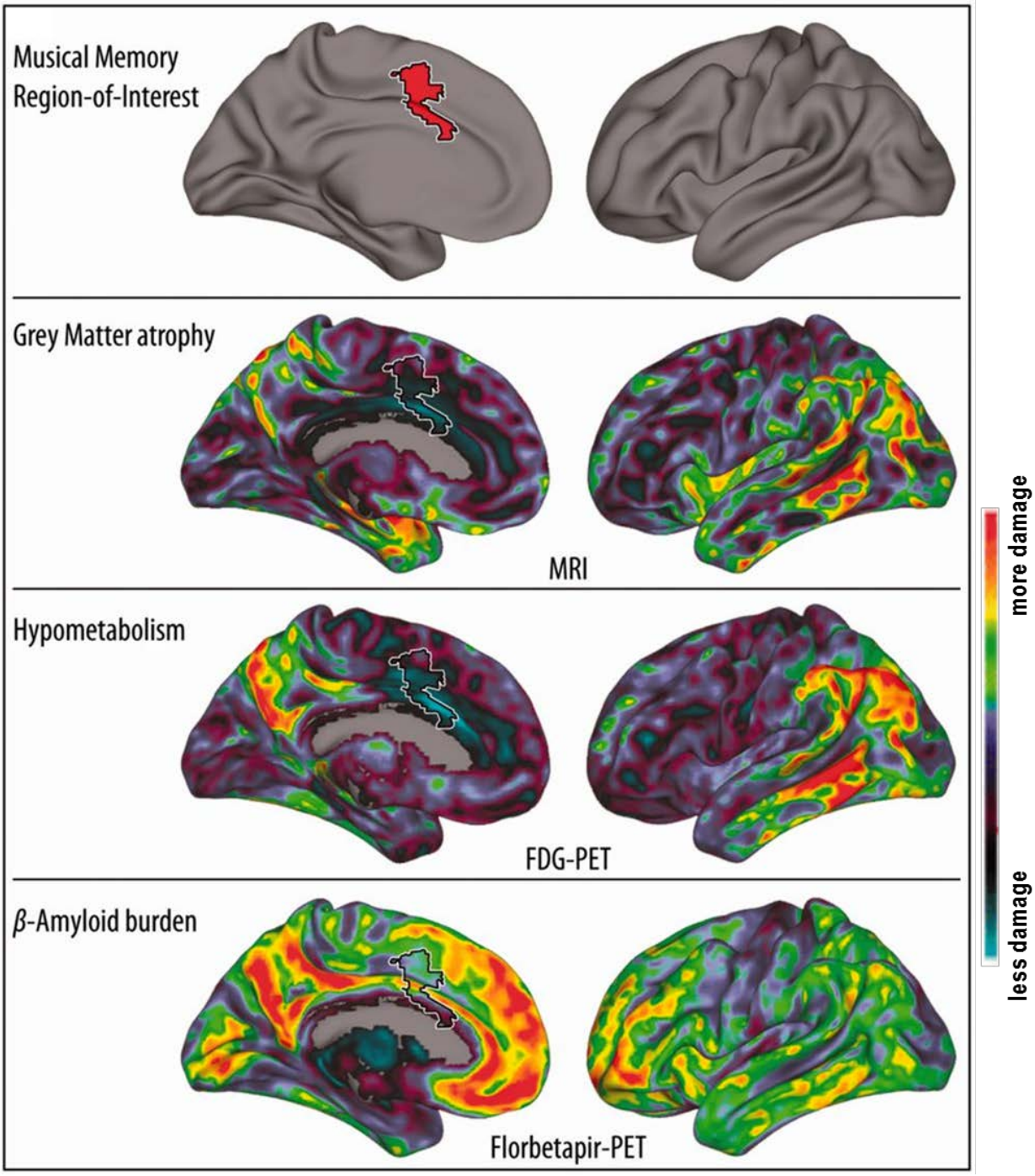

Fig. 5.2. The "musical memory area" (MMA) and its resilience to Alzheimer's disease. Each row shows a picture of the left hemisphere of the brain from the inside (left) and outside (right) views. The top row displays the MMA in red; it is also present in a symmetrically located region in the right hemisphere. In the second, third, and fourth rows, the MMA is denoted with a white or black border (you'll have to take a closer look to see it). The second, third, and fourth rows show the amount of 3 different kinds of damage in the brains of AD patients: cortical or grey matter tissue atrophy, glucose hypometabolism, and β-amyloid plaque buildup. The MMA shows

the least damage from grey matter tissue atrophy and glucose hypometabolism, compared to other brain regions. Adapted from [5] with permission.

A1.4. Conclusion

Identification of the MMA provides key neuroscientific evidence to explain the brain basis of lasting musical memory traces in the face of amnesia and dementia. As beautifully described by Dr. Sacks [1], individuals who have lost their long-term memories may appear to be stuck in the present, having lost access to their autobiographies and their ability to think about a future time, but they can amaze us with their memories for music. If musical memories can outlast damage to the hippocampus and a network of temporal, frontal, and other regions, they must be different than other long-term memories. This explains why Dr. Sacks was able to witness "mute, isolated, confused individuals warm to music, recognize it as familiar and start to sing and bond" [1, Ch. 29], and is testament to the strange yet magical power of music.

A1.5. References

1. Sacks, O. *Musicophilia: Tales of Music and the Brain.* New York: Vintage Books (2008).
2. Dickerson, B. C. & Eichenbaum, H. The episodic memory system: Neurocircuitry and disorders. *Neuropsychopharmacology*, 35, 86–104 (2010). doi:10.1038/npp.2009.126
3. Jagust, W. Dad can do time travel… but grandpa can't! *Front. Young Minds* 2 (2014). doi:10.3389/frym.2014.00018
4. Benzinger, T. L. S. et al. Regional variability of imaging biomarkers in autosomal dominant Alzheimer's disease. *PNAS* 110, e4502-4509 (2013). doi:10.1073/pnas.1317918110
5. Jacobsen, J. H., Stelzer, J., Fritz, T. H., Chételat, G., La Joie, R. & Turner, R. Why musical memory can be preserved in advanced Alzheimer's disease. *Brain* 138, 2438–2450 (2015). doi:10.1093/brain/awv135

Acknowledgements: We extend huge thanks to Dr. Robert Knight. We would also like to thank Keith Johnson, Rafael Nadal, and the BNP Paribas Open 2016 for inspiring the conceptualization of this article. This work was supported by the Nielsen Corporation (F.F.) and NIH Grant R37NS21135 (E.L.J.).

Appendix 2: Task-Evoked Pupillometry Provides a Window into the Development of Short-Term Memory Capacity

Citation: Johnson, E. L., Miller Singley, A. T., Peckham, A. D., Johnson, S. L. & Bunge, S. A. Task-evoked pupillometry provides a window into the development of short-term memory capacity. *Front. Psychol.* 5 (2014).

The capacity to keep multiple items in short-term memory (STM) improves over childhood and provides the foundation for the development of multiple cognitive abilities. The goal of this study was to measure the extent to which age differences in STM capacity are related to differences in task engagement during encoding. Children (n = 69, mean age = 10.6 years) and adults (n = 54, mean age = 27.5 years) performed two STM tasks: the forward digit span test from the Wechsler Intelligence Scale for Children (WISC) and a novel eyetracking digit span task designed to overload STM capacity. Building on prior research showing that task-evoked pupil dilation can be used as a real-time index of task engagement, we measured changes in pupil dilation while participants encoded long sequences of digits for subsequent recall. As expected, adults outperformed children on both STM tasks. We found similar patterns of pupil dilation while children and adults listened to the first six digits on our STM overload task, after which the adults' pupils continued to dilate and the children's began to constrict, suggesting that the children had reached their cognitive limits and that they had begun to disengage from the task. Indeed, the point at which pupil dilation peaked at encoding was a significant predictor of WISC forward span, and this relationship held even after partialing out recall performance on the STM overload task. These findings indicate that sustained task engagement at encoding is an important component of the development of STM.

A2.1. Introduction

The ability to maintain information for a short period of time, known variably as short-term memory (STM) or the storage component of working memory, increases over childhood (for meta-analysis see Simmering and Perone, 2013). STM capacity is tied to the ability to perform complex cognitive tasks, such as reading and math (Baddeley, 1992; Cowan et al., 2011), and the development of STM capacity partially governs age-related gains in higher-order cognitive functions (Bayliss et al., 2005; Magimairaj and Montgomery, 2012). The goal of the present study was to gain mechanistic insights into developmental changes and individual differences in STM capacity.

One of the most commonly used indices of STM in children is the digit span task, a measure of verbal STM (Bayliss et al., 2005; Cowan et al., 2005). The digit span task requires the encoding and immediate serial recall of a list of numbers presented aurally, and the length of an individual's span depends on how well s/he can attend to, rehearse, and subsequently repeat back the stimuli. The ability to remember long lists in simple span tasks has been validated as a robust correlate of higher-order cognitive functions as measured by complex span tasks in children (Cowan et al., 2005) and adults (Unsworth and Engle, 2007a,b). Age-related changes and individual differences in digit span could in theory reflect differences in cognitive resource allocation at encoding, rehearsal, and/or recall. Here, we sought to assess the extent to which age-related changes and individual differences in STM capacity could be explained by

differences in cognitive effort during stimulus encoding, as measured via the task-evoked pupillary response to cognitive load (Hess and Polt, 1964; Beatty, 1982; Beatty and Lucero-Wagoner, 2000; Karatekin, 2007; Laeng et al., 2012).

Pupil size is governed both by ambient light levels and physiological arousal (Kahneman, 1973; Beatty, 1982; Beatty and Lucero-Wagoner, 2000; Karatekin, 2007; Laeng et al., 2012). Pupil dilation related to physiological arousal is mediated by the simultaneous activation of sympathetic pathways and inhibition of parasympathetic pathways (Beatty and Lucero-Wagoner, 2000), and evidence suggests that task-evoked pupil dilation results from cortical inhibition of the parasympathetic oculomotor nucleus (Wilhelm et al., 1999; Steinhauer et al., 2004). During a state of heightened attention, neurons in the locus coeruleus fire rapidly, supplying high levels of noradrenaline to numerous targets throughout the body, including both the eyes and brain. In the eye, this neurotransmitter mediates pupil dilation; in the brain, it regulates attention through its modulatory effects on brain activity (see Gilzenrat et al., 2010; Laeng et al., 2012; Donner and Nieuwenhuis, 2013; Eldar et al., 2013).

Task-evoked pupil dilation in well-controlled experimental settings has been referred to variably as a peripheral marker of heightened attention, mental effort, or allocation of cognitive control when the task prompts focus or conscious engagement. Kahneman (1973) described it as reflecting the "intensive aspect" of attention; more recently, Gilzenrat et al. (2010) have described task-evoked pupillary dilation as reflecting task engagement. Indeed, a large body of research provides compelling evidence that task-evoked pupil dilation is sensitive to cognitive load (Beatty, 1982; Beatty and Lucero-Wagoner, 2000). Beginning with Kahneman and Beatty (1966), researchers have consistently shown that adults' pupils dilate incrementally with each digit encoded in a digit span task until the length of the digit sequence exceeds STM capacity, at which point pupil size begins to plateau or diminish (Kahneman et al., 1968; Peavler, 1974; Granholm et al., 1996, 1997; Cabestrero et al., 2009). Pupils also tend to constrict during recall as items are offloaded from STM (Kahneman and Beatty, 1966; Cabestrero et al., 2009). These findings are consistent with the idea that cognitive resources are dedicated in a manner proportionate to the cognitive load.

Pupil dilation patterns have also been used to examine individual differences in cognitive functioning among adults. Ahern and Beatty (1979, 1981) showed that cognitively higher-functioning adults—as defined based on their scores on the Scholastic Aptitude Test—exhibited consistently smaller dilation amplitudes on STM, mental multiplication, and sentence comprehension tasks than lower-functioning adults. These patterns of pupil dilation were interpreted as indices of mental effort, suggesting that performance of the same cognitive task was less challenging for higher-functioning adults. Taken together, the results of prior studies validate pupil dilation as a measure of task engagement, with pupils dilating as cognitive effort is expended.

Simmering and Perone (2013) have argued that the field of cognitive development would benefit from research linking theory to real-time behavior; specifically, they call for approaches that combine evidence from "micro-behavior"—i.e., indices of mechanisms underlying cognitive processes—and "macro" measures such as performance accuracy. We propose that task-evoked pupillometry represents a "micro" index of mental effort that can be used to probe developmental

changes in task engagement. Given its high temporal resolution, well-validated use in studies of adult cognition, and non-invasive nature, task-evoked pupillometry has the potential to provide important insights with regard to cognitive development (cf. Karatekin, 2007; Laeng et al., 2012).

Thus far, there have been only a few studies of task-evoked pupillometry involving children (Boersma et al., 1970; Karatekin, 2004, 2007; Karatekin et al., 2007a,b; Chatham et al., 2009), and only one of these studies involved a digit span task (Karatekin, 2004). In this study, 10-year-olds (n = 15) and young adults (n = 21) performed a digit span task in which they listened to sequences of 4, 6, and 8 digits. Although the 10-year-olds did not perform as well as the adults on either the 6- or 8-digit sequences, their patterns of pupil dilation differed only when they encoded the 8-digit sequences (Karatekin, 2004). On these long sequences, children exhibited shallower mean rates of dilation per digit than did adults, which the authors interpreted as indicating that they allocated fewer cognitive resources to the task.

Here, we sought to more closely examine the relationships between task engagement at encoding and developmental changes and individual differences in STM capacity. To this end, we measured pupil diameter continuously as participants encoded digit sequences that exceeded typical STM capacity, i.e., an STM overload task. If, as the results of Karatekin (2004) suggest, children are unable to recruit cognitive resources sufficient to encode at high loads, then their pupils should stop dilating (Cabestrero et al., 2009) and/or constrict (Peavler, 1974; Granholm et al., 1996) earlier in the sequence as compared to adults. Seeking to explore the relationship between these task-evoked pupillary responses and differences in STM capacity, we also administered the forward span task from the Digit Span subtest of the Wechsler Intelligence Scale for Children (Wechsler, 2003) to both children and adults. We hypothesized that if the point at which pupil diameter asymptotes is related to the amount of information encoded into STM, then this value should be related to STM capacity.

A2.2. Methods

A2.2.1. Participants

Sixty-nine healthy children (36 males, 33 females; ages 7.5–14.0 years, mean 10.6 ± 1.1 years) and 54 healthy adults (27 males, 27 females; ages 18.3–60.8 years, mean 27.5 ± 10.8 years) participated in this study.[1] Children were recruited through the Berkeley Chess School outreach program at public schools in Oakland, CA, or surrounding San Francisco Bay Area communities, and thanked via a classroom gift by request of the school administration. Adults were recruited from the University of California, Berkeley, or the San Francisco Bay Area via advertisements, and received monetary compensation or—for UC Berkeley students in the Research Participation Pool—course credit. All participants had normal or corrected-to-normal vision and hearing, and were fluent in English.

[1] Three adults and one child who reported having taken medications on the day of testing were excluded from the current sample. Two adults took an antihistamine and one took Flomax; the child's medication is not known. Six of the young adults recruited through the UC Berkeley Research Participant Pool did not provide their exact ages.

A2.2.2. Behavioral Forward Digit Span

To assess STM capacity, we used the forward span task in the Digit Span subtest on the Wechsler Intelligence Scale for Children— Fourth Edition (WISC-IV; Wechsler, 2003). The forward span task is a commonly used behavioral measure of verbal STM in multiple populations (Kane et al., 2004; Bayliss et al., 2005; Cowan et al., 2005; Alloway et al., 2009). The Digit Span subtest procedure is identical in the children and adult Wechsler test batteries; we chose to use the WISC subtest across age groups to keep the digit lists constant. Participants are read a series of digits (e.g., "9, 4, 2") at a rate of one digit per second and are asked to repeat the digits back to the experimenter in the same serial order presented. Two trials are presented at each span length, starting with two digits per trial. If the participant repeats at least one of the two trials of the same sequence length successfully, the experimenter presents two trials of a sequence that is one digit longer. This procedure continues until the participant misses both trials of a particular span length or completes the trials with the maximum 9-digit span.

In tests of verbal STM, healthy adults remember an average of seven digits, plus or minus two (Miller, 1956); children tend to remember fewer digits than adults (Simmering and Perone, 2013). An individual's STM span is calculated as the length of the longest sequence of digits successfully repeated back to the experimenter, for a maximum of 9. The forward total score reflects the number of trials each participant completed correctly, for a maximum of 16.

A2.2.3. STM Overload Task

Following administration of the WISC forward digit span, participants completed a computerized STM overload task while undergoing eyetracking. Our task was adapted from Peavler (1974), Granholm et al. (1996, 1997), Karatekin (2004), and Cabestrero et al. (2009). As in the WISC task, participants heard a sequence of digits, presented at the rate of one digit per second, and were asked to repeat them back immediately in the same order presented (Wechsler, 2003). In our adaptation of the task, participants completed a total of four trials, all involving the same number of digits. Children were asked to encode sequences of nine digits, whereas adults were asked to encode sequences of 11 digits (the same nine digits as for the children, with two additional digits added at the end of the sequence). These digit sequence lengths were chosen because they exceed average WISC forward spans, allowing us to examine pupillary responses once participants surpassed their individual encoding limitations (Granholm et al., 1996, 1997; Karatekin, 2004; Cabestrero et al., 2009). For the present purposes, we were interested in average pupil dilation and subsequent serial recall accuracy for each digit.

All participants were informed that they would hear a series of numbers. They were instructed to remember the digits as presented and then do their best to recall the full sequence of digits in the correct order. Each trial began with a 1-s auditory cue ("memorize"), alerting participants to the beginning of a trial. After the last digit for the trial was presented, the word "recall" signaled the participant to repeat the numbers back; as in the WISC forward digit span, the recall phase was self-paced. Participants completed all four trials irrespective of recall accuracy. The experimenter manually recorded participants' responses during the recall phase.

Both children and adults completed the same two practice trials before the experimental trials: a 3-digit trial followed by a 5-digit trial. They were permitted to repeat this round by request. After practice, participants underwent a 5-point eyetracking calibration procedure, and then began the experimental trials. Within each age group, all participants completed the same four experimental trials, with the order of trials randomized.

Participants were instructed to look at a 1×1inch fixation cross in the middle of the screen, presented in white on a black background, throughout the computer task. This design permitted the recording of pupil data at fixed luminance for the duration of the task, ensuring that pupillary responses were independent of pupillary light reflexes (Beatty, 1982; Beatty and Lucero-Wagoner, 2000). To allow participants' pupil diameters to return to a neutral baseline before the start of each trial (e.g., Cabestrero et al., 2009; van der Meer et al., 2010), we programmed the task in such a way that it proceeded automatically to the next trial only after the eyetracker had captured 2 s of continuous data.

A2.2.4. Eyetracking Apparatus

Stimuli were presented using the Tobii E-Prime Software Extensions (Psychology Software Tools, Pittsburgh, PA), which syncs the timing of stimulus presentation with a second computer that records pupil data. Participants were seated comfortably in front of the Tobii T120 Eye Tracker (17-inch monitor, 1280×1024 pixel resolution); distance was calibrated individually so that each participant focused on the middle of the screen, within a range of 50–80 cm. The Tobii T120 built-in camera captures data with a temporal resolution of 120 Hz, producing a data point every 8.3 ms, and average spatial resolution of 0.3◦ of visual angle. Because the camera can automatically compensate for small head movements (within a 30×22cm area at 70cm distance), participants' heads were not restrained. The camera simultaneously recorded the pupil diameter of the left and right eyes.

A2.2.5. Data Analyses

Nineteen children and eight adults were excluded from the sample due to insufficient recording of eyetracking data, yielding data from 69 children and 54 adults. We considered recordings insufficient if pupil data were absent across all four trials of at least one digit or while hearing the "memorize" cue (i.e., the cue period), or if less than 25% of data remained overall after cleaning the data to remove artifacts (adapted from Granholm et al., 1996; Siegle et al., 2011). These were cases of either technical error or excessive blinking or head motion on the participant's part, and so using such stringent cutoffs permitted us to perform analyses without need for interpolating data points to fill gaps in data collection.

Data were cleaned using a local fit procedure. We manually inspected graphic displays of a subset of data in each group sample for artifacts (e.g., partial eyelid closures, apparent changes in diameter resulting from motion), and then implemented a computer algorithm to automate this process for all subjects. A local regression model was applied to the full datasets (loess model; Cleveland et al., 1992), such that data points were removed from analysis if they fell out of the range of five standard errors above or below the locally defined, weighted mean. We applied this process separately to the raw pupil diameter of each eye, fitting locally over 400-ms segments of

data around each diameter data point.[2] Because subjects' heads were not restrained, we also applied this procedure to the mean distance between subjects' eyes and the camera. We used a more conservative fit based on 200 ms around each distance data point in order to pick up artifacts due to abrupt changes in head position. Overall, data were discarded if they fell out of range in either eye based on pupil diameter, or based on distance; fewer than 4% of data points were removed in this procedure.

To measure pupil dilation during encoding, we calculated the average pupil diameter across both eyes at each remaining data point (8.3 ms). Data for one eye were used when data for both were not available. We then calculated the mean diameter over each second, time-locked to the presentation of each stimulus, averaged across the four experimental trials. This procedure yielded one data point for the "memorize" cue, and either nine or eleven data points for the digit sequence, depending on whether the participant was a child or an adult.

The absolute diameter of the pupil at rest is known to decrease from childhood into adulthood. This age-related change is posited to reflect a gradual decrease over childhood in the influence of the sympathetic branch concurrent with a decrease in central inhibition of the parasympathetic pathway (Karatekin et al., 2007a). Thus, to compare patterns of pupil dilation between children and adults, it is necessary to control for these differences in baseline pupil diameter.

Task-evoked pupil dilation was defined as the percentage of dilation at each digit, over 1 s, relative to the mean pupil diameter over the 1-s cue period, i.e., $\text{dilation}_{digit} = (\text{diameter}_{digit} - \text{diameter}_{cue}) / \text{diameter}_{cue}$ (Karatekin, 2004; also Hess and Polt, 1964; Beatty and Lucero-Wagoner, 2000). Pupil dilation data were submitted to a mixed-model, repeated-measures analysis of variance (ANOVA), with digit as the within-subjects factor and age group as the between-subjects factor. Planned post-hoc comparisons between dilation at each digit and the next consecutive digit in the sequence were performed within each age group.

Recall accuracy was defined as the proportion of digits correctly recalled as a function of serial position on the STM overload task (Cowan et al., 2005). If a participant correctly recalled the first digit on all four trials, s/he was given an accuracy of 1 on the first digit. If, however, a participant correctly recalled a digit on three of the four trials, and missed it or recalled it incorrectly on one trial, s/he was given an accuracy of 0.75 for that digit. This procedure yielded values of 1, 0.75, 0.5, 0.25, or 0 for each digit. We conducted a mixed-model, repeated-measures ANOVA, and performed post-hoc comparisons between each digit and the next digit in the sequence within each age group. We also conducted regression analyses to further explore the relationships between measures of STM capacity and pupillary dilation at encoding, controlling for age group.

[2] A wider range of data points, up to 700 ms on pupil diameter and 500 ms on distance, was used on datasets with fewer recorded data points, as required by the loess model.

A2.3. Results

A2.3.1. Age-Related Differences in STM

First, we tested for group differences in STM capacity on the WISC digit span test and on our computerized STM overload test. As expected, adults had significantly higher WISC forward spans and scores than children, $t_{span(115.1)} = 7.6$, $t_{score(117.8)} = 7.9$; both $p < 0.001$ (Table A2.1). On our STM overload task, adults recalled more digits than children (Fig. A2.1, Table A2.1). A 9 (digit: 1 through 9) × 2 (age group) ANOVA revealed significant main effects of digit, $F_{(8, 960)} = 258.92$, $MSE = 0.03$, $p < 0.001$, $\eta_p^2 = 0.68$, and age group, $F_{(1, 120)} = 68.87$, $MSE = 0.15$, $p < 0.001$, $\eta_p^2 = 0.37$.

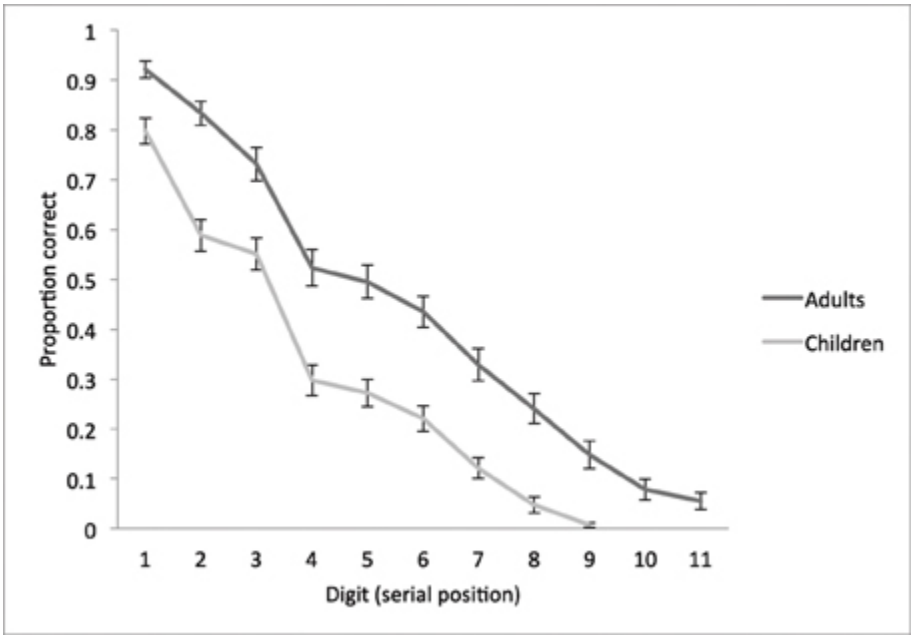

Fig. A2.1. Behavioral performance on the STM overload task. Mean proportion of digits correctly recalled as a function of serial position, plotted separately for children and adults. Error bars represent standard mean error.

	Adults	Children	Group differences
	Mean (*SD*)	Mean (*SD*)	
WISC FORWARD SPAN TASK			
Span	7.19 (1.23)	5.47 (1.26)	$t_{(115.11)} = 7.56$, $p < 0.001$
Total score	11.41 (2.12)	8.24 (2.33)	$t_{(117.84)} = 7.86$, $p < 0.001$
STM OVERLOAD TASK			
Encoding phase			
Mean pupil diameter in mm			
Cue	3.81 (0.57)	3.99 (0.47)	
Digit 1	3.88 (0.60)	4.06 (0.48)	
Digit 2	3.93 (0.63)	4.08 (0.52)	
Digit 3	3.94 (0.62)	4.13 (0.51)	
Digit 4	3.98 (0.61)	4.15 (0.53)	
Digit 5	4.01 (0.62)	4.19 (0.51)	
Digit 6	4.09 (0.63)	4.21 (0.51)	
Digit 7	4.12 (0.64)	4.20 (0.52)	
Digit 8	4.14 (0.65)	4.17 (0.51)	
Digit 9	4.14 (0.66)	4.15 (0.54)	
Digit 10	4.15 (0.66)	n/a	
Digit 11	4.13 (0.67)	n/a	
Digit-at-peak dilation	7.65 (1.81)	6.10 (2.02)	$t_{(118.73)} = 4.46$, $p < 0.001$
Recall phase			
Total correct	4.79 (1.35)	2.90 (1.13)	
Proportion correct	0.44 (0.12)	0.32 (0.13)	$t_{(114.82)} = 4.99$, $p < 0.001$

WISC and recall phase data were missing for one child. Digit-at-peak dilation computations are based on data from digits 1 to 9. Independent-samples t-tests were performed on variables that were standardized for comparison across age groups.

Table A2.1. Descriptive statistics for WISC, pupillary, and recall accuracy data by age group.

Both groups exhibited a primacy effect, such that proportion of correctly recalled digits was high at the beginning of the digit sequence and diminished with each additional digit (i.e., serial position), consistent with prior research on immediate serial recall (Kane et al., 2004; Unsworth and Engle, 2007a,b). In adults, there were significant incremental decreases from positions 1 to 2, 2 to 3, 3 to 4, 6 to 7, 7 to 8, 8 to 9, and 9 to 10 [all $t_{(53)} > 3.0$, $p < 0.01$]; and in children, from positions 1 to 2, 3 to 4, 6 to 7, 7 to 8, and 8 to 9 [all $t_{(67)} > 2.7$, $p < 0.01$]. A follow-up one-way

ANOVA showed that adults were significantly more accurate than children on all digits, all $p < 0.001$, and an independent samples *t*-test confirmed that adults recalled 12% more digits than children overall ($p < 0.001$, see Table A2.1). This finding is consistent with prior literature on the development of STM, showing that capacity increases with age from childhood into adulthood (Simmering and Perone, 2013).

Next, we used partial correlation analyses to test whether the standardized WISC digit span subtest and our STM overload task elicited similar behaviors, controlling for age group. This analysis showed that recall accuracy on the STM overload task was significantly, albeit modestly, correlated with WISC score after controlling for group [$r_{(119)} = 0.19$, $p < 0.04$]. The partial correlation between recall accuracy and WISC span, however, did not retain significance [$r_{(119)} = 0.14$, $p < 0.12$].

These findings suggest that the cognitive factors that contribute to performance on our STM overload task overlap partially with those of the standard digit span task, in which the length of the test sequence increases only after mastery is demonstrated at a particular sequence length. Indeed, behavioral performance on a memory test reflects the combined outcome of cognitive processes operating during encoding, maintenance, and retrieval. Given the high temporal resolution of pupillometry, by contrast, it is possible to examine measurements taken during a specific task phase. Here, we probe the relationships between STM capacity and pupil dilation during the encoding phase of our STM overload task.

A2.3.2. Age-Related Differences in Pupil Dilation at Encoding

In accordance with our research aim of investigating the relationship between task-evoked pupillary responses and STM capacity, we tested for group differences in dilation relative to the cue period immediately prior to task. Consistent with prior work (Karatekin, 2004; also Beatty and Lucero-Wagoner, 2000), children had larger pupils at all timepoints than adults (Table A2.1); thus, we plotted pupil dilation in terms of percentage change from the cue period (Fig. A2.2).

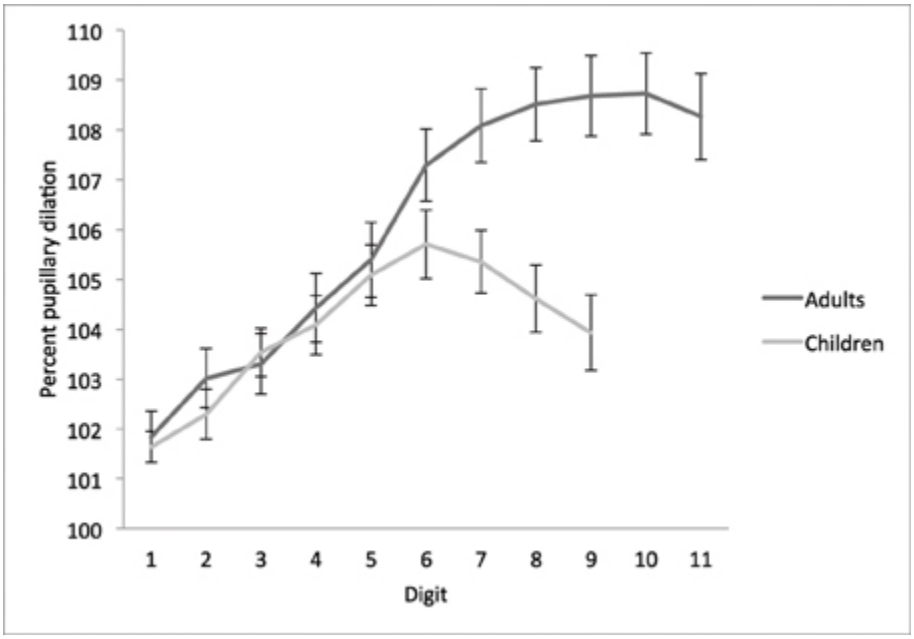

Fig. A2.2. Temporal dynamics of pupil dilation and constriction on the STM overload task. Mean percentage of pupil dilation for each digit relative to mean pupil diameter over the cue period (set to a starting point of 100%; Karatekin, 2004), by age group. Adults encoded four sequences of 11 digits each, and children encoded four sequences of 9 digits each. Error bars represent standard mean error.

A 9 (digit) × 2 (age group) ANOVA revealed significant main effects of digit, $F_{(8, 968)} = 59.24$, $MSE = 7.23$, $p < 0.001$, $\eta_p^2 = 0.33$, and age group, $F_{(1, 121)} = 4.09$, $MSE = 168.03$, $p < 0.05$, $\eta_p^2 = 0.03$, and a significant digit × group interaction, $F_{(8, 968)} = 13.51$, $MSE = 7.23$, $p < 0.001$, $\eta_p^2 = 0.10$. Both age groups demonstrated an increase in pupil dilation as a function of digit, to a point. Adults' pupils showed incremental increases from cue to digit 1, and digits 1 to 2, 3 to 4, 4 to 5, 5 to 6, and 6 to 7 [all $t(53) > 2.8$, $p < 0.01$], and continued to dilate until almost digit 9 on average (8.7 ± 2.2). Children's pupils dilated until digit 6 on average (6.1 ± 2.0), with incremental increases from cue to digit 1, digit 2 to 3, and digit 4 to 5, [all $t_{(68)} = 2.7$, $p < 0.01$], and a marginally significant increase from digit 1 to 2 [$t_{(68)} = 2.0$, $p = 0.05$]. In contrast, a significant decrease was observed from digit 7 to 8, $t_{(68)} = 2.1$, $p < 0.05$.

A one-way ANOVA with age group as the between-subjects factor confirmed that adults' pupils were significantly more dilated than children's while encoding digits 7, 8, and 9 (all $p < 0.01$), indicating that where adults' pupil diameters continued to dilate or reached a stable plateau, children's pupils reached an asymptote or began to constrict. The age groups did not differ significantly in pupil dilation on digits 1 through 6 (all $p > 0.12$), suggesting a similar rate of dilation within the constraints of STM capacity.

To directly compare the latency to peak pupil dilation—i.e., digit-at-peak—between groups, we also conducted a planned comparison based on the digit (1–9) at which pupils reached maximum dilation. Adults' maximum pupil dilation occurred on average at digit 7.7 ± 1.8, which was

significantly greater than children's maximum at digit 6.1 ± 2.0, $t_{(118.7)} = 4.5$, $p < 0.001$ (Table A2.1).

A2.3.3. Relationship Between Pupil Dilation and STM

The correspondence between average digit-at-peak values (7.7 and 6.1 for adults and children, respectively) and average WISC spans (7.2 and 5.5) hints at a relationship between STM capacity and the dynamics of pupil dilation during STM encoding. To test this hypothesis directly, we first conducted linear regression analyses between the pupillary measure of digit-at-peak dilation and each behavioral STM measure: recall accuracy on the STM overload task, WISC span, and WISC score. Digit-at-peak was significantly correlated with all three measures ($\beta_{recall} = 0.30$, $\beta_{span} = 0.38$, $\beta_{score} = 0.37$; all $p \leq 0.001$). The correlation between digit-at-peak and each WISC measure retained significance after partialing out recall accuracy on the STM overload task [$r_{span(119)} = 0.30$, $r_{score(119)} = 0.29$; both $p = 0.001$].

Next, we measured the extent to which individual variability in digit-at-peak explained individual differences in STM capacity, controlling for age group. In a multiple regression analysis, we modeled STM capacity as a function of digit-at-peak and group. This analysis revealed a strong effect of group on all three STM measures, as expected, as well as an independent contribution of digit-at-peak to each measure, $p < 0.05$ (see Table A2.2 for full results). These results indicate that cognitive resource allocation at encoding, as measured by the point of maximal pupil dilation on our STM overload task, can explain individual differences in STM capacity on a standard digit span task.

	B	*SE B*	β
WISC FORWARD SPAN			
Digit-at-peak	0.14	0.06	0.19*
Group	−1.50	0.24	−0.50**
WISC FORWARD TOTAL SCORE			
Digit-at-peak	0.24	0.10	0.18*
Group	−0.28	0.43	−0.51**
STM OVERLOAD RECALL ACCURACY			
Digit-at-peak	0.01	0.01	0.18*
Group	−0.10	0.02	−0.35**

*$p < 0.05$, **$p < 0.001$

Table A2.2. Multiple regression analyses for WISC score, WISC span, and recall accuracy

A2.4. Discussion

Consistent with decades of prior research in adults, the present results corroborate a close link between cognitive demands imposed by the digit span task and task-evoked pupil dilation (Kahneman and Beatty, 1966; Kahneman et al., 1968; Peavler, 1974; Granholm et al., 1996, 1997; Cabestrero et al., 2009), and show that children also exhibit this link (also Karatekin,

2004). Our findings extend prior work in two ways. First, we provide evidence that the children disengaged from the task as soon as the cognitive load surpassed their STM capacity, whereas adults stayed engaged while encoding additional items beyond their span. Second, we show that the point at which pupil dilation peaks is related to STM capacity—independent of age, and even after partialing out recall accuracy on the STM overload task.

With our STM span overload paradigm, we obtained similar trajectories of pupil dilation for children and adults until the sixth digit, after which the age groups diverged. Whereas adults showed dilation during encoding up to the ninth digit and then exhibited a plateau in pupil diameter until the end of the 11-digit sequence, children's pupils plateaued from digit 6 to 7, constricted from 7 to 8, and then plateaued until the end of the 9-digit sequence. In contrast to Karatekin (2004), who showed that children exhibited shallower dilation than adults during encoding of an 8-digit sequence, this finding shows children and adults dilate at similar rates up to digit 6, after which the groups' dilation patterns diverge.

Analyses focused on digit-at-peak revealed a significant relationship between the ordinal number corresponding to the digit at which maximal pupil dilation was reached on digits 1–9 and STM capacity, as reflected in our STM task and the WISC Digit Span subtest. That is, individual children or adults whose pupils peaked later in the encoded sequence were more likely to have a higher STM span, as reflected in multiple measures. This pupil-behavior relationship, observed independently of age group, is all the more noteworthy because performance on our STM overload task was not significantly related to WISC forward span after partialing out the effect of group. Thus, pupillometry reveals a relationship between encoding on one task and recall on another that would not have been detected via comparison of behavioral performance on the two tasks. These findings suggest that the allocation of cognitive resources—what Kahneman (1973) called the "intensive aspect" of attention—during encoding of information at high cognitive loads is an important contributor to the development of STM.

However, the group difference in STM performance suggests that attention is not the only factor. The groups exhibited the same rate of dilation for digits 1 through 6, indicating a similar level of cognitive effort on those digits, yet adults outperformed children on recall for all digits, not just digits 7 and higher. Thus, similar levels of cognitive resource allocation in children and adults could not fully account for the group difference in recall performance (also Karatekin, 2004). Success on the digit span task requires participants to maintain encoded digits in STM while additional digits are presented, as well as during the recall phase. Attention, echoic memory, rehearsal, and mnemonic strategies are all components of maintenance that contribute to STM performance, and it is likely that each of these cognitive components contributes to the more global measure offered by the task-evoked pupillary response. Further, STM capacity is operationalized in the digit span task as the number of digits that one can accurately recall in the right order via verbal report. This number is likely to be smaller than the number of digits in a sequence that one could accurately identify as "old" on a test of recognition memory (e.g., Unsworth and Engle, 2007b). Pupillometry has been employed in the context of long-term recognition memory (for review see Goldinger and Papesh, 2012), and given the relationship we have found between peak pupil dilation and STM span, it would be of interest to examine how the dynamics of pupil dilation and constriction at encoding relate to subsequent recognition memory as well as recall.

In summary, this study provides insight into the unique relationship between task engagement at encoding and STM capacity, and highlights the role that pupillometry can play in elucidating developmental changes and individual differences in cognition. This work supports Simmering and Perone's (2013) thesis that measures of "micro-behaviors" combined with "macro" performance measures can inform research on cognitive development. Our results further highlight the potential of pupillometry to address inquiries that extend well beyond the study of prototypical adult cognition.

The methodological approach reported here also has practical applications. Our STM overload task could provide insights regarding the cognitive deficits observed in specific patient populations (e.g., in amnesics, Laeng et al., 2007)—or, perhaps in the future, in individual patients. More generally, the task-evoked pupillary response could in theory be used to evaluate the effectiveness of a targeted cognitive intervention, pinpointing precisely at what stage(s) of a task the intervention influences cognitive processing.

A2.5. References

1. Ahern, S. K., and Beatty, J. (1979). Physiological signs of information processing vary with intelligence. *Science* 205, 1289–1292. doi:10.1126/science.472746
2. Ahern, S. K., and Beatty, J. (1981). "Physiological evidence that demand for processing capacity varies with intelligence," in *Intelligence and Learning*, eds M. Friedman, J. P. Dos, and N. O'Connor (New York, NY: Plenum Press), 121–128.
3. Alloway, T. P., Gathercole, S. E., Kirkwood, H., and Elliott, J. (2009). The cognitive and behavioral characteristics of children with low working memory. *Child Dev.* 80, 606–621. doi:10.1111/j.1467-8624.2009.01282.x
4. Baddeley, A. D. (1992). Working memory. *Science* 255, 556–559. doi:10.1126/science.1736359
5. Bayliss, D., Jarrold, C., Baddeley, A. D., Gunn, D., and Leigh, E. (2005). Mapping the developmental constraints on working memory span performance. *Dev. Psychol.* 41, 579–597. doi:10.1037/0012-1649.41.4.579
6. Beatty, J. (1982). Task-evoked pupillary responses, processing load, and the structure of processing resources. *Psychol. Bull.* 91, 276–292. doi:10.1037/0033-2909.91.2.276
7. Beatty, J., and Lucero-Wagoner, B. (2000). "The pupillary system," in *Handbook of Psychophysiology*, 2nd Edn., eds J. Cacioppo, L. Tassinary, and G. Berntson (Cambridge, UK: Cambridge University Press), 142–162.
8. Boersma, F., Wilton, K., Barham, R., and Muir, W. (1970). Effects of arithmetic problem difficulty on pupillary dilation in normals and educable retardates. *J. Exp. Child Psychol.* 9, 142–155. doi:10.1016/0022-0965(70)90079-2
9. Cabestrero, R., Crespo, A., and Quiros, P. (2009). Pupillary dilation as an index of task demands. *Percept. Mot. Skills* 109, 664–678. doi:10.2466/pms.109.3.664-678
10. Chatham, C. H., Frank, M. J., and Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proc. Natl. Acad. Sci. U.S.A.* 106, 5529–5533. doi:10.1073/pnas.0810002106
11. Cleveland, W. S., Grosse, E., and Shyu, W. M. (1992). "Local regression models," in *Statistical Models* in S, eds J. M. Chambers and T. J. Hastie (Pacific Grove, CA: Wadsworth and Brooks/Cole), 309–376.

12. Cowan, N., Elliott, E. M., Scott Saults, J., Morey, C.C., Mattox, S., Hismjatullina, A., et al. (2005). On the capacity of attention: Its estimation and its role in working memory and cognitive aptitudes. *Cogn. Psychol.* 51, 42–100. doi:10.1016/j.cogpsych.2004.12.001
13. Cowan, N., Morey, C. C., AuBuchon, A. M., Zwilling, C. E., Gilchrist, A. L., and Scott Saults, J. S. (2011). "New insights into an old problem: distinguishing storage from processing in the development of working memory," in *Cognitive Development and Working Memory: a Dialogue Between neo-Piagetian and Cognitive Approaches*, eds V. Gaillard and P. Barrouillet (Hove; New York: Psychology Press), 137–150.
14. Donner, T. H., and Nieuwenhuis, S. (2013). Brain-wide gain modulation: the rich get richer. *Nat. Neurosci.* 16, 989–990. doi:10.1038/nn.3471
15. Eldar, E., Cohen, J. D., and Niv, Y. (2013). The effects of neural gain on attention and learning. *Nat. Neurosci.* 16, 1146–1153. doi:10.1038/nn.3428
16. Gilzenrat, M. S., Nieuwenhuis, S., Jepma, M., and Cohen, J. D. (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cogn. Affect. Behav. Neurosci.* 10, 252–269. doi:10.3758/CABN.10.2.252
17. Goldinger, S. D., and Papesh, M. H. (2012). Pupil dilation reflects the creation and retrieval of memories. *Psychol. Sci.* 21, 90–95. doi:10.1177/0963721412436811
18. Granholm, E., Asarnow, R. F., Sarkin, A. J., and Dykes, K. L. (1996). Pupillary responses index cognitive resource limitations. *Psychophysiology* 33, 457–461. doi:10.1111/j.1469-8986.1996.tb01071.x
19. Granholm, E., Morris, S. K., Sarkin, A. J., Asarnow, R. F., and Jeste, D. V. (1997). Pupillary responses index overload of working memory resources in schizophrenia. *J. Abnorm. Psychol.* 106, 458–467. doi:10.1037/0021-843X.106.3.458
20. Hess, E. H., and Polt, J. M. (1964). Pupil size in relation to mental activity during simple problem-solving. *Science* 143, 1190–1192. doi:10.1126/science.143.3611.1190
21. Kahneman, D. (1973). *Attention and Effort.* Englewood Cliffs, NJ: Prentice-Hall.
22. Kahneman, D., and Beatty, J. (1966). Pupil diameter and load on memory. *Science* 154, 1583–1585. doi:10.1126/science.154.3756.1583
23. Kahneman, D., Onuska, L., and Wolman, R. E. (1968). Effects of grouping on the pupillary response in a short-term memory task. *Q. J. Exp. Psychol.* 20, 309–311. doi:10.1080/14640746808400168
24. Kane, M. J., Hambrick, D. Z., Tuholski, S.W., Wilhelm, O., Payne, T.W., and Engle, R. W. (2004). The generality of working memory capacity: a latent-variable approach to verbal and visuospatial memory span and reasoning. *J. Exp. Psychol. Gen.* 133, 189–217. doi:10.1037/0096-3445.133.2.189
25. Karatekin, C. (2004). Development of attentional allocation in the dual task paradigm. *Int. J. Psychophysiol.* 52, 7–21. doi:10.1016/j.ijpsycho.2003.12.002
26. Karatekin, C. (2007). Eye tracking studies of normative and atypical development. *Dev. Rev.* 27, 283–348. doi:10.1016/j.dr.2007.06.006
27. Karatekin, C., Marcus, D. J., and Couperus, J. W. (2007a). Regulation of cognitive resources during sustained attention and working memory in 10- year-olds and adults. *Psychophysiology* 44, 128–144. doi:10.1111/j.1469-8986. 2006.00477.x
28. Karatekin, C., Marcus, D. J., and White, T. J. (2007b). Oculomotor and manual indices of incidental and intentional spatial sequence learning in middle childhood and adolescence. *J. Exp. Child Psychol.* 96, 107–130. doi:10.1016/j.jecp.2006.05.005

29. Laeng, B., Sirois, S., and Gredeback, G. (2012). Pupillometry: a window to the preconscious? *Perspect. Psychol. Sci.* 7, 18–27. doi:10.1177/174569161 1427305
30. Laeng, B., Waterloo, K., Johnsen, S. H., Bakke, S. J., Låg, T., Simonsen, S. S., et al. (2007). The eyes remember it: oculography and pupillometry during recollection in three amnesic patients. *J. Cogn. Neurosci.* 19, 1888–1904. doi:10.1162/jocn.2007.19.11.1888
31. Magimairaj, B. M., and Montgomery, J. W. (2012). Children's verbal working memory: relative importance of storage, general processing speed, and domain-general controlled attention. *Acta Psychol.* 140, 196–207. doi:10.1016/j.actpsy.2012.05.004
32. Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychol. Rev.* 63, 81–97. doi:10.1037/h0043158
33. Peavler, W. (1974). Pupil size, information overload, and performance differences. *Psychophysiology* 11, 559–566. doi:10.1111/j.1469-8986.1974.tb01114.x
34. Siegle, G. J., Steinhauer, S. R., Friedman, E. S., Thompson, W. S., and Thase, M. E. (2011). Remission prognosis for cognitive therapy for recurrent depression using the pupil: utility and neural correlates. *Biol. Psychiatry* 69, 726–733. doi:10.1016/j.biopsych.2010.12.041
35. Simmering, V. R., and Perone, S. (2013). Working memory capacity as a dynamic process. *Front. Psychol.* 3:567. doi:10.3389/fpsyg.2012.00567
36. Steinhauer, S. R., Siegle, G. J., Condray, R., and Pless, M. (2004). Sympathetic and parasympathetic innervation of pupillary dilation during sustained processing. *Int. J. Psychophysiol.* 52, 77–86. doi:10.1016/j.ijpsycho.2003.12.005
37. Unsworth, N., and Engle, R.W. (2007a). On the division of short-term and working memory: an examination of simple and complex span and their relation to higher order abilities. *Psychol. Bull.* 133, 1038–1066. doi:10.1037/0033-2909. 133.6.1038
38. Unsworth, N., and Engle, R. W. (2007b). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychol. Rev.* 114, 104–132. doi:10.1037/0033-295X.114.1.104
39. van der Meer, E., Beyer, R., Horn, J., Foth, M., Bornemann, B., Ries, J., et al. (2010). Resource allocation and fluid intelligence: insights from pupillometry. *Psychophysiology* 47, 158–169. doi:10.1111/j.1469-8986.2009. 00884.x
40. Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—4th Edn., Administration and Scoring Manual.* San Antonio, TX: Harcourt Assessment, Inc.
41. Wilhelm, B., Wilhelm, H., and Lüdtke, H. (1999). "Pupillography: principles and applications in basic and clinical research," in *Pupillography: Principles, Methods and Applications,* eds J. Kuhlmann and M. Böttcher (München: Zuckschwerdt Verlag), 1–11.

Acknowledgements: This study was supported by a James S. McDonnell Foundation Scholar Award and a Tourette Syndrome Association research grant to Silvia A. Bunge. The Bunge lab's first eyetracking study could not have been completed without the help of numerous individuals. We thank Jordan Tharp and Galen Mancino for assistance with programming and set-up of the eyetracker, Jesse Niebaum, Jordan Tharp, and Farida Valji for assistance collecting and compiling data, Se Ri (Sally) Bae for writing up preliminary results for an undergraduate honor's thesis, Dr. Robert DiMartino for helpful discussions, and Marcus Stoiber and Dr. Davide Risso for invaluable consultation regarding data preprocessing.

Appendix 3: Development of Neural Networks Supporting Goal-Directed Behavior

Citation: Johnson, E. L., Munro, S. E. & Bunge, S. A. Development of neural networks supporting goal-directed behavior, in *Minnesota Symposia on Child Psychology: Developing Cognitive Control Processes: Mechanisms, Implications, and Interventions* (eds. P. D. Zelazo & M. Sera) (John Wiley & Sons, Inc., 2014).

Improvements in cognitive control over childhood and adolescence have long been attributed to the "coming online" of prefrontal cortex (PFC). However, fMRI studies reveal that children can engage PFC for cognitive control just as much as adults do—if not more so. We argue that gains in cognitive control are linked most closely to strengthening of the anatomical pathways that enable PFC to communicate efficiently and reliably with distant brain regions.

A3.1. Introduction

One of the most salient facets of cognitive and socioemotional development is the emergence of goal-directed thought and behavior, a concept referred to as cognitive control, executive function, or self-regulation (Bunge & Wright, 2007; Luna, Padmanabhan, & O 'Hearn, 2010; Somerville & Casey, 2010). We begin life with a broad focus of attention, registering incoming stimuli without attempting to filter them. Over childhood, we begin to selectively attend to, manipulate, and act on goal-relevant information; in so doing, we exhibit greater volitional control over our attention, memory, action, and emotions (Bunge & Crone, 2009; Munakata, Snyder, & Chatham, 2012).

Implications of the healthy development of cognitive control are far-reaching. Self-control has been linked to health, wealth, and public safety on a population level (Moffitt et al., 2011). Moffitt and colleagues (2011) followed a cohort of 1,000 children from birth to age 32, and found that physical health, substance dependence, personal finances, and criminal offense in adulthood were predicted by a gradient of self-control in childhood, even after accounting for IQ and social class origins. Recent evidence has also linked failure of self-regulation to a host of neuropsychiatric problems—including attention deficit hyperactivity disorder, addiction, risk-taking behaviors, and conduct disorders—as well as suboptimal performance in school (Fjell et al., 2012). However, the mechanisms by which this important ability develops are not well understood.

A3.2. Cognitive Control in the Developing Brain

Crucial to understanding goal-directed behavior is an understanding of what it means to be goal-directed; it is deliberate, volitional, and governed by top-down processing. We have a limited capacity for processing information, and successful control depends on our ability to focus on relevant information and filter out irrelevant information. The prefrontal cortex (PFC) is integrally involved in top-down processing, as discussed in an influential theoretical review by Miller and Cohen (2001). Following on the theoretical work of Desimone and Duncan (1995), Miller and Cohen proposed that cognitive control rests on the active maintenance of patterns of PFC activity that represent goals, which signal "bias" to other brain structures and permit mappings between inputs, outputs, and internal states appropriate to achieve those goals.

Goal-directed behavior is subserved by widespread neural networks. Several frontal regions—including the anterior cingulate cortex (ACC), inferior frontal gyrus (IFG), ventrolateral PFC (VLPFC), dorsolateral PFC (DLPFC), and frontal eye fields—and regions outside the PFC, including the posterior parietal cortex (PPC), striatum, thalamus, and cerebellum, are associated with inhibitory control and working memory (Hwang & Luna, 2013).

The protracted development of the PFC and parietal cortex into young adulthood (Gogtay et al., 2004) underscores the extended developmental trajectory of cognitive control from early childhood through adolescence. Until now, literature on the development of goal-directed behavior has focused on the PFC; we propose, along with others (Hwang & Luna, 2013), that changes in interactions between PFC and other cortical regions are at least as important as changes in the PFC itself.

The adult state is implicitly viewed as the ideal in developmental research (Poldrack, 2010). Developmental cognitive neuroscience (DCN) asks questions like: Do we perform a particular task better as adults than as children because of increased efficiency of one or more cognitive processes? Or, do we perform the task better as adults because an additional, or different, cognitive process is involved? (Bunge, 2008). For instance, many developmental theories have assumed that in the temporal dynamics of how goal representations are activated, children are like adults, only less skilled (Munakata et al., 2012). However, recent work suggests that children use a qualitatively different, reactive form of cognitive control, which is recruited on an as-needed basis (Andrews-Hanna et al., 2011; Chatham, Frank, & Munakata, 2009) and is supported by neural networks that are differentiable from those supporting the proactive control more likely to be observed in adults (Madsen et al., 2010).

Basic processes underlying cognitive control are evident early in development; gains in goal-directed thought appear to be linked to the reliable engagement of specific processes that fi ne-tune cognitive control (Hwang & Luna, 2013). Maturational changes in white and gray matter enhance the ability of the brain to integrate function between the PFC and other distributed cortical and subcortical regions, which is critical for processing complex information. Underlying these improvements in functional integration is the coupling of neural synchrony across neuronal assemblies. Prefrontally guided top-down connectivity continues to strengthen through early adulthood, supporting flexible executive control of behavior.

Executive functions develop most rapidly during preschool years and undergo another period of relative plasticity in the transition to adolescence (Zelazo & Carlson, 2012). Children become more flexible in attentional control during the preschool years, and at as young as 4 years of age can switch between two sets of rules in a card sorting task when instructed (Hanania & Smith, 2010). However, a hallmark of early executive function is externally driven behavior; without being told what rule to switch to, children tend to perseverate, demonstrating an immature capacity for generating goals internally (Kharitonova & Munakata, 2011; Munakata et al., 2012).

Bunge and Zelazo (2006) proposed that gains in flexible rule use reflect the growth of regions in lateral PFC. Specifically, the ability to represent hierarchical rule systems depends on the development of an increasingly complex hierarchical network of PFC regions. The order of acquisition of rule types—first one rule, then two rules, then two incompatible pairs of rules—

corresponds to the order in which each of the implicated brain regions matures, with the orbitofrontal cortex (OFC) maturing earliest, and the dorsolateral (DLPFC) and rostrolateral PFC (RLPFC) maturing last. Bunge and Zelazo (2006) observed that 8-year-olds showed greater lateral PFC activation for bivalent than univalent rules, and that age differences in the pattern of PFC activation across rule conditions differs between 8- and 12-year-olds, 13- and 17-year-olds, and young adults. Other brain regions showed mature patterns of activation across age-groups, suggesting that age-related gains in rule use between 8 years of age and adulthood are associated with development of the lateral PFC.

Regions within the PFC also interact with multiple brain systems through complex networks, and development marks a period of mass neural network shifts. If we hope to understand why behavior becomes increasingly goal-directed over childhood and adolescence, it is imperative that we discover how the specific brain networks that regulate cognitive and emotional processes emerge over development (Fjell et al., 2012; Somerville & Casey, 2010; Stevens, 2009), and how they are shaped by genetic and environmental factors (Johansen-Berg, 2010). DCN calls for research on the underlying anatomy of pathways that support the development of cognitive control (Fjell et al., 2012). One likely suspect is myelination; reliable and timely transmission of signals is necessary to influence activity in a distal brain region, as is the case in PFC modulation of the parietal cortex and other regions. The child and adolescent brain undergoes substantial myelination and white matter growth (Fields, 2008; Giedd, 2008; Hua et al., 2009; Tamnes, Østby, Fjell, et al., 2010). Another likely suspect is the reorganization of local circuitry, achieved in part via cortical thinning (Gogtay & Thompson, 2010), enabling long-range fiber connections to play a greater role in altering local network dynamics.

A3.3. Developmental Shift from Reactive to Proactive Control

Flexible rule-guided behavior develops gradually, and is essential for success in life (Munakata et al., 2012); it entails the ability to remember rules, switch between them as needed, and implement them in the face of competing information (Wendelken, Munakata, Baym, Souza, & Bunge, 2012). In a recent review, Munakata and colleagues outlined three key developmental transitions toward more flexible behavior (Munakata et al., 2012). First, children develop an increasing ability to overcome habits by engaging cognitive control in response to environmental signals. At first, infants rapidly detect regularities in their environments to bring order to what they see and hear (Romberg & Saffran, 2010), but show limitations in breaking out of habitual behaviors or responding flexibly to novel situations. Young children begin to show signs of overcoming this inflexibility (Hanania & Smith, 2010). Second, children shift from recruiting cognitive control reactively, as needed in the moment, to recruiting cognitive control proactively, in preparation for needing it. Goal representations also provide top-down support to speed responses and to support generalization to new situations. Finally, children become more self-directed. As their capacity for active maintenance increases across development—linked to maturation of lateral PFC regions (Bunge & Zelazo, 2006)—it increasingly becomes sufficient to support proactive control (Munakata et al., 2012).

Chatham and colleagues tested the temporal dynamics of cognitive control in 3.5- versus 8-year-old children on the AX-Continuous Performance Task using high-resolution pupillometry (Chatham et al., 2009). In this task, participants provide a target response to a frequent sequential

pair of stimuli, and a nontarget response to all other pairs. The authors found that 8-year-old children resembled adults in their proactive use of cognitive control, whereas 3.5-year-olds exhibited a qualitatively different, reactive form of cognitive control, responding to events only as they unfold and retrieving information from memory as needed in the moment. These contrasting approaches were evident in distinct profiles of errors, reaction times, and pupillometric indices of mental effort. Three-year-olds exerted more effort—indexed by pupil diameter (Beatty & Lucero-Wagoner, 2000)—after the second stimulus in a pair was presented, an effect consistent with reactive engagement of cognitive control in that moment, whereas 8-year-olds exerted more effort after the first stimulus was presented, an effect consistent with proactive maintenance of this information until it is needed.

With regard to neural networks, proactive control is associated with sustained and/or anticipatory activation of lateral PFC, which reflects the active maintenance of task goals (Braver, 2012). This activity serves as a source of top-down bias that can facilitate processing of expected upcoming events that have a high cognitive demand (Miller & Cohen, 2001). By contrast, reactive control should be reflected in transient activation of lateral PFC, along with a wider network of additional brain regions (Bunge, 2004). In addition, the two control mechanisms should differ in terms of the involvement of the dopaminergic system (Braver, 2012; also Ezekiel, Bosma, & Morton, 2013), which changes through adolescence.

Building on Bunge and Zelazo ’s (2006) account of hierarchical rule representation, a recent study from our group probed the development of arbitrary rule representation (Wendelken et al., 2012). We collected task-related functional magnetic resonance imaging (fMRI) in children ages 8 to 13 years and young adults performing our so-called Nemo task, in which participants must switch flexibly from one task rule to another. One rule involves an arbitrary response mapping—“press the left button if the character is blue”—and the other a nonarbitrary response mapping—“press the left button if the character is facing left.” The task involves three distinct manipulations: (1) rule type: a manipulation of rule representation, comparing arbitrary with nonarbitrary stimulus-response mappings, (2) switching: whether the rule switches or repeats, and (3) incongruency: whether a stimulus would elicit the same response or a different response depending on whether participants are required to make a judgment based on the color or the orientation of the stimulus.

Children performed fairly well overall on the Nemo task, although they were less accurate than adults and exhibited slower responses (Wendelken et al., 2012). Two key questions were addressed: First, are overlapping or distinct networks taxed by these manipulations? Second, are there differences between children and adults in the networks engaged for these various aspects of rule use? Several regions—left DLPFC, left posterior parietal cortex (PPC), and presupplementary motor area (pre-SMA)—were recruited by both the rule representation and the rule-switching manipulations, across age-groups. This is consistent with theoretical frameworks that emphasize the role of task-relevant activation in supporting a variety of executive functions, including shifting, inhibition, updating, and monitoring (Ezekiel et al., 2013; Miller & Cohen, 2001). However, evidence from functional selectivity and temporal dynamics also suggested that adults engage in more proactive control processing while children engage in more reactive control processing.

Fig. A3.1 shows a hypothesized progression of an arbitrary rule trial in the Nemo task, accompanied by brain regions associated with proactive versus reactive control (Fig. A3.1), and by brain regions differentially engaged by adults and children (Fig. A3.2). Importantly, we found a difference in the networks engaged by better and worse performers that applied not only to developmental differences, but also individual differences. Top performers recruited the left frontoparietal network, especially the ventrolateral PFC (VLPFC; Bunge, 2004), posited to be associated with proactive control, while worse performers recruited the right cingulo-operculum network, posited to be associated with reactive control (Dosenbach, Fair, Cohen, Schlaggar, & Petersen, 2008; Madsen et al., 2010).

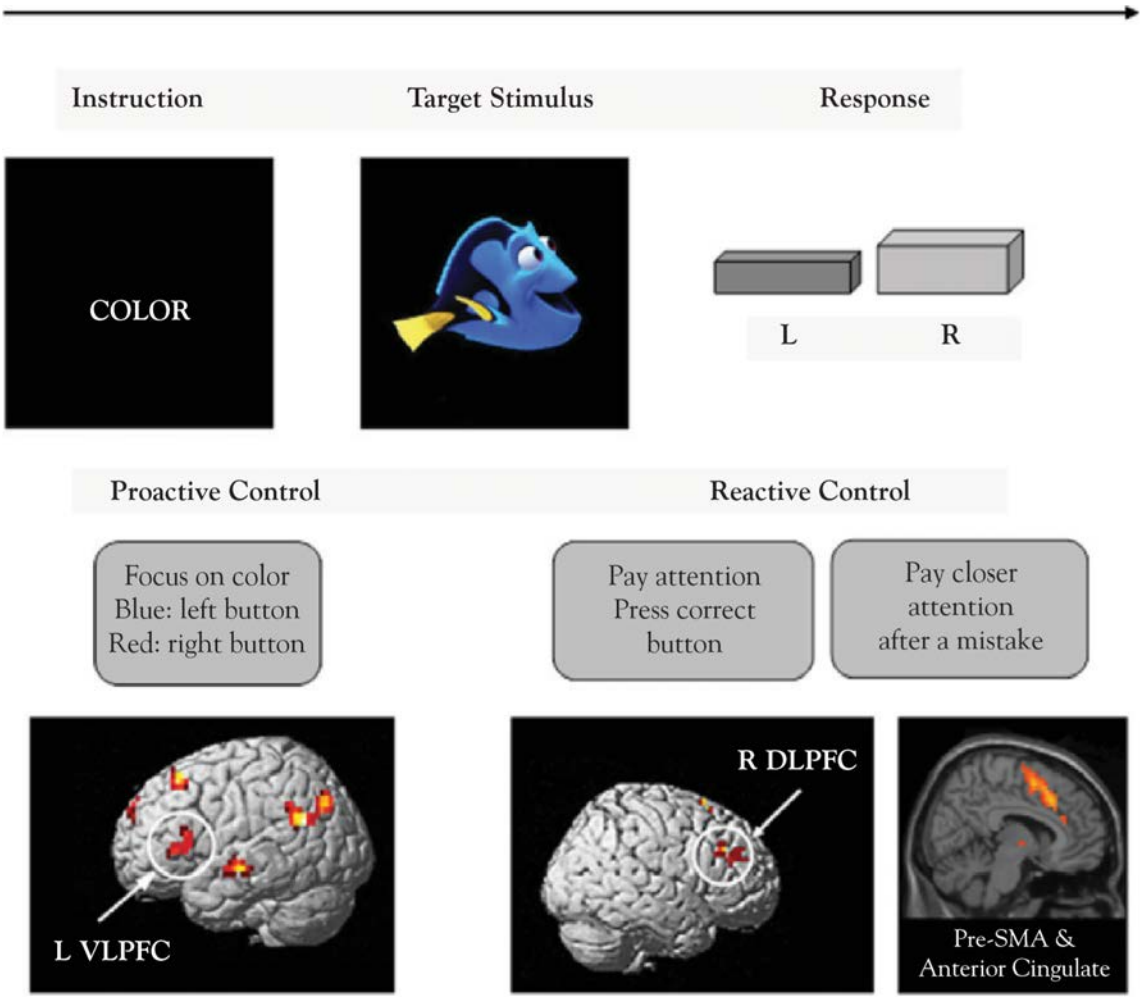

Fig. A3.1. Hypothesized progression of a single trial. From the cognitive control task: the "Nemo" task. Participants view the task instruction, followed by an image of a fish from Finding Nemo (protected by copyright), and respond by pressing one of two buttons. Best-performing participants engaged a brain network previously associated with rule representation, whereas worst-performing participants engaged a network associated with performance monitoring and conflict detection. We hypothesize that the best performers engaged in proactive control, retrieving and holding in mind the currently relevant rule prior to the onset of the target stimulus, whereas the worst performers engaged in reactive control, exhibiting heightened monitoring at the time that a response is required.

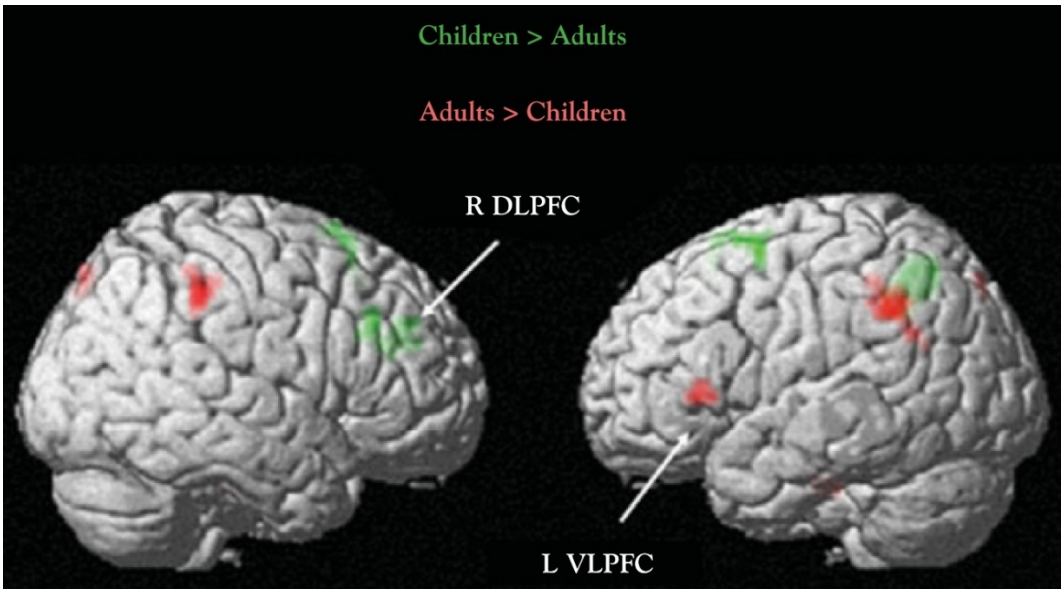

Fig. A3.2. Age-related differences in PFC activation. The left arrow points to areas where activation was greater in children than adults; the right arrow points to areas that were more activated in adults than children. From the cognitive control task: the Nemo task.

The time course of activation in left DLPFC also suggested that the children were more reactive, updating task rules more slowly than adults. In children, but not in adults, activation at the beginning of each trial reflected the previous trial 's rule, rather than the current rule. This finding is consistent with developmental transitions observed in the temporal dynamics of cognitive control (Andrews-Hanna et al., 2011; Chatham et al., 2009; Finn, Sheridan, Kam, Hinshaw, & D'Esposito, 2010).

Employing a hybrid block/event-related fMRI Stroop design in conjunction with self-report measures, Andrews-Hanna and colleagues investigated the shift to proactive control in a cohort of individuals aged 14 to 25 years (Andrews-Hanna et al., 2011). They found that adolescents under-activated a set of brain regions implicated in proactive top-down control, especially left posterior DLPFC (Bunge & Crone, 2009; Gogtay et al., 2004; Paus, 2005). Furthermore, adolescents who exhibited greater activation of the inferior frontal junction (IFJ; which includes posterior DLPFC) exhibited better self-report composite measures of impulse control, foresight, and resistance to peer pressure, and tended toward more successful Stroop task performance. Although no differences in activation were observed between adults and adolescents for the ACC, heterogeneous patterns of ACC/pre-SMA activation within the adolescent group suggest a compensatory reactive response.

In contrast, adults exhibited the opposite relationship with Stroop performance, such that adults who activated IFJ to a lesser degree trended toward more successful Stroop behavior (Andrews-Hanna et al., 2011). The relationship between prefrontal activity and age was curvilinear, peaking approximately at age 21 years and decreasing thereafter.

Taken all together, these results suggest that children are more likely than adults to maintain prior rule information when it is no longer relevant, and to retrieve current rule information

reactively rather than maintaining it proactively (Munakata et al., 2012). Results in adolescents suggest that maturation of cognitive control may be partly mediated by earlier development of neural systems supporting reactive control and delayed development of systems supporting proactive control (Andrews-Hanna et al., 2011). Future work is needed to determine whether the developmental progression from reactive to proactive control is governed by qualitatively distinct mechanisms or by a continuous developmental shift (Chatham et al., 2009).

A3.4. Structural Development

Recent findings from developmental neuroimaging studies suggest that the enhancement of cognitive processes during development results from a fi ne-tuning of the structural and functional organization of brain. But the mechanisms by which this takes place are not yet understood. Looking at regional topological properties and inter-regional connectivity, Khundrakpam and colleagues found a “time window of plasticity” during late childhood, which they suggested may accommodate the changes that come with pubertal development (Khundrakpam et al., 2012). Importantly, they report early maturation of primary sensorimotor regions and protracted development of higher-order association and paralimbic regions, which have been linked to cognitive control (Hwang & Luna, 2013). However, studies that have compared structural and functional MRI measures of brain development have not provided evidence for a simple relationship between them (Lu et al., 2009).

A3.4.1. Developmental Changes in Cortical Thickness

Cortical thickness is the distance from the gray matter/white matter boundary to the outer surface of the brain (i.e., the pial surface), which likely reflects the number of cells within cortical columns (Fjell et al., 2012). Thanks to the development of powerful analytic tools for measuring longitudinal changes in brain structure, we now have detailed information about within-person changes in cortical thickness over development (Gogtay & Thompson, 2010; Tamnes, Østby, Fjell et al., 2010). These data reveal piecemeal cortical thinning over childhood and adolescence, with association cortices—including but not limited to PFC—maturing later than primary sensory cortices. Within PFC, medial and ventral regions undergo thinning most quickly, such that DLPFC matures later than other prefrontal subregions. In another analysis, Fjell et al. (2012) found that developmental gains in cognitive control, as measured by a flanker task in a cohort of 725 individuals ages 4 to 21, were associated with changes in surface area in the anterior cingulate (ACC), an area that has been previously linked to impulse, attention, and executive problems across a range of neurodevelopmental disorders. Specifically, surface area of the right caudal ACC accounted for a significant proportion of the variance in cognitive performance.

Cortical thinning is likely to reflect multiple changes at the cellular level, including decreased gray matter as a result of synaptic pruning and increased white matter as a result of myelination and/or increased axon diameter (Giedd, 2008; Tamnes, Østby, Fjell et al., 2010). Indeed, recent structural MRI analyses by Gogtay and Thompson (2010) and Hua et al. (2009) suggest that there is white matter growth underlying areas of thinning gray matter. Longitudinal studies of individuals ages 3 to 30 years have demonstrated general patterns of peaks of gray matter in childhood followed by declines in adolescence, increases in long-range structural and functional connectivity, and a shift of activation from limbic and subcortical regions to the frontal lobe in

cognitive tasks (Giedd, 2008; also Finn et al., 2010). Histological studies have shown that sensory areas may develop first, followed by a longer trajectory in frontal executive regions linked to the late development of executive function (Casey, Tottenham, Liston, & Durston, 2005).

The functional significance of these changes in cortical thickness is not yet clear. While it is possible to find evidence for positive and/or negative relationships between cortical thickness and cognitive performance, recent work from our lab suggests a positive relationship between performance on a complex reasoning task and thinning in the inferior parietal lobule (IPL) across children and adolescents (Wendelken, O 'Hare, Whitaker, Ferrer, & Bunge, 2011). Karama and colleagues previously found a positive relationship between cognitive ability and thinning across most multimodal association areas (Karama et al., 2009). These brain-behavior relationships can also be influenced by age, gender, and their interaction (Christakou et al., 2009), although such differences are not always linked to differences in cognitive performance (Lenroot & Giedd, 2010).

A3.4.2. Developmental Changes in White Matter Pathways

The development of goal-directed behavior partly reflects the maturation of white matter pathways; for instance, white matter maturation has been shown to explain additional variance in cognitive control beyond cortical thinning (Fjell et al., 2012). Importantly, tracts that provide connections between frontal and other cortical and subcortical regions demonstrate the most protracted developmental trajectory, occurring in parallel with changes in gray matter (Hwang & Luna, 2013). The large number of reciprocal connections between PFC and other cortical regions position it nicely to coordinate, influence, and integrate information needed for executive control. As these connections develop to enhance neuronal transmission, the ability to form functional networks to support complex function improves, consistent with gains in executive control.

The advent of diffusion tensor imaging (DTI) has made it possible to measure within-individual changes in white matter tracts over development, and their relationship to changes in cognition. DTI provides an indirect measure of white matter tracts in vivo in the human brain (Fields, 2008). A diffusion-weighted MRI scan is sensitive to the movement of protons in the brain, particularly those within water molecules. Water molecules in white matter diffuse preferentially along axon bundles because the myelin sheath surrounding the axons impedes their diffusion across a tract. Water molecules that have high directionality are said to exhibit anisotropic diffusion. Two DTI measures are reported below: fractional anisotropy (FA), a widely used measure of white matter coherence, and perpendicular diffusivity, thought to be sensitive to levels of myelination.

Although white matter maturation takes place throughout the brain, it is possible to link cognitive performance to the strength of specific tracts (Johansen-Berg, 2010; Madsen et al., 2010; Niogi, Mukherjee, Ghajar, & McCandliss, 2010; Olson et al., 2009; Uddin, Supekar, Ryali, & Menon, 2011). Johansen-Berg (2010) found that, independent of age, white matter anatomy was linked with cognitive skills in healthy adults.

As shown in Fig. A3.3, McCandliss and colleagues provided evidence of a triple dissociation in the brain-behavior relationships of three white matter tracts (Niogi et al., 2010). They examined performance across three dissociable functional components of attention—alerting, orienting, and conflict processing—identified by the Attention Network Task (ANT). The ANT is a simple computerized task that measures the speed and accuracy with which a participant can press one of two buttons to indicate whether the central arrow in a row of visually presented arrows points left or right. This task has three critical manipulations that tax different aspects of attention. To measure the alerting response, researchers measure how much better participants perform when the stimulus array is preceded by a visual warning cue versus when it is not. To measure the orienting response, they measure how much better they respond when the visual cue indicates where on the screen the stimulus array will appear versus when it is not spatially predictive. Finally, to measure executive/conflict processing, researchers measure how much better the participant performs when the arrows in the stimulus array all point in the same direction versus when the flanking stimuli point in the opposite direction from the central, target arrow.

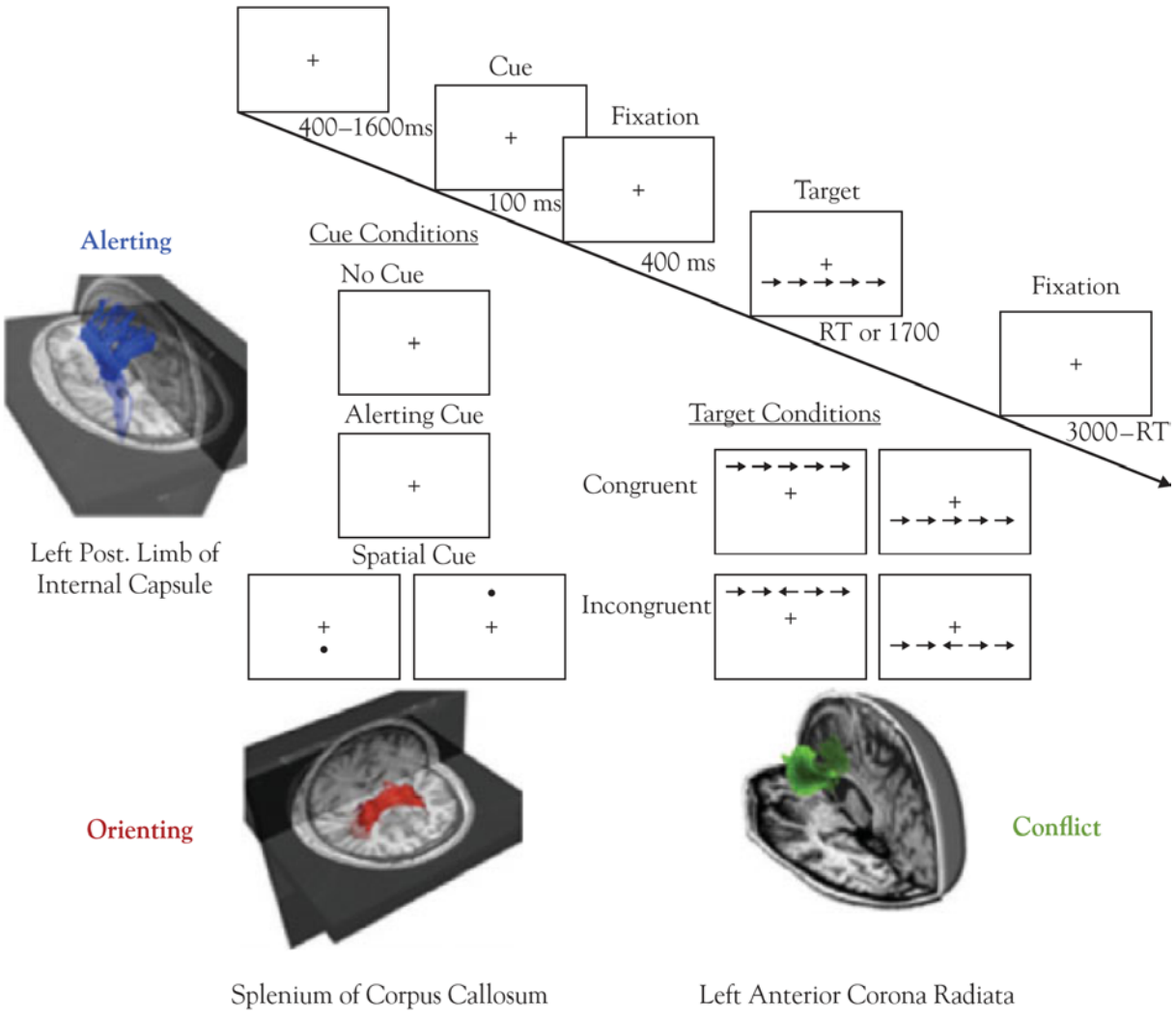

Fig. A3.3. A triple dissociation is shown in the interindividual relationships between white matter integrity of three tracts and cognitive performance on the three components of the

Attention Network Task (ANT): alerting, orienting, and executive function (specifically, conflict resolution). Fig. modified with permission from Niogi et al., 2010.

Interindividual performance on each functionally distinct component was differentially correlated with coherence in a distinct set of white matter tract regions. Correlations were found between alerting and the left anterior limb of the internal capsule, orienting and the splenium of the corpus callosum, and conflict and the left anterior corona radiata. Analyses revealed a triple dissociation providing evidence of three anatomically and functionally separable networks.

It stands to reason, then, that the development of these and other white matter tracts may be partly accountable for the development of cognitive skills such as goal-directed behavior. Whole-brain analyses in individuals aged 9 to 23 years revealed that success on a delay discounting task—that is, less impulsive performance—was associated with increased FA in tracts in and across the bilateral frontal and temporal lobes (Olson et al., 2009). In a stop-signal task, faster response inhibition was associated with higher FA in the right inferior frontal gyrus (IFG) and right presupplementary motor area (pre-SMA) in typically developing children (Madsen et al., 2010). Furthermore, individuals with high verbal abilities may show accelerated white matter development compared to the steadier and prolonged development observed in their average-ability counterparts (Tamnes, Østby, Walhovd, et al., 2010).

Lebel and colleagues recently reported the results of longitudinal studies of white matter maturation that demonstrate clear age-related microstructural changes throughout the brain in individuals aged 5 years through adulthood (Lebel et al., 2012; Lebel & Beaulieu, 2011). The corpus callosum and fornix mature in early childhood, reaching peak FA by young adulthood. In contrast, they found that major tracts that connect PFC with posterior regions and have been hypothesized to support cognitive control, namely the cingulum, uncinate fasciculus, and superior longitudinal fasciculus, develop slowly. Because these changes in FA were driven by perpendicular diffusivity, results suggest that they result from changes of myelination and/or axonal density (Lebel et al., 2012). The researchers also found volume increases in several association tracts post-adolescence (Lebel & Beaulieu, 2011). Corroborating evidence for the prolonged maturation of association tracts comes from an earlier cross-sectional study (Lebel, Walker, Leemans, Phillips, & Beaulieu, 2008). Other findings provide converging anatomical data suggesting that developmental gains in goal-directed behavior in adolescence may be associated with structural changes enhancing long-distance connections, coupled with synaptic pruning in the cortex (Giorgio et al., 2010).

A3.5. Developmental Changes in Functional Networks

While techniques like DTI help us to characterize the development of white matter tracts, research on patterns of correlated brain activation provide a complementary picture of developing cortical networks. Functional connectivity analyses identify regions with strongly correlated patterns of functional MRI activation over time, either during performance of a cognitive task or at rest. Brain regions that are not directly connected to one another via white matter tracts may nonetheless act in concert as part of a distributed network. Conversely, two brain regions that are anatomically connected may not yet be fully integrated into a shared network (Barnes et al., 2012; Biswal et al., 2010; Supekar et al., 2010). Therefore, a promising

approach is to integrate these complementary measures of brain connectivity (Rykhlevskaia, Gratton, & Fabiani, 2008) to address inquiries regarding how and why changes in the "wiring" of neural networks promote the development of cognitive control (Stevens, 2009).

A3.5.1. Developmental Shift from Local to Long-Range Functional Connectivity

Imaging the brain during rest (i.e., in the absence of task demands) reveals low-frequency fluctuations in the fMRI signal that are temporally correlated across regions presumed to be functionally related. Reproducibility across resting-state fMRI datasets suggests that the human functional connectome has a common architecture, yet each individual 's connectome is unique; age and gender are notable as significant determinants (Biswal et al., 2010). Over the past few years, there have been a number of studies characterizing changes in network connectivity in typically and atypically developing populations (Fan et al., 2011; Gao et al., 2009; Supekar et al., 2010; Thomason et al., 2008; for review, see Stevens, 2009).

These developmental changes in resting-state functional connectivity are sufficiently robust that multivariate pattern analysis can be used to predict an individual 's age with a fairly high degree of accuracy (Dosenbach et al., 2010). Wang and colleagues (2012) found that these age-related changes in interregional functional connectivity exhibited spatially and temporally specific patterns over the lifespan (Wang, Su, Shen, & Hu, 2012). Functional connectivity decreased linearly in the sensorimotor system, increased linearly in the emotion system, and followed a quadratic trajectory—with increases through childhood and early adulthood, followed by decreases later in life—in systems associated with higher-order cognition, from childhood through old age.

In a recent functional connectivity MRI study, Barnes and colleagues (2012) demonstrated that cortical parcellation of the left lateral parietal cortex—part of a left lateral frontoparietal network correlated with proactive control (Wendelken et al., 2012)—in school-aged children resembled that of adults. However, age-related differences were found in its functional connectivity with other brain regions, suggesting that structure and function in this region mature along different developmental trajectories, with functional connectivity following a prolonged trajectory (Barnes et al., 2012). In contrast, Supekar and colleagues (2010) found that functional connectivity in children can reach adult-like levels despite immature structural connectivity. They proposed that the prolonged maturation of the posterior cingulate cortex (PCC)-medial PFC structural connectivity may be linked with development of the self-related and social-cognitive functions that emerge during adolescence (Supekar et al., 2010).

The maturation of executive function is dually supported by functional specialization—regional neural support of specific processes—and functional integration—large-scale neural network support (Hwang & Luna, 2013). One of the central developmental findings in recent functional connectivity work is the progression from short-range connections within cortical areas to longer-range cortico-cortico connections (Fair et al., 2009; Jolles, van Buchem, Crone, & Rombouts, 2011; Stevens, Pearlson, & Calhoun, 2009). As children mature, short-range functional connections become weaker, and long-range connections strengthen (Church et al., 2009). At first, the distributed network is composed of many weak connections, but as children move into adolescence and adulthood, functional connections tend to become stronger but

sparser, reflecting the increasing specificity of emerging functional networks (Fair et al., 2009; Supekar et al., 2010). Jolles and colleagues investigated whole-brain functional connectivity in middle-aged children and young adults (Jolles et al., 2011). Interestingly, they found that while children and adults displayed similar patterns of functionally connected regions, the size (number of voxels) and strength (correlation value) of connectivity differed between brain regions into young adulthood. Regions associated with higher cognitive and emotional functions became more tightly coupled with age, while connectivity between regions linked to basic visual and sensorimotor functions showed the opposite effect (Jolles et al., 2011).

A study by Barber and colleagues (2013) showed developmental differences between late childhood and adulthood in the "default mode" network, comparing task-positive and task-negative regions, that is, the regions that are engaged and disengaged during task performance relative to when the participant is asked to rest, respectively (Barber, Caffo, Pekar, & Mostofsky, 2013). They found that task-positive regions showed greatest age-related discrepancy in the left DLPFC, a region strongly implicated in cognitive control. In contrast, task-negative regions, posited to play a role in social cognition and self-referential thought, showed greatest age-related differences in medial PFC and right parahippocampal gyrus. Connections between the task-positive and task-negative regions also displayed developmental differences. Importantly, there was a significant relationship between anticorrelations—that is, the extent of inverse correlation between activations in task-positive versus task-negative regions—in a swath of task-negative regions (right anterior insula, right IFG, right PCC, and bilateral parietal cortex), exhibited in adults, and successful inhibitory control performance on Go/No-Go tasks. This result suggests that the development of certain anticorrelations into adulthood supports mature inhibitory control (Barber et al., 2013).

A3.5.2. Functional Connectivity in the Development of Cognitive Control

A lateral frontoparietal neural network underlies goal-directed behavior across diverse contexts (Cole, Yarkoni, Repovs, Anticevic, & Braver, 2012; Wendelken et al., 2012). Cole et al. (2012) have pointed to global connectivity—that is, the mechanisms by which regions in the frontoparietal network might coordinate other cortical networks—as the driving force of cognitive control. They found that a lateral PFC region 's activity exhibited global connectivity and predicted performance in a high-demand working memory task. Furthermore, individual differences in lateral PFC global connectivity were related to individual differences in fluid intelligence (Cole et al., 2012).

Work from our lab has shown age-related changes in the temporal dynamics of DLPFC activation, such that children appeared to update rules more slowly than adults, engaging in reactive rather than proactive control (Wendelken et al., 2012; described earlier). Ezekiel and colleagues (2013) recently investigated a similar possibility, consistent with our finding of an age-related shift from the cingulo-operculum to the frontoparietal network (see Fig. A3.1): Age is associated with changes in the functional integration of lateral PFC with a larger cognitive control network (Ezekiel et al., 2013). They tested middle-school-aged children and adults on a card sort/switch cognitive control task. Results demonstrated that adults engaged regions within a "cognitive control" network, including bilateral DLPFC, right IFG, ACC/medial PFC, inferior parietal cortex, and the ventral tegmental area (VTA). Children showed engagement of a

different network; regions included anterior frontal gyri, bilateral rostrolateral PFC, right anterior insula, and left posterior temporal cortex. These findings are consistent with observations that children and adults may both have relatively efficient systems for cognitive processing, but that they solve problems in different ways (Fair et al., 2009).

Signals for updating representations in lateral PFC—a crucial aspect of a switch task like this one and our Nemo task (Wendelken et al., 2012)—originate in dopamine neurons in ventral tegmental area (VTA), which Ezekiel et al. (2013) found to be more strongly connected to lateral PFC in adults than children. They hypothesized that functional connectivity between lateral PFC and VTA is associated with the speed with which rule representations may be updated in lateral PFC (Ezekiel et al., 2013). Taken together, it is suggested that complex cognitive operations may be emergent products of rapid bidirectional interactions between functionally specialized brain regions; and shift or improve concurrently with various aspects of neural development.

Broadly, the differences between children and adults in patterns of functional connectivity are consistent with the trajectories of gray and white matter development. At the same time that local functional connections within the cortex are weakening, cortical gray matter is thinning—and, as long-range white matter tracts are getting stronger, so, too, is long-range functional connectivity. Supekar et al. (2010) found that some, but not all, changes in functional connectivity had obvious anatomical correlates. Thus, these structural and functional measures provide valuable and complementary views of brain development.

A3.6. Developmental Cognitive Neuroscience and the Study of Cognitive Control

As discussed earlier, developmental cognitive neuroscience (DCN) has begun to uncover the neural mechanisms governing the development of goal-directed behavior. First, fMRI has revealed that patterns of activation observed in the performance of cognitive control tasks in children versus adults are consistent with those observed in reactive versus proactive control. Second, structural MRI has shown that cortical thinning in association cortices is correlated with gains in behavioral performance. Third, DTI and functional connectivity analyses have shown that projections between the PFC and other cortical regions shift from local to long-range with development. Taken together, DCN has enabled us to show that gains in cognitive control observed with age are subserved by a confluence of factors in the developing brain.

The field of cognitive neuroscience, which straddles the gap between the mind and brain, is the right level at which to begin to understand how cognitive developmental trajectories are influenced by such important factors as genetic variants, hormonal changes during puberty, schooling, and socioeconomic and cultural contexts. Indeed, behavioral findings have often been considered controversial until accompanied by discovery of an underlying biological mechanism (Diamond & Amso, 2008). Neuroimaging has allowed, for instance, the delineation of how task-related and resting-state brain networks develop through strengthening and weakening of functional connections (e.g., Fair et al., 2009).

In both the temporal and spatial domains, the scale at which we examine brain development is at once an important strength and an important limitation of DCN. In the spatial domain, the fact that we can take neural measurements across the entire brain means that we can identify brain

networks: sets of tightly coupled brain regions that underlie specific cognitive processes. Until recently, the bulk of DCN research has focused on specific brain regions of a priori interest—for example, studies focusing on maturational changes in the PFC that underlie improvements in cognitive control. Current DCN research is focusing more and more on analyses of neural networks. In the temporal domain, the fact that we can measure brain structure and function in vivo at multiple times in the life of an individual child means that we can examine true developmental change, rather than merely extrapolating from comparisons between samples of children from different age-groups. The possibility for longitudinal brain research is a clear advantage of noninvasive imaging techniques that has yet to be used to its full potential. We briefly discuss next promising future directions for DCN in the study of the development of goal-directed behavior.

A3.6.1. Early Brain Development

Although most of the behavioral literature on cognitive development has focused on the period of rapid changes observed during early childhood, most of the DCN studies to date have, for practical reasons, focused on older children and adolescents (Poldrack, 2010). In recent years, researchers have refined pediatric imaging protocols that make it possible to obtain high-quality structural and functional MRI data from infants (Dehaene-Lambertz et al., 2010; Fan et al., 2011; Gao et al., 2009) and young children (Cantlon, Pinel, Dehaene, & Pelphrey, 2011; Nordahl et al., 2008). This advance makes it possible to measure the functional organization of the newborn brain, and to examine the neural changes that support the emergence of new cognitive abilities over early childhood.

DCN has revealed that even at 2 weeks of age, infants exhibit spontaneous resting-state activity in some of the same regions as adults (Gao et al., 2009). At 1 month, infants already display modular functional connectivity (Fan et al., 2011), and by 2 years of age, children have the beginnings of adult-like connectivity and modular organization (Fan et al., 2011; Gao et al., 2009). Friedman, Miyake, Robinson, and Hewitt (2011) showed, in a longitudinal study of 950 twins, that self-restraint in toddlers predicted individual differences in three executive functions—inhibition, updating, and shifting—in late adolescence (Friedman, Miyake, Robinson, & Hewitt, 2011). Self-restraint was measured by showing the children an attractive toy and instructing them not to touch it for 30 seconds; capacity for self-restraint was measured by how long the toddler waits before touching the toy. The twin models indicated relative contributions of genetics and environmental factors to gains in these components of executive function; results suggested a biological relationship between individual differences in self-restraint and executive functions that begins in infancy and persists through adolescence.

A3.6.2. Genetic, Hormonal, and Environmental Influences on Brain Development

An important next step in DCN is the elucidation of genetic, hormonal, and environmental factors that interact to influence brain and cognitive development. There has been research on gene x environment influences on behavior during development (Wiebe et al., 2009). Until recently, this work has left the brain out of the equation, but DCN is beginning to examine genetic and/or environmental influences on brain structure and function (Casey, Soliman, Bath, & Glatt, 2010; Chiang et al., 2009; Hackman & Farah, 2009; Lenroot et al., 2009; Thomason et

al., 2010), and is also beginning to look at the influence of changing pubertal hormone levels (Blakemore, Burnett, & Dahl, 2010).

Links between cognitive control and dopamine suggest that the neurodevelopment of cognitive control should interact with age-related variability in the dopaminergic system (Braver, 2012; Ezekiel et al., 2013). Indeed, gene × environment effects—including variation in a polymorphism related to D2 dopamine receptor transmission—have been shown to differentially impact slow-developing functions such as self-regulation depending on the developmental period (Wiebe et al., 2009). Another study showed that typical white matter development was related to a common genetic variant in the dopamine signaling pathway, COMT, that influences dopamine levels in PFC (Thomason et al., 2010). Using a visual working memory task, Dumontheil et al. (2011) reported age × genotype interactions in the effects of the COMT genotype observed in the intraparietal sulcus (IPS), with greater gray matter volumes bilaterally and greater right IPS activation in the Val/ Val group compared with the Met carriers. Genetic maps have also revealed the complex heritability of white matter integrity, cortical thickness, and even IQ (Chiang et al., 2009; Lenroot et al., 2009). Chiang and colleagues (2009) found that white matter integrity was highly heritable in a number of regions, including bilateral frontal, bilateral parietal, and left occipital lobes, and that common genetic factors mediated the correlation between IQ and white matter integrity. Lenroot et al. (2009) investigated gender differences in brain volume, and found, among other things, that males and females exhibited dissociable patterns of activation on cognitive tasks, without differences in performance.

A3.6.3. Longitudinal Research

To examine—and interrelate—developmental trajectories for cognition, brain structure, and brain function, it is necessary to acquire data at multiple time-points per individual. Longitudinal research can provide important insights regarding typical and atypical cognitive development (Reichenberg et al., 2010). Although there are few published longitudinal MRI studies of children (Giedd et al., 2009; Gogtay & Thompson, 2010), and even fewer that include functional as well as structural measures (Fan et al., 2011; Shaw et al., 2009), a number of research groups are conducting this important work now.

Durston and colleagues conducted the first combined cross-sectional and longitudinal fMRI study on the development of cognitive control (Durston et al., 2006). They directly compared between-group measurements of brain activation with within-person changes in brain function during performance of a Go/No-Go task. These two analyses yielded somewhat different results in the lateral PFC, with only longitudinal findings showing attenuated activation in DLPFC areas and increased activation in focal VLPFC areas. These data underscored the need for further longitudinal brain imaging studies.

The past few years have seen several promising results from longitudinal behavioral and brain imaging research involving children and adolescents. Childhood self-control has been found to predict physical health, substance dependence, personal finances, and criminal behavior in adulthood (Moffitt et al., 2011), and early childhood self-restraint to predict executive functions in late adolescence (Friedman et al., 2011). Within-person tracking of brain structure and function shows peaks and dips in gray matter volume and white matter integrity over

development (Giedd, 2008; Giorgio et al., 2010; Lebel & Beaulieu, 2011; Lebel et al., 2012), and a shift in neural networks engaged during performance of cognitive tasks (Finn et al., 2010). Finally, Moriguchi and Hiraki (2011) demonstrated how PFC engagement interacted with performance on cognitive shifting tasks in children studied at age 3 and again at 4 years of age. They found that better-performing children at age 3 showed significant activation of right inferior PFC, and that better-performing children at age 4 showed this activation bilaterally. These intriguing results underscore the importance and future potential of the longitudinal method to address the link between cognitive and neural development (Moriguchi & Hiraki, 2011).

A3.7. Considerations and Future Directions in Developmental Cognitive Neuroscience

As neuroimaging methods have grown more sophisticated, so, too, has DCN. Many new investigators have joined the field, including individuals with strong statistical and computational backgrounds. Researchers have developed procedures that facilitate pediatric fMRI data acquisition and analysis (Fonov et al., 2011; Ghosh et al., 2010), and have addressed many of the basic questions and concerns surrounding pediatric MRI methodology (Church, Petersen, & Schlaggar, 2010; Luna, Velanova, & Geier; 2010). Many have also moved on to tackle greater challenges, such as the acquisition of longitudinal MRI data (Fan et al., 2011; Lebel et al., 2012; Moffitt et al., 2011; Moriguchi & Hiraki, 2011; Shaw et al., 2009) and the integration of multiple measures in the study of brain development (Fjell et al., 2012; Paus, 2010; Thomason et al., 2010).

We have learned a lot over the past few years about the typical developmental trajectory of cortical thickness and white matter tracts. However, we still know little about how these changes relate to developmental changes or individual differences in brain function or behavior, and imaging the developing brain continues to pose unique challenges. With regard to functional connectivity MRI, Power and colleagues recently revealed that subjects' head motion in the scanner causes systematic but spurious correlations between brain regions (Power, Barnes, Snyder, Schlaggar, & Petersen, 2012). Specifically, they found that subject motion produces substantial changes in time-course data; many long-distance correlations may appear less robust than they are and short-distance correlations may appear more robust than they are. This is especially pertinent for DCN, not only because functional network analyses are integral to the study of cognitive development, but also because motion artifacts are more pronounced in children than adults. The authors also noted explicitly that connections between lateral PFC and the anterior cingulate—which carry import for successful goal-directed behavior, as described in this chapter—may appear distorted from subject motion (Power et al., 2012).

This confound is important to consider, both because motion artifacts are more pronounced in children than adults, and because long-distance connections are thought to mature more slowly than short-distance ones. It is precisely the slow development of long-distance connections that has been linked to delayed maturation of the ability of one region to influence neural activity in distal regions. Thus, DCN research focused on functional brain networks must address this potential confound.

It is also important to note that most of the structural data purporting to characterize typical brain development are based on a fairly homogeneous sample of children, despite efforts to diversify the samples. It is necessary to cast a wider net to determine how well such findings generalize to children across a wide range of intellectual abilities, and from a variety of socioeconomic and cultural backgrounds.

There is still much to be discovered regarding the interplay of external and internal factors on cognitive and brain development, and an endeavor of this level of complexity necessitates a multidisciplinary approach with large research teams, large sample sizes, and data collection at multiple time points per individual. At the same time, it will be important in the coming years for DCN to strike the right balance between data-driven research—so-called discovery science (Biswal et al., 2010)—and hypothesis-driven research grounded in theories of cognitive development.

A3.8. References

1. Andrews-Hanna, J. R., Mackiewicz Seghete, K. L., Claus, E. D., Burgess, G. C., Ruzic, L., & Banich, M. T. (2011). Cognitive control in adolescence: Neural underpinnings and relation to self-report behaviors. *PLoS ONE,* 6 (6), e21598.
2. Barber, A. D., Caffo, B. S., Pekar, J. J., & Mostofsky, S. H. (2013). Developmental changes in within- and between-network connectivity between late childhood and adulthood. *Neuropsychologia*, 51, 156–167.
3. Barnes, K. A., Nelson, S. M., Cohen, A. L., Power, J. D., Coalson, R. S., Miezin, F. M., ... Schlaggar, B. L. (2012). Parcellation in left lateral parietal cortex is similar in adults and children. *Cerebral Cortex*, 22, 1148–1158.
4. Beatty, J., & Lucero-Wagoner, B. (2000). The pupillary system. In J. Cacioppo, L. Tassinary, & G. Berntson (Eds.), *Handbook of psychophysiology* (2nd ed., pp. 142–162). New York, NY: Cambridge University Press.
5. Biswal, B. B., Mennes, M., Zuo, X.-N., Gohel, S., Kelly, C., Smith, S. M., ... Milham, M. P. (2010). Toward discovery science of human brain function. *Proceedings of the National Academy of Sciences, USA*, 107 (10), 4734–4739.
6. Blakemore, S.-J., Burnett, S., & Dahl, R. E. (2010). The role of puberty in the developing adolescent brain. *Human Brain Mapping*, 31, 926–933.
7. Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, 16, 106–113.
8. Bunge, S. A. (2004). How we use rules to select actions: A review of evidence from cognitive neuroscience. *Cognitive, Affective, & Behavioral Neuroscience*, 4, 564–579.
9. Bunge, S. A. (2008). Changing minds, changing brains. *Human Development,* 51, 162–164.
10. Bunge, S. A., & Crone, E. A. (2009). Neural correlates of the development of cognitive control. In J. Rumsey & M. Ernst (Eds.), *Neuroimaging in developmental clinical neuroscience* (pp. 22–37). Cambridge, UK: Cambridge University Press.
11. Bunge, S. A., & Wright, S. B. (2007). Neurodevelopmental changes in working memory and cognitive control. *Current Opinion in Neurobiology*, 17, 243–250.
12. Bunge, S. A., & Zelazo, P. D. (2006). A brain-based account of the development of rule use in childhood. *Current Directions in Psychological Science,* 15, 118–121.

13. Cantlon, J. F., Pinel, P., Dehaene, S., & Pelphrey, K. A. (2011). Cortical representations of symbols, objects, and faces are pruned back during early childhood. *Cerebral Cortex*, 21, 191–199.
14. Casey, B. J., Soliman, F., Bath, K. G., & Glatt, C. E. (2010). Imaging genetics and development: Challenges and promises. *Human Brain Mapping*, 31, 838–851.
15. Casey, B. J., Tottenham, N., Liston, C., & Durston, S. (2005). Imaging the developing brain: What have we learned about cognitive development? *Trends in Cognitive Sciences*, 9, 104–110.
16. Chatham, C. H., Frank, M. J., & Munakata, Y. (2009). Pupillometric and behavioral markers of a developmental shift in the temporal dynamics of cognitive control. *Proceedings of the National Academy of Sciences, USA*, 106 (14), 5529–5533.
17. Chiang, M.-C., Barysheva, M., Shattuck, D. W., Lee, A. D., Madsen, S. K., Avedissian, C., ... Thompson, P. M. (2009). Genetics of brain fiber architecture and intellectual performance. *Journal of Neuroscience*, 29, 2212–2224.
18. Christakou, A., Halari, R., Smith, A. B., Ifkovits, E., Brammer, M., & Rubia, K. (2009). Sex-dependent age modulation of frontostriatal and temporo-parietal activation during cognitive control. *NeuroImage*, 48, 223–236.
19. Church, J. A., Fair, D. A., Dosenbach, N. U. F., Cohen, A. L., Miezin, F. M., Petersen, S. E., & Schlaggar, B. L. (2009). Control networks in paediatric Tourette Syndrome show immature and anomalous patterns of functional connectivity. *Brain*, 132, 225–238.
20. Church, J. A., Petersen, S. E., & Schlaggar, B. L. (2010). The "'task b problem'" and other considerations in developmental functional neuroimaging. *Human Brain Mapping*, 31, 852–862.
21. Cole, M. W., Yarkoni, T., Repovs, G., Anticevic, A., & Braver, T. S. (2012). Global connectivity of prefrontal cortex predicts cognitive control and intelligence. Journal of Neuroscience, 32, 8988–8999.
22. Dehaene-Lambertz, G., Montavont, A., Jobert, A., Allirol, L., Dubois, J., Hertz-Pannier, L., & Dehaene, S. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain & Language*, 114, 53–65.
23. Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, 18, 193–222.
24. Diamond, A., & Amso, D. (2008). Contributions of neuroscience to our understanding of cognitive development. *Current Directions in Psychological Science*, 17, 136–141.
25. Dosenbach, N., Nardos, B., Cohen, A., Fair, D., Power, J. D., Church, J. A., ... Schlaggar, B. L. (2010). Prediction of individual brain maturity using fMRI. *Science*, 329, 1358–1361.
26. Dosenbach, N. U. F., Fair, D. A., Cohen, A. L., Schlaggar, B. L., & Petersen, S. E. (2008). A dual-networks architecture of top-down control. *Trends in Cognitive Sciences*, 12, 99–105.
27. Dumontheil, I., Roggeman, C., Ziermans, T., Peyrard-Janvid, M., Matsson, H., Kere, J., & Klingberg, T. (2011). Influence of the COMT genotype on working memory and brain activity changes during development. *Biological Psychiatry*, 70, 222–229.
28. Durston, S., Davidson, M. C., Tottenham, N., Galvan, A., Spicer, J., Fossella, J. A., & Casey, B. J. (2006). A shift from diffuse to focal cortical activity with development. *Developmental Science*, 9, 1–8.

29. Ezekiel, F., Bosma, R., & Morton, J. B. (2013). Dimensional change card sort performance associated with age-related differences in functional connectivity of lateral prefrontal cortex. *Developmental Cognitive Neuroscience*, 5, 40–50.
30. Fair, D. A., Cohen, A. L., Power, J. D., Dosenbach, N. U. F., Church, J. A., Miezin, F. M., ... Petersen, S. E. (2009). Functional brain networks develop from a "local to distributed" organization. *PLoS Computational Biology*, 5 (5), e1000381.
31. Fan, Y., Shi, F., Smith, J. K., Lin, W., Gilmore, J. H., & Shen, D. (2011). Brain anatomical networks in early human brain development. *NeuroImage*, 54, 1862–1871.
32. Fields, R. D. (2008). White matter in learning, cognition and psychiatric disorders. *Trends in Neurosciences*, 31, 361–370.
33. Finn, A. S., Sheridan, M. A., Kam, C. L. H., Hinshaw, S., & D'Esposito, M. (2010). Longitudinal evidence for functional specialization of the neural circuit supporting working memory in the human brain. *Journal of Neuroscience*, 30, 11062–11067.
34. Fjell, A. M., Walhovd, K. B., Brown, T. T., Kuperman, J. M., Chung, Y., Hagler, D. J., ... Pediatric Imaging, Neurocognition, and Genetics Study. (2012). Multimodal imaging of the self-regulating developing brain. *Proceedings of the National Academy of Sciences, USA*, 109 (48), 19620–19625.
35. Fonov, V., Evans, A. C., Botteron, K., Almli, C. R., McKinstry, R. C., Collins, D. L., & Brain Development Cooperative Group. (2011). Unbiased average age-appropriate atlases for pediatric studies. *NeuroImage*, 54, 313–327.
36. Friedman, N. P., Miyake, A., Robinson, J. L., & Hewitt, J. K. (2011). Developmental trajectories in toddlers' self-restraint predict individual differences in executive functions 14 years later: A behavioral genetic analysis. *Developmental Psychology*, 47, 1410–1430.
37. Gao, W., Zhu, H., Giovanello, K. S., Smith, J. K., Shen, D., Gilmore, J. H., & Lin, W. (2009). Evidence on the emergence of the brain's default network from 2-week-old to 2-year-old healthy pediatric subjects. *Proceedings of the National Academy of Sciences, USA*, 106 (16), 6790–6795.
38. Ghosh, S. S., Kakunoori, S., Augustinack, J., Nieto-Castanon, A., Kovelman, I., Gaab, N., ... Fischl, B. (2010). Evaluating the validity of volume-based and surface-based brain image registration for developmental cognitive neuroscience studies in children 4 to 11 years of age. *NeuroImage*, 53, 85–93.
39. Giedd, J. N. (2008). The teen brain: Insights from neuroimaging. *Journal of Adolescent Health*, 42, 335–343.
40. Giedd, J. N., Lalonde, F. M., Celano, M. J., White, S. L., Wallace, G. L., Lee, N. R., & Lenroot, R. K. (2009). Anatomical brain magnetic resonance imaging of typically developing children and adolescents. *Journal of the American Academy of Child and Adolescent Psychiatry*, 48, 465–470.
41. Giorgio, A., Watkins, K. E., Chadwick, M., James, S., Winmill, L., Douaud, G., ... James, A. C. (2010). Longitudinal changes in grey and white matter during adolescence. *NeuroImage*, 49, 94–103.
42. Gogtay, N., Giedd, J. N., Lusk, L., Hayashi, K. M., Greenstein, D., Vaituzis, A. C., Thompson, P. M. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. *Proceedings of the National Academy of Sciences, USA*, 101 (21), 8174–8179.

43. Gogtay, N., & Thompson, P. M. (2010). Mapping gray matter development: Implications for typical development and vulnerability to psychopathology. *Brain and Cognition*, 72, 6–15.
44. Hackman, D. A., & Farah, M. J. (2009). Socioeconomic status and the developing brain. *Trends in Cognitive Sciences*, 13, 65–73.
45. Hanania, R., & Smith, L. B. (2010). Selective attention and attention switching: Towards a unified developmental approach. *Developmental Science*, 13, 622–635.
46. Hua, X., Leow, A. D., Levitt, J. G., Caplan, R., Thompson, P. M., & Toga, A. W. (2009). Detecting brain growth patterns in normal children using tensor-based morphometry. *Human Brain Mapping*, 30, 209–219.
47. Hwang, K., & Luna, B. (2013). The development of brain connectivity supporting prefrontal cortical functions. In D. Stuss & R. Knight (Eds.), *Principles of Frontal Lobe Function* (2nd ed., pp. 164–184). Oxford, UK: Oxford University Press.
48. Johansen-Berg, H. (2010). Behavioral relevance of variation in white matter microstructure. *Current Opinion in Neurobiology*, 23, 351–358.
49. Jolles, D. D., van Buchem, M. A., Crone, E. A., & Rombouts, S. A. R. B. (2011). A comprehensive study of whole-brain functional connectivity in children and young adults. *Cerebral Cortex*, 21, 385–391.
50. Karama, S., Ad-Dab'bagh, Y., Haier, R., Deary, I., Lyttelton, O., Lepage, C., ... Brain Development Cooperative Group. (2009). Positive association between cognitive ability and cortical thickness in a representative US sample of healthy 6 to 18 year-olds. *Intelligence*, 37, 145–155.
51. Kharitonova, M., & Munakata, Y. (2011). The role of representations in executive function: Investigating a developmental link between flexibility and abstraction. *Frontiers in Psychology*, 2, 347.
52. Khundrakpam, B. S., Reid, A., Brauer, J., Carbonell, F., Lewis, J., Ameis, S., ... Brain Development Cooperative Group. (2012). Developmental changes in organization of structural brain networks. *Cerebral Cortex*. doi:10.1093/cercor/bhs187
53. Lebel, C., & Beaulieu, C. (2011). Longitudinal development of human brain wiring continues from childhood into adulthood. *Journal of Neuroscience*, 31, 10937–10947.
54. Lebel, C., Gee, M., Camicioli, R., Wieler, M., Martin, W., & Beaulieu, C. (2012). Diffusion tensor imaging of white matter tract evolution over the lifespan. *NeuroImage*, 60, 340–352.
55. Lebel, C., Walker, L., Leemans, A., Phillips, L., & Beaulieu, C. (2008). Microstructural maturation of the human brain from childhood to adulthood. *NeuroImage*, 40, 1044–1055.
56. Lenroot, R. K., & Giedd, J. N. (2010). Sex differences in the adolescent brain. *Brain and Cognition*, 72, 46–55.
57. Lenroot, R. K., Schmitt, J. E., Ordaz, S. J., Wallace, G. L., Neale, M. C., Lerch, J. P., ... Giedd, J. N. (2009). Differences in genetic and environmental influences on the human cerebral cortex associated with development during childhood and adolescence. *Human Brain Mapping*, 30, 163–174.
58. Lu, L. H., Dapretto, M., O 'Hare, E. D., Kan, E., McCourt, S. T., Thompson, P. M., ... Sowell, E. R. (2009). Relationships between brain activation and brain structure in normally developing children. *Cerebral Cortex*, 19, 2595–2604.

59. Luna, B., Padmanabhan, A., & O 'Hearn, K. (2010). What has fMRI told us about the development of cognitive control through adolescence? *Brain and Cognition*, 72, 101–113.
60. Luna, B., Velanova, K., & Geier, C. F. (2010). Methodological approaches in developmental neuroimaging studies. *Human Brain Mapping*, 31, 863–871.
61. Madsen, K. S., Baaré, W. F. C., Vestergaard, M., Skimminge, A., Ejersbo, L. R., Ramsøy, T. Z., . . . Jernigan, T. L. (2010). Response inhibition is associated with white matter microstructure in children. *Neuropsychologia*, 48, 854–862.
62. Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.
63. Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, USA*, 108 (7), 2693–2698.
64. Moriguchi, Y., & Hiraki, K. (2011). Longitudinal development of prefrontal function during early childhood. *Developmental Cognitive Neuroscience*, 1, 153–162.
65. Munakata, Y., Snyder, H. R., & Chatham, C. H. (2012). Developing cognitive control: Three key transitions. *Current Directions in Psychological Science*, 21, 71–77.
66. Niogi, S., Mukherjee, P., Ghajar, J., & McCandliss, B. D. (2010). Individual differences in distinct components of attention are linked to anatomical variations in distinct white matter tracts. *Frontiers in Neuroanatomy*, 4, 2.
67. Nordahl, C. W., Simon, T. J., Zierhut, C., Solomon, M., Rogers, S. J., & Amaral, D. G. (2008). Brief report: Methods for acquiring structural MRI data in very young children with autism without the use of sedation. *Journal of Autism and Developmental Disorders*, 38, 1581–1590.
68. Olson, E. A., Collins, P. F., Hooper, C. J., Muetzel, R., Lim, K. O., & Luciana, M. (2009). White matter integrity predicts delay discounting behavior in 9- to 23-year-olds: A diffusion tensor imaging study. *Journal of Cognitive Neuroscience*, 21, 1406–1421.
69. Paus, T. (2005). Mapping brain maturation and cognitive development during adolescence. *Trends in Cognitive Sciences*, 9, 60–68.
70. Paus, T. (2010). Population neuroscience: Why and how. *Human Brain Mapping*, 31, 891–903.
71. Poldrack, R. A. (2010). Interpreting developmental changes in neuroimaging signals. *Human Brain Mapping*, 31, 872–878.
72. Power, J. D., Barnes, K. A., Snyder, A. Z., Schlaggar, B. L., & Petersen, S. E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *NeuroImage*, 59, 2142–2154.
73. Reichenberg, A., Caspi, A., Harrington, H., Houts, R., Keefe, R. S. E., Murray, R. M., ... Moffitt, T. E. (2010). Static and dynamic cognitive deficits in childhood preceding adult schizophrenia: A 30-year study. *American Journal of Psychiatry*, 167, 160–169.
74. Romberg, A. R., & Saffran, J. R. (2010). Statistical learning and language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 1, 906–914.
75. Rykhlevskaia, E., Gratton, G., & Fabiani, M. (2008). Combining structural and functional neuroimaging data for studying brain connectivity: A review. *Psychophysiology*, 45, 173–187.
76. Shaw, P., Lalonde, F., Lepage, C., Rabin, C., Eckstrand, K., Sharp, W., ... Rapoport, J. (2009). Development of cortical asymmetry in typically developing children and its

disruption in attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, 66, 888–896.
77. Somerville, L. H., & Casey, B. J. (2010). Developmental neurobiology of cognitive control and motivational systems. *Current Opinion in Neurobiology*, 20, 236–241.
78. Stevens, M. C. (2009). The developmental cognitive neuroscience of functional connectivity. *Brain and Cognition*, 70, 1–12.
79. Stevens, M. C., Pearlson, G. D., & Calhoun, V. D. (2009). Changes in the interaction of resting-state neural networks from adolescence to adulthood. *Human Brain Mapping*, 30, 2356–2366.
80. Supekar, K., Uddin, L. Q., Prater, K., Amin, H., Greicius, M. D., & Menon, V. (2010). Development of functional and structural connectivity within the default mode network in young children. *NeuroImage*, 52, 290–301.
81. Tamnes, C. K., Østby, Y., Fjell, A. M., Westlye, L. T., Due-Tønnessen, P., & Walhovd, K. B. (2010). Brain maturation in adolescence and young adulthood: Regional age-related changes in cortical thickness and white matter volume and microstructure. *Cerebral Cortex*, 20, 534–548.
82. Tamnes, C. K., Østby, Y., Walhovd, K. B., Westlye, L. T., Due-Tønnessen, P., & Fjell, A. M. (2010). Intellectual abilities and white matter microstructure in development: A diffusion tensor imaging study. *Human Brain Mapping*, 31, 1609–1625.
83. Thomason, M. E., Chang, C. E., Glover, G. H., Gabrieli, J. D. E., Greicius, M. D., & Gotlib, I. H. (2008). Default-mode function and task-induced deactivation have overlapping brain substrates in children. *NeuroImage*, 41, 1493–1503.
84. Thomason, M. E., Dougherty, R. F., Colich, N. L., Perry, L. M., Rykhlevskaia, E. I., Louro, H. M., ... Gotlib, I. H. (2010). COMT genotype affects prefrontal white matter pathways in children and adolescents. *NeuroImage*, 53, 926–934.
85. Uddin, L. Q., Supekar, K. S., Ryali, S., & Menon, V. (2011). Dynamic reconfiguration of structural and functional connectivity across core neurocognitive brain networks with development. *Journal of Neuroscience*, 31, 18578–18589.
86. Wang, L., Su, L., Shen, H., & Hu, D. (2012). Decoding lifespan changes of the human brain using resting-state functional connectivity MRI. *PLoS ONE*, 7 (8), e44530.
87. Wendelken, C., Munakata, Y., Baym, C., Souza, M., & Bunge, S. A. (2012). Flexible rule use: Common neural substrates in children and adults. *Developmental Cognitive Neuroscience*, 2, 329–339.
88. Wendelken, C., O 'Hare, E. D., Whitaker, K. J., Ferrer, E., & Bunge, S. A. (2011). Increased functional selectivity over development in rostrolateral prefrontal cortex. *Journal of Neuroscience*, 31, 17260–17268.
89. Wiebe, S. A., Espy, K. A., Stopp, C., Respass, J., Stewart, P., Jameson, T. R., ... Huggenvik, J. (2009). Gene–environment interactions across development: Exploring DRD2 genotype and prenatal smoking effects on self-regulation. *Developmental Psychology*, 45, 31–44.
90. Zelazo, P. D., & Carlson, S. M. (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, 6, 354–360.

Acknowledgments: The authors thank Henna Mishra, Carter Wendelken, Kirstie Whitaker, Jessica Church, and Mario Bunge for contributions to the review. The authors were supported by

a National Science Foundation Predoctoral Fellowship (S.M.), a grant from the National Institute of Neurological Disorders and Stroke, and a MacArthur Law & Neuroscience Project grant.

Appendix 4: Hemispheric Differences in Relational Reasoning: Novels Insights Based on an Old Technique

Citation: Vendetti, M. S.*, Johnson, E. L.*, Lemos, C. J. & Bunge, S. A. Hemispheric differences in relational reasoning: Novel insights based on an old technique. *Front. Hum. Neurosci.* 9 (2015). *equal contribution

Relational reasoning, or the ability to integrate multiple mental relations to arrive at a logical conclusion, is a critical component of higher cognition. A bilateral brain network involving lateral prefrontal and parietal cortices has been consistently implicated in relational reasoning. Some data suggest a preferential role for the left hemisphere in this form of reasoning, whereas others suggest that the two hemispheres make important contributions. To test for a hemispheric asymmetry in relational reasoning, we made use of an old technique known as visual half-field stimulus presentation to manipulate whether stimuli were presented briefly to one hemisphere or the other. Across two experiments, 54 neurologically healthy young adults performed a visuospatial transitive inference task. Pairs of colored shapes were presented rapidly in either the left or right visual hemifield as participants maintained central fixation, thereby isolating initial encoding to the contralateral hemisphere. We observed a left-hemisphere advantage for encoding a series of ordered visuospatial relations, but both hemispheres contributed equally to task performance when the relations were presented out of order. To our knowledge, this is the first study to reveal hemispheric differences in relational encoding in the intact brain. We discuss these findings in the context of a rich literature on hemispheric asymmetries in cognition.

A4.1. Introduction

Relational reasoning is a cognitive process that requires the joint consideration of relations in order to generate an inference to support a conclusion. Although there is a wide range of theoretical models for relational reasoning (for review, see Goodwin and Johnson-Laird, 2005; Knowlton et al., 2012), all of these models present relational reasoning as a unitary system. However, work from neuropsychological and neuroimaging literatures indicates that some cognitive functions may be supported by multiple, redundant systems in the brain (Roser and Gazzaniga, 2004; Marinsek et al., 2014). Here, we sought to test whether one hemisphere displays an advantage over the other during relational encoding, or whether this function can be carried out equally well by each hemisphere.

Hints of a possible left-hemisphere advantage in relational reasoning have emerged over the course of a number of neuroimaging experiments (e.g., Goel and Dolan, 2004; Green et al., 2006; Bunge et al., 2009; Wendelken et al., 2011). Importantly, similar patterns have been observed for tasks involving either verbal or non-linguistic/pictorial stimuli, suggesting that the observed differences are not entirely stimulus-driven and do not completely overlap with regions supporting language (Monti and Osherson, 2012). However, the conclusions we can draw from these fMRI studies about lateralization of function are limited in several ways. Namely, brain imaging provides correlational rather than causal evidence, and results depend on the specific contrasts used as well as the choice of statistical threshold. All of these factors can mask whether both hemispheres are indicated as being involved in a particular task, and thus, any conclusions about localization should converge with experimental findings using multiple approaches.

The neuropsychological literature also hints at possible hemispheric differences in contributions to reasoning. Much of the early work investigating differential hemispheric contributions to cognitive function came from work on split-brain patients (e.g., Sperry et al., 1969). These studies indicated an improved ability for hypothesis testing during problem solving in the left relative to the right hemisphere (LeDoux et al., 1977) and has led to the idea of the left hemisphere being an "interpreter" of events – i.e., the hemisphere with a major role of integrating newly acquired perceived information with previously constructed theories (Gazzaniga, 2000; Marinsek et al., 2014).

Following the seminal work of Gazzaniga et al. (1962) indicating how cognitive function differed in the two hemispheres following sectioning of the commissures, hemispheric asymmetries in cognition have alternately been characterized as a dichotomy between local and global (van Kleeck, 1989), categorical and coordinate (Kosslyn, 1987; van der Ham et al., 2014), or serial and parallel (e.g., Cohen, 1973) processes (for review, see Bradshaw and Nettleton, 1981). In the present study, we did not set out to evaluate these competing accounts of hemispheric specialization; rather, we sought to characterize the contribution of each hemisphere to performance of a relational reasoning task adapted from one used in a prior fMRI study from our group (Wendelken and Bunge, 2010).

There is not a consistent pattern relating relational reasoning ability to damage in a particular hemisphere. Neuropsychological work on relational reasoning has demonstrated the necessity of prefrontal and posterior parietal regions during transitive inference (Waltz et al., 1999; Krawczyk et al., 2008; Waechter et al., 2012), analogical reasoning (Morrison et al., 2004; Krawczyk et al., 2008), and matrix reasoning (Baldo et al., 2010; Woolgar et al., 2010). Additionally, studies employing voxel-based lesion symptom mapping to investigate relationships between patterns of brain damage and resulting cognitive deficits in fluid intelligence (Barbey et al., 2014) have suggested that damage to the right hemisphere plays a more critical role. However, Baldo et al. (2010) demonstrated that patients who have incurred strokes in the left hemisphere have been shown to also have significant deficits in a visuospatial relational reasoning task; therefore, more research is needed to provide a better understanding of each hemisphere's role in relational reasoning.

We designed the current study to test the role of each hemisphere in relational encoding through the use of a visual half-field stimulus presentation procedure. This paradigm was originally developed for use in split-brain patients, who have either minimal or no connection between the two hemispheres (e.g., Gazzaniga et al., 1962). Here, our participants were healthy adults whose hemispheres are presumed to interact closely in the coordination of task performance (Weissman and Banich, 2000). Nevertheless, we sought to test for differences in response times and/or accuracy when relational information is initially encoded by the left or the right hemisphere. This visual half-field stimulus presentation procedure allowed us to test whether left and right hemispheres differentially support relational encoding.

In the present study, we used a transitive inference task adapted from an fMRI task that we have used previously (Wendelken and Bunge, 2010). When reasoning using transitive inference, the logical conclusion is deduced through transferring relational inferences among terms expressed in the premises (e.g., if A > B and B > C, then A must be greater than C). On this task, shown in

Fig. A4.1, participants view a new set of relations on every trial and are expected to integrate them in working memory. There has been a rich literature on this form of reasoning (e.g., Halford, 1984; Cohen et al., 1997; Andrews and Halford, 1998; Greene et al., 2001). Importantly, this form of relational reasoning bears only a passing resemblance to transitive inference paradigms that involve learning paired associations over many trials (e.g., Acuna et al., 2002; Zalesak and Heckers, 2009; Koscik and Tranel, 2012; for discussion, see Wendelken and Bunge, 2010). The major difference between our transitive inference paradigm and those based on learning paired associations is that our task does not rely on remembering associations to be transferred; instead, participants must infer the spatial relationship based on the relations from the most recent trial only. Having to perform this inference anew each trial reduces any tendency to assume an object-order relationship when attempting to solve the task.

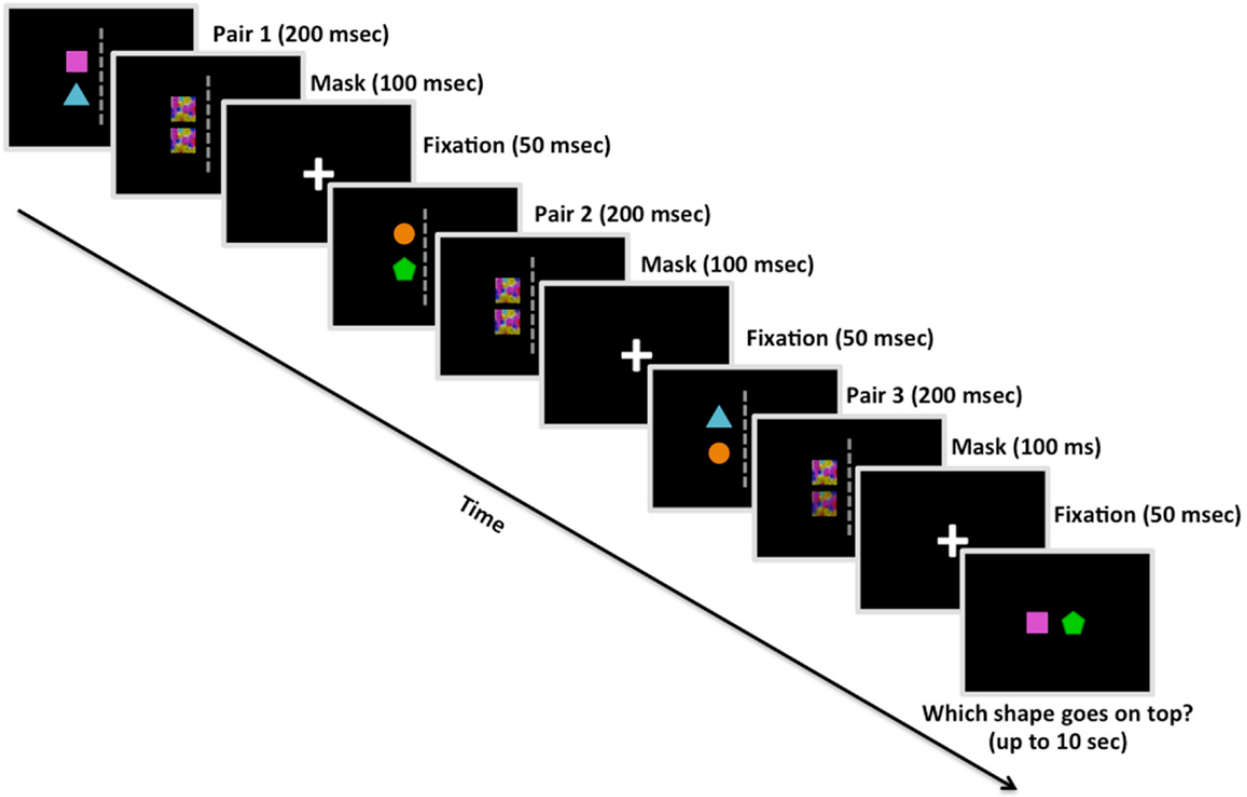

Fig. A4.1. Example trial from Study 2 (including the visual mask). Participants were shown three pairs of colored shapes. Following each pair, participants were shown a visual mask overlaying the previous shapes, and then a fixation cross. After the third pair was presented in a given trial, participants had up to 10 s to decide the correct linear order of two shapes based on the spatial relationships observed among the pairs. This is an example of a reordered trial, in which participants would presumably have to manipulate their memory of the pairs in order to deduce that the square goes on top of the pentagon. Study 1 was similar in design except for the absence of the visual mask presentations.

Inspired by neuropsychological research demonstrating that prefrontal patients have difficulty with transitive inference when the relations are presented out of order (e.g., “Sam is taller than Roy,” “James is taller than Sam”; Waltz et al., 1999; Krawczyk et al., 2008), we manipulated the

sequence of presentation of the three relations. On half of the trials, the relations were ordered (A > B; B > C; C > D), and on the other half, they were reordered (A > B; C > D; B > C or C > D, A > B, B > C).We hypothesized that manipulating encoding in this manner would have an influence on the downstream integration process, and sought to test for hemispheric differences in performance on trials whose relations could be integrated readily (ordered trials) and those that could not (reordered trials).

A4.2. Materials and Methods

A4.2.1. Participants

Experiment 1: Twenty-three healthy adults (14 female, aged 18–34 years; *M* ± *SD* age, 22 ± 3.08 years). *Experiment 2:* Thirty-one healthy adults (24 female, aged 18–25 years; *M* ± *SD* age, 20 ± 1.80 years). All participants attended the University of California, Berkeley, and participated in either Experiment 1 or 2 for partial fulfillment of a course requirement. All participants had normal or corrected-to-normal vision, were right-handed, and were fluent in English. Participants had no reported history of neurological or psychiatric disorders. All participants gave their informed consent to participate in the study, which was approved by the Committee for Protection of Human Subjects at the University of California, Berkeley.

A4.2.2. Design

We ran two studies with a similar design except for the addition of brief visual masks immediately following presentation of each object pair (100 ms) and an additional 48 trials, both of which were implemented in Experiment 2. We chose to insert the visual masks in Experiment 2 to reduce any after-image perceptual influences on decision making, in effect making the participant's deduction solely based on information stored and manipulated in working memory (Kim and Blake, 2005). The task designs were identical with the exception of these additions in Experiment 2; therefore, all of the information below applied to both studies unless explicitly stated. The stimulus set consisted of four colored shapes: blue triangle, orange circle, green pentagon, and pink square. On each trial, three sets of relations – pairs of shapes arranged vertically, with one colored shape positioned directly above another colored shape – were presented in sequence (Fig. A4.1). One-third of the transitive inference trials involved ordered problems, in which the source relations were presented in order (e.g., A > B, B > C, C > D; A – D?); the other two-thirds involved reordered problems, in which the middle relation was presented last (e.g., A > B, C > D, B > C; A – D? or C > D, A > B, B > C; A – D?). Placing the middle relation last instead of the final relation of the sequence assured that participants could not rely on simple memory for the most recent pair when making their decision.

Prior to the onset of each trial, white arrows appeared coming from the four corners of the screen for 400 ms in order to direct eyegaze to the center of the screen. Trials began with a white central fixation cross displayed on screen for 50 ms. Each pair of shapes was presented in the left or right visual hemifield for 200 ms, followed by a visual mask for 100 ms (Experiment 2 only) and a central fixation inter-stimulus interval (ISI) for 50 ms, and then a different pair of shapes in either the same or opposite visual hemifield for 200 ms. After being shown three pairs individually, participants were asked to deduce the correct linear order of two items (e.g., square

and pentagon) based on the spatial relations presented in the sequence of object pairs (e.g., square above triangle, triangle above circle, and circle above pentagon). Participants had ≤10 s to make their decision regarding the correct linear order of two colored shapes (i.e., which of the two objects would be on top following the spatial relations represented in the trial).

A4.2.3. Procedure

Participants placed their heads in a chinrest affixed at arm's length from the screen, and were instructed to maintain their gaze on a central fixation cross. Vertical pairs of shapes were displayed between 4° and 6° of visual angle from central fixation (Buschman et al., 2011).

In Experiment 1, the task included 96 trials total: 24 in which all three shape pairs were presented to the left hemisphere (LLL), 24 in which they were presented to the right hemisphere (RRR), 24 in which they were presented to alternating hemispheres (12 LRL and 12 RLR trials), and 24 in which they were presented to opposite hemispheres but did not alternate (12 LRR and 12 RLL trials). The LRL, RLR, LRR, and RLL trials were inserted so that participants could not reliably predict where the second and third pairs would be presented. Experiment 2 included an additional 48 trials, but the balance of trial types was consistent with Experiment 1. Trials were evenly counterbalanced by hemispheric presentation and ordering condition, and the trial order was fully randomized.

The final prompt displayed two shapes next to each other and participants were instructed to indicate via key press which shape should "go on top" based on the information in the three pairs of relations. The "z" key corresponded to the shape on the left and the "?/" key to the shape on the right; participants were instructed to keep their left hand on the "z" key and right hand on the "?/" key throughout the trials. In half the trials, the correct answer appeared on the left and half on the right. Participants were given a short break at the mid-point of the task. Experiment 2 contained a third block of trials, so participants were given a second break.

A4.3. Results

A4.3.1. Fully Lateralized Trials

We first investigated whether the small differences in task design between Experiments 1 and 2 would lead to any reliable differences in the results. A three-way mixed effects analysis of variance (ANOVA) with experiment number as the between-subjects variable, and hemispheric presentation (LH versus RH) and ordering condition (ordered versus reordered) as within-subjects variables indicated neither a main effect of experiment nor any interaction with other factors, F's < 1, p's > 0.54. Thus, all subsequent reported effects were generated from models collapsing across studies.[3] We analyzed accuracy and response time data in separate two-way repeated measures ANOVAs, with hemispheric presentation and ordering condition as within-

[3] Including gender as a factor in the full model, we found that the males in this study were more accurate than the females. Given the large gender imbalance in our relatively small sample, this result should not be over-interpreted. Notably, both males and females exhibited higher accuracy when the relations were presented to the left hemisphere than to the right hemisphere.

subjects factors. In this first section, we discuss only those trials that were solely presented to the left or right hemisphere. Behavioral results are presented in Fig. A4.2.

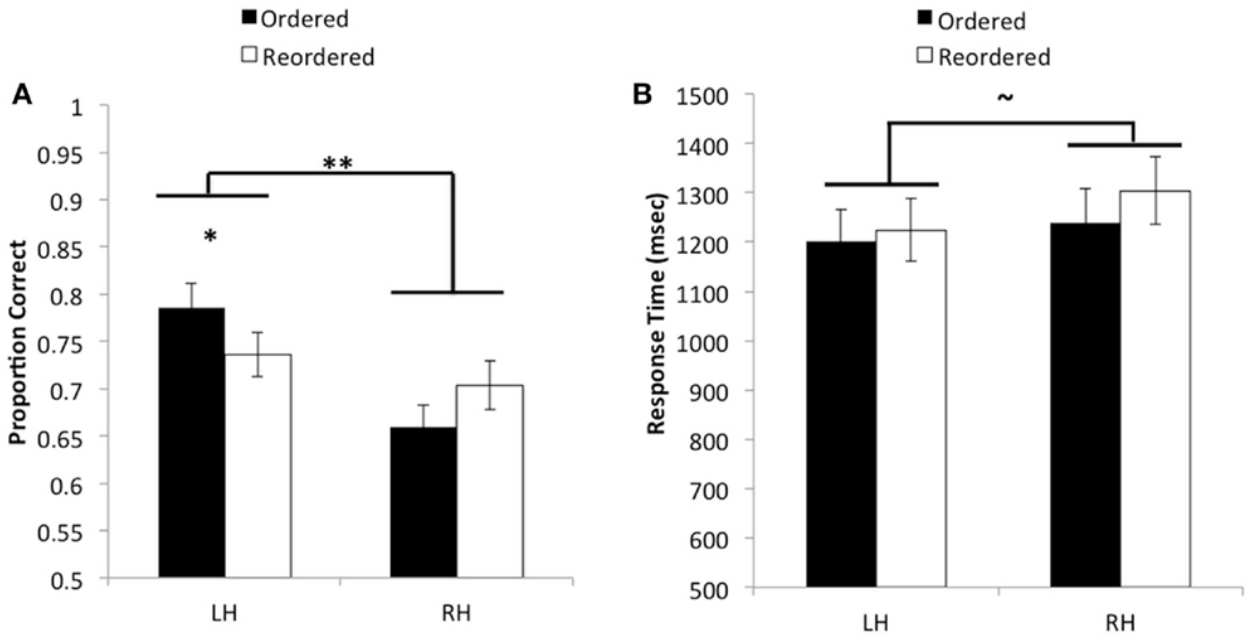

Fig. A4.2. (A) Average proportion correct as a function of hemisphere and ordering condition. A significant interaction was found such that when pairs of objects were presented in order, performance was significantly better when information was initially presented to the left versus the right hemisphere. However, no reliable difference was observed between hemispheres when pairs needed to be reordered in memory. Additionally, an overall main effect was found indicating that accuracy improved when pairs were initially encoded by the left hemisphere as opposed to the right hemisphere. (B) Average response time in milliseconds as a function of hemisphere and ordering condition, for correct trials. No reliable differences were observed for response time. $^{**}p < 0.01$.

The ANOVA revealed a significant main effect of hemisphere on accuracy, $F(1, 53) = 27.15$, $MSE = 0.012$, $p < 0.01$, $\eta_p^2 = 0.34$, such that participants performed better when relational information in the reasoning problem was initially encoded by the left hemisphere ($M = 0.76$, $SD = 0.17$) as compared to the right hemisphere ($M = 0.68$, $SD = 0.16$). A significant interaction between hemispheric presentation and ordering condition was also observed, $F(1, 53) = 8.2$, $MSE = 0.013$, $p < 0.01$, $\eta_p^2 = 0.13$. Post hoc *t*-tests using Bonferroni correction showed that participants were significantly more accurate when ordered pairs were presented to the left hemisphere ($M = 0.79$, $SD = 0.19$) as compared to the right hemisphere ($M = 0.66$, $SD = 0.16$), $t(53) = 6.02$, $p < 0.001$, $\eta_p^2 = 0.41$. By contrast, no significant differences were found in accuracy between the left hemisphere ($M = 0.74$, $SD = 0.17$) and right hemisphere ($M = 0.70$, $SD = 0.19$) on reordered trials, $t(53) = 1.51$, $p > 0.13$, $\eta_p^2 = 0.04$. We could also describe this interaction by looking at differences between trial types within each hemisphere. Although neither of these comparisons passed Bonferroni correction, in the left hemisphere, performance on ordered trials was better than on reordered trials, whereas the opposite was true in the right hemisphere. These results suggest that, although performance was best when stimuli were presented in order to the

left hemisphere, both hemispheres performed similarly when relations were not presented in an order that is conducive to integration before solving the transitive inference problem.

When including response times from correctly performed trials as the dependent variable, the ANOVA produced a marginally significant effect of hemispheric presentation, such that participants were faster to produce the correct decision on trials that were presented to the left hemisphere (M = 1218.41, SD = 433.56) as compared to the right hemisphere (M = 1273.10, SD = 448.26), $F(1, 53) = 3.93$, $MSE = 41115.21$, $p = 0.053$, $\eta_p^2 = 0.07$. No other effects in relation to response time were found to be statistically significant, F's < 1.26, p's > 0.26. These results suggest that the left-hemisphere boost in performance was not due to a speed-accuracy tradeoff; rather, when object pairs were presented to the left hemisphere, participants tended to respond faster than they would have if information had been presented to the right hemisphere.

A4.3.2. All Trials

In this section, we describe analyses investigating performance across both fully lateralized and mixed hemisphere trials (Fig. A4.3). We ran 4×2 repeated measures ANOVAs with number of times in the left hemisphere (0, 1, 2, 3) and order (ordered versus reordered) as within-subject factors, predicting accuracy and response time scores in separate models.

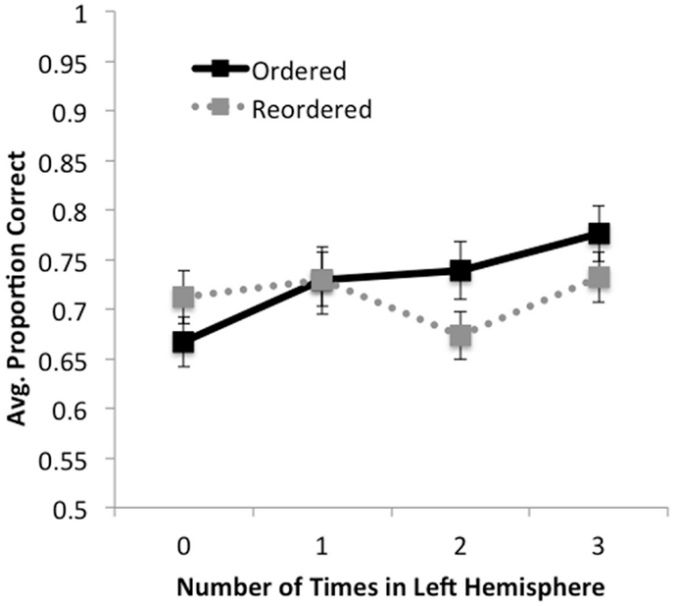

Fig. A4.3. Accuracy as a function of ordering condition and number of times premise was presented in the left hemisphere (0, 1, 2, 3). For ordered trials, accuracy increased monotonically with the number of times a premise was presented in the left hemisphere. For reordered trials, a simple pattern was not observed; rather, accuracy decreased when premises were presented in the left hemisphere two times (i.e., on LRL and RLL trials) relative to one or three times. No effects were observed for response times.

No significant effects were found for response times, F's < 1.8, p's > 0.18. In terms of accuracy, we found a significant main effect of number of times in the left hemisphere, $F(3,159) = 8.79$,

MSE = 0.013, $p < 0.001$, $\eta_p^2 = 0.14$, such that greater accuracy was observed the more often premises were presented in the left hemisphere. We also observed a trend for the effect of order, such that accuracy on ordered trials (*M* = 0.74, *SD* = 0.16) was marginally higher than on reordered trials (*M* = 0.72, *SD* = 0.15), $F(1,53) = 3.45$, *MSE* = 0.017, $p < 0.07$, $\eta_p^2 = 0.06$. We observed a significant interaction between number of times in the left hemisphere by order, $F(3, 159) = 5.55$, *MSE* = 0.013, $p < 0.001$, $\eta_p^2 = 0.1$. We found that for ordered trials there was a significant monotonic increase in accuracy as premises were presented to the left hemisphere, $F(1, 53) = 38.11$, *MSE* = 0.011, $p < 0.001$, $\eta_p^2 = 0.42$. For reordered trials, no such linear trend was observed, $F(1,53) < 1$, $p > 0.5$. These results suggest that when information is already ordered, increases in accuracy can be significantly predicted by how many times the premises are presented in the left hemisphere, and support our finding that participants performed better when ordered trials were presented only to the left hemisphere than to the right.

A4.3.3. Follow-Up Analyses

In testing for hemispheric differences in performance on this transitive inference task, we sought to ensure that participants were performing this task in the manner expected. When three relations are presented in order, it is possible to produce the correct response even without integrating multiple relations (Bryant and Trabasso, 1971). In our design, this simpler, non-integrative strategy could be undertaken by paying attention only to the top item in the first premise rather than encoding all premises and integrating the relations between them. If participants were to take this strategy, they would be expected to achieve roughly 100% accuracy on ordered trials, but only around 50% accuracy on reordered trials (because the first item of the first premise only appeared in the final prompt on two-thirds of the trials). Six out of 54 participants exhibited a pattern consistent with the use of this strategy. The findings reported here hold even when excluding these six participants.

A4.4. Discussion

Inspired by findings from the neuroimaging and neuropsychological literatures, we tested whether healthy young adults' performance on a reasoning task would differ on whether the stimuli were presented to the left or right hemisphere. By designing a transitive inference task with visual half-field stimulus presentation, we were able to show differences in reasoning performance as a function of the hemisphere that initially encoded the sets of visuospatial relations. Given that the two hemispheres communicate freely in the intact brain, we had expected only modest differences in response times for left- versus right-hemifield stimulus presentation. As such, we were surprised by the magnitude of the behavioral difference elicited by visual half-field presentation in this study, with an average difference in accuracy of 11% between left-lateralized and right-lateralized ordered trials. Although claims of inter-hemispheric differences in cognition have been made for many years (Gazzaniga et al., 1962; Cohen, 1973), our study is the first to demonstrate hemispheric differences in relational encoding in neurologically intact participants.

Although task performance (i.e., accuracy) improved overall when participants encoded the visuospatial relations in the left hemisphere, this effect was driven by performance on the ordered trials. That is, we observed a left-hemisphere advantage when the relations were ordered

linearly and, therefore, could be integrated directly, but not when it was necessary to rearrange the relations before integrating them. For right-hemisphere trials, participants did not show the predicted pattern of worse performance for reordered versus ordered trials. This pattern was unexpected, and warrants further investigation. Surprisingly, given that reordered trials are hypothesized to require additional processing relative to ordered trials (Waltz et al., 1999; Krawczyk et al., 2008), left-hemisphere encoding of reordered relations was superior even to right-hemisphere encoding of ordered relations. These results suggest that the left hemisphere excels at relational encoding.

The present results fit well with neuroimaging studies that have pointed toward a left-hemisphere specialization in relational reasoning (Wendelken et al., 2008; Bunge et al., 2009; Green et al., 2010). In light of these findings, it is interesting to consider a recent resting-state functional connectivity study showing that the left-hemisphere interacts more exclusively with itself, whereas the right hemisphere demonstrates connectivity patterns associated with both hemispheres (Gotts et al., 2013). This result suggests that the left hemisphere may operate independently, whereas the right hemisphere functions, at least partly, with assistance from the left hemisphere. Given these findings, we would predict a left-hemisphere advantage if relational encoding hinges more on intra-hemispheric interactions, and indeed this prediction was supported by our analysis including the mixed trials.

A4.4.1. A Left-Hemisphere Advantage for Relational Encoding

The behavioral improvement observed in our study does not indicate that the right hemisphere cannot encode relational information, but rather suggests that relational encoding may be processed more effectively in the left hemisphere. Although the stimuli were visuospatial in nature, they nonetheless were easily identifiable verbally (e.g., circle, square, pentagon). Given how quickly premises were presented, it does not seem feasible that very many participants would have had enough time to verbally label objects while they solved the task; however, we cannot conclusively rule out this possibility. The present study establishes a paradigm that could be used for further examination of the necessity of verbal labeling for relational reasoning.

Numerous dichotomies have been used to explain hemispheric asymmetries in cognitive functioning (Bradshaw and Nettleton, 1981), and so we do not claim that the left-hemisphere advantage observed in our study is unique to relational encoding, per se. Beyond the verbal/non-verbal distinction (Gazzaniga et al., 1962), other theories have focused on local versus global (van Kleeck, 1989), serial versus parallel (Cohen, 1973), holistic versus analytic (Nebes, 1978; Cooper and Wojan, 2000), categorical versus coordinate (Kosslyn, 1987), or syntactical versus intuitive/“gist” (Bogen, 1975; Phelps and Gazzaniga, 1992) processing, to name a few. Such dichotomies are useful in that they demonstrate how a higher level cognitive task such as reasoning might be represented as a combination of lower order cognitive processes. Our transitive inference task could be construed as being syntactical, serial, and analytic, and previous work focusing on these distinctions has consistently demonstrated a left-hemispheric specialization (for review, see Bradshaw and Nettleton, 1981). Additionally, encoding spatial relations in the premises categorically (e.g., identifying the square as above the triangle) would also fit with previous work demonstrating a left-hemispheric advantage for categorical encoding of spatial relations (Kosslyn, 1987; van der Ham et al., 2012).

A4.5. Conclusion and Future Directions

Our results shed light on cognitive theories of relational reasoning, as they provide evidence for differential processing of relations by the two hemispheres. Specifically, we found that participants performed better on our transitive inference task when the premises were presented to the left hemisphere. This effect was driven by an interaction such that there was a greater difference in performance when the premises were ordered than when participants presumably had to reorder the premises before making their conclusion. Theories describing a unitary mechanism of relational reasoning (e.g., Hummel and Holyoak, 2003; Goodwin and Johnson-Laird, 2005) may need to incorporate multiple components in order to fully represent interhemispheric differences used during relational reasoning.

The present results are consistent with theoretical predictions concerning hemispheric specialization of cognitive functions. Specifically, participants are expected to perform better when information is presented to the left hemisphere for tasks that could be solved using a stepwise and analytical strategy. Our findings extend previous work given that our transitive inference task not only exemplifies these types of strategies but also relies on the comparison of relational information between premises in order to arrive at a solution.

These behavioral results warrant further investigation with neuroscientific techniques. First, functional imaging techniques could be used to measure the dynamic interplay between hemispheres during performance of this lateralized transitive inference task. Second, transcranial direct current stimulation could be used to increase or reduce cortical excitability within a hemisphere and test whether relational reasoning performance in each hemisphere changes as a function of cortical excitability (Nitsche and Paulus, 2001; Ardolino et al., 2005). Finally, patients with unilateral brain injuries could be tested on this lateralized task to assess whether relational encoding is primarily a left-hemisphere function, or whether the right hemisphere could specialize in this function after left-hemisphere damage. Thus, reapplying this well-established stimulus presentation procedure in these multiple contexts will help us to better understand the underlying mechanisms required for processing relational information during reasoning.

A4.6. References

1. Acuna, B. D., Eliassen, J. C., Donoghue, J. P., and Sanes, J. N. (2002). Frontal and parietal lobe activation during transitive inference in humans. *Cereb. Cortex* 12, 1312–1321. doi:10.1093/cercor/12.12.1312
2. Andrews, G., and Halford, G. S. (1998). Children's ability to make transitive inferences: the importance of premise integration and structural complexity. *Cogn. Dev.* 13, 479–513. doi:10.1016/S0885-2014(98)90004-1
3. Ardolino, G., Bossi, B., Barbieri, S., and Priori, A. (2005). Non-synaptic mechanisms underlie the after-effects of cathodal transcutaneous direct current stimulation of the human brain. *J. Physiol.* 568, 653–663. doi:10.1113/jphysiol.2005.088310
4. Baldo, J.V., Bunge, S. A., Wilson, S. M., and Dronkers, N. F. (2010). Is relational reasoning dependent on language? A voxel-based lesion symptom mapping study. *Brain Lang.* 113, 59–64. doi:10.1016/j.bandl.2010.01.004

5. Barbey, A. K., Colom, R., Paul, E. J., and Grafman, J. (2014). Architecture of fluid intelligence and working memory revealed by lesion mapping. *Brain Struct. Funct.* 219, 485–494. doi:10.1007/s00429-013-0512-z
6. Bogen, J. E. (1975). *Educational Aspects of Hemispheric Specialization.* UCLA Educator.
7. Bradshaw, J. L., and Nettleton, N. C. (1981). The nature of hemispheric specialization in man. *Behav. Brain Sci.* 4, 51–91. doi:10.1017/S0140525X00007548
8. Bryant, P. E., and Trabasso, T. (1971). Transitive inferences and memory in young children. *Nature* 232, 456–458. doi:10.1038/232456a0
9. Bunge, S. A., Helskog, E. H., and Wendelken, C. (2009). Left, but not right, rostrolateral prefrontal cortex meets a stringent test of the relational integration hypothesis. *Neuroimage* 46, 338–342. doi:10.1016/j.neuroimage.2009.01.064
10. Buschman, T. J., Siegel, M., Roy, J. E., and Miller, E. K. (2011). Neural substrates of cognitive capacity limitations. *Proc. Natl. Acad. Sci. U. S. A.* 108, 11252–11255. doi:10.1073/pnas.1104666108
11. Cohen, G. (1973). Hemispheric differences in serial versus parallel processing. *J. Exp. Psychol.* 97, 349–356. doi:10.1037/h0034099
12. Cohen, N. J., Poldrack, R. A., and Eichenbaum, H. (1997). Memory for items and memory for relations in the procedural/declarative memory framework. *Memory* 5, 131–178. doi:10.1080/741941149
13. Cooper, E. E., and Wojan, T. J. (2000). Differences in the coding of spatial relations in face identification and basic-level object recognition. *J. Exp. Psychol. Learn. Mem. Cogn.* 26, 470–488. doi:10.1037/0278-7393.26.2.470
14. Gazzaniga, M., Bogen, J., and Sperry, R. (1962). Some functional effects of sectioning the cerebral commissures in man. *Proc. Natl. Acad. Sci. U. S. A.* 48, 1765–1769. doi:10.1073/pnas.48.10.1765
15. Gazzaniga, M. S. (2000). Cerebral specialization and interhemispheric communication: does the corpus callosum enable the human condition? *Brain* 123, 1293–1326. doi:10.1093/brain/123.7.1293
16. Goel, V., and Dolan, R. J. (2004). Differential involvement of left prefrontal cortex in inductive and deductive reasoning. *Cognition* 93, B109–B121. doi:10.1016/j. cognition.2004.03.001
17. Goodwin, G. P., and Johnson-Laird, P. N. (2005). Reasoning about the relations between relations. *Q. J. Exp. Psychol.* 59, 1047–1069. doi:10.1080/ 02724980543000169
18. Gotts, S. J., Jo, H. J., Wallace, G. L., Saad, Z. S., Cox, R.W., and Martin, A. (2013). Two distinct forms of functional lateralization in the human brain. *Proc. Natl. Acad. Sci. U. S. A.* 110, E3435–E3444. doi:10.1073/pnas.1302581110
19. Green, A. E., Fugelsang, J. A., Kraemer, D. J. M., Shamosh, N. A., and Dunbar, K. N. (2006). Frontopolar cortex mediates abstract integration in analogy. *Brain Res.* 1096, 125–137. doi:10.1016/j.brainres.2006.04.024
20. Green, A. E., Kraemer, D. J. M., Fugelsang, J. A., Gray, J. R., and Dunbar, K. N. (2010). Connecting long distance: semantic distance in analogical reasoning modulates frontopolar cortex activity. *Cereb. Cortex* 20, 70–76. doi:10.1093/cercor/bhp081
21. Greene, A. J., Spellman, B. A., Dusek, J. A., and Eichenbaum, H. B. (2001). Relational learning with and without awareness: transitive inference using nonverbal stimuli in humans. *Mem. Cognit.* 29, 893–902. doi:10.3758/BF03196418

22. Halford, G. S. (1984). Can young children integrate premises in transitivity and serial order tasks? *Cogn. Psychol.* 16, 65–93. doi:10.1016/0010-0285(84)90004-5
23. Hummel, J. E., and Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychol. Rev.* 110, 220–264. doi:10.1037/ 0033-295X.110.2.220
24. Kim, C. Y., and Blake, R. (2005). Psychophysical magic: rendering the visible 'invisible'. *Trends Cogn. Sci.* 9, 381–388. doi:10.1016/j.tics.2005.06.012
25. Knowlton, B. J., Morrison, R. G., Hummel, J. E., and Holyoak, K. J. (2012). A neurocomputational system for relational reasoning. *Trends Cogn. Sci.* 16, 373–381. doi:10.1016/j.tics.2012.06.002
26. Koscik, T. R., and Tranel, D. (2012). The human ventromedial prefrontal cortex is critical for transitive inference. *J. Cogn. Neurosci.* 24, 1191–1204. doi:10.1162/ jocn_a_00203
27. Kosslyn, S. M. (1987). Seeing and imagining in the cerebral hemispheres: a computational approach. *Psychol. Rev.* 94, 148–175. doi:10.1037/0033-295X. 94.2.148
28. Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., et al. (2008). Distraction during relational reasoning: the role of prefrontal cortex in interference control. *Neuropsychologia* 46, 2020–2032. doi:10.1016/j.neuropsychologia.2008.02.001
29. LeDoux, J. E., Risse, G. L., Springer, S. P., Wilson, D. H., and Gazzaniga, M. S. (1977). Cognition and commissurotomy. *Brain* 100, 87–104. doi:10.1093/brain/ 100.1.87
30. Marinsek, N., Turner, B.O., Gazzaniga, M., and Miller, M.B. (2014). Divergent hemispheric reasoning strategies: reducing uncertainty versus resolving inconsistency. *Front. Hum. Neurosci.* 8:839. doi:10.3389/fnhum.2014.00839
31. Monti, M. M., and Osherson, D. N. (2012). Logic, language and the brain. *Brain Res.* 1428, 33–42. doi:10.1016/j.brainres.2011.05.061
32. Morrison, R. G., Krawczyk, D. C., Holyoak, K. J., Hummel, J. E., Chow, T. W., Miller, B. L., et al. (2004). A neurocomputational model of analogical reasoning and its breakdown in frontotemporal lobar degeneration. *J. Cogn. Neurosci.* 16,260–271. doi:10.1162/089892904322984553
33. Nebes, R. D. (1978). "Direct examination of cognitive function in the right and left hemispheres," in *Asymmetrical Function of the Brain*, ed. M. Kinsbourne (Cambridge: University Press), 99–140.
34. Nitsche, M. A., and Paulus, W. (2001). Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. *Neurology* 57, 1899–1901. doi:10.1212/WNL.57.10.1899
35. Phelps, E. A., and Gazzaniga, M. S. (1992). Hemispheric differences in mnemonic processing: the effects of left hemisphere interpretation. *Neuropsychologia* 30, 293–297. doi:10.1016/0028-3932(92)90006-8
36. Roser, M., and Gazzaniga, M. S. (2004). Automatic brains–interpretive minds. *Curr. Dir. Psychol. Sci.* 13, 56–59. doi:10.1111/j.0963-7214.2004.00274.x
37. Sperry, R.W., Gazzaniga, M. S., and Bogen, J. E. (1969). "Interhemispheric relationships: the neocortical commissures; syndromes of hemisphere disconnection," in *Handbook of Clinical Neurology*, eds P. J. Vinken and G.W. Bruvn (Amsterdam: North-Holland Publishing Company), 273–290.

38. van der Ham, I. J. M., Postma, A., and Laeng, B. (2014). Lateralized perception: the role of attention in spatial relation processing. *Neurosci. Biobehav. Rev.* 45, 142–148. doi:10.1016/j.neubiorev.2014.05.006
39. van der Ham, I. J. M., van Wezel, R. J. A., Oleksiak, A., van Zandvoort, M. J. E., Frijns, C. J. M., Kappelle, L. J., et al. (2012). The effect of stimulus features on working memory of categorical and coordinate spatial relations in patients with unilateral brain damage. *Cortex* 38, 737–745. doi:10.1016/j.cortex.2011.03.002
40. van Kleeck, M. H. (1989). Hemispheric differences in global versus local processing of hierarchical visual stimuli by normal subjects: new data and a meta-analysis of previous studies. *Neuropsychologia* 27, 1165–1178. doi:10.1016/0028-3932(89) 90099-7
41. Waechter, R. L., Goel, V., Raymont, V., Kruger, F., and Grafman, J. (2012). Transitive inference reasoning is impaired by focal lesions in parietal cortex rather than rostrolateral prefrontal cortex. *Neuropsychologia* 51, 464–471. doi:10.1016/j.neuropsychologia.2012.11.026
42. Waltz, J. A., Knowlton, B. J., Holyoak, K. J., Boone, K. B., Mishkin, F. S., deMeneze Santos, M., et al. (1999). A system for relational reasoning in human prefrontal cortex. *Psychol. Sci.* 10, 119–125. doi:10.1111/1467-9280.00118
43. Weissman, D. H., and Banich, M. T. (2000). The cerebral hemispheres cooperate to perform complex but not simple tasks. *Neuropsychology* 14, 41–59. doi:10.1037/0894-4105.14.1.41
44. Wendelken, C., and Bunge, S. A. (2010). Transitive inference: distinct contributions of rostrolateral prefrontal cortex and the hippocampus. *J. Cogn. Neurosci.* 22, 837–847. doi:10.1162/jocn.2009.21226
45. Wendelken, C., Chung, D., and Bunge, S. A. (2011). Rostrolateral prefrontal cortex: domain-general or domain-sensitive? *Hum. Brain Mapp.* 33, 1952–1963. doi:10.1002/hbm.21336
46. Wendelken, C., Nakhabenko, D., Donohue, S. E., Carter, C. S., and Bunge, S. A. (2008). "Brain is to thought as stomach is to ??": investigating the role of rostrolateral prefrontal cortex in relational reasoning. *J. Cogn. Neurosci.* 20, 682–693. doi:10.1162/jocn.2008.20055
47. Woolgar, A., Parr, A., Cusack, R., Thompson, R., Nimmo-Smith, I., Torralva, T., et al. (2010). Fluid intelligence loss linked to restricted regions of damage within frontal and parietal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 107, 14899–14902. doi:10.1073/pnas.1007928107
48. Zalesak, M., and Heckers, S. (2009). The role of the hippocampus in transitive inference. *Psychiatry Res.* 30, 24–30. doi:10.1016/j.pscychresns.2008.09.008

Acknowledgements: This work was made possible by a James S. McDonnell Foundation Scholar Award to Silvia A. Bunge. We thank Farida Valji and Kiana Modavi for their assistance with data collection.

Appendix 5: Eyegaze Patterns Reveal Optimal Strategies During Analogical Reasoning

Citation: Vendetti, M. S., Johnson, E. L., Modavi, K. & Bunge, S. A. Eyegaze patterns reveal optimal strategies during analogical reasoning (in revision).

Analogical reasoning is a complex cognitive process in which inferences are made based on the relational similarity between two domains. Although analogical reasoning has been the focus of inquiry for many years, most models rely on measures that cannot capture participants' thought processes moment by moment. Measuring participants' eye movements provides a way to investigate reasoning strategies in real time. Previous eyetracking research studying analogical reasoning has identified eyegaze patterns that participants make most often while reasoning, but it is still unclear which strategies are optimal for solving analogy problems in the face of distracting information. The current study used patterns of eye movements to infer participants' strategies while solving visual propositional analogy problems (A:B::C:D), and then identified which strategy was associated with best overall performance. Analysis of eye movements revealed that participants were initially drawn to perceptual lures when looking at response choices, spent more time fixating among analogy terms and the target relative to other response choices, and made more AB saccades relative to other saccades patterns. We then used participants' gaze sequences to classify each trial as representing either a project-first, structure-mapping, or semantic-constraint strategy and found that the more often participants used a project-first strategy, the higher their overall analogy accuracy. These findings provide new insight into the role of strategic processing while solving analogy problems.

A5.1. Introduction

Analogical reasoning – the process of generating inferences based on relational correspondences between two domains – is ubiquitous in most forms of thought (Hofstadter & Sanders, 2013). Numerous models have been put forth to explain the processes involved in analogical reasoning (see Gentner & Forbus, 2011). Although most of these models share many core processes (e.g., mapping items based on shared roles), what differentiates them is what information is considered most useful to compare while generating inferences.

Project-first models (e.g., Sternberg, 1977) stem from the psychometric tradition of using propositional analogies (i.e., A:B::C:? – see Fig. A5.1) to study fluid intelligence. In this model, analogies are solved by first generating a rule that relates the A and B terms before mapping the A and C terms, and finally applying a similar rule that generates D. According to this model, when presented with the analogy GLOVE : HAND :: SOCK : FOOT, the reasoner would first identify a rule relating glove and hand (e.g., a glove covers a hand for warmth). Then s/he would identify a rule that relates glove and sock (e.g., they are both articles of clothing). After mapping the rule between the two domains, a solution is produced (e.g., a sock covers a foot for warmth). Thus, the project-first model focuses on generating a rule between the A and B terms to guide subsequent judgments.

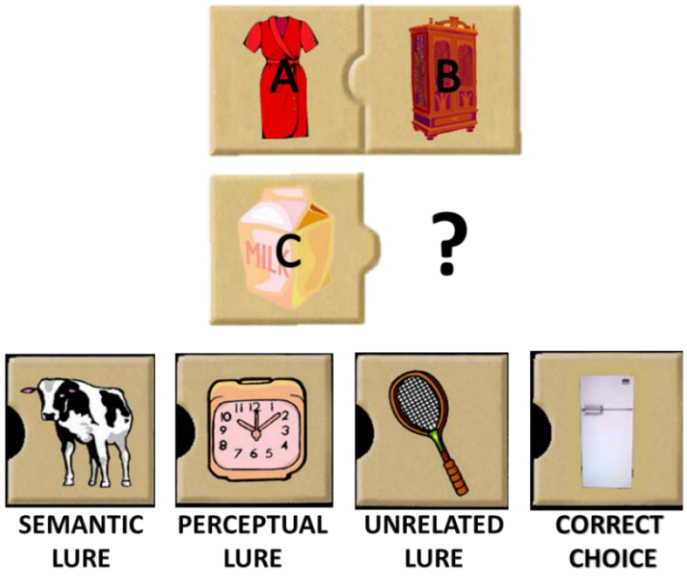

Fig. A5.1. Example analogy trial. Participants were asked to choose which item best fills the position of the question mark. Each trial contained four response choices: the correct choice, a perceptual lure, a semantic lure, and an unrelated lure. See main task for detailed instructions. The position of each response choice was randomized across trials. The letters and the response choice labels are for illustrative purposes only.

Structure-mapping models (e.g., Falkenhainer, Forbus, & Gentner, 1989; Gentner, 1983; 2010) assume that knowledge is structured hierarchically with relations connecting items, and that analogies are solved by mapping items from one structured representation to another. Using the same analogy example, structure-mapping models would propose that generating a rule to relate glove and sock would be the integral step; other inferences would then be guided by how well the two structures correspond (Gentner & Forbus, 2011).

A third model is based on the idea that there could be a large number of rules that relate the A and B terms in an analogy problem. Using the example analogy, a glove and hand may be related as, “clothing used to warm body parts”. However, many other relations may exist between these two items depending on the context, such as “things that are smaller than a breadbox”. Therefore, many possible solutions for solving the analogy could influence one’s ultimate decision. This model, referred to hereafter as the Semantic-constraint Model (e.g., Chalmers, French & Hofstadter, 1992; Glady, Thibaut, French, & Blaye, 2012; Thibaut, French, Missault, Gerard, & Glady, 2011), assumes that mapping relations between one’s search space and the C term in the analogy is the most useful step for solving an analogy.

Most models rely on behavioral findings, including comparing word passages (e.g., Day & Gentner, 2007; Gick & Holyoak, 1983), complex visual scenes (e.g., Markman & Gentner, 1993; Richland, Morrison, & Holyoak, 2006), and propositional analogies (e.g., Cho, Holyoak, & Cannon, 2007; Krawczyk, et al., 2008; Sternberg, 1977). One potential limitation of behavioral data is its low temporal specificity: recording a behavioral response can provide a measurement for the outcome of a trial, but it does not capture the possible strategy used to arrive at that response.

One approach that allows researchers to gain insight into strategies is eyetracking (e.g., Hayes, Petrov, & Sederberg, 2011; Hodgson, Bajwa, Owen, & Kennard, 2000; Salvucci & Anderson, 2001; von der Malsburg & Vasishth, 2011). This approach records eye movements around the screen space to infer what participants were thinking based on where they were looking while solving a task (e.g., Just & Carpenter, 1976; Rayner, 2009; Yarbus, 1967). In our study, we were interested in several eyetracking measures, delineated below.

The first fixation after trial onset reflects immediate information processing, and thus suggests what participants might deem most important when trying to glean information from an analogy problem (Antes, 1974). The amount of time spent fixating in an area of interest (AOI), or fixation duration, represents which aspects of the visual scene are most informative (Mackworth & Morandi, 1967). Finally, first-order fixation transitions measure where participants are likely to fixate next given their current fixation location, and can provide a reliable marker of relational processing based on the idea that transitioning between two objects is likely to represent shared information (e.g., Gordon & Moser, 2007).

Each of the three strategies would predict different patterns of eye movements that would be more informative for solving analogy problems (see Fig. A5.2). According to the project-first strategy, solving analogy problems involves looking earlier at the A and B terms of the analogy problem, making 1st-order fixation transitions between the AB terms, and then making 1st-order fixation transitions between the C term and the Target. The structure-mapping strategy would predict earlier fixations on the A and C terms, more saccades between the A and C terms, and between the B term and the Target. Finally, the semantic association strategy would predict earlier fixations on, and more saccades between, the C term and response choices.

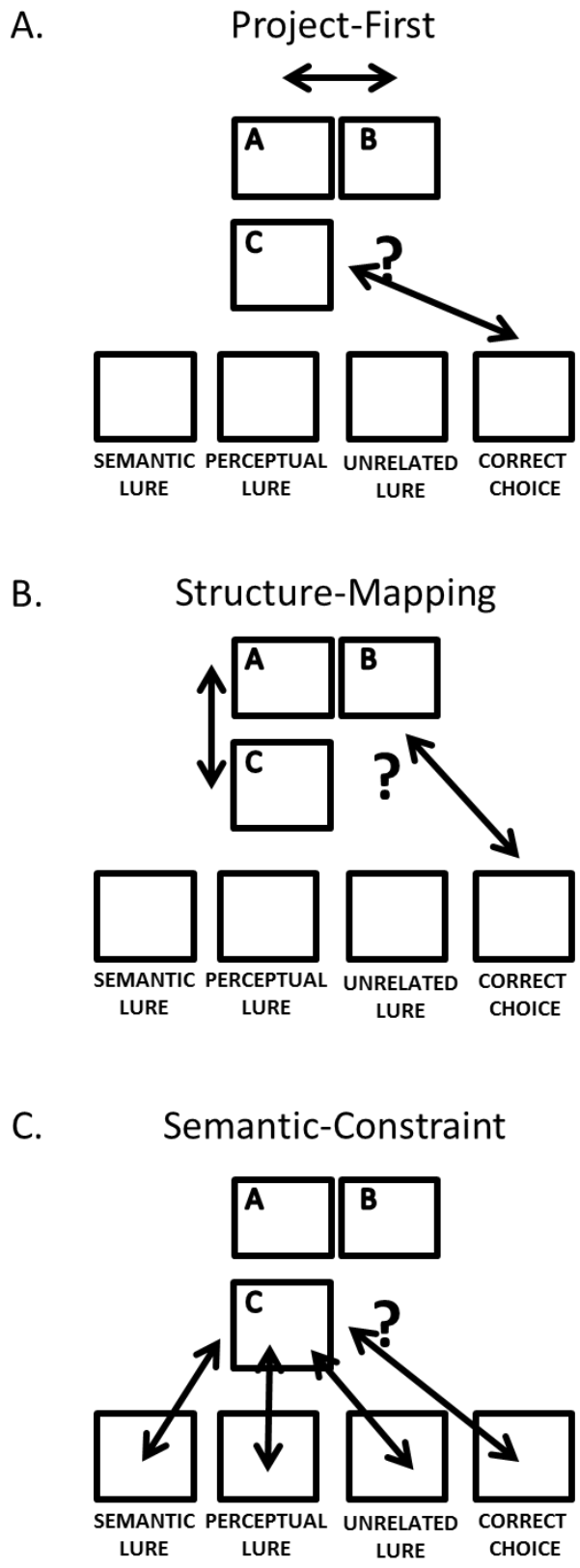

Fig. A5.2. Patterns of eye movements as predicted by (A) Project-first, (B) Structure-mapping, and (C) Semantic-constraint strategies.

Recently, several studies used eyetracking to identify where participants fixated while solving analogies (e.g., Glady, French & Thibaut, 2014; Gordon & Moser, 2007; Thibaut et al., 2011). These studies found that participants tend to make more mappings between the A and B terms, especially towards the beginning of a trial. Thus, these studies have demonstrated a preference for a project-first strategy while solving analogy problems. In the current study, we used eyetracking to describe what participants attended to while solving propositional analogy problems in the face of distracting information. We then used patterns of eye movements to classify participants' trials as resembling the project-first, structure-mapping, or semantic-constraint strategy. Finally, we tested whether using a particular strategy more often would predict overall analogy accuracy.

A5.2. Methods

A5.2.1. Participants

Twenty-eight healthy young adults (17 female, aged 18-25 years; $M \pm SD$ age, 20.39 ± 2.01 years) participated in the study. All participants attended the University of California at Berkeley, had normal vision, were fluent in English, had no reported history of neurological or psychiatric disorders, and gave their informed consent to participate in the experiment for partial fulfillment of a course requirement. The study was approved by the Committee for Protection of Human Subjects at the University of California at Berkeley. We excluded two participants: one whose proportion correct was more than two standard deviations below our sample's average performance and another whose eyetracking data was incomplete due to an error in data collection.

A5.2.2. Eyetracking Apparatus

Stimuli were presented using the Tobii E-Prime Software Extensions (Psychology Software Tools, Pittsburgh, PA), which synchronizes stimulus presentation timing with a second computer that records eyegaze position. Participants were seated comfortably in front of the Tobii T120 Eye Tracker (17-inch monitor, 1280×1024 pixel resolution). Distance was calibrated individually so that each participant focused on the middle of the screen, within a range of 50-80 cm. The Tobii T120 built-in camera captures data with a temporal resolution of 120 Hz and average spatial resolution of 0.3° of visual angle. The camera can automatically compensate for small head movements (within a 30 x 22 cm area at 70 cm distance); thus, participants' heads were not restrained. The camera independently recorded eyegaze position of the left and right eyes.

A5.2.3. Materials

Our lab designed the analogy trials using Adobe Photoshop, and made use of line drawings from "The Big Box of Art: 1 Million". All stimuli were pictures of common objects. Analogy problems consisted of an incomplete propositional analogy (i.e., A:B::C:?) above a row of four items. For example, in Fig. A5.1, participants were asked to indicate with a button press which of the four bottom pictures best completes the analogy. Participants were told that there may be several pictures that they think go with the C term, but that they should choose the picture that goes with the C term in the same way that the A term goes with the B term. Responses consisted

of the correct response, a perceptual lure (i.e., an item that shared a similar shape and color to the C-term), a semantic lure (i.e., an item whose meaning was associated with the C-term, but did not match the relation shared between the A- and B- terms), and a lure that was designed to be unrelated both perceptually and semantically. The location of each response type was randomized across trials.

A5.2.4. Procedure

Participants completed seven practice trials with feedback followed by 40 experimental trials without feedback. Experimental trials were split evenly into two blocks of trials. Each trial began on a black screen with a white central fixation cross. After 1000 ms, an analogy problem was presented and remained on the screen until a response was made, or until the trial timed out after 10 seconds. Participants pressed a button on the keyboard to reflect which of the four response choices they thought best completed the analogy problem, and were instructed to respond as quickly and accurately as possible. The stimulus display then disappeared and participants were immediately presented with the next trial.

A5.3. Results

A5.3.1. Behavioral Results

The average overall proportion correct (M = 0.91, SD = 0.05) and response times on correct trials (M = 3078 ms, SD = 563 ms) were similar to performance levels obtained in a previous study using similar analogy problems (e.g., Wright, Matlen, Baym, Ferrer, & Bunge, 2008). Although participants performed well overall, when they did make errors they were most likely to choose the semantic lure (~64%), followed by perceptual (~26%) and unrelated lures (~10%). This pattern of errors was significantly different than what would be expected if error types were equally likely, $X^2(2) = 89.13$, $p < 0.001$.

A5.3.2. Eyetracking Results

In this section, we report analyses concerning the eyetracking measures described above. Looking at trials in which participants correctly solved the analogy problem, we first tested whether there were differences in the frequency of the eyetracking measures, and then examined how these patterns predicted individual differences in overall analogy accuracy.

First fixations among analogy terms. One potentially important component underlying participants' strategy is where they decide to look first when presented with the analogy problem. Fig. A5.3A shows average first fixations to analogy terms (i.e., those areas containing the A, B, C terms – see Fig. A5.1). There was a main effect of location, $F(2, 50) = 43.64$, $p < 0.001$, $\eta_p^2 = 0.64$, such that participants made significantly more first fixations to the C term in the analogy than to any other term.

First fixations among response choices. We were also interested in testing whether participants were influenced by a particular lure when beginning their search among response choices. In our dataset, participants were more likely to first look at the analogy terms (i.e., A, B, or C) before

looking at the response choices (see Fig. A5.3D for example scan sequences); therefore, we only included participants' first fixations among response choices that followed fixating at least one of the analogy terms. The pattern of first fixations among response choices is shown in Fig. A5.3A. There was a significant difference among the number of first fixations within the response choice AOIs, $F(3,75) = 39.24$, $p \leq 0.001$, $\eta_p^2 = 0.61$, such that participants made significantly more first fixations to the perceptual lure ($M = 12.96$, $SD = 2.49$) than to the target ($M = 7.92$, $SD = 1.62$), semantic lures ($M = 7.58$, $SD = 1.96$), or unrelated lures ($M = 7.54$, $SD = 1.90$).

Proportion of time spent fixating in AOIs. We measured the proportion of time that participants fixated at each AOI (i.e., Analogy terms: A, B, C; and response choices: Target, Perceptual Lure, Semantic Lure, Unrelated Lure). As shown in Fig. A5.3B, participants spent significantly more time fixating on the A ($M = 0.20$, $SD = 0.11$), B ($M = 0.16$, $SD = 0.09$), and C ($M = 0.21$, $SD = 0.11$) terms as well as the Target ($M = 0.20$, $SD = 0.08$), relative to the perceptual ($M = 0.09$, $SD = 0.08$), semantic ($M = 0.08$, $SD = 0.07$), and unrelated lures ($M = 0.05$, $SD = 0.05$), all t's$(25) > 2.5$, p's < 0.01, $d > 0.47$.

First-order fixation transitions. Following from previous research investigating first-order fixation transition patterns while solving analogy problems (e.g., Gordon & Moser, 2007), we investigated where participants were likely to fixate next given their current fixation location (i.e., first-order fixation transitions, or saccades). As described earlier, the three strategies would predict different fixation transitions to be most useful while solving our analogy problems. Given the large number of possible saccades, we calculated participants' number of saccades between A ←→ B, A ←→ C, and C ←→ response choices relative to all other possible saccades. Our score reflected the proportion of these transitions between AOIs rather than absolute number of transitions. Participants were significantly more likely to produce AB 1st-order transitions ($M = 0.17$, $SD = 0.04$) as compared to both AC 1st-order transitions ($M = 0.11$, $SD = 0.03$) and C ←→ response choices ($M = 0.05$, $SD = 0.01$), $F(2, 50) = 87.57$, $p < 0.001$, $\eta_p^2 = 0.79$ (Fig. A5.3C).

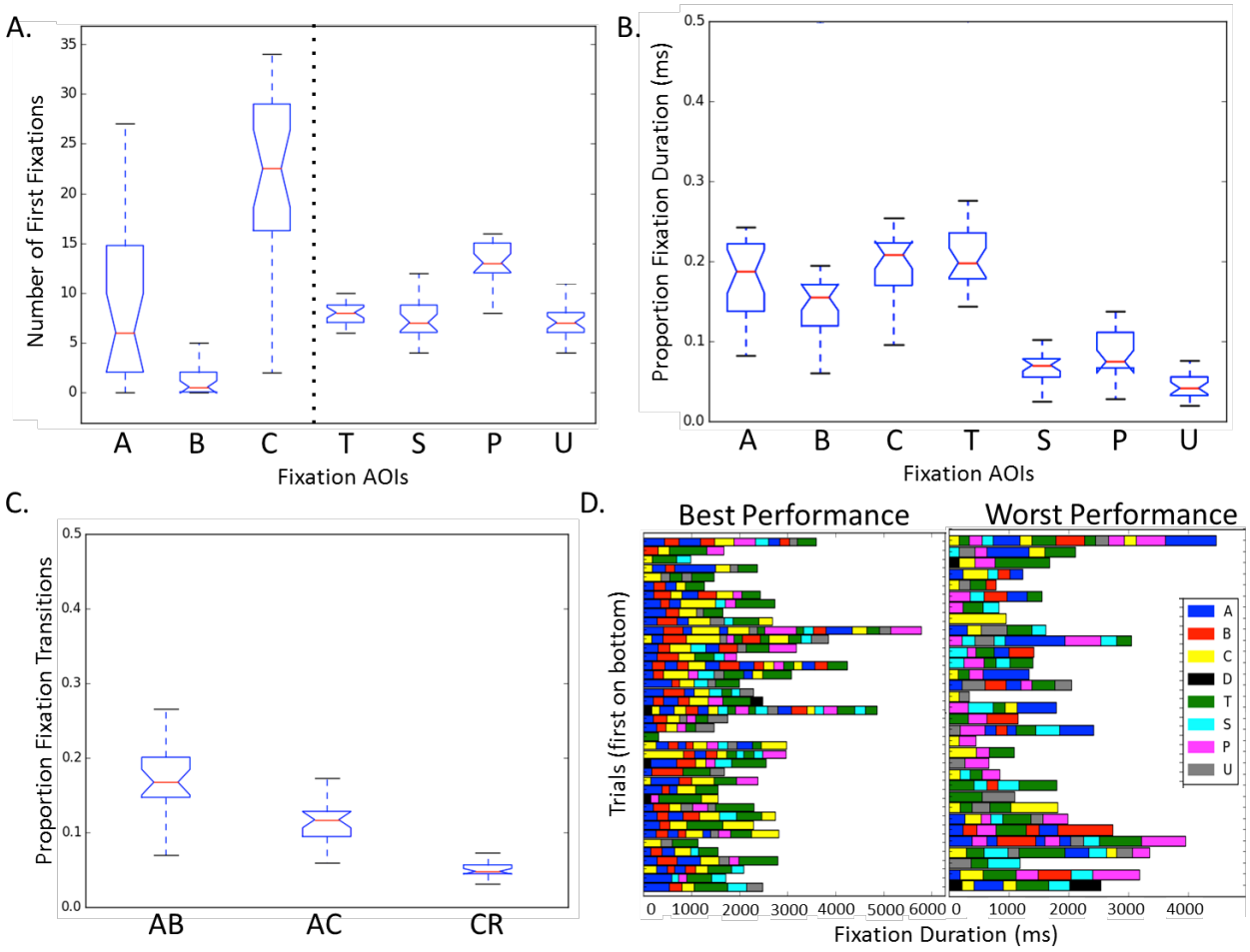

Fig. A5.3. (A) Left: Number of first fixation for analogy terms A, B, & C. Participants made more first fixations to the C term than either A or B terms of the analogy. Right: Number of first fixation for response choices conditional on having already looked at the analogy terms. Participants made significantly more first fixations to the perceptual lure than to any other response choice. T = Target, S = Semantic Lure, P = Perceptual Lure, U = Unrelated Lure. (B) Proportion Fixation Durations for analogy terms and Response Choices. Participants spent significantly more time looking at the analogy terms and the target than for the lures. (C) Proportion number of Fixation Transitions for AB, AC and C to Response Choices. Participants made significantly more AB transitions than the other types. (D) Plot showing sequences of fixations among fixation AOIs across trials, beginning at the bottom. Fixation durations are in milliseconds. Left: Fixation sequence for participant with the highest analogy accuracy. Right: Fixation Sequence for participant with the lowest analogy accuracy.

Using eyegaze sequences to identify optimal strategies. For this analysis we used participants' fixation sequences and corresponding fixation durations to calculate a score that was used to classify a participant's strategy for each trial (see Supplementary Materials for further information). We then calculated the number and proportion of trials classified according to each strategy, and used this information to describe the distribution of strategy use across all trials and to predict participants' overall analogy accuracy. We found that strategies were not classified

equally, $X^2(2) = 151.63$, $p < 0.001$, such that the majority were classified as project-first trials (n = 446), followed by structure-mapping (n = 307) and semantic-constraint (n = 145).

We then correlated participants' proportion of trials classified as each strategy with analogy accuracy and found that the proportion of trials classified as project-first was positively correlated with accuracy ($r = 0.624$, $p < .001$). The proportion of structure-mapping trials was not significantly related to accuracy ($r = –0.30$, $p = 0.14$), and the proportion of semantic-constraint trials was negatively correlated with accuracy ($r = –0.50$, $p < 0.01$). Using the cocor statistical package (Diedenhofen & Musch, 2015) we compared the correlation values for these analyses, accounting for the dependent nature of the correlations (Zou, 2007), and found that the correlation coefficient between project-first and analogy accuracy was significantly different than the correlation values for the other two strategies' relationships with accuracy (95% confidence interval for the difference = 0.30 – 1.37). Thus, our analysis indicates that even though a majority of the participants used the project-first strategy, those who used it more often had greater analogy accuracy, and those who used the semantic-constraint strategy more often had worse performance.

A5.4. Discussion

The current study used participants' eyegaze to identify optimal strategies for solving visual analogy problems in the face of distracting information. Thus, we were interested in how eyetracking could reveal how perceptual and relational similarity guides our analogical judgments, what types of eye movements and fixations were used most often, and whether patterns of eye movements could predict one's overall analogy accuracy. Our strategies were based on previous theories describing how participants might solve analogical reasoning problems: 1) project-first (e.g., Sternberg, 1997), in which information between items in the source domain are used to generate a solution for the target domain; 2) structure-mapping (e.g., Gentner, 1983; 2010), in which comparing items between source and target domains is most fruitful for generating an analogy; and 3) semantic-constraint (e.g., Chalmers et al., 1992; Thibaut, et al., 2010), in which participants rely on the semantic associations between the C term and response choices to guide their interpretation of the shared relationship in the source domain. Each of these strategies predicts different steps taken while solving analogy problems, and we leveraged participants' eye movements as a proxy of their strategies.

How might perceptual and semantic distracting information influence one's process of solving analogy problems? Previous work manipulated the number of semantic distractors while solving analogy problems and found that participants made more fixations to semantic distractors as the number of distractors increased (e.g., Glady et al., 2014; Gordon & Moser, 2007; Thibaut et al., 2010). The current study extended this work by including both perceptual and semantic lures on each trial, thus allowing us to examine whether participants were influenced by perceptual similarity, or whether they have learned to ignore this information and instead map items using relational information (i.e., semantic associations).

Theories of analogical reasoning development propose that the similarity processes used to guide one's analogical decision may either be purely due to insufficient knowledge of a particular domain (i.e., relational primacy hypothesis; Goswami & Brown, 1990), or is a function of both

object and relational similarity (i.e., relational shift hypothesis; Rattermann & Gentner, 1998). According to the relational primacy hypothesis, once someone has sufficient knowledge of a domain they should not be influenced by perceptual similarity; this could still be the case according to the relational shift hypothesis. Using eyetracking we observed that participants' first fixations among response choices were predominantly on the perceptual lure. Contrary to participants' behavior, which indicated more semantic than perceptual errors – and this eyetracking result indicates that perceptual similarity can still influence initial similarity judgments, even if relational information ultimately guides one's analogy decision. As such, this finding demonstrates that even adults can indicate – with eye movements – the types of similarity judgments that children demonstrate behaviorally.

We found that participants spent more time looking at the analogy terms and the target AOIs, with much less time spent fixating on the lure response choices, even the semantic lure. Contrary to the semantic-constraint hypothesis (e.g., Chalmers et al., 1992; Thibaut et al., 2010), participants did not allocate time equally among the response choices to identify a relation to constrain possible relationships between the A and B terms of the analogy. Moreover, our strategy classification analysis indicated that those participants who used the semantic-constraint strategy more often had lower overall accuracy. This pattern of results resonates with previous work indicating that younger children – who are less experienced with solving analogy problems – use the semantic-constraint strategy (Glady et al., 2012; Thibault et al., 2011). Thus, the semantic-constraint strategy represents a suboptimal approach to solving analogy problems.

Ultimately, we were interested in which strategy was optimal for solving analogy problems. Using both the average number of 1st-order fixation transitions across trials and our strategy classification analysis within trials, we found that participants were much more likely to use the project-first strategy. Even though the project-first strategy was the most common, many trials were also classified as representing the structure-mapping strategy across participants, and even across trials within participants. After we identified the proportion of trials classified as each strategy type within individual subjects, we used this information to show that the project-first strategy was optimal for solving the analogy problems.

In conclusion, we have presented work that used eyetracking measures to identify what types of strategies participants used while solving visual analogy problems in the face of distracting information. By taking an individual differences approach, we were able to show which patterns were correlated with higher overall accuracy scores, thus indicating which patterns of eye movements were most useful for solving the analogies correctly. Finally, by using our strategy classification analysis, we were able to show that the project-first strategy was the optimal strategy participants could use, and the semantic-constraint strategy was the worst strategy. Measuring eye movements allows researchers to go beyond the final behavioral decision and into the realm of real-time strategies used while thinking analogically. Future work in this area will allow greater precision for the steps necessary for effective reasoning and thus can be incredibly useful for informing theories of analogical reasoning as well as interventions aimed at improving one's reasoning ability.

A5.5. References

1. Antes, J. R. (1974). The time course of picture viewing. *Journal of Experimental Psychology*, 103(1), 62-70.
2. Chalmers, D. J., French, R. M., & Hofstadter, D. R. (1992). High-level perception, representation, and analogy: A critique of artificial intelligence methodology. *Journal of Experimental & Theoretical Artificial Intelligence*, 4(3), 185-211.
3. Cho, S., Holyoak, K. J., & Cannon, T. D. (2007). Analogical reasoning in working memory: Resources shared among relational integration, interference, resolution, and maintenance. *Memory & Cognition*, 35(6), 1445-1455.
4. Day, S. B., & Gentner, D. (2007). Nonintentional analogical inference in text comprehension. *Memory & Cognition*, 35(1), 39-49.
5. Diedenhofen, B., & Musch, J. (2014). Cocor: A comprehensive solution for the statistical comparisons of correlations. *PLoS ONE*, 10(4), 1-12.
6. Falkenhainer, B., Forbus, K. D., & Gentner, D. (1989). The structure-mapping engine: Algorithm and examples. *Artificial intelligence*, 41(1), 1-63.
7. Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science*, 7(2), 155-170.
8. Gentner, D. (2010). Bootstrapping the mind: Analogical processes and symbol systems. *Cognitive Science*, 34(5), 752-775.
9. Gentner, D., & Forbus, K. D. (2011). Computational models of analogy. *Cognitive Science*, 2 (3), 266-276.
10. Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology*, 15(1), 1-38.
11. Glady, Y., French, R., & Thibaut, J. P. (2014). Adults' eye tracking search profiles and analogy difficulty. In *Proceedings of the Thirty-Sixth Annual Meeting of the Cognitive Science Meeting* (pp. 535-540).
12. Glady, Y., Thibaut, J. P., French, R. M., & Blaye, A. (2012). Explaining children's failure in analogy making tasks: A problem of focus of attention? In *Proceedings of the Thirty-Fourth Annual Meeting of the Cognitive Science Meeting* (pp. 384-389).
13. Gordon, P. C., & Moser, S. (2007). Insight into analogies: Evidence from eye movements. *Visual Cognition*, 15(1), 20-35.
14. Goswami, U., & Brown, A. L. (1990). Higher-order structure and relational reasoning: Contrasting analogical and thematic relations. *Cognition*, 36, 207-226.
15. Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2011). A novel method for analyzing sequential eye movements reveals strategic influence on Raven's Advanced Progressive Matrices. *Journal of Vision*, 11(10), 1-11.
16. Hodgson, T. L., Bajwa, A., Owen, A. M., & Kennard, C. (2000). The strategic control of gaze direction in the Tower of London task. *Journal of Cognitive Neuroscience*, 12(5), 894907.
17. Hofstadter, D. & Sander, E. (2013). *Surfaces and Essences.* New York, NY: Basic Books.
18. Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review*, 110, 220-264.
19. Just, M. A., & Carpenter, P. A. (1976). Eye fixations and cognitive processes. *Cognitive Psychology*, 8, 441-480.

20. Krawczyk, D. C., Morrison, R. G., Viskontas, I., Holyoak, K. J., Chow, T. W., Mendez, M. F., Miller, B. L., & Knowlton, B. J. (2008). Distraction during relational reasoning: The role of prefrontal cortex in interference control. *Neuropsychologia*, 46, 2020-2032.
21. Mackworth, N. H., & Morandi, A. J. (1967). The gaze selects informative details within pictures. *Perception & Psychophysics*, 2(11), 547-552.
22. Markman, A. B., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, 25(4), 431-467.
23. Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from http://www.pstnet.com
24. Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, 62(8), 1457-1506.
25. Rattermann, M. J., & Gentner, D. (1998). More evidence for a relational shift in the development of analogy: Children's performance on a causal-mapping task. *Cognitive Development*, 13(4), 453-478.
26. Richland, L. E., Morrison, R. G., & Holyoak, K. J. (2006). Children's development of analogical reasoning: Insights from scene analogy problems. *Journal of Experimental Child Psychology*, 94(3), 249-273.
27. Salvucci, D. D., & Anderson, J. R. (2001). Integrating analogical mapping and general problem solving: The path-mapping theory. *Cognitive Science*, 25(1), 67-110.
28. Sternberg, R. J. (1977). A componential theory of analogical reasoning. In A. W. Melton (Ed.), *Intelligence, information processing, and analogical reasoning* (pp. 134-172). Hillsdale, NJ: Lawrence Erlbaum Associates.
29. Thibaut, J. P., French, R., & Vezneva, M. (2010). The development of analogy making in children: Cognitive load and executive functions. *Journal of Experimental Child Psychology*, 106(1), 1-19.
30. Thibaut, J. P., French, R. M., Missault, A., Gerard, Y., & Glady, Y. (2011). In the eyes of the beholder: What eye-tracking reveals about analogy-making strategies in children and adults. In *Proceedings of the Thirty-Third Annual Meeting of the Cognitive Science Society* (pp. 1-6).
31. von der Malsburg, T., & Vasishth, S. (2011). What is the scanpath signature of syntactic reanalysis? *Journal of Memory and Language*, 65(2), 109-127.
32. Wright, S. B., Matlen, B. J., Baym, C. L., Ferrer, E., & Bunge, S. A. (2008). Neural correlates of fluid reasoning in children and adults. *Frontiers in Human Neuroscience*, 1(8), 1-8.
33. Yarbus, A. L. (1967). *Eye movements and vision*, New York, NY: Plenum Press.
34. Zou, G. Y. (2007). Toward using confidence intervals to compare correlations. *Psychological Methods*, 12, 399-413.

Acknowledgements: This work was funded by a James S. McDonnell Foundation Scholar Award to S.A.B. The authors thank Carter Wendelken, Alison Miller Singley, and Yana Fandakova for advice related to eyetracking analyses, and Keith Holyoak for his insights regarding an early version of the manuscript.

A5.6. Supplementary Methods

Expanded description of strategy classification analysis. In this analysis, we used AOI location, fixation duration, and position within the fixation sequence to generate a score reflecting participants' strategy within each trial. The score is based on the assumption that fixations that occur earlier in a trial and those that occur longer are more representative of AOIs that are more useful for solving the analogy problem. Therefore, for each fixation in the fixation sequence for a trial, we calculated the score as the product of its duration and inverse of its location in the sequence. Thus, longer fixations occurring earlier get higher scores and shorter fixations occurring later in the trial get lower scores. Although the idea to weight scores by position in the fixation sequence was the authors' idea, the notion that longer durations be more representative of deliberate thinking follows from previous eyetracking research investigating fixation sequences (e.g., von der Malsburg & Vasishth, 2011).

Each of the three strategies has differential predictions about which fixations and first-order transitions would be most informative while solving analogy problems. The project-first model (e.g., Sternberg, 1979) states that initial fixations and transitions within the source domain (i.e., the A and B terms) are most useful for generating a relation to be applied between the C term and the target. Therefore, the project-first equation would be the sum of the scores associated with these eye-gaze components:

Eq. 1. $Project\text{-}First\ Strategy = Score_{initial\ fix.A} + Score_{initial\ fix.B} + Score_{AB} + Score_{BA} + Score_{CT}$

Where initial fix. corresponds to the first fixation occurrence in the sequence *for the specified AOI* (i.e., not necessarily the first fixation within the sequence), AB corresponds to 1st-order fixation transitions from the A to the B terms, and T stands for the target in the analogy problem. Under the project-first model, transitions are reciprocal between the A and B terms, but are directed from the C term to the target.

As described in the main text, the structure-mapping model (e.g., Gentner, 1983) posits that initial fixations to and transitions between the A and C terms are most informative for generating a relation that could be applied between the B term and the Target. The structure-mapping equation was scored as follows:

Eq. 2. $Structure\text{-}Mapping\ Strategy = Score_{initial\ fix.A} + Score_{initial\ fix.C} + Score_{AC} + Score_{BT}$

Finally, the semantic-constraint model (Chalmers et al., 1992) claims that relations between the C term and the response choices can be used to constrain the appropriate relation shared between the A and B terms in the analogy. Therefore, initial fixations to the C term and response choices, as well as bidirectional transitions would be most informative when solving the analogy problem. Thus, the semantic-constraint equation was scored as follows:

Eq. 3. $Semantic\text{-}Constraint\ Strategy = Score_{initial\ fix.C} + Score_{initial\ fix.T} + Score_{initial\ fix.S} + Score_{initial\ fix.P} + Score_{initial\ fix.U} + Score_{CS} + Score_{SC} + Score_{CP} + Score_{PC} + Score_{CT} + Score_{TC}$

Where S, P, and U correspond to the semantic, perceptual, and unrelated lure response choices, respectively.

Once a score was calculated for each strategy based on the information in a trial, we classified each trial as belonging to a particular strategy if its respective score was greater than the score generated by either of the other two strategies. Therefore, for a trial to be classified as project-first, the score in Eq. 1 had to be greater than both the score from Eq. 2 and the score from Eq. 3. If the scores happened to be equal, no strategy classification would be given. Such an occurrence was rare, occurring in only 36 out of 934 trials included in the analysis. Therefore, for each participant we calculated the number of trials that were classified as each strategy.

We then calculated the proportion of each strategy for each participant to account for the different number of trials included in the analysis. From here, we took these proportions and correlated them with analogy accuracy to investigate whether using a particular strategy would allow us to predict a participant's analogy accuracy. We found that project-first was positively correlated with accuracy ($r = 0.624$, $p < 0.001$), structure-mapping was not significantly related ($r = -0.30$, $p = 0.14$), and semantic-constraint was negatively related with analogy accuracy ($r = -0.50$, $p < 0.01$). Thus, from our results, project-first was the most optimal strategy.

For access to the code used to generate the scores for the strategy classification, please follow the link below:

https://github.com/msv0915/bungelab_eyetracking/blob/master/VisAn_eyetracking_supplemental.py