Derivation of Idiographic Model-based Parameters from Resting State Neuroimaging Data

Patrick J. Rice

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Andrea Stocco, Chair

Susan Joslyn

Christian Lebiere

Program Authorized to Offer Degree:

Psychology

University of Washington

Abstract

Derivation of Idiographic Model-based Parameters from Resting State Neuroimaging Data

Patrick J. Rice

Chair of the Supervisory Committee:

Andrea Stocco

Department of Psychology

To work towards a complete and holistic understanding of human cognition, an essential task is to relate idiographic, or individual, parameters representing cognitive functioning to interactions between connected brain networks identified by neuroimaging methods. Here, using the ACT-R cognitive architecture, two models of cognitive strategy under two separate conditions of a working memory N-back are implemented. Individual participants who completed the N-back task are modeled by the determination of a set of parameter values which result in accurate model prediction of a given participant's response time and accuracy under the task. These parameter-based characterizations, in turn, demonstrate reasonable predictions of individual behavior, and are interpreted as reflecting differences between individuals in the contribution of multiple cognitive components to working memory function. A method of determining a subset of parameters that are critical for prediction of behavior is applied before

the resulting parameters are related to functional connectivity measures in individual resting-state fMRI data. It is observed that ACT-R model parameters interpreted to represent goal-directed attention, memory decay, and processing speed each reflect the coupling between the task-negative default mode system and a set of task-positive brain networks; however, coupling between different regions within these networks was specific to individual parameters. This approach seeks to validate previously reported associations between behavior, cognitive constructs, and brain activity while simultaneously extending these associations to account for differences in behavior and ability between individuals.

Table of Contents

Figure Reference

Table Reference

Chapter 1. Introduction

As humans, we possess a number of characteristics and traits that provide a commonality in our shared experiences. From the joy of welcoming a new person into the world to the grief of losing a loved one, humans respond to a wide array of experiences in shared manners. A factor that may be shared more commonly than any other, however, is the uniqueness, or *individuality*, with which we perceive and interact with the world. Two individuals may share an experience and even identify with each other's response to that experience, but unique aspects in the manner in which these individuals process this experience enforce a distinction in perspective between the two. These separate perspectives can result in divergences in behavior within and in response to these experiences. Differences in the processing of experience between individuals are observable in a wide array of behaviors, such as reading ability (Daneman & Carpenter, 1980), the learning of a second language (Skehan, 1991), or propensity for risk-taking (Bromiley & Curley, 1992). Furthermore, this variability in behavior across individuals is associated with differences in constructs designed to capture aspects of psychological functioning that underpin the behavior. For example, performance in reading comprehension is associated with measures of an individual's capacity in working memory (Cain et al., 2004; Daneman & Carpenter, 1980), a cognitive ability that supports the online processing of recent, task-relevant information in a goal-directed manner (A. Baddeley, 1992). A conclusion that stems from this observation is that the psychological processes that underlie an individual's ability in reading comprehension may involve working memory (at least in part), and differences between individuals in reading comprehension may be attributable to differences in the capacity of their working memory systems. Alongside this relationship with behavior, various psychological constructs are also

observed to be related to one another - the same measure of working memory capacity that is predictive of reading ability is also correlated to measures of fluid intelligence (Engle et al., 1999; Michael J. Kane et al., 2005), a construct that aims to capture reasoning and problem-solving ability (Cattell, 1963); this relationship is also observed when alternative measures of working memory function are considered (Jaeggi et al., 2008). Constructs like these are employed by models of cognition to formalize the constituent elements of a given psychological process; these models of cognition instantiate explanatory theory in a manner that allows for further prediction and potential support for the constructs and relationships considered by the model. In other words, cognitive models have the potential to posit an explanation for *why* differences in behavior are observed, and provide leverage to examine the veracity of the model's claims. Finally, these individual differences in behavior and constructs of psychological functioning have, in turn, been associated with individual differences in brain functioning (Dubois & Adolphs, 2016). Continuing the example of working memory, increased connectivity between anterior cingulate cortex (a region of the brain implicated in attentional regulation of cognitive processing (Botvinick et al., 2004; Bush et al., 2000)) and the left inferior frontal gyrus (observed to be involved in a wide range of functions, including working memory, response inhibition, and semantic/phonological processing (Liakakis et al., 2011)) is observed in individuals that display relatively greater working memory ability (Osaka et al., 2004). This observation could be used to support or refute a given model of working memory function on the basis of the model's predicted relationship between constructs that are associated with anterior cingulate cortex and the left inferior frontal gyrus. From another perspective, the model's formalisms allow for the increased understanding of the contribution of specific brain regions and activity to cognition and behavior. In this manner, the combination of model-based

characterization of behavior with measures of neurological functioning promotes refined understanding of the constructs that are considered by the model, as well as the neural substrates that may underpin these psychological processes.

The present work aims to extend efforts on delineating the relationships between individual differences in behavior, model-based characterization of individuals, and neurophysiology at the individual level. To accomplish this aim, two computational models of cognitive processes that contribute to behavior in two different conditions of the N-back task (a behavioral task commonly applied in cognitive neuroscience research on working memory) were implemented in ACT-R, a popular and broadly successful cognitive architecture (John R. Anderson & Lebiere, 2014). ACT-R includes empirically-supported formalisms/psychological interpretations of the parameters utilized in the modeling effort, allowing for the evaluation of the contribution of these constructs to observable differences in working memory function. Model-parameterized characterizations of individual performance under the task were produced by determining a set of parameter values for each participant that minimize error between a given participant's behavior under the task and the model's prediction of the same. Of these parameters, those most critical to a model's predictive efficacy are further examined for relationships to the observed functional connectivity of an individual's brain at rest, providing a link between individual differences in brain function, working memory ability, and behavior. This work will serve to further understanding of how individual differences in brain functioning are related to individual differences in ability, while simultaneously examining the validity of ACT-R as a cognitive modeling architecture by demonstrating the correspondence between specific components of the architecture and individual differences in brain functioning and behavior.

1.1 Working Memory

1.1.1 What is working memory?

"Memory", defined as a process or set of processes by which information is stored and accessed over a period of time, is a ubiquitous and familiar concept in day-to-day life. We all experience memory to varying degrees, such as remembering to (accessing the information related to) grab the coffee you put on the roof of your car, before driving away; remembering that when you get to the intersection down the road, you want to make a left instead of your usual right, as you need to go to the post office; or remembering that you sent a letter to your friend two weeks ago, and that you should check the mail in expectation of receiving a reply. This ubiquity of memory has resulted in the investigation of processes that support memory under multiple disciplines, including neuroscience and psychology (A. D. Baddeley, 2004; Milner et al., 1998). While these approaches take different perspectives on memory, each seek to determine a set of processes by which information can be reliably stored and accessed over a period of time. As suggested by this definition and the above examples, a critical component to these processes of memory is *time*, both in the retrospective sense (how much time has passed since a specific piece of information has been encoded into memory) as well as in the prospective sense (how far into the future that current information, eligible to be encoded into memory, is expected to be necessary or useful). The passage of time has been a consideration since the genesis of the modern study of memory. Herman Ebbinghaus, commonly referred to as a "pioneer" of the study of memory, investigated his own acquisition, rehearsal, and subsequent memory of lists of "nonsense" syllables over varying periods of time (Ebbinghaus, 1913). Ebbinghaus made note of an initial, early memory of the sequence under study, but also that the experience of this early memory did not necessarily guarantee that the sequence would be able to be recalled at a later

time. Around the same time as Ebbinghaus' experiments, a distinction between initial memory (termed "primary memory" at the time) and an individual's body of knowledge (termed "secondary memory") was made (James, 1890). "Primary memory" was proposed as the limited set of information held in the conscious mind at the present moment, while "secondary memory" was relegated to the knowledge that had been previously learned and utilized over the lifetime prior to the "present moment". These early observations and definitions of the form or quality of memory at different times provided early reasoning for the separation of memory into separate components.

The modern view divides the psychological construct of memory into three, rather than two, major stages: long term memory (LTM), short term memory (STM), and working memory (WM) (Nelson Cowan, 2008). LTM can be ascribed to James' (1890) proposal of "secondary memory" - a large collection of knowledge that was learned in the past and is available to an individual through both volitional as well as involuntary recall. One characteristic of this long-term memory recall is the relatively large amount of time required, when compared to other forms of human information processing (J. R. Anderson, 1983). Conversely, STM can be related to James' (1890) proposal of "primary memory" as a time- and capacity-limited temporary store of information. Compared to LTM, the information held in STM has a heightened "availability" to parallel cognitive processes that reduces the time necessary for retrieval; simultaneously, this short-term storage has strong limitations on the amount of information that can be held, as well as the duration for which this limited amount of information can be represented (Nelson Cowan, 2008). As such, there seems to be a "cost" to the heightened availability of this information in terms of these limitations. These views of short-term and long-term memory suggest a dual role for these systems in cognition, in which STM provides highly accessible, but limited, storage of

information that is expected to be behaviorally relevant in the current task or setting (N. Cowan, 2001; G. A. Miller, 1956), while LTM provides high-capacity, long-duration storage of information that, while not currently relevant, may become so in the future (Ericsson & Kintsch, 1995).

But what about working memory? There have been at least nine definitions proposed (Nelson Cowan, 2017), and furthermore, many resources conflate the terms "short term memory" and "working memory". However, most of these definitions and resources are in agreement that WM consists of multiple components, of which STM is one. The broadest distinction is that WM involves both short-term storage of information (through STM) and processing of this information by parallel cognitive domains in support of ongoing behavior. This view adds an active, intentional element to the idea of WM; WM capitalizes on a STM store in order to manipulate the information contained within (Daneman & Carpenter, 1980). Others adopt this "functional" view of WM, but emphasize a role of attention on information in STM as a necessary component to the processing of this information (N. Cowan, 1988; Engle et al., 1999; M. J. Kane et al., 2001). Attention in WM may serve to maintain currently active representations of information and "protect" these representations from concurrent processing of separate information, which could be either task-relevant or distracting, and can be broken down into the concepts of "control of attention" and "scope of attention". The scope of attention is potentially related to the capacity of working memory, and may underlie observed differences in the capacity for the short-term storage of passive information compared to the short-term storage of information undergoing active processing (Vecchi et al., 2005). Note that this WM capacity may be separate from STM capacity - the scope of attention may influence the number of items that can be actively processed ("attended to") concurrently in WM, compared to the passive

storage of information in STM, which is independent of attention. The control of attention, conversely, could be related to both the entry of information into WM as well as the active processing of this information once entered. As WM is of limited capacity and subserves behaviorally-relevant information processing, the entry of "distracting" or task-irrelevant information into WM could result in potentially deleterious effects on behavior. Control of attention to task-relevant information may assist in "filtering" distracting information before it is entered into WM. This control over attention to specific information may extend to processing of information within WM itself - irregardless of whether a single item is processed at a time (while others are maintained), or if multiple items are processed in parallel, control of attention would serve to limit interference between these actively-changing representations.

To add to the "multicomponent" view of WM, it has also been proposed that the modality of information to be represented relies on separate systems. Participants who perform pairwise combinations of a verbal memory task, a visual memory task, and a perceptuomotor tracking task show little to no disruption of performance on either of the tasks performed simultaneously (Cocchini et al., 2002), suggesting that at least two separate systems can operate in parallel to support working memory performance in separate sensory domains. However, it has also been observed that participants performing an auditory WM task and a visual WM task in tandem have degraded performance in each task relative to performing one of these tasks alone (Klingberg, 1998). These participants underwent measurement of regional cerebral blood flow (through PET imaging) while performing these tasks, and it was observed that the tasks activated overlapping areas in the prefrontal, cingulate, and inferior parietal cortex when each of the two tasks were performed alone as well as in the dual-task condition. This observation led to the proposal that competing activity supporting task performance in these shared cortical areas was a

source of interference between the tasks, resulting in the observed degraded performance in the dual-task condition. The results of these two studies may initially seem to be at odds, but can be consistent if one considers the WM system as a set of modality-specific "stores" of memory that are manipulated through a shared attentional mechanism (Conway et al., 2005). In this case, some dual-tasks may allow attention to be effectively "switched" between these stores in a manner that leads to robust performance on each, while other dual-tasks may demand attention to multiple stores simultaneously, thereby introducing interference in the application of processing, with reduction in performance as a result.

Efficacy in dividing attention between multiple items in multiple stores may be partially driven by the availability, or "level of activation", of these representations. This "activation" can be interpreted in working memory terms as the fidelity of the representation of the item, or equivalently in neural terms as the strength and cohesiveness of the networked neural activity involved in the item representation (Nelson Cowan, 2008). Attention towards a given item may increase the activation of shared information through the spreading of networked activation, and rehearsal or repeated retrieval of a given representation results in stronger encoding and increased availability (John R. Anderson, 1983; J. R. Anderson, 1983). Interference, or the reduction of activation, between items may stem from feature overlap in item representation, with shared networks of neural populations divided between the representation of two or more items simultaneously (Klingberg, 1998). Another manner in which information can decrease in accessibility or "level of activation" is in a time-dependent fashion (Portrat et al., 2008). If the activation of a given representation is not increased through attention, retrieval, or shared representation, this activation declines until the item becomes no longer available to be attended to or recalled. This concept of "decay" therefore provides a source of forgetting in WM.

Rehearsal, or repeated sequential attention to the individual elements of information held in WM, serves to "boost" this level of activation and maintain accessibility. However, there is disagreement as to whether evidence for temporal decay in WM exists (Barrouillet et al., 2011; Lewandowsky & Oberauer, 2009; Oberauer & Lewandowsky, 2013, 2014; Portrat et al., 2008). The discussion consists of two sides: one posits that the phenomenon of "forgetting" information held in WM can be attributed (at least in part) to a time-related decay of the representation with WM, while the other claims that forgetting is sourced not in decay, but rather interference between competing representations. Both agree that a mechanism such as rehearsal provides the opportunity to restore the availability of the memory - in the case of decay, by "refreshing" the level of activation of the the memory; in the case of destructive interference, by expanding representations of information with additional features, or by re-coding representations to minimize feature overlap. In an interesting link to the potential role of attention in WM, both views also consider attention as a necessary component of this "restoration", in that attention to items in WM is necessary to either "refresh" the activation (in the face of decay) or to "restore" a degraded representation by re-coding.

1.1.2 Models of working memory

A number of models of working memory, or models of memory that include a WM component, have been proposed over the years. One early and highly influential model was the multi-store model of memory (Atkinson & Shiffrin, 1968). While Atkinson and Shriffin's model did not include a WM component, it does consider a number of concepts and elements that became adopted by subsequent WM literature, and so is worthwhile introducing as a "foundation" or "precursor" of models of WM. The foremost concept of the Atkinson and Shriffin model is that human memory is composed of three separate components: a sensory

register, where information derived from sensory processing exists very briefly; a short-term store, which encodes information from the sensory register in a primarily auditory format to be represented over a short period of time; and a long-term memory store, which can encode information from the short-term memory store into a more indefinite, semantic form, and subsequently retrieve this stored information back into the short-term memory store for use. It was proposed that each of these three components differ in the manner in which they encode information, their capacity for information, and the duration for which information can be held in the store.

Another core aspect of the multi-store model of memory is the specification of processes that control the movement of information between these separate stores. Under this model, attention to information in the sensory register is necessary for the encoding of the attended information into the short-term store. The majority of information that enters the sensory register is not attended to, and decays or becomes overwritten very quickly. The information in the sensory register that is attended to is entered into the short-term store, where it can be conserved for a limited time through maintenance rehearsal, in which the information is repeated without further processing for meaning or association. Information can be lost from the short-term store through either decay (when it is not rehearsed), or through displacement (when the short-term store is at capacity, and newly-entered information takes precedence over old information). To be encoded into the long-term store, information in the short-term store must undergo elaborative rehearsal. This rehearsal process differs from maintenance rehearsal in that it requires processing of the information in the short-term store to a level that allows it to be associated in a meaningful way with information already present in the long-term store. If this occurs, the information in the short-term store is transferred to the long-term store, where it can subsequently be retrieved

back to the short-term store for use in the future. The Atkinson and Shriffin model utilizes a number of concepts that have been carried forward into conceptualizations of working memory: that multiple components are necessary to model observed features of memory; that these components interact through different mechanisms; that representation of information may be modality-specific; capacity and duration limitations in storage of information; the processes of rehearsal and retrieval; and the role of attention in the successful representation of memory.

While the Atkinson and Shriffin model was successful in modeling a number of memory phenomena, one critique was that the unitary short-term store was too simple to describe the myriad behavioral phenomena associated with memory. For instance, the Atkinson and Shriffin model predicts that information stored in long-term memory is "gated" by processing in the short-term memory store; if there is dysfunction in or excessive demand on the short-term memory store, then the ability to transfer information to the long-term store in an effective manner should be compromised. Additionally, the Atkinson and Shriffin model posits that the short-term memory store acts as WM through this "gating" mechanism, and as such, dysfunction/excessive demand of the short-term store should have a negative impact on cognitive performance. In fact, patients with deficits in short-term memory capacity demonstrate normal long-term memory (Shallice & Warrington, 1970), and while participants who are asked to perform cognitive tasks while under progressively demanding memory loads decline in the speed at which they perform, error rates in the cognitive tasks remain steady as WM load is manipulated (A. D. Baddeley & Hitch, 1974).

To address these discrepancies in the predictions of the Atkinson and Shriffin model, Alan Baddeley and Graham Hitch proposed a multi-component model of working memory (A. D. Baddeley & Hitch, 1974). In their model, the singular short-term memory store of the

previous model is replaced by a system comprised of three components - two separate stores which interact with a third “central executive” component. The two separate stores, a “phonological loop” and “visuospatial sketchpad”, account for the proposed division in the representation of modality-specific information - the phonological loop represents information that was presented in spoken or written form, while the visuospatial sketchpad handles spatial and object-feature information derived from the visual scene. These stores, assumed to be of limited capacity, implement a “rehearsal” procedure across the information contained in the store, reiterating the information contained within in order to maintain the fidelity of the representation in the face of an assumed decay of this information over time. The third component, the “central executive”, implements attention-based control processes that govern the operation of the two stores, alongside the manipulation of the information contained therein. As they are based in attention, these control processes include those dedicated to focusing attention on particular representations, as well as those that allow for divided attention across multiple representations, thereby reflecting the “control” and “scope” of attention, respectively (Alan Baddeley, 2007). Correspondingly, the central executive itself is not thought of as a singular entity, but rather a synthesis of multiple, co-occurring executive functions. Later, a fourth component, termed the “episodic buffer”, was included in the model alongside the initial three components (A. Baddeley, 2000). This inclusion was necessitated by the observation of amnesiac patients who nevertheless were capable of short-term recall of an amount of information thought to exceed the capacity of the individual stores (Alan Baddeley & Wilson, 2002). The episodic buffer serves to integrate information between the short-term memory stores (phonological loop/visuospatial sketchpad), perception, and long-term memory in a manner that imposes a chronological sequencing (hence the term “episodic” buffer) and makes these

representations available to the conscious mind. These subsystems of the multicomponent model function in an interactive fashion to represent and manipulate information on a short-term basis. As such, the multicomponent model carries forward many of the core tenets of working memory as described by Ebbinghaus, Atkinson and Shriffin, and the like, including modality-specific representation, decay and rehearsal of these representations, and the role of attention in controlling the processing of information represented in stores. The major advancement that the multicomponent model introduces is the conceptualization of working memory as a set of interacting stores and processes that rely on the combined contributions of multiple cognitive domains operating in parallel, rather than a sequential series of stores with passive representation of information in each location. From this perspective, ability in working memory can be attributed to the efficacy of operation within each of the involved cognitive domains alongside an individual's ability to synthesize these processing streams in an effective and goal-directed manner.

One additional conceptualization of working memory is useful to introduce as an alternative perspective to that of Baddeley and Hitch, although there are multiple others that have found success (such as the Time-based Resource Sharing model (Barrouillet & Camos, 2020) and neurobiological models such as (E. K. Miller et al., 2018)). Nelson Cowan's Embedded-Processes model (Nelson Cowan, 1999) proposes that working memory is composed of a limited-capacity attentional system that serves to focus on activated representations in long-term memory. Cowan establishes a set of principles of the model which each consider correspondences between attention and memory. The first of these principles describes the model's formalization of working memory information; the contents of working memory are the areas of long-term memory that have been sufficiently activated through parallel processing of

shared features in other cognitive domains, and subsequently "focused" on by attention. This primary principle absolves the brain from a dedicated representation of working memory information, and instead conceptualizes working memory information as existing long-term memory representations that are more activated than other LTM representations and have become the focus of attention (partially enabled by this increased activation). This is a sequential process; it is assumed that information in LTM cannot be attended to unless that information had first become activated to some degree. Note that Cowan makes a distinction between attention to representation in LTM and attention to representation elsewhere; it is assumed that this increase in activation of particular representations in LTM is sources from processing of shared features represented in other cognitive domains, with attention to those representations enhancing the resulting LTM activation. The second principle treats the idea of limits in WM, and establishes two sources of these limitations: that the focus of attention is capacity limited (the "scope of attention" idea), and that the activation of LTM as a critical component of WM is time-limited (the idea of decay in representation). Cowan also suggests that these limitations become especially relevant in situations in which additional factors may increase cognitive load, such as when interference between items with overlapping features occurs in WM (Nelson Cowan, 1999). Under Cowan's model, interference would occur when two separate representations with shared features become co-active in LTM. When this occurs, attention may not be able to be appropriately or efficiently focused on the intended representation, resulting in slowed processing of the correct representation, or even processing of an incorrect representation. This co-activation likely also affects the decay of these representations; while the rate of decay may be slowed by the processing of shared features that initially elicited the co-activation (through spreading activation; these two co-activated representations may even spread activation to each

other through the shared features), an increase in the rate of decay could also result from improper attentional allocation, sourced in the aforementioned interference. The third principle of the Embedded-Processes model assumes that the focus of attention reflects the combined influence of voluntary processes (here, Cowan explicitly names a central executive system, akin to Baddeley's central executive component) and involuntary processes, which Cowan ascribes to an automatic attentional orienting system. The orienting system allocates attention to newly presented or changed stimuli on the basis of change, but can become habituated to a repeatedly presented stimulus that does not undergo change. When attention to habituated stimuli is demanded by goal-directed behavior, the central executive system enables attention to be voluntarily oriented in an effortful manner. Cowan's fourth principle considers how this habituation affects the activation of representation in LTM; the model assumes that perceptual processing of these habituated stimuli still spread some amount of activation to LTM, even if they are not processed to a depth that results in conscious awareness of the stimuli. Attention to these perceptual representations (and subsequent "awareness" of them) is associated with an increase in the activation of related LTM representations. This is the final, fifth principle of the Embedded-Processes model of working memory; awareness enhances the depth of processing, resulting in increased activation of LTM. Awareness in perceptual processing leads to increased fidelity in the features of the stimulus being processed, which in turn elicits stronger activation of LTM representations that share these features. Once these LTM representations have become active enough, subsequent allocation of attention to these items increases the activation of related episodic representations in LTM, again providing a source of increased activation and availability.

Cowan's Embedded-Processes model shares many of the same aspects that the "multi-store" (Atkinson and Shriffin) and "multicomponent" (Baddeley and Hitch) models rely on to describe the cognitive ability labeled "working memory". Each of these models emphasize a role of attention in the representation of information in working memory, and in each, attention provides a similar function of defining the information that is held in working memory. In Atkinson and Shriffin's multi-store model, attention to information in the sensory register "gates" entry of this information to the short-term store, but has little influence afterwards; instead, the depth of processing that occurs to information in the short-term store governs encoding to LTM. The multicomponent and Embedded-Processes models expand the role of attention to the control of processes that are applied to or in conjunction with this information, and define sources of this attention; both of these later models consider a "central executive" source of voluntary, consciously directed attention, while Cowan's model also includes automatic orientation of attention to changing features of the environment. While the Embedded-Processes model does not define individual components of WM as distinctly as the multi-component model, both assume that the construct of working memory relies of a set of subsystems that act in cohesion to support the ability to dynamically represent and manipulate goal-relevant information sourced in both the environment as well as previous memory. These theoretical models of working memory provide a framework which can be used to evaluate behavior thought to rely on working memory ability.

1.1.3 Measures of working memory

There are two major behavioral measures of working memory ability that are regularly employed in the working memory literature - "span" tasks and the "N-back" task. "Simple" span tasks solely measure storage and maintenance of information without additional processing

demands, and are therefore better suited to measure STM instead of WM. As an example of a simple span task, a "digit span task" presents participants with a certain number of digits one at a time, before the participant is asked to recall the presented digits in either forward or reverse order. If the participant is successful in recalling the list, the length of the next list to be presented is increased by one; if not, the length of the list stays the same, and the task ends when the participant is unable to correctly recall a list of a given length twice in a row. The digit span task produces a measure of the STM capacity of an individual through the maximum length of correctly-recalled lists. While participants must rely on memory of the digits in the order they appeared (possibly supported through rehearsal), the digit span task does not demand that the participant perform a cognitively demanding task *while* maintaining this information, and so it is a poor measure of *working* memory.

Working memory "span" tasks extend the simple span tasks to include the simultaneous processing of additional or unrelated information alongside the memory component ("complex" span tasks) (Conway et al., 2005). These additional processing demands require the engagement of attention in order for a participant to be able to successfully engage in all aspects of the task, thereby necessitating a system such as WM. For example, in the "operation span task" (Turner & Engle, 1989), participants are presented with a series of words to be remembered (similar to the digit span task), but these presentations are interspersed with the presentation of "mathematical operation strings". Each presented mathematical operation string consists of an equation with two arithmetic operations on one side and a stated solution on the other. Half of the presented solutions are correct; on each presentation, the participant is asked to verify the shown solution. Presentation of a mathematical operation string is followed by the presentation of a word to be remembered; once all words are presented, the participant's recall of this list of words is

examined. While this is a singular example, WM span tasks largely follow this format of items to-be-remembered interleaved with items to-be-processed, and produce a capacity measure of WM, similar to the simple span tasks. It is important to note that simple span tasks usually produce a different estimate of memory capacity compared to complex span tasks - approximately 7 ± 2 items under simple span tasks compared to approximately 4 ± 1 items under complex span tasks (Mathy et al., 2018). These types of WM span tasks are predominantly utilized in behavioral studies investigating cognition and individual differences in WM (Redick & Lindsey, 2013). Additionally, complex span tasks are eminently reliable and valid measures of WM function: they are more strongly associated with other memory tasks that require information processing compared to memory tasks that do not (Engle et al., 1999); they have been found to correlate with measures of intellectual ability (such as fluid intelligence), which is theorized to be dependent on WM processes (Conway et al., 2005; Daneman & Carpenter, 1980; Engle et al., 1999; Engle & Kane, 2004); and shortened versions of complex span tasks reliably measure WM capacity along with demonstrating the same relationship with fluid intelligence (Felez-Nobrega et al., 2018; Foster et al., 2015).

The N-back task, on the other hand, is more commonly applied in cognitive neuroscience studies focusing on the neural mechanisms that support WM (Redick & Lindsey, 2013). In the N-back task, participants are presented with a series of stimuli (akin to the span tasks), and are asked to report whether the currently presented stimulus is identical to the stimulus that occurred N presentations ago. The value of N is manipulated to make the task more or less difficult to perform; performance usually degrades as the value of N increases (Jaeggi et al., 2010; K. M. Miller et al., 2009). A special case of the task occurs when N=0; here, a cued stimulus is presented before the presentation of a block of trials begins, and the participants are asked to

report whether each of the presented stimuli are a match to this cue. Compared to complex span tasks, the N-back task also requires active processing while information is held in memory, but one major difference is that the N-back task requires that that processing be done *with* the information in memory, rather than in parallel to the information in memory. To accurately perform the task, participants must maintain a "window" of at least the last N stimuli; memory of the Nth stimulus is necessary in order to compare against the current stimulus, while the stimuli presented between the Nth and the current stimulus need to be maintained in order to update this "window" when a new stimulus is presented. As such, the processing demanded by the N-back task manipulates the information held in WM: there is a comparison process between an item in WM and the current stimulus; the items held in the "window" are "updated" to reflect new positions in the sequence; old items may have to be actively "discarded" to minimize interference with ongoing performance. Despite the shared focus on combining storage demands with processing demands, the N-back task demonstrates relatively low convergent validity as a WM *capacity* measure, when compared to complex span tasks. N-back performance has been observed to show little to no relationship to complex span measures in multiple studies (Michael J. Kane et al., 2007; K. M. Miller et al., 2009; Redick & Lindsey, 2013), and also demonstrates low behavioral reliability (Jaeggi et al., 2010), which indicates that the N-back task may not serve well in capturing individual differences in working memory capacity.

This is not to say that the N-back task does not capture *some* aspect of WM ability in a reliable fashion, only that it may not measure WM *capacity* as well as complex span tasks. There is evidence that brain activation during the N-back task (as measured by fMRI) is reliable across participants (Drobyshevsky et al., 2006) as well as time (Caceres et al., 2009). It has also been observed that training on variants of the N-back task increases measures of fluid intelligence (Au

et al., 2015; Jaeggi et al., 2008), and N-back and complex span measures explain unique variance in fluid intelligence (Michael J. Kane et al., 2007). This suggests that the N-back and complex span tasks may be capturing two different aspects of WM, each of which contribute to general fluid intelligence; complex span tasks may more accurately probe an individual's capacity in information retention while performing cognitive processing in parallel, while the N-back task may capture an individual's ability to efficiently perform processing on a subset of this retained information in a manner that does not interfere with simultaneous representations that are not currently undergoing processing. This perspective is supported by the observation that N-back accuracies are related to measures of processing speed (K. M. Miller et al., 2009), and individuals with multiple sclerosis, who demonstrate slowed processing speed in multiple domains (Litvan et al., 1988; Rao et al., 1989), are impaired in an N-back-like task (Lengenfelder et al., 2006).

1.1.4 Individual differences in working memory

Variation between individuals in behavior that is thought to rely on working memory ability has been a topic of interest for as long as working memory itself, with research consistently considering how differences between individuals inform understanding of working memory (Daneman & Carpenter, 1980; Jarrold & Towse, 2006; Unsworth & Engle, 2007). Hermann Ebbinghaus remarked on these individual differences in his seminal work, noting that some individuals retain and recall memory more effectively than others (Ebbinghaus, 1913). Understanding of how individuals differ in their behavior under task paradigms that are thought to require a contribution of working memory can provide leverage in the determination of the cognitive mechanisms that contribute to working memory ability.

Working memory capacity, evaluated through complex span tasks, is one measure of working memory ability for which inter-individual variation is consistently observed across materials and tasks (N. Cowan, 2001). As introduced in section 1.1.3, capacity measures under complex span tasks result in capacity estimations of approximately 4 ± 1 items; however, there is also evidence that multiple items can be "chunked" together, consolidating the representation of a number of related concepts (N. Cowan, 2001). It has been proposed that differences in working memory capacity between individuals could arise from two components of working memory: attention-based processes and encoding-based processes (Barrett et al., 2004; Oberauer et al., 2007; Unsworth & Engle, 2007). Variance between individuals in the ability to allocate attention to multiple items (the "scope" of attention concept) could account for difference in capacity limits; the observed "working memory capacity" is actually a capacity in the number of items that can be attended to in a manner that allows for adequate task performance. This view is usually adopted by models of working memory that view working memory representation as a subset of LTM that is activated above some threshold and available for interaction with attentional processes (Nelson Cowan, 1999; Unsworth & Engle, 2007); however, Baddeley and Hitch's multicomponent model also considers the central executive component (to which attentional processes, among other executive functions, are attributed) to be limited in ability to allocate attention. Differences in the ability to encode multiple item representations in parallel may be another source of observable capacity differences (Jarrold & Towse, 2006; Unsworth & Engle, 2007). This account of differences in capacity assumes that there is a limited set of resources that are available for the representation of items in working memory; the capacity of the system is determined by the efficiency with which these resources can be utilized for representation. Shared features between items in working memory may interfere with each other,

taxing resources and reducing capacity; however, shared features may also promote chunking where appropriate, potentially having a bidirectional effect on observable capacity, dependent on the informational demands of the behavioral task. Individuals could potentially differ in the amount of interference that occurs between two given representations, or the degree to which that interference demands additional attention or representational resources. However, it is likely that both of these potential explanations of capacity in working memory contribute to observable differences; greater ability to allocate attention may enhance encoding of item representation, reducing interference between representation or allowing for chunking of appropriate items to occur.

Efficacy in processing of working memory information is another measure of working memory ability in which individuals differ (Jaeggi et al., 2010; K. M. Miller et al., 2009; Redick & Lindsey, 2013). Complex span tasks require participants to hold in working memory a set of items over time, but do not require processing to be done using these item representations - instead, processing is performed on an alternative set of information that does not have to be stored during ongoing performance. As a result, complex span tasks do not examine the ability to perform processing using the same information that must be maintained. The N-back task addresses this discrepancy by requiring continual updating of items in working memory, enforcing the need to perform processing on these items. Accuracy in the N-back task provides a measure of this processing efficacy, assuming that greater ability to perform updating processing on working memory information results in greater success in distinguishing N-back targets from lures and nontargets. Differences in accuracy over varying values of N are observed within individuals, indicating that this updating process becomes more difficult to perform successfully as the number of items necessary to update increases. A potential explanation for this increase in

difficulty as N increases is that it requires the consideration of item representations that have decayed to a considerable degree (as relatively larger values of N require representation of items that occurred further back in the past). However, differences between individuals within a given value of N are also observed, indicating that individuals vary in their ability to perform this updating procedure. While differences in capacity may partially contribute to differences in N-back accuracy at relatively larger values of N (such as N=3), this would not explain observable behavioral differences at lower values of N. Instead, differences in efficacy of processing may be sourced in individual differences in the ability to control attention. During updating, it is thought that attention must be focused on each item in a sequential manner (Kessler & Oberauer, 2015). Before it is updated, an item's relevant features may have to be interpreted to determine the intended outcome of updating for the item. The ability of attention to be allocated to this item may be partially dependent on the activation of the item, relating decay of representation to efficacy in updating. The item itself must then be updated in accordance with this intended outcome, the process of which may also be dependent on the item's decay, as well as proving the potential of interference with parallel representations that have either already been updated or are awaiting updating. Alternatively, the intended outcome of updating may be determined for all items before updating of each item occurs. In either case, greater ability in the control of attention would enhance the updating process by allowing for more efficient or thorough evaluation of the items to be updated, as well as the application of the transformation of these item representations during the course of updating.

1.1.5 Neural correlates of working memory in humans

While there is ongoing and lively discussion as to what exactly working memory *is* (that is, how the concept of working memory can be formalized), there is no question that working

memory processes are produced by the physiology of the brain (E. K. Miller et al., 2018). The concept of localization of function suggests that particular regions of the brain should be activated in a reliable and consistent manner when working memory is engaged; however, the "multicomponent" view that the predominant models of working memory adopt indicates that there is likely not one particular area of the brain dedicated to the implementation of working memory processes. Rather, a number of regions or circuits are likely recruited to varying degrees as demanded by the task - whether that be through differing modalities of task-relevant information, or varying demands on the number of items to be remembered (demands on capacity), duration over which these items are relevant (demands on decay of representation and maintenance processes), division of attention over these items (the control of attention), or the form of processing of these items required by the task (whether items are maintained or actively transformed through processing). These regions or circuits are not involved in working memory alone - instead, they encode independent processes, useful across a range of behaviors, that are recruited in support of working memory function. Because of the potential for a myriad of brain regions to be involved in working memory function, as well as the potential for these regions to be involved in functioning not related to working memory, care must be taken in the identification of brain regions thought to contribute to working memory. Investigation of the neural substrates of working memory function has taken place across the spatial and temporal scales relevant to brain measurement, from the activity of single neurons (Fuster & Alexander, 1971; Goldman-Rakic, 1995) to the functioning of distributed cortical networks across the entire brain (D'Esposito et al., 1999, 2000; Pessoa et al., 2002). While these perspectives each contribute unique and valuable information concerning the functioning of the brain, this work will solely review findings from research performed through functional magnetic resonance

imaging (fMRI), as these results best correspond to the features of the theoretical (e.g. the multicomponent or Embedded-Processes models) and computational models (e.g. the ACT-R cognitive architecture; described below) considered here. fMRI images the blood-oxygenation-level-dependent (BOLD; a measure of the degree of oxygenation of the blood supply to the brain) signal in a 3D array of cubic voxels, which encompass the volume of the brain. While the BOLD signal is not a direct measure of neuronal activity, it has been demonstrated that this signal is a strong predictor of underlying neuronal activity (J. H. Lee et al., 2010; Siero et al., 2014), indicating that the BOLD signal at a given position in the brain can be related to the physiology of the neuronal populations at the same location. If these neurons are involved in processing task-related information, then this activity is likely to be reflected in the BOLD signal.

The primary manner of determining brain-behavior correspondences using fMRI is to have participants perform a working memory task paradigm such as a complex span or N-back task in the scanner, while the brain is being imaged. This approach directly examines increases and decreases in regional activity that co-occur with behavior under the task. However, another approach in evaluating brain-behavior relationships with fMRI is to relate the activity of the brain at a resting state during one period of time to the behavior of an individual performing a task at another point in time. This approach seeks to determine if there are intrinsic, fundamental aspects of brain functioning that influence brain activity both at rest and during task behavior (M. H. Lee et al., 2013). It has been demonstrated that resting-state activity predicts task-based fMRI activation (Cole et al., 2016; Tavor et al., 2016), indicating that the physiology of the brain at rest shares features with the brain in a task-based state - these shared features may be relatable to cognitive processes that are both task-general and active at rest (i.e., those that contribute to

"internal tasks"). By comparing observations made between task-based and resting-state fMRI, two general sets of brain networks (sets of brain regions which co-activate under certain circumstances) have been observed, which are usually labeled as "task-positive" (those found to exhibit increased activity during active task behavior, and decreased activity at rest) and "task-negative" (those that exhibit the opposite tendency) (Chai et al., 2012; Fox et al., 2005; Power et al., 2011). Task-positive networks have been identified to include the dorsal attention system, the fronto-parietal task control system, and the cingulo-opercular task control system, while the task-negative system is mainly composed of the default mode network (DMN) (Power et al., 2011). The DMN is composed of regions found to show increased activity at rest over task-based states (Raichle et al., 2001). While they are anticorrelated, these task-positive and task-negative networks interact both at rest and during task performance; the nature of these interactions between opposing systems is likely to influence cognitive processing in both states. As resting-state activity is predictive of task-based activation, the resting-state dynamics between regions within a network, as well as between regions across networks, are likely to be informative of the same dynamics under task performance, providing a window to estimate task-influential factors on the basis of activity at rest. To quantify these dynamics, the functional connectivity between brain regions can be calculated (Rogers et al., 2007; van den Heuvel & Hulshoff Pol, 2010). Functional connectivity is usually defined as the correlation between fMRI activations of two brain regions over a specified period of time; a "functional connectome" can be produced by computing the pairwise correlations between all regions of a given parcellation of the brain. The measure can be interpreted as reflecting how predictive the activity of one region is of the activity of another over the specified duration, but does not indicate that the two regions are directly connected in a causal manner.

These approaches to the evaluation of brain-behavior relationships have been leveraged to investigate the contribution of specific brain regions and networks to working memory ability in a number of studies. A consistent set of brain regions (and by extension, cognitive networks that include these regions) have been identified through this work. A common set of brain regions identified across studies include those attributed to a fronto-parietal task control network, including the lateral prefrontal cortex (PFC) and parietal cortex (Barch et al., 2013; D'Esposito et al., 2000; Gazzaley et al., 2004; Lamichhane et al., 2020; Owen et al., 2005). Performance in the N-back task has been related to lateral PFC activity that occurs during behavior (Barch et al., 2013; Lamichhane et al., 2020), and resting state functional connectivity of the lateral PFC region has been observed to be predictive of N-back task accuracy (Lamichhane et al., 2020). Other studies have further divided lateral PFC into dorsal and ventral subregions, determining that while both are involved in maintenance of WM information, dorsolateral PFC is particularly active during manipulation of WM information (D'Esposito et al., 1999, 2000; Gazzaley et al., 2004). Rostral PFC, adjacent to lateral PFC in the anterior direction, has also been implicated in a meta-analysis of WM neuroimaging studies. Regions of other task-positive networks that have been found to be related to working memory function include the anterior cingulate cortex (ACC) (Barch et al., 2013; Owen et al., 2005), a critical region of the cingulo-opercular task-control network, as well as the intraparietal sulcus of the dorsal attention network (Gazzaley et al., 2004). However, alongside these brain regions that participate in task-positive networks, a number of those associated with the task-negative DMN have also been identified as correlates of WM function. Deactivation of regions of the DMN, including medial prefrontal cortex, posterior cingulate, and the occipital-parietal junction has been observed in a large cohort of individuals who were scanned while performing the N-back task (Barch et al., 2013) and the

extent of this deactivation of the DMN, as measured through functional connectivity, has been related to individual task performance under the N-back task (Esposito et al., 2009). Functional connectivity can also relate the activity of task-positive networks to that of the DMN - it has been observed that individuals with relatively stronger negative correlations between DMN and a "WM network" (composed of regions belonging to the task-positive networks described above) in both resting-state and task conditions exhibit better behavioral performance in an N-back task (Sala-Llonch et al., 2012). As such, the results of these studies indicate not only the involvement of task-positive brain regions associated with executive task control and attention, but also their activity relative to task-negative regions. This indicates that resting-state activity, particularly the correspondence between the DMN and task-positive networks at rest, may be predictive of working memory ability. The degree of anticorrelation between specific task-positive brain regions and the DMN may reflect differences in the contribution of multiple cognitive components to working memory function; differences in these relationships between individuals may underlie differences in behavior under working memory paradigms. To further investigate this, characterization of individual differences in the contribution of multiple cognitive components is necessary. One approach to this characterization is computational cognitive modeling.

1.2 ACT-R Cognitive Modeling Architecture

Adaptive Control of Thought - Rational (ACT-R) is a theory of the architecture and process of human cognition, instantiated within a computational framework (John R. Anderson & Lebiere, 2014). ACT-R combines both symbolic components (in the form of an IF-THEN system that treats information represented as "chunks", described below) and subsymbolic components (in the form of equations and parameter values) in a manner such that the

subsymbolic equations contribute to the operation of the symbolic processes. In other words, specific elements of the symbolic processes utilized by ACT-R are operationally defined as mathematical entities or relationships. ACT-R theory encompasses a wide range of human cognitive functions, including memory, utility learning, and perception in multiple modalities. This flexibility and depth allow for an individual, or "modeler", to instantiate specific theories, or "models", of cognitive functioning in specific circumstances. This is most commonly applied to task-oriented cognition, including topics such as mental arithmetic, operation of a motor vehicle, and skill acquisition (Gunzelmann et al., 2011; Lebiere, 1999; Tenison & Anderson, 2016); more broadly, ACT-R provides a cognitive framework for the modeling of human decision-making (Dimov et al., 2020; Gonzalez et al., 2003; Marewski & Mehlhorn, 2011). However, ACT-R has also been applied to mind-wandering and other resting state conditions (van Vugt et al., 2018, 2015), demonstrating broad capabilities in capturing the dynamics of human cognition.

1.2.1 ACT-R's formalization of the internal and external environment

In order to be capable of effectively modeling human cognition under this broad array of situations, ACT-R represents both the external world and the "internal" theorized facets of a cognitive mind. Additionally, ACT-R maintains a "meta-process" which overarches the external and internal components, instituting a simulation timeline upon which the meta-process schedules and executes events at specific times. These events can occur in either the external or internal environments, and are executed sequentially in the temporal order in which they are scheduled. In this manner, ACT-R includes the passage of time as a central component of the interaction between cognition and an external environment. Two essential components of ACT-R's "external environment" are the visual and motor modalities. The visual modality is

modeled by the *visicon*, which makes all features of all objects of the current visual scene available to the ACT-R model. Visual objects are composed of features, both of which are defined by the modeler and presented on the visicon (scheduled through the meta-process) for the model to detect. The motor modality provides a pair of "hands" that an ACT-R model can use to manipulate objects in the external world. These "hands" are considered as being placed above a horizontal plane in front of the model on which items can be placed to be manipulated. By default, an ACT-R model is provided with a keyboard and mouse, although it is possible to extend ACT-R to include other manipulable items. Movements of these "hands" (and therefore potential manipulation of items) requires specification of a type of movement and a specific effector (hand, or finger on a hand) to make that movement. Examples of movements include moving the hands to new locations, moving items to new locations, or use of individual fingers to interact with items (such as "punching" a key on the keyboard with a specific finger).

For the internal cognitive mind, ACT-R's computational framework consists of a set of *modules*, which at the theoretical level are assigned to a general cognitive function that is usually associated with the function of a specific brain region. Examples of modules in ACT-R include (but are not limited to): a declarative memory module, which provides a mechanism for the storage and retrieval of declarative knowledge; an imaginal module, which allows for the maintenance of contextually relevant information that is needed in the short-term (that is, without retrieval from declarative memory); a goal module, which acts similarly to the imaginal module, but maintains state-based information; a visual module, which supports visual attention through "what" and "where" subsystems; and a motor module, which mediates the production and execution of motor actions in a representation of the physical world. Note that some of these modules, such as the visual and motor modules, interface with the external environment in order

to allow the model to sense and interact with elements of that environment, while other modules, such as the goal module or the imaginal module, are purely internal. The modules of ACT-R contain *buffers*, which allows a given module to be interfaced with (by other modules of the architecture, or by a human modeler). A module's buffer(s) represents information currently in use by the module in a manner that is available to all other modules, and in addition is used to parse queries about the current state of the buffer/module or requests for the module to perform some action. States and actions of modules are specific to a given module; in an example of a module state, the visual module possesses the state "`scene-change`", which is flagged as being true or false and represents whether a change in the visual scene has been detected in the recent past. An example of a module action is a request to the declarative module to attempt to retrieve a specified representation of information (discussed below).

1.2.2 Knowledge representation in ACT-R

ACT-R's set of modules interact with and are utilized by ACT-R's knowledge representations. One core assumption of ACT-R theory is that knowledge is represented in two different forms, declarative and procedural. Declarative knowledge represents semantic facts such as, "the stimulus was a red square", or "two comes before three, and after one". In ACT-R, these forms of declarative knowledge are represented in the form of *chunks*, which are a set of attributes termed *slots*, where each attribute has a single associated value. An example of the "stimulus" fact given above may be:

```
STIMULUS-1
     ISA       stimulus
     category  colored-shape
     color     red
     shape     square
```

where "ISA" is a special slot that give the type of chunk (or "chunk-type") , while the slots "category", "color", and "shape" define attributes of the chunk-type stimulus, and each of these slots have a corresponding value that characterizes the particular stimulus that this example chunk is representing. The name of the chunk, "STIMULUS-1", can be used to reference the chunk, but from the perspective of ACT-R theory is not "part" of the chunk - the chunk itself is the collection of attributes and values (not including the special "ISA" chunk-type). As a comparative example, the "number" fact described above could be formed as:

```
TWO
     ISA       number
     previous  one
     number    two
     next      three
```

Here, the slots "previous", "number", and "next" are attributes of the chunk-type "NUMBER", which provide information on the quantities that occur "previous" to and "next" to the specified "number". These slot values themselves reference additional chunks, such as a "THREE" referencing the chunk:

```
THREE
     ISA       number
     previous  two
     number    three
     next      four
```

and so on. Again, note that the text "THREE", both as a slot-value in the "TWO" chunk as well as the name of the "THREE" chunk itself, only references the collection of attributes and associated slot values, and is not considered to be part of the chunk itself - and so, the "THREE" chunk references itself in it's "number" slot.

Chunks representing declarative knowledge may be instantiated as any combination of slots without first having to specify an appropriate chunk-type, and slots can be added to,

removed from, or updated in chunks as the information that the chunk represents changes. For instance, if the red-square stimulus above were to turn green, the chunk could be updated to:

```
STIMULUS-1
     ISA       stimulus
     category  colored-shape
     color     green
     shape     square
```

while if it were to become devoid of color entirely, the "`color`" slot (and associated value) could be removed. If it were to then gain a new attribute, such as making a "chirp" sound, a new slot could be added representing this information:

```
STIMULUS-1
     ISA       stimulus
     category  colored-shape
     shape     square
     sound     chirp
```

The collection of chunks that are either instantiated with the creation of the model or created by the model while it operates are stored in ACT-R's declarative module, which was mentioned above as providing both this storage capability as well as a retrieval mechanism for these chunks. Chunks may be specified by a modeler as being added to the model's declarative memory when the model itself is defined - the model is "born" with this knowledge. When created by the model itself during its operation, chunks are instantiated in the buffer of the module that is producing the chunk. For example, when a new display is updated onto the visicon, the model will produce a chunk presenting the information contained in the visicon, and place that chunk in the visual buffer of the visual module, where this visual information is now accessible to the other modules. When this information is checked by the model as a condition for action by a production, the model will automatically "clear" the information from the buffer as a part of the subsequent action that is performed (known as *strict harvesting*). When a chunk is cleared from

any buffer, a representation of that chunk is instantiated in the declarative memory module. If a matching chunk already exists in declarative memory, the two chunks are "merged" together, and the result is a single chunk that is made more available to the model through the reference to the chunk gained by the merging process. In this manner, ACT-R stores the chunks that it either has been given or has produced/used in the course of behavior into a system of memory.

The second manner of knowledge representation in ACT-R, procedural knowledge, represents specific actions or operations that the model knows how to perform. This system of procedural knowledge is implemented by the *procedural module*. The procedural module provides the ability to specify actions or operations, termed *productions*, which can be internally- or outwardly-oriented (in the sense that the production utilizes and manipulates internal or external information), or positioned at the interface between the two domains. Productions take the form of IF-THEN statements (another term would be a "condition-action pair"), with a "left hand side" (LHS; IF/condition) that specifies the conditions that the model should meet in order to execute the actions specified in the "right hand side" (RHS; THEN/action). An example of a simple ACT-R production is:

```
(P new-goal
     =goal>
          state     new-goal
          goal      nil
     =visual>
          goal      =goal
==>
     =goal>
          goal      =goal
          state     satisfy-goal
)
```

In this example, a production named "`new-goal`" uses information in the visual buffer in order to update the model's representation of some abstract "goal" that the model would like to

achieve. Here, the LHS of the production (the indented text that comes before the "==>" symbol) checks that the goal buffer contains a chunk that possesses a "`state`" slot with a value of "`new-goal`", and that the chunk contained by the goal buffer *does not* possess a slot named "`goal`" (specified by the special value of "`nil`", representing both the boolean false and null). Additionally, the LHS checks that the visual buffer contains a chunk that possesses a "`goal`" slot, the contents of which are bound to the variable "`=goal`". Variable binding in productions allows for information to be compared between slots or buffers on the LHS of the production (for instance, if the production should fire when chunks in two different buffers each contain a slot that should have the same value), or as in this example, used in the actions specified by the RHS of the production. The RHS of this example production (the indented text that comes after the "==>" symbol) then creates a "`goal`" slot in the chunk contained by the goal buffer, sets the value of this slot to the value represented by the "`=goal`" variable, and updates the value of the "`state`" slot in this same chunk to the value of "`satisfy-goal`".

The procedural module also provides a pattern matcher to determine whether the model's current state (across the current contents of the model's buffers, as well as the states of those buffers and the modules they belong to) matches the condition specified by any given production. A production is "selected" to be "fired" when the conditions specified on the LHS are found to match the model's current state, and is subsequently "fired", at which point all actions specified on the RHS are executed. These actions usually consist of modification to chunks in buffers (that is, manipulation of information represented by certain areas) or requests made to specific modules (which can otherwise be considered as the performance of some cognitive function by a specific area, prompted by the existence of the state specified by the LHS). In this manner, the operation of the procedural module subserves the interaction between

all other modules considered by the ACT-R theory and provides the theory with the ability to interact with and manipulate an evolving set of information.

One essential function for this manipulation of dynamical information is managing declarative memory, including the ability to retrieve these memories for active use (Alan Baddeley, 2003; Nelson Cowan, 2008). It was discussed earlier that ACT-R stores the chunks that it represents in a declarative memory module; but how does ACT-R "remember" a chunk when it becomes necessary? That is, how is information retrieved from declarative memory in ACT-R? Productions can specify a *retrieval request* of the declarative module, in which a (potentially partial) description of a chunk is provided to the declarative module. In response to this retrieval request, the declarative module will attempt to find a chunk that matches the description and place it in the declarative module's buffer (termed the *retrieval buffer*), thereby making the retrieved chunk available to the rest of the model. An example of a production that makes a retrieval request could be:

```
(P retrieve-number-fact
     =goal>
          state     retrieve-number-fact
          count     =count
     ?retrieval>
          state     free
==>
     +retrieval>
          number    =count
     =goal>
          state     nfact-retrieved
)
```

This production is named "`retrieve-number-fact`" which, when it is fired, attempts to retrieve a chunk representing a number from the declarative module's long-term memory. Here, the LHS of the production checks for the chunk in the goal module's buffer to have a slot named "`state`" with the value "`retrieve-number-fact`", as well as a slot named "`count`" for

which the value is assigned to the variable "=count". Additionally, it checks that the state of the retrieval module is "free" (that is, it is not currently busy performing some other requested action). If the model state described by the LHS of the production is satisfied, the production is selected and fired, upon which the actions specified on the RHS are executed. In this case, a retrieval request (+retrieval>) is made of the declarative module to attempt to retrieve a chunk that has a "number" slot that contains a value that is the same as the value contained in the goal buffer's "count" slot (through the "=count" variable binding). For any given retrieval request, if declarative memory contains a chunk that is adequately "available" to be retrieved (the determination of this "availability" is discussed in detail below), it will be placed in the declarative memory module's retrieval buffer after some delay (discussed below). For the duration of this delay, the declarative memory module's state is marked as "busy"; once the retrieval request has been satisfied and the retrieval buffer full, the declarative memory module's state is marked as "free". If, however, there are no chunks in declarative memory that are adequately "available", the retrieval request will fail, and nothing will be put into the retrieval buffer. In this case, the declarative memory module's state is marked as "busy" for a duration before the retrieval request "fails", and the declarative memory module's state is set to "error".

1.2.3 Availability of knowledge in ACT-R

The availability of a chunk to a model's attempt at retrieval can be determined through two different methods. The first method, termed *symbolic matching*, relies on comparing the description of the chunk that has been requested to the chunks in the model's declarative memory directly. Only chunks which match this description exactly are able to be retrieved, and chunks are retrieved immediately (i.e., it takes no time to retrieve a chunk); if more than one chunk is

matched, then one of the matched chunks is chosen on either a random basis or through a deterministic process that ensures that the same chunk will be chosen whenever the same set of chunks match the retrieval request.

However, this method does not take advantage of ACT-R's subsymbolic capabilities, and is not used in the models presented in this work. Instead, the models examined in this work utilize the second method, *subsymbolic computation*. In this method, the availability of a given chunk is dependent on a quantity termed *activation*, and each chunk possesses its own activation value. An ACT-R chunk's activation value can be considered as a formalization of the availability, or level of activity, of an item's representation in a cognitive module that is representing or actively processing the item. These activation values are computed upon every attempt at retrieval, and the set of chunks with activation values greater than the retrieval threshold $\boldsymbol{\tau}$ (defaults to 0.0) become eligible to be retrieved - any chunks below this threshold can be considered to not be contained in memory. Of the set of chunks that possess activation values that exceed the retrieval threshold, the chunk with the largest computed activation value is the one that is entered into the declarative memory module's buffer as a result of the retrieval request. There are a number of history- and context-driven factors that affect the computation of a chunk's activation value, increasing or decreasing the likelihood that a given chunk is retrieved in accordance with these factors. A given chunk i's activation value A_i has four different components, and is computed according to the equation:

$$A_i = B_i + S_i + P_i + \epsilon_i$$

where B_i is the base-level activation of the chunk; S_i is the spreading activation attributed to the chunk; P_i is the partial matching value computed for the chunk; and ϵ_i is a noise value that can potentially contain both a transient and permanent noise component. Each of these components

will be described below; while there are many factors that can affect the computation of each of these components of the activation value of a chunk, only those that are relevant to the functioning of the models that will be introduced will be covered in detail.

The base-level activation B_i of chunk *i* is a scalar value that reflects both the "age" (time since creation) of the chunk as well as the frequency of use of the chunk over time. It is representative of the degree of availability of a particular semantic representation that can be attributed to: 1. the time that has passed since the representation was first instantiated; and 2. any "refreshing" of this representation that may occur when it is referenced by the operation of cognitive processes subsequent to its genesis. In ACT-R, there are multiple options for the calculation of this quantity, but as time is a critical factor in the definition of the base-level activation, each of these options accounts for a decay in the activation of a chunk over a period of time. The method of calculation used in the present work strikes a balance between simplicity and inclusion of these essential factors:

$$B_i = \ln(n/(1 - d)) - d * \ln(L) + \beta_i$$

where *n* is the total number of presentations of chunk *i*; *d* is a static decay parameter that contributes to the reduction of B_i as time passes; *L* is the time since creation of chunk *i*; and β_i represents a constant offset in the base-level activation. Simplicity is favored as this method does not require a record of the complete history of a chunk, including the exact time intervals between "uses" of the chunk; however, the equation does include the time-based factors *n*, *L*, and *d*, and so it does model a decrease in the availability of semantic information as time progresses. The equation above can be broken down to examine the contribution of these factors to the calculation of B_i. The first term, $\ln(n/(1 - d))$, reflects the amount of base-level activation attributed to the number of "uses" of the chunk that the model has experienced at the time of

calculation of B_i. As n increases, the resulting value of B_i also increases; however, the natural log ensures that this increase in B_i plateaus with larger values of n. This could be interpreted as the model becoming "maximally familiar" with the information represented by the chunk, where subsequent "uses" of the information contribute very little to the availability of the information. The second term, $d * \ln(L)$, reflects the *decrease* in base-level activation that can be attributed to the passage of time, as it is subtracted from the first term. Larger decay values will emphasize this decrease; similar to the first term, the natural log ensures that the second term approaches a "maximum" decrease in activation due to the passage of time in an asymptotic manner. Finally, in this work the default value of 0.0 for β_i was used, and so it can be ignored. This description of the calculation of B_i reveals that it is composed of two competing processes, both dependent on the passage of time: an increase in the base-level activation due to increasing experience with the information (represented by n, which is dependent on time in order to accrue more experiences) and a decrease in the base-level activation due to increasing time since the initial formation of the chunk. The fewer the number of "uses" (i.e., lower values of n) since creation, the more the second term dominates the calculation, and as a result, the chunk will have low base-level activation; the greater the number of uses since creation, the more the first term dominates, resulting in a larger base-level activation.

The spreading activation S_i that a chunk i receives is a scalar value that takes into account potential "boosts" to the total activation A_i of the chunk that are sourced from the local processing of similar/identical information in parallel cognitive domains (John R. Anderson, 1983; Collins & Loftus, 1975). This local processing must co-occur with the attempt to retrieve information in order to provide spreading activation and increase the likelihood of retrieval. This component of a chunk's activation level can also be thought of as a method of "attentional weighting", in which

information that is being actively processed by the model increases the likelihood of retrieval of similar information (John R. Anderson & Lebiere, 2014). In ACT-R, each module corresponds to the local processing done by a particular cognitive unit, and each of these modules are able to provide differing amounts of spreading activation as "sources", represented by the module's specific *source activation value*. As such, the spreading activation S_i is computed through the equation:

$$S_i = \sum_k \sum_j W_{kj} S_{ji}$$

where the elements *k* being summed over are each of the buffers in the model which, at the time of the retrieval request, (1) currently contain a chunk; and (2) have a nonzero source activation value; the elements *j* being summed over are the slots of the chunk contained in the buffer indexed by *k*; W_{kj} is the amount of spreading activation provided to chunk *i* from slot *j* in buffer *k*; and S_{ji} is the strength of association between the contents of that slot *j* and the chunk *i*. By default, W_{kj} is calculated as the source activation value of buffer *k* divided by the number of slots in the chunk contained by buffer *k*, while S_{ji} is calculated on the basis of how much information is shared between the contents of slot *j* and the chunk *i*. In the present work, the default calculation of W_{kj} is used, but a different manner of computing S_{ji} is implemented, the details of which are reported below in the Methods section. As a result of the above equation, chunks that are contained in declarative memory can receive differing amounts of spreading activation from multiple modules, dependent on an individual module's source activation value, the number of slots in the chunk contained by a module's buffer, and the amount of information shared between that chunk and a given chunk in declarative memory. From this, it is straightforward that the

greater a module's source activation value, and the greater the amount of information the chunk in that module's buffer shares with a chunk in declarative memory, the greater the amount of spreading activation that declarative memory chunk receives; this amount of spreading activation is "diluted" by an increasing amount of slots in the source's buffer's chunk.

By default, retrieval requests in ACT-R are only able to return a chunk that exactly matches the specification given by the request. For example, in the (`retrieve-number-fact`) production above, a default retrieval request can only result in the retrieval of a number fact that exactly matches the number specified in the `count` slot, by the `=count` variable (so, if "two" was specified, only chunks that contain a `number` slot with the value "two" could be retrieved). ACT-R allows this specificity to be "toned-down" and for chunks that are not perfect matches to the retrieval request to be retrieved. The partial matching P_i captures the process of determining how much a chunk in declarative memory "matches" the specification of the retrieval request, which influences the likelihood of retrieval through an effect on A_i. This process involves comparing the values specified by the retrieval request to the values of the slots in a given chunk *i* through their *similarities*. By default, the range of potential similarities between two values (i.e., slot contents/retrieval specifications) is from -1 to 0, where -1 indicates a maximal difference between the two values, while 0 indicates a maximal similarity; a value has maximal similarity to itself, and maximal difference to all other values. Due to this behavior, similarity acts as a penalty on chunks in memory that have slot values that are different to what was specified by the retrieval request. Similarities between two specific values can be set by the modeler to be within the allowable range, thereby representing two concepts that, while not identical, are also not completely dissimilar. The partial matching P_i component is calculated as such:

$$P_i = \sum_k PM_{ki}$$

where *k* are the slot values specified by the retrieval request; *P* is a scaling parameter (this defaults to 1.0, which was used in this work); and M_{ki} is the similarity between the slot value *k* in the retrieval request and the value in the corresponding slot of chunk *i*. Therefore, the greater the number of slots in a given chunk *i* that contain values that are not maximally dissimilar to values specified by the retrieval request, the lower the partial matching penalty given to that chunk *i* during the calculation of it's activation value. A chunk that exactly matches the retrieval request will receive no penalty; any chunk that does not match exactly will receive a varying amount of penalty to its activation value, dependent on the number of slots specified by the retrieval request, the scaling parameter *P*, and the setting of the similarity M_{ki} between the request-specified values and the values in the corresponding slots of those chunks. Partial matching allows for the case in which a chunk that doesn't match the retrieval request whatsoever to be retrieved, as long as the activation that chunk receives from the other three sources in the activation value equation (base-level, spreading, and noise) is great enough. However, this supplemental activation would have to be enough not only to eclipse the penalty applied by partial matching, but to also make it so that this "non-matching" chunk possesses the greatest activation of all eligible chunks at the time of retrieval.

Finally, the noise component ϵ_i provides a random element to the computation of a chunk's activation value. This noise itself is composed of two sources: a permanent component, which is associated with each chunk as it is instantiated in declarative memory; and a transient component, which is computed for each chunk each time a retrieval request is made. Each source of noise, when utilized, is generated from a logistic distribution with a mean of zero and a scale

parameter set to the value supplied as the noise parameter for the given source. These two sources of noise are simply summed together in order to compute the ϵ_i value.

Once the activation values for each chunk in declarative memory have been calculated, the chunk with the largest activation value is selected for retrieval. As described above, the activation value is dependent on historical (the base-level component), contextual (the spreading activation and partial matching components), and random (the noise component) factors. Accordingly, the result of any given retrieval request will be different depending on the chunks that are in declarative memory, the time since the creation of each of those chunks, how many times they have been referenced, how much information in each of those chunks is shared by information represented by other buffer at the time of retrieval, how much the information contained in the declarative memory chunks resembles the specification of the retrieval request, and the amount of noise determined for each chunk at the time of retrieval. Once the chunk-to-be-retrieved has been selected, a final process determines the time required for the chunk to be retrieved and placed into the retrieval buffer of the declarative memory module. This process is dependent on the activation level of the chunk that has been selected. For successful retrievals, this retrieval time *RT* is calculated as such:

$$RT = Fe^{-(f * A_i)}$$

where *F* is the latency factor scale parameter; *f* is the latency factor exponent parameter; and A_i is the activation level of the chunk that has been selected to be retrieved. In this manner, the speed at which the retrieval request is satisfied is dependent on the activation level of the chunk that is selected to be retrieved; the more active the chunk is, the faster it is placed into the retrieval buffer and made available to the rest of the model. On the other hand, if no chunks are eligible for retrieval (that is, the largest calculated activation value is less than the retrieval threshold $\boldsymbol{\tau}$),

then the retrieval will fail, and the time that it takes for this to be determined is calculated by the equation:

$$RT = Fe^{-(f*\tau)}$$

The only difference between the retrieval time calculation for retrieval failures is that it is dependent on the retrieval threshold **τ** rather than the activation level A_i of the selected chunk (as no chunk was eligible to be selected). In both retrieval success and and failure equations, the scale parameters *F* and *f* function to increase the retrieval time. The scale parameter *F* simply scales the computed retrieval time in a multiplicative manner, while the exponent parameter *f* scales the activation level or retrieval threshold in a multiplicative manner before the exponential function is applied. As the quantities ($f * A_i$) and ($f *$ **τ**) are negated (multiplied by -1) before the exponential function is applied, increasing values of *f*, A_i, or **τ** result in the exponential term asymptotically approaching zero. Therefore, chunks with relatively larger activation values are retrieved more quickly than those with relatively smaller activation values. Finally, the scale parameter *F* represents the time to retrieve a chunk when the quantities ($f * A_i$) or ($f *$ **τ**) are zero - this is especially relevant in retrieval failure, as by default the value of **τ** is zero (John R. Anderson & Lebiere, 2014). When this is the case, *F* is the sole determinant of the time for retrieval failure.

1.2.4 Correspondences between components of ACT-R and the activity of specific brain regions

A number of components of the ACT-R cognitive architecture have been identified as correlates of the activity of specific brain regions. These correspondences have been established through both fMRI- (John R. Anderson et al., 2008; Borst & Anderson, 2013, 2017) as well as

EEG-based research (Cassenti et al., 2011; van Vugt, 2012, 2014), providing converging evidence on the validity of these relationships. Activity of ACT-R's declarative memory module, responsible for the retrieval of information from a long-term store on the basis of the level of activation of a chunk representation of the information, has been associated with the lateral inferior prefrontal cortex (LIPFC) (John R. Anderson et al., 2007, 2008) alongside the anterior cingulate cortex (ACC) (Borst & Anderson, 2013). The LIPFC has previously been associated with retrieval processes in neuroimaging studies (Buckner et al., 1999; Cabeza et al., 2002) and lesions to this area can result in memory deficits (Baldo & Dronkers, 2006). In studies which specifically examined the relationship between ACT-R model dynamics and fMRI activity, the LIPFC has been observed to be associated with declarative memory retrievals (Borst & Anderson, 2013) as well as the long-term storage of problem states in an ACT-R model based analysis of task-based fMRI activity (Borst et al., 2011), while ACC activity has been observed to increase upon successful retrieval of previously-learned problem solutions (John R. Anderson et al., 2009).

However, activity of the ACC has also been associated with ACT-R's goal module, which provides a manner of representing state-based and behavioral goal information (John R. Anderson et al., 2007, 2008). ACC reactivity to the retrieval of problem solutions could be associated with the updating of behavioral goals oriented toward the application and evaluation of the retrieved solution. Additionally, a meta-analysis of model-based fMRI applied to five ACT-R cognitive models revealed that activity of the ACC was associated not only with declarative memory retrievals but also working memory updates (Borst & Anderson, 2013)

ACT-R's imaginal module, which provides a resource for the short-term maintenance of behaviorally relevant information, has been associated with activity of the posterior parietal

cortex (John R. Anderson et al., 2008; Borst & Anderson, 2013). This correspondence is in agreement with fMRI studies of the neurophysiological substrates of short-term and working memory function, which also implicate the posterior parietal cortex as a dynamic representational area (Jonides et al., 1998; Oztekin et al., 2009). When compared directly to ACT-R model functioning, the activity of the posterior parietal cortex reflects the contribution of the imaginal buffer to the temporary representation of problem states (Borst et al., 2011) as well as transformation of task-relevant information (John R. Anderson et al., 2007). The same meta-analysis that observed an association between activity of the ACC and both declarative memory retrieval/working memory updates revealed that activity of the posterior parietal cortex was also associated with model functions that could be construed as analogues of working memory update mechanisms (Borst & Anderson, 2013). However, a region of the intraparietal sulcus was observed to exclusively reflect working memory updates, independent of memory retrieval. This indicates that ACC may serve to mediate the transfer or updating of working memory on the basis of recently retrieved information, which is consistent with the ACT-R goal buffer's role in controlling the interaction between the architecture's other buffers (such as the declarative memory and imaginal buffers).

1.2.5 ACT-R models of working memory

As ACT-R is a commonly adopted architecture for the modeling of human cognition, it has previously been applied in the modeling of working memory (Borst & Anderson, 2013; Daily et al., 2001; Juvina & Taatgen, 2007; Nijboer et al., 2016). Examining these efforts provides insight regarding the relation of individual ACT-R components to elements of working memory theory, as well as providing a basis for further development of ACT-R models which capture behavior under working memory task paradigms. For example, ACT-R modeling has

found support for a multicomponent perspective of working memory in a dual tasking paradigm, in which participants were asked to perform the n-back task both alone as well as in conjunction with a tracking task and a tone-counting task (dual task conditions) (Nijboer et al., 2016). In this study, two ACT-R models were compared - one which represented a centralized or unitary conceptualization of working memory, and another that considered multiple components. The authors found that the multicomponent model, which included attentional focus, declarative memory retrieval, and rehearsal components, better explained both behavioral and neuroimaging data than did the unitary model. Other studies have utilized ACT-R modeling in conjunction with fMRI in order to identify relationships between brain activity and ACT-R models of working memory. In a meta-analysis of five of these studies (Borst & Anderson, 2013), it was determined that model-based declarative memory retrievals (that occurred in support of behavior during working memory tasks) were associated with neuronal activity in the inferior frontal gyrus and the anterior cingulate, which have previously been identified by working memory neuroimaging studies that did not employ ACT-R modeling (Barch et al., 2013; Lamichhane et al., 2020). These regions were also observed to be associated with working memory updating, along with inferior parietal cortex, another region previously implicated in working memory function (Borst & Anderson, 2013). The convergence of the ACT-R model based results with the results of neuroimaging studies suggests that the contribution of individual components of ACT-R models of working memory are relatable to human brain activity under a working memory task.

Other ACT-R modeling efforts have investigated the potential for individual differences in cognitive strategies applied to the control of working memory (Daily et al., 2001; Juvina & Taatgen, 2007). While rehearsal strategies are commonly considered to play a role in the maintenance of working memory information at a level that allows for comparison to current

stimuli, decay-based familiarity judgements may also subserve working-memory based recognition (Juvina & Taatgen, 2007). These authors presented two ACT-R models of working memory function during performance of the two- and three-back conditions of the N-back task, which they termed the "high control" and "low control" models. The "high control" model implemented a classic rehearsal mechanism in which the model maintained a rehearsal window of size N by subvocalizing each new stimulus in order to add it to the ACT-R auditory module's history of recently "heard" items. In conjunction with this subvocalization, the model would attempt to retrieve the stimulus from declarative memory, and the result of this retrieval would be compared to the contents of the auditory history. If a match is found, the stimulus is determined to be a target. When not actively attempting to respond to the task, the model attempts to subvocalize the N most recent items as a rehearsal procedure. The "low control" model instead assumes that participants are not rehearsing, and instead make judgements of recency based on a learned estimation of the temporal lag between a target and the preceding presentation of that stimulus. When a new stimulus is presented, the model attempts to retrieve the stimulus in a manner identical to the "high control" model. When the result of retrieval matches the current stimulus, the model determines the time lag between the two presentations through use of ACT-R's temporal module. If it is equal to a duration learned through experience (of the delay between initial presentation and subsequent target presentation), the stimulus is determined to be a target. The predictions of these two models on N-back performance were used to classify a group of participants as either "high control" or "low control" based on the fit between a given participant's behavior and each model's prediction. Almost half of the participants were associated with the "high control" model, while the remainder assigned to the "low control" group. To validate these assignments, the grouping variable was used to predict a measure of

Stroop interference, an independent measure of cognitive control, and it was observed that "high control" participants demonstrated significantly lower amounts of interference than "low control" individuals (Juvina & Taatgen, 2007). This effort demonstrates the ability of ACT-R to capture differences in strategy between individuals. However, ACT-R parameter values for the models were fixed, and as assignment to the two groups was based on fit to each model's predictions, it is unknown whether modulation of these parameters would result in differences in assignment.

Parameterization of ACT-R provides another perspective on individual differences. Here, rather than investigating differences in strategy, differences in the contribution of cognitive factors to execution of a cognitive strategy are considered. These cognitive factors are captured by ACT-R parameters such as those discussed in section 1.2.3; if individuals differ in how these factors contribute to their cognition in a task environment, it would be expected that parameterized ACT-R models that reflect these differences would be capable of making predictions that distinguish individual behavioral patterns. Previous work that has taken this approach suggests that individual variation in working memory ability under a modified digit span task may be captured through ACT-R's goal buffer spreading activation parameter (Daily et al., 2001). In this work, an ACT-R model of behavior under the task was fit to participant behavior through variation of solely the goal buffer spreading activation parameter. The model's goal buffer represented one of two task-related goals during task performance: the goal to articulate stimuli as they are presented, and the goal to recall the list of relevant stimuli from the current trial. These goals provide spreading activation to chunks in declarative memory for which sequential retrieval attempts are made during both rehearsal and recall. This model was capable of closely predicting individual participant measures of accuracy over varying set sizes

as well as individual serial position curves (in which items in the middle of a set are less likely to be remembered than those that occur at the beginning or end of the set). In all, these former ACT-R modeling efforts demonstrate the viability of examining individual differences in working memory through the ACT-R cognitive modeling framework.

1.3 Study Aims and Approach

1.3.1 Aims

The present study aims to further the characterization of relationships between observable individual differences in working memory ability, core features of the ACT-R cognitive modeling architecture, and the neurophysiological underpinnings of general working memory ability. To do so, a subset of the Human Connectome Project (HCP) Young Adult dataset (Van Essen et al., 2013), will be leveraged to model individual behavior under the N-back working memory task, and subsequently relate model-parameter-based characterizations of individual working memory function to their observed resting-state fMRI functional connectivity. As the ACT-R cognitive architecture contains a number of components that have previously been related to both working memory function and the activity of specific regions of the brain, model-based characterizations of participant cognitive functioning and behavior under a working memory task can serve as a bridge between behavior and brain activity. In the course of this approach, a number of types of associations can be examined, including those between a measure of working memory ability and specific components of ACT-R, as well as those between components of ACT-R and the resting activity of specific brain networks. These sets of comparisons then allow for the determination of relationships between brain activity and characterizations of working memory ability. This effort will further understanding of the

neurological underpinnings of working memory ability by relating critical components of a broadly applied and well-supported cognitive architecture to individual behavior and neurological function.

1.3.2 Approach

The Human Connectome Project is a large-scale effort to systematically examine the relationship between human brain structure/function and individual behavior under a broad array of measures (Van Essen et al., 2013). Over 1200 participants were recruited for a two-day visit to Washington University - Minnesota where they underwent a battery of behavioral, cognitive, and neuroimaging measures. Behavioral/cognitive measures included those drawn from the NIH Behavioral Toolbox as well as a number of non-Toolbox measures, including (but not limited to) those of sustained attention, visual-spatial processing, and fluid intelligence (Barch et al., 2013). Participants performed an additional set of task paradigms (eight total: a working memory task; a recognition memory task (embedded in the WM task); an incentive processing task; a motor task; a language processing task; a social cognition task; a relational processing task; and an emotion processing task) while undergoing functional magnetic resonance imaging (fMRI), and a resting-state fMRI condition was also collected. To complete the aims of this study, the resting-state fMRI and working memory task data of a subset of HCP participants was selected to be used. The working memory task applied by the HCP protocol was the N-back task, and participant behavior under the zero-back and two-back conditions was collected.

Two ACT-R models were implemented in order to model participant behavior under this task - one which implements a cognitive strategy to complete the zero-back condition of the N-back task, and another which does the same for the two-back condition. These ACT-R models utilize components of the ACT-R cognitive architecture that have previously been associated

with both working memory function as well as specific cortical regions. A set of parameters for each model was fit to individual participants in the subset of data used by minimizing the error between measures of the participant's behavior in the model-specific condition and the model's predictions of those measures. The two models shared the individually-fit parameters W_g and W_i, (W_{kj} in the ACT-R spreading-activation equation for the goal and imaginal buffers, respectively) d (the decay parameter found in the ACT-R base-level activation equation), and F (the latency factor scale parameter found in the ACT-R retrieval time equations), while the zero-back model considered an additional similarity parameter c (the value of M_{ki} in the ACT-R partial matching equation, in certain circumstances). These parameter estimates capture individual differences in multiple aspects of the participant's working memory function. The spreading activation components W_g and W_i can be viewed as capturing an attentional element of WM function - in the implemented models, they assist in the gating of information to the WM system by boosting activation of information that is related to both current task demands as well as long-term task goals, which in turn promotes the processing of the correct information during response and memory updating procedures. The decay rate parameter d has a fairly straightforward interpretation as the time-dependent decay of task-relevant information in memory. As the potential role of decay in working memory functioning is controversial in the literature (Barrouillet et al., 2011; Lewandowsky & Oberauer, 2009), a competing explanation of interference between representations in WM is represented in these models through the contribution of the partial matching component of the ACT-R chunk activation equation. Partial matching allows the model to mistake the representation of a given element of knowledge as that of another, negatively impacting the model's ability to maintain the correct representation at a level of activation that allows for appropriate use of the representation in the generation of

task-based behavior. The contribution of spreading activation (attention) has the potential to counteract both decay-based and partial matching (interference)-based reduction in the availability of a given representation, reflecting a role of attention that has been included in numerous models of WM (N. Cowan, 1988; Engle et al., 1999; M. J. Kane et al., 2001; Lewandowsky & Oberauer, 2009; Portrat et al., 2008). While a parameter-based measure of individual differences in interference is only estimated for the zero-back model (the parameter *c*), this partial matching mechanism is included in the N-back model and provides a noteworthy contribution to the model's ability to perform the task. Finally, the latency scale parameter *F* is a reflection of the accessibility of the information considered by the model's working memory system, as higher levels of activation need to be countered by increasingly larger values of *F* in order to maintain retrieval times that promote reasonable response times as well as rates of updating and rehearsal.

The importance of these parameters to the model's efficacy in predicting individual differences in task behavior (and therefore capturing individual differences in working memory function) was determined through a unique procedure in which the weakest parameter in the set was iteratively determined and dropped out until only two parameters were left to compare. Of the parameters that resulted from this procedure, those that were significantly predictive of participant behavior were investigated for relationships to the participant's resting-state functional connectivity. This functional connectivity was established through determining the pair-wise partial correlations between the time-varying resting state fMRI activity of a set of predefined brain regions that are each associated with a canonical brain network (Power et al., 2011). A regression-based method was applied to determine the set of resting-state functional connections that best predict an individual's ACT-R parameter estimate. This method estimates

the number of regressors that minimize model complexity while maintaining the regression model's ability to predict the criterion variable. As the functional connectome is used as the set of predictors, this reduction in regressors has the effect of isolating the resting-state functional connections that best predict these model-based characterizations of WM function. Regression-based predictions of individual model parameter values from resting state functional connectivity were compared to the parameter estimates derived from task-based behavior to evaluate the validity of the identified functional connections as predictors of WM function. The brain regions and networks involved in these identified functional connections were then examined to determine relationships between the model-based characterizations of WM function and the resting-state activity of specific brain regions and networks, providing further information on the correspondence between individual brain function and individual differences in working memory ability.

Chapter 2. Methods

2.1 Participants

Data from 178 participants of the HCP 1200 Subjects release was utilized in the present work. This subset included individuals aged 22-36+ (the HCP behavioral data only includes a general age range for each participant; in the present work's subset of individuals, 27 participants were aged 22-25; 81 participants were aged 26-30; 69 participants were aged 31-35; and one participant was aged 36+. Due to this manner of age reporting, mean and standard deviation of the ages of the participants included in this work cannot be determined. See (Van Essen et al., 2013) for a discussion of why ages were reported in this manner by the HCP investigators), with 99 female and 79 male participants. The majority of the HCP participant pool was recruited from Missouri-based sibships (groups of offspring sharing two parents) using recruiting techniques to ensure that the demographics of the recruited individuals are reflective of the ethnic and racial composition of the U.S. population (Van Essen et al., 2013). Exclusion criteria for sibships extended to those with severe neurodevelopmental, neuropsychiatric, or neurologic disorders, as well as individuals with diabetes or high blood pressure. Additionally, individuals born prematurely (prior to 34 weeks gestation for twins, and 37 weeks gestation for non-twins) were excluded (Van Essen et al., 2013).

Recruited participants visited Washington University - Minnesota for a two-day visit to complete the HCP protocol. Informed consent was reviewed and acquired at the beginning of the first day (Van Essen et al., 2013; WU-Minn, 2017). Otherwise, each day was scheduled in roughly the same manner, in which participants first underwent a mock scanner practice (to

become familiar/comfortable with the scanner environment and receive training on how to minimize head movement while in the scanner before the first scanning session was administered. On the first day, this initial scanning session collected structural imaging, while the initial scanning session on the second day collected diffusion imaging (Van Essen et al., 2013). After these initial scanning sessions, participants completed a battery of behavioral assessments, with non-NIH Toolbox methods being administered on the first day, and NIH Toolbox measurements on the second day (additional details regarding these behavioral assessments can be found in (Barch et al., 2013)). After a break, participants then proceeded to a second scanner session on both days. This scanner session first included a 30-minute resting-state fMRI phase in which two 15-minute "runs" were collected. During these runs, participants were instructed to keep their eyes open and fixated on a bright cross-hair on a dark background (WU-Minn, 2017). Additional details regarding the resting-state fMRI image acquisition and preprocessing can be found in section 2.3. The resting-state phase was followed by a 30-minute task-based fMRI phase in which participants completed seven task protocols, including working memory, incentive processing, motor, language processing, social cognition, relational processing, and emotional processing tasks (Barch et al., 2013). After the second scanner session, each of the two days of the visit ended with a recognition task based on the stimuli seen during the working memory task in the scanner, and the end of the second day also included drug/alcohol tests and a satisfaction survey (WU-Minn, 2017). In total, the participants spent approximately seven hours each day engaged in the HCP protocol. The two scanning sessions required two hours total on each day (not including set-up/practice), while the behavioral assessments required about 1.5 hours on the first day, and two hours on the second (WU-Minn, 2017).

The HCP recruitment procedures and informed consent documents were approved by the Washington University Institutional Review Board (Glasser et al., 2016). Permission was obtained by the author to use the HCP Open Access data for the present work. A protocol detailing the use of the HCP Open Access data was filed with the University of Washington Institutional Review Board and found to be exempt.

2.2 Human Connectome Project Working Memory Task

The HCP employed a version of the N-back task as the working memory task paradigm, which participants performed as part of the task-based fMRI phase on each of the two days of their visit. Each of the two runs of the task consisted of the presentation of eight blocks of ten trials each. Within each run, half of the blocks present the two-back version of the N-back task (in which the participants are asked to identify stimuli that are the same as the stimulus that occurred two presentations ago as a "target", and all other stimuli as "nontargets"), while the other half of the blocks present the zero-back version of the N-back task (in which a cue stimulus is presented at the beginning of each block, and participants are asked to identify any identical stimulus as a "target", and all other stimuli as "nontargets") (Barch et al., 2013). Each of these two subsets of blocks within a run presented one of four different stimulus categories (pictures of faces, places, tools, and body parts (non-mutilated, with no nudity included)) as the stimuli for a block, so that two blocks were presented for each stimulus category (one of the two-back condition and one of the zero-back condition) (Barch et al., 2013). In the run held on the first day, the N-back condition over the first four blocks was alternated as two-back/zero-back/two-back/zero-back, with each of these four blocks presenting a unique stimulus category. The N-back condition of the second half of the blocks of the first day's run was alternated as zero-back/two-back/zero-back/two-back, again with each block presenting a

unique stimulus category. In the run held on the second day, the N-back condition over the first four blocks was again alternated as two-back/zero-back/two-back/zero-back, with each of these four blocks presenting a unique stimulus category; however, in the second half of the blocks presented on the second day's run, the N-back condition was presented in the order of two-back/two-back/zero-back/zero-back, while each of the four blocks still presented a unique stimulus category. Additionally, four fixation blocks (15 s duration for a single fixation block) were presented after every second task block in each run (Barch et al., 2013).

At the beginning of each block, a cue is presented for 2.5 s, displaying the N-back condition (zero-back or two-back), and in the case of the zero-back condition, the cue (or "target") stimulus. Each of the ten trials of the block are displayed for 2 s, with a 0.5 s intertrial interval held between trials, leading to a total duration of 27.5 s per block. The full duration of a run, including the task presentation, fixation duration, and a participant preparatory phase, was five minutes and one second (Barch et al., 2013).

This paradigm produces three different types of stimuli, from the "perspective" of the current stimulus: targets, nontargets, and lures. In the HCP, targets for the zero-back condition were defined as stimuli that matched the block's cue stimulus, while targets for the two-back condition were defined as stimuli that matched the stimulus that occurred two presentations previously. Nontargets for the zero-back condition were defined as stimuli that both did not match the cue stimulus and were novel (that is, having not been presented previously within the block), while nontargets for the two-back condition were defined as stimuli that did not match any of the last three presentations. Lures for the zero-back condition were defined as (previously) nontarget stimuli that had occurred at least once before in the block, while lures for the two-back

condition were defined as stimuli that matched the stimulus that occurred either one or three presentations previously (WU-Minn, 2017).

2.3 Resting-state fMRI Image Acquisition and Preprocessing

The HCP resting-state functional neuroimaging data was acquired with a Siemens 3T "Connectome Skyra", using a 32-channel Siemens head coil and gradient-echo echo-planar-imaging (GE EPI) (Smith, Beckmann, et al., 2013; Van Essen et al., 2013). Custom modifications to this scanner were made by Siemens to account for the limited space available at Washington University - Minnesota that increased the maximum gradient strength of the hardware to 100 mT/m. These customizations resulted in an inner bore diameter of 56 cm, requiring placement of the patient table in a non-standard position. This placement led to larger-than-normal gradient distortions, which have been corrected in the preprocessing pipeline (Van Essen et al., 2013) (described below). The HCP implemented slice-accelerated multiband acquisition, a technique that increases the amount of data acquired on each pass by imaging multiple brain slices in parallel. These slices are separated from one another using the spatial sensitivity profiles of the 32-channel head coil. Optimal pulse sequence parameters were determined through an extensive piloting effort (Smith, Beckmann, et al., 2013; Uğurbil et al., 2013; Van Essen et al., 2013). These parameters include a multiband factor of 8, a TR of 720 ms, a TE of 33.1 ms, a flip angle of 52 degrees, a field of view of 208 x 180 mm (read-out x phase-encoding directions) and a spatial resolution of 2.0 mm isotropic voxels, resulting in a collection of 72 2.0 mm thick slices (Van Essen et al., 2013; WU-Minn, 2017). The two 15-minute runs of each 30-minute resting-state fMRI acquisition session were acquired with opposite phase encoding directions, left-to-right and right-to-left. During most sessions, real-time head position information was acquired with an infrared motion tracking camera system, and

cardiac and respiratory signals were collected using a Siemens pulse oximeter respiratory belt (Van Essen et al., 2013). In addition, two spin echo EPI images with geometrical, echo spacing, and phase encoding directions that are identical to the gradient echo EPI images are collected in each functional session. These images are intended to be used to produce a field map to aid EPI distortion correction, and spin echo EPI was utilized as this imaging protocol is faster than gradient echo EPI and therefore less susceptible to motion distortion (Glasser et al., 2013).

Raw data produced by the scanning process was first converted into standard DICOM images before undergoing another conversion to the standard NIFTI format, which are then defaced to remove identifiable facial features (Uğurbil et al., 2013; Van Essen et al., 2013). These NIFTI images are considered to be the "unprocessed" HCP fMRI data. The HCP has developed multiple "minimal preprocessing" pipelines, which are intended to provide the data of the HCP in a cohesive, systematic manner that addresses the complexities and nuances of the dataset, such as the larger-than-normal gradient distortions described above (Glasser et al., 2013; Van Essen et al., 2013). Described here is the minimal preprocessing pipeline that was applied to the fMRI data used in the present study (the *fMRIVolume* pipeline), but there is at least one other major preprocessing pipeline that has been applied to the HCP fMRI data (Glasser et al., 2013). The fMRIVolume pipeline is composed of six major steps, and requires the completion of the HCP structural pipelines (*PreFreeSurfer*, *FreeSurfer*, and *PostFreeSurfer*) (Glasser et al., 2013).

The first step of the fMRIVolume pipeline is the correction of gradient-nonlinearity-induced distortion, which addresses the larger-than-normal distortions sourced in the custom modifications to the scanner. In this step, the gradient distortions present in the data are corrected by producing a model of the spatial distortion of the image, which will subsequently be used to "unwarp" the image. (Glasser et al., 2013). This model is based on a

spherical harmonic expansion for which a "warpfield" is calculated on the basis of a proprietary Siemens gradient coefficient file and the mm coordinate space of the image.

The second step of the fMRIVolume pipeline is to correct for motion of the participant's head through realignment of the timeseries. This is done through an FSL ("FMRIB Software Library"; FSL is a collection of analysis tools for MRI-derived imaging data) FLIRT (FMRIB's Linear Image Registration Tool) rigid body transformation with six degrees of freedom (DOF; corresponding to the three axes of motion (x, y, and z) and the derivatives of these axes). Each frame of the gradient echo EPI image is aligned to the single-band reference image, which is a full-tissue contrast structural image that is acquired during a reference scanning procedure that occurs just before commencement of the fMRI scanning process (Glasser et al., 2013). This procedure produces a motion transformation matrix for each timepoint in the fMRI image.

The third step in this minimal preprocessing pipeline functions to account for the distortion that occurs in the phase encoding direction of EPI fMRI images, and is the reason why the pair of spin echo EPI with opposite phase encoding directions were acquired in each functional imaging session. This pair of images is used to estimate the distortion field through the FSL toolbox "topup" (Glasser et al., 2013). A similar registration to that in step two then occurs, in which the distorted spin echo EPI images are aligned to the distorted gradient echo EPI single-band reference image using another six DOF FLIRT registration procedure. The outcome of this procedure is combined with the estimated distortion field before being applied to the single-band reference image with spline interpolation, thereby correcting the phase encoding direction-based distortion in the reference image (Glasser et al., 2013).

In step four of the fMRIVolume pipeline, the corrected single-band reference image is registered to the T1w structural image produced by the HCP structural pipeline by again

applying a six DOF FLIRT registration. However, the registration in this step also includes first the use of a BBR (boundary-based registration) cost function and the the use of FreeSurfer's BBRegister (FreeSurfer is another collection of software for analyzing MRI images), which results in a highly accurate registration between the fMRI data and the structural data (as the undistorted single-band reference image is now registered to the preprocessed T1w structural image) (Glasser et al., 2013). The BBR cost function and BBRegister tool applied in this step utilizes white-matter boundaries sourced from the preprocessed T1w structural image to inform the registration of the EPI image. The intensity of pairs of values that bridge these boundaries should have a large difference when the registration between the structural and functional images is accurate, allowing for the calculation of a cost function that informs the fitting of the functional image to the structural image (Greve & Fischl, 2009).

In the penultimate (fifth) step of the preprocessing pipeline, each transformation produced by the previous four steps (the "warpfield" for the correction of gradient-nonlinearity-induced distortion, the transformation matrix for the correction of motion, the combined distortion field/spin echo-to-gradient echo single-band reference image registration, and the corrected single-band reference image-to-preprocessed T1w image registration, respectively) are combined together alongside a native-space T1w to MNI coordinate system transform in order to apply a single nonlinear transformation in a single resampling step using spline interpolation. This results in the direct transformation of every frame in the raw timeseries to a corrected image positioned in MNI space (Glasser et al., 2013).

In the final, sixth step of the fMRIVolume preprocessing pipeline, the bias field (a low frequency, smooth signal produced by inhomogeneities of the magnetic field produced by the MRI machine (Juntu et al., 2005)) calculated from the structural images is removed, and the

functional images are masked by the final brain mask produced in the PostFreeSurfer pipeline (Glasser et al., 2013). In brief, the bias field is calculated from the T1w and T2w structural images by taking the square root of the product of the T1w and T2w images, while the final brain mask is created by binarizing a cortical parcellation scheme produced by the FreeSurfer pipeline (additional details can be found in (Glasser et al., 2013)). Once this bias field removal and masking is complete, the functional images are normalized to a 4D whole brain mean of 10,000, and the result of the pipeline is output in the standard NIFTI file format.

One additional processing step was applied to the minimally preprocessed data after it was retrieved from the HCP database. Spatial smoothing of each frame of the 4D functional timeseries was performed by applying a Gaussian kernel with a full width at half maximum of 8.0 mm (corresponding to four voxels) to each voxel in the frame. This process reduces noise over the spatial domain in an attempt to increase the SNR of each frame of the timeseries; however, smoothing also has the effect of "sharing" data between voxels, potentially resulting in an overestimation of the functional connectivity between two pairs of voxels (the calculation of functional connectivity for this dataset is discussed in section 2.5.1; while this is an important issue to acknowledge, it is of minimal impact in this work due to the application of a large-scale brain parcellation before computation of the resting-state functional connectivity) (Chen & Calhoun, 2018; Liu et al., 2017).

2.4 ACT-R Modeling of HCP WM Task

2.4.1 Task devices

In order for the ACT-R models (described in section 2.4) to be fit to the participant data, modified versions of the n-back task utilized by the HCP (Barch et al., 2013) were implemented

in the LISP programming language. A task device interfaces with a given ACT-R model to present the task elements and allow the model to make responses accordingly. The zero-back and two-back conditions of the HCP N-back task were implemented into separate task devices so that a given model would only be exposed to one of these conditions. The zero-back task device and the two-back task device each present two sessions of the task, with each session containing four blocks and each block containing ten trials. The four stimulus categories (faces, places, tools, and body parts) were presented by each task device in an alternating manner across blocks within each session, and the task devices presented 15 s fixation periods after every other block (for a total of four fixation blocks across the two sessions in each device) in order to mimic the timing of the HCP task. To accomplish the successive presentation of the elements of the tasks, the task devices maintained a record of the current session, block, trial, and "phase" (stimulus presentation, inter-trial interval presentation, or fixation presentation) of the task, as well as the times of presentation associated with each of these "phase" elements. On each "step" of the task (in which the next "phase" is presented for the specified period of time), the ACT-R visicon was updated with the relevant information of the current task "phase" before the next visicon update is scheduled to occur N seconds in the future, where N is the duration of presentation associated with the current task "phase". After this scheduling, the task device's record of the current session, block, trial, and "phase" was updated accordingly. The task device also maintains a behavioral record of the model's keypress responses (storing the time that the keypress occurred alongside the specific key that was pushed), as well as a record of the chunks retrieved from declarative memory by the model, to be used for the evaluation of model performance.

The order of stimuli presented within each block and session was identical across all HCP participants included in this work, and so this order of presentation was encoded into each task

device. Each block of 10 trials in the task presented two "target" stimuli and two or three "lure" stimuli, with the remainder being non-targets. At the beginning of a given block, a 2.5 s cue indicated the N-back condition (either "zero" or "two"), and in the case of the zero-back condition, the "target" stimulus for the block was also identified. On each trial, a stimulus is presented for 2 s, followed by a 500 ms blank-screen ITI. As in the HCP version, the task device required a response to be made within the 2 s duration of the stimulus presentation by indicating that the current stimulus is either a "target" or "nontarget". If a response was not made by the model within this period, a "no response" was logged in the model's behavioral record.

While all relevant aspects of the HCP zero-back and two-back tasks were maintained by the task devices, there were some differences that should be noted. Rather than displaying task information on an experimental window, the task devices directly updated the ACT-R visicon by first removing all existing information from the visicon before adding a new set of information representative of the current task display. Instead of updating the visicon with the actual images of the category-specific stimuli as in the HCP task, the task devices displayed the cue (in the case of the zero-back task) and stimulus (both tasks) image filenames as a text string, as well as an additional text string to represent the stimulus category ("face", "place", "tool", or "body" for body parts) that was not present in the HCP version of the task. Additionally, an oval (one of the fundamental shapes that ACT-R is already programmed to recognize) was presented to the visicon during the fixation periods in place of the standard fixation cross used in the HCP task. Finally, in the HCP, the fMRI sessions were held over the course of two days, with the first session occurring on the first day and the second session occurring approximately 24 hours later (WU-Minn, 2017); as a simplification, the task device modeled the elapsed time between sessions as exactly 24 hours for each run of a given model.

The task devices implemented a number of functions (aside from the presentation of the task itself) that were essential for the operation of the model or the estimation/evaluation of the parameter values. Instead of the standard manner of computing the strength of association (S_{ji}) between two chunks, a linear activation spread function was implemented and used by the models through the ACT-R `:sji-hook` parameter, which allows a custom function to be provided in place of the standard calculation. This linear activation spread function computes the S_{ji} between two chunks (a "source", present in a non-retrieval buffer in the model, and a "target", present in the model's declarative memory) by determining the number of times the source chunk appears in the slots of the target chunk, which the function will return as the S_{ji} value; if the source and target chunks are identical, the function will return a value of 1; if no information is shared between the chunks, a value of zero is returned. The task devices also implement functionality to allow for the ACT-R task device and model, which reside in a LISP environment, to interact with a Python environment so that parameter values can be estimated on the basis of comparison of the model's predictions to a participant's responses. This interface allows an ACT-R model to be run from the Python environment with the appropriate parameters set to specified values, before the model's behavioral record and list of retrieved chunks are returned to the Python environment to be used subsequently.

In addition to the zero-back and two-back conditions that were performed by the HCP participants, task devices for putative one-back and three-back conditions were also implemented. These additional conditions will be used in later sections to provide further tests of the model's psychological validity. Both of these task devices functioned identically to the zero-back and two-back devices, except rather than encoding a series of stimuli as used by the HCP protocol, the order of presentation of stimuli were customized in order to capture the

dynamics of these N-back conditions (such that there were “one-back” targets in the one-back task device, and “three-back” targets in the three-back task device). Identical strings were utilized to represent stimuli across all four task devices. Each block of the one-back device presents two one-back targets, two two-back lures, and the remainder of trials (six total) as nontargets; the three-back device presents two three-back targets, two two-back lures, and the remainder as nontargets (due to the relatively small number of trials per block, four-back lures were not modeled in the three-back device).

2.4.2 Zero-back model

The general strategy of the zero-back model is, for each trial, to attempt to retrieve the block’s cued stimulus. The information retrieved in this effort is compared against the current stimulus to determine a response - if the information matches the current stimulus, the model responds as though the stimulus was a target; when the retrieved information and the current stimulus do not match, the model responds as though the stimulus was a nontarget. A flowchart depicting the general actions that the model performs to complete the task can be found in Figure 1; each of these aspects of the model are described in detail below.

The zero-back model, implemented in the LISP programming language, has a number of general ACT-R parameters that are set before the model performs the task. The `:auto-attend` parameter is enabled (set to “t”), ensuring that when a visual location is found by the model in the visicon, a request for the visual buffer to move attention to the found location

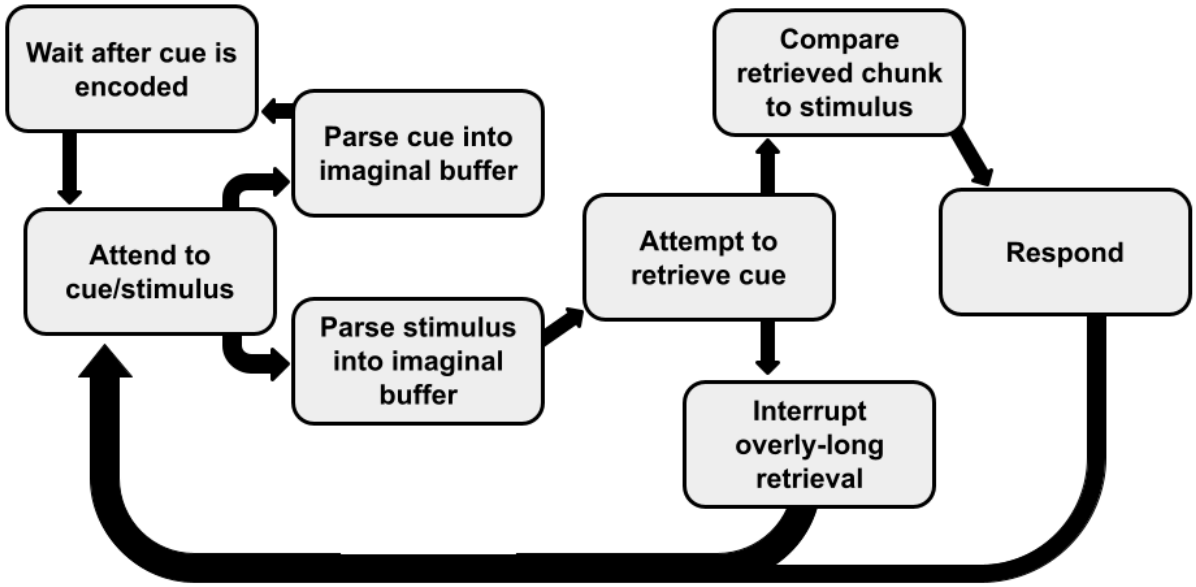

Figure 1. Zero-back model flowchart. Each rounded box represents a stage in the model's strategy to perform the task; a single stage may reflect multiple ACT-R model productions. Black arrows between rounded boxes indicate potential transitions between stages.

is automatically generated. The `:esc` parameter is enabled (set to "t"), specifying that the modules utilized by the model should use their full subsymbolic processing (relevant only for retrieval requests). The maximum associative strength parameter `:mas` is set to 1.0, enabling the use of the spreading activation calculation. The partial matching parameter `:mp` is also set to 1.0, enabling ACT-R's partial matching functionality and setting the value of the match scale parameter in the partial matching equation to 1.0. The `:ans` parameter is set to 0.2, thereby setting the scale parameter of the logistic function used to generate transient noise to the same value. This value, which is the default value in ACT-R, is similar to that found to best capture the expected stochastic noise in human memory performance (Lebiere et al., 2003). Finally, the `:sji-hook` and `:retrieved-chunk-hook` parameters are set to utilize the `(linear-activation-spread)` and `(record-retrieved-chunk)` functions, which are established in the task device and described in section 2.4.1.

The zero-back model uses a limited set of chunk-types to represent the declarative knowledge necessary to perform the task. A `parse-item` chunk-type contains the slots `state`, `nback`, and `target`, where `state` encodes the current general state of the model, `nback` encodes the fact that the nback condition is "zero", and `target` is intended to encode the fact that the target of retrieval attempts in this condition is a "cue". When the model is instantiated, before the task begins, this chunk is entered into the goal buffer with only the `state` slot containing the information that the current state of the model is to "find" a visual location; the other two slots are set to `nil` and technically do not exist until instantiated by the model at a later time during performance of the task. A `stimulus` chunk-type contains the slots `category`, `item`, and `kind`, while a `cue` chunk-type contains the same three slots plus an

`nback` slot. These chunk-types are implemented in order to represent the zero-back stimulus and cue information that is presented to the model; the `category` slot contains a representation of the category of a cue/stimulus (face, place, tool, or "body" for body parts); the `item` slot contains the string representation of a stimulus/cue image; the `kind` slot contains a representation of whether the information in the chunk was sourced from a cue or a stimulus; and for cues, the `nback` slot represents the fact that the nback value associated with the cue in this condition is zero. Finally, a "placeholder" chunk-type `blank` is included in order to represent a blank inter-trial interval screen.

The zero-back model contains a total of nine productions which can be grouped into three general categories: four productions necessary to parse the information in the visicon in a manner that makes that information accessible to the model; two productions designed to handle the retrieval of cue information from the declarative memory module on the basis of the information parsed from the visicon (on the current trial as well as during the block's cue period); and three productions that utilize the retrieved information alongside the current representation of the stimulus in order to make a keypress response.

To perform the task, the zero-back model first finds the location of information presented on the visicon. As there is only ever at most one entity in the visicon (either a cue, a stimulus, or a fixation cue), the process of finding this location automatically results in a request for the visual buffer to move attention to the found location (as ACT-R's `:auto-attend` parameter is enabled). Once this location has been found, the model does not move attention from it; as such, this process only occurs in the cue period of the first block of the first session. Information in the attended location is automatically entered into the visual buffer, after which one of two productions are eligible to be selected. If the chunk in the visual buffer contains an `nback` slot

with an associated value, a “cue parsing” production will move the chunk contained in the visual buffer into the imaginal buffer using ACT-R’s “buffer overwrite” mechanism. Additionally, this production will instantiate the `nback` and `target` slots in the goal buffer chunk with values derived from the visual buffer representation of the cue information - in the case of this condition, these values are “zero” and “cue” respectively. If the chunk in the visual buffer does not contain an `nback` slot, a “stimulus parsing” production will instead be selected to be fired. This production also transfers the chunk contained in the visual buffer to the imaginal buffer using the “buffer overwrite” mechanism, but does not update any information in the chunk contained by the goal buffer.

Once information has been parsed from the visicon into the visual buffer and subsequently transferred to the imaginal buffer, the model reaches a branching point. In the case that the chunk contained in the imaginal buffer has a `kind` slot with an associated value of “cue”, it will choose to wait for the next task event by reverting to being prepared to parse visual information - all necessary actions in the cue period of a block have been completed. If instead the chunk in the imaginal buffer has a `kind` slot with an associated value of “stimulus”, the model will proceed to attempt to retrieve a chunk from declarative memory. This retrieval request specifies a `kind` slot with an associated value of “cue” and a `category` slot with a value equal to the value of the `category` slot in the imaginal buffer chunk. When this retrieval request occurs, the chunks stored in declarative memory have the potential to receive spreading activation from two general sources: the goal buffer and the imaginal buffer. The `nback` and `target` slots in the goal buffer chunk, containing the values “zero” and “cue” respectively, will provide spreading activation to chunks that also have slots containing these values - namely, chunks representing cue information, which have `nback` and `kind` slots that carry these values.

In parallel, the `kind`, `category`, and `item` slots of the imaginary buffer chunk have the potential to provide spreading activation to both cue and stimulus chunks. The `kind` slot, with the value "stimulus", will provide spreading activation to all stimulus chunks in declarative memory, but not cue chunks. The `category` slot, with a value equal to the category of the current stimulus, will provide spreading activation to all cue and stimulus chunks that share this category. Finally, the `item` slot, with a value of the current stimulus, will provide spreading activation to all cue and stimulus chunks that also represent this item.

If this retrieval request is pending for enough time that the task proceeds into the trial's inter-trial interval, the model will detect this transition and select/fire a production which interrupts the pending retrieval request. This is done by making a retrieval request which is guaranteed to fail by having the request specify a chunk which cannot exist in declarative memory. This retrieval request also has two special parameters set which alter the functioning of this specific retrieval request only, and do not affect any other retrieval request made by the model. The first, `:mp-value`, temporarily changes ACT-R's partial matching parameter, and is set to "nil". This disables partial matching for this specific retrieval request, which ensures a retrieval failure will occur (otherwise, partial matching would allow for a chunk that does not fully match the request specification to be retrieved, and the request would succeed). The second, `:rt-value`, temporarily changes the declarative module's retrieval threshold value, and is set to 10. Due to the equation governing retrieval times for retrieval failures, this parameter value will ensure an essentially instantaneous retrieval time. The result is the model recognizing that it has lost its opportunity to respond on the trial, and "snapping back" from attempting to retrieve a cue to being prepared to receive the next trial's stimulus. When this occurs, a "no response" is logged in the model's behavioral record.

In the majority of cases, however, the retrieval request for a cue chunk with a category equal to that represented in the imaginary buffer either succeeds or fails before the inter-trial interval occurs. The model will proceed differently according to whether cue information was successfully retrieved or not. If the retrieval request failed, the model has nothing to compare the current stimulus information to, and concludes that it cannot make a response. The model reverts to being prepared to receive the next trial's stimulus, and a "no response" is logged in the model's behavioral record. If the retrieval request for cue information is successful, the model will then compare this retrieved information to the information represented by the chunk in the imaginal buffer. If the value of the `item` slot in the retrieved chunk matches the value of the `item` slot of the chunk in the imaginal buffer, the model concludes that the current stimulus is a match to the block's cue, and responds by making a request of the motor module for a "punch" action with the index finger of the model's left hand, which results in a keypress of the "f" key of the virtual keyboard. If, however, these two `item` slots do not match, the model instead concludes that the current stimulus is *not* a match to the block's cue, and instead makes a request of the motor module for the same action using the right hand instead, resulting in a keypress of the "j" key of the virtual keyboard. In either case, the model's keypress response and time of response is recorded by the `(collect-response)` function of the task device.

Five parameters were chosen to be estimated for the zero-back model, due to their potential to affect the ability of the model to perform the task: the cue-stimulus similarity value *c*; the goal buffer spreading activation value W_g, the imaginal buffer spreading activation value W_i, the base-level activation decay parameter *d*, and the latency factor scale parameter *F*. The cue-stimulus similarity value *c* represents the degree of similarity between the model's representation of cue information versus stimulus information, and affects the computation of the

partial matching component P_i of a chunk's activation value by setting the value of M_{ki} in the partial matching equation when a stimulus chunk is considered for retrieval in response to a request that specifies a cue chunk (such as when the `kind` slot is specified to be associated with the value "cue" in the model above). The goal and imaginal buffer spreading activation values W_g and W_i (respectively) model the increase in activation of information in memory due to concurrent representation of that information in other cortical domains, as these parameters specify the amount of spreading activation that the two buffers can provide to a chunk eligible for retrieval. This affects the value of W_{kj} in the spreading activation S_i equation, as W_{kj} is calculated by dividing the spreading activation value of the buffer k by the number of slots of the chunk contained by that buffer. The base-level activation decay parameter d represents the decrease in availability of a memory due to the passage of time, and is involved in the computation of the base-level activation component B_i of a chunk's activation value, as described in section 1.2.3. These four parameters largely influence the model's ability to retrieve the correct representation of cue information from declarative memory, and therefore the model's ability to correctly respond to a given stimulus in the task. However, they also influence the retrieval time of this information (and subsequently, the response time of the model) by ultimately affecting the chunk's activation value A_i in the retrieval time equation. The last estimated parameter, the latency factor scale parameter F, models a basal duration for the retrieval of a specific memory, and determines the amount of time necessary to determine a failure in retrieval as well as modulating the retrieval time of a chunk by setting the value of F in the retrieval time and retrieval failure time equations. A visualization of the model's ability to retrieve the chunk representation of the block's cue can be seen in Figure 2.

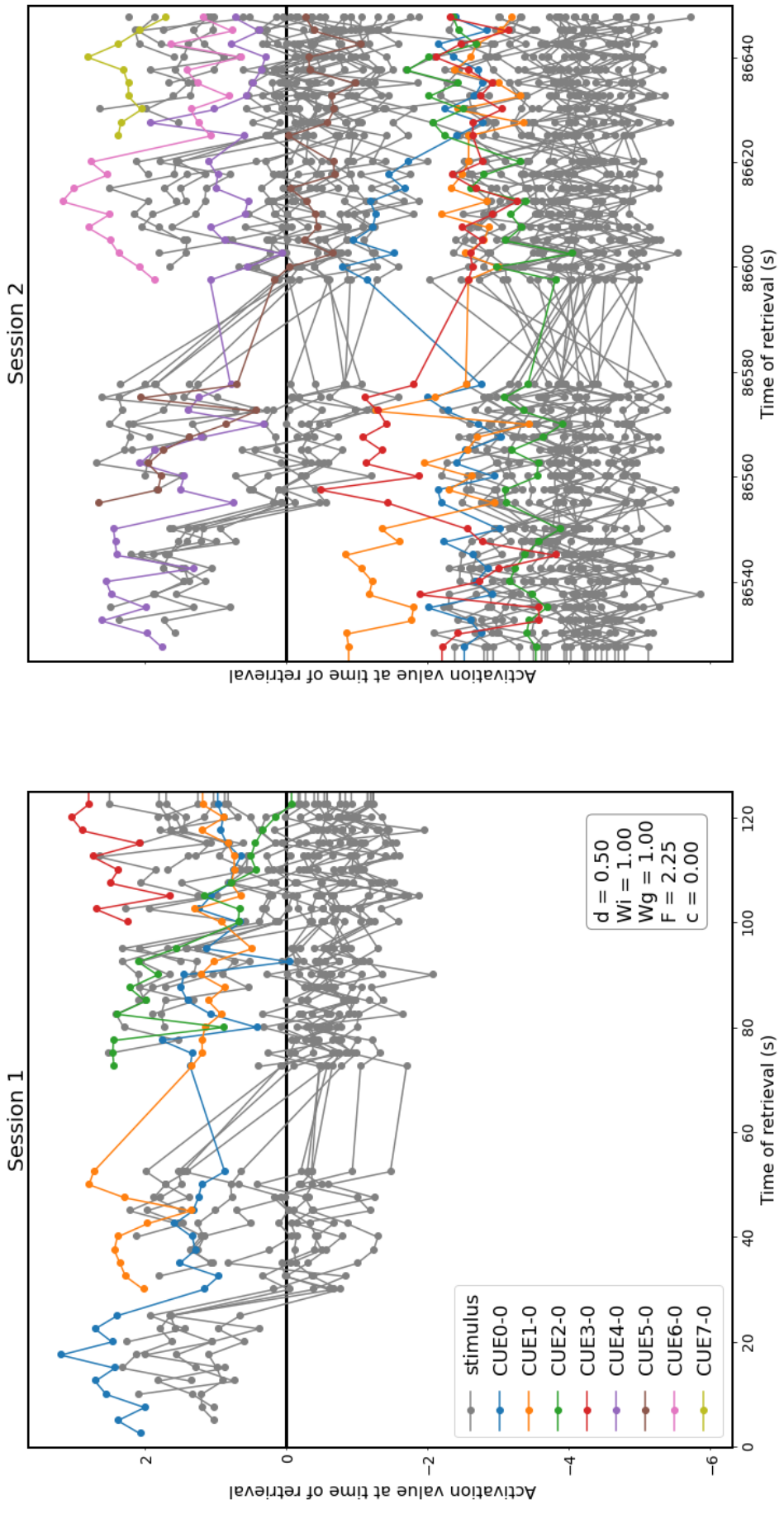
Session 1
Session 2
Activation value at time of retrieval
Time of retrieval (s)
stimulus
CUE0-0
CUE1-0
CUE2-0
CUE3-0
CUE4-0
CUE5-0
CUE6-0
CUE7-0
d = 0.50
Wi = 1.00
Wg = 1.00
F = 2.25
c = 0.00

Figure 2. Chunk activation at time of retrieval for the zero-back model at default parameter settings. Left plot shows chunk activations in session 1 of the task; the plot on the right shows chunk activations during session 2. Plotted on the x-axis in each is the time of retrieval (relative to the beginning of the task), while the y-axis of each plot represents the activation value of a given chunk at the time of retrieval. Colored points denote the activation value of chunks that represent block cues, while grey points denote the activation value of stimulus chunks. Horizontal black line in each plot represents an activation value of zero. Inset in the left plot's bottom left is a legend with the mapping of each color to each chunk type; inset in the left plot's bottom right are the parameter values that were used to generate the presented values.

2.4.3 N-back model

The general strategy of the N-back model is to attempt to maintain a "stimulus-window" of the last *n+1* stimulus presentations (where *n* is the cued N-back value) and use the information contained in that window to attempt to determine the stimulus that occurred *n* presentations ago through retrieval. As in the zero-back model, if the retrieved stimulus matches the current stimulus, the model responds as though the stimulus was a target; when the retrieved information and the current stimulus do not match, the model responds as though the stimulus was a nontarget. When the model is not busy responding to the task or updating this "stimulus-window", it will sequentially rehearse the stimuli contained in this window, increasing the likelihood of retrieval of these stimuli on later trials. A flowchart depicting the general actions that the model performs to complete the task can be found in Figure 3; each of these aspects of the model are described in detail below.

The model, also implemented in the LISP programming language, enables the same general ACT-R parameters as the zero-back model, and uses the same parameter settings/values for these general parameters as described in section 2.4.2. The set of chunk-types used to represent declarative knowledge for the two-back task are similar to that of the zero-back model. A `parse-item` chunk-type contains the slots `state` and `nback`, identical to the zero-back model, but does not include a `target` slot and instead includes `stage`, `count`, `next-up`, `category`, and `lastitem` slots. The `state` and `nback` slots function identically to those of the zero-back `parse-item` chunk-type. The `stage` slot encodes the current stage of the model's cognitive strategy - "interact" while engaging with a cue/stimulus and making a

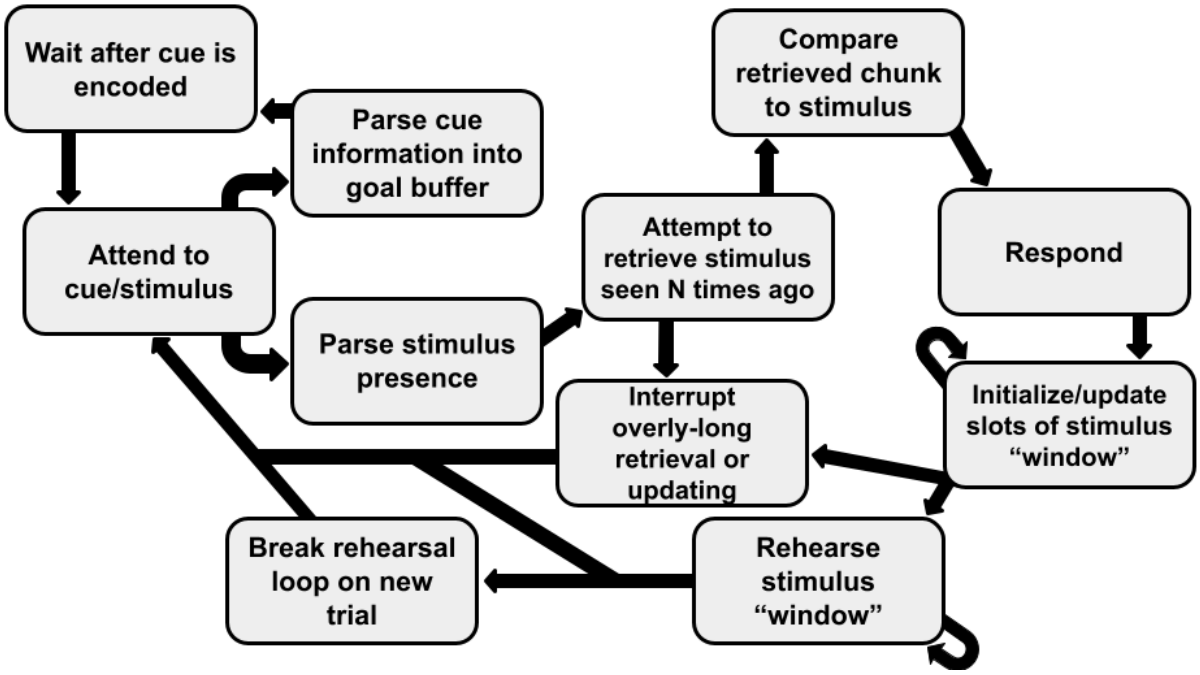

Figure 3. N-back model flowchart. Each rounded box represents a stage in the model's strategy to perform the task; a single stage may reflect multiple ACT-R model productions. Black arrows between rounded boxes indicate potential transitions between stages.

response; "updating" when updating the model's representation of the last *n+1* stimuli that were presented; and "rehearsal" while rehearsing this stimulus-window representation. The `count` slot tracks the position of the stimulus in the stimulus-window that is currently being updated or rehearsed, while the `next-up` slot is used to temporarily maintain the model's goal-directed attention on the *next* slot to be updated once the updating of the current slot is complete. The `category` slot encodes the category of the stimuli ("face", "place", "tool", or "body") that are being shown in the current block. The `lastitem` slot maintains a representation of the last stimulus that was seen by the model, but is not accessed directly when the model makes a response, and instead passively provides spreading activation to the chunks in declarative memory that represent the last item.

Another critical chunk that the model utilizes is not pre-established as a specific chunk-type, but is instead created by the model as it performs the task. This chunk can be considered as a list representation of the stimuli that occurred in the last *n+1* presentations, and contains the slots `one`, `two`, `three` (and in the case of the three-back condition, `four`), and `category`. The named number slots represent the identity of the stimulus that the model believes occurred the specified number of presentations ago (so, the stimulus identity associated with the `two` slot is thought of by the model as having occurred two presentations ago), while the `category` slot represents the category of the stimuli represented by this window.

The `stimulus` and `blank` chunk-types used in the zero-back model are also used to represent stimulus information and the inter-trial interval screen (respectively) for the two-back model; because there is no cue-stimulus associated with the cue of two-back blocks, the zero-back model's `cue` chunk-type was not necessary in the two-back model. The last chunk-type used by the two-back model are `number` chunk-types, which contain the slots

`number`, `next`, and `previous`. These chunk-types represent number-line information such that a numerical value *x* is represented by the `number` slot (e.g., "two") while the `next` and `previous` slots represent the numerical values *x+1* and *x-1*, respectively. These `number` chunks enable the model to "count" through a sequence of numbers, or increment/decrement a number, by retrieving the chunk with the current value associated with the `number` slot, and then using the `next` and `previous` slots. Five `number` chunks representing the values "one", "two", "three", "four", and "five" were added to the model's declarative memory upon initialization of the model, and the `:reference-count` parameter for each of these chunks was set to a value of 10,000, thereby setting the value of *n* in the base-level activation equation. This has the effect of making these chunks exceedingly "accessible" to retrieval by the model, and represents the extensive familiarity that individuals accrue with the basic sequence of single-digit values by the time they are young adults (as participants in this study were).

The two-back model contains a total of thirty productions which can be grouped into six general categories: two productions necessary to perform numerical incrementing/decrementing operations when necessary; five productions necessary to parse the information in the visicon in a manner that makes that information accessible to the model; three productions designed to handle the retrieval of n-back stimulus information from the declarative memory module on the basis of information regarding previously presented stimuli, which is held in the imaginal buffer; three productions that utilize the retrieved information alongside the current representation of the stimulus in order to make a keypress response; ten productions to handle the initialization and updating of a "stimulus-window", which represents the last *n+1* stimuli that were presented; and seven productions which provide the model with the ability to rehearse the stimuli represented by this "stimulus-window" after it has been updated on a given trial.

To perform the task, the two-back model functions similarly to the zero-back model, with the additional steps of creating/updating and rehearsing a "stimulus-window" of the last *n+1* stimuli. The two-back model's process of finding visual information in the visicon and attending to this information is identical to that of the zero-back model. If the chunk in the visual buffer contains an `nback` slot with an associated value, a "cue parsing" production will instantiate the `nback` and `lastitem` slots in the goal buffer chunk - the `nback` slot will be filled with the value instructed by the cue, while the `lastitem` slot will be filled with "nothing" (representing the fact that until the second trial of the block, "nothing" has occurred previously). If the chunk in the visual buffer represents a stimulus, one of two "stimulus parsing" productions will instead be selected to be fired. The first of these productions will fire when the `lastitem` slot of the goal buffer is "nothing" (that is, on the first trial), and will update the `category` slot of the goal buffer chunk to represent the category of the stimuli being presented in the current block. The second of these productions only serves to detect when a new stimulus has appeared on the screen, and prompts the model to attempt to retrieve a representation of the stimulus that occurred *n* presentations ago.

This retrieval request specifies a `category` slot with a value equal to the value of the `category` slot in the goal buffer chunk, and an `item` slot with a value equal to the value of the named `nback` slot in the imaginal buffer (so, the imaginal buffer contains a chunk with the slots `one`, `two`, and `three` (and in the case of the three-back condition, `four`); with each slot potentially containing the identity of the stimulus that occurred that many presentations ago; the value of the goal buffer `nback` slot indexes the name of this slot in the imaginal buffer, and the stimulus associated with this slot becomes the target of retrieval). Upon this retrieval request, the chunks in declarative memory receive spreading activation from the same two sources as the

equivalent retrieval request in the zero-back model - the goal and imaginal buffers. The value of the `category` slot of the goal buffer chunk provides spreading activation to all declarative memory chunks that share this category, and the `lastitem` slot of the goal buffer chunk provides spreading activation to declarative memory chunks that represent the specified stimulus. From the imaginal buffer, the chunks contained in the `one`, `two`, and `three` slots provide spreading activation to the chunks in declarative memory that represent the stimuli named by the values associated with these slots, while the `category` slot also provides spreading activation to all declarative memory chunks that share the specified category. If "nothing" is contained in the n-back named slot of the imaginal buffer, the model recognizes that there is no stimulus that occurred *n* presentations ago, does not attempt to retrieve anything, and proceeds to response.

As in the zero-back model, if this retrieval request is pending for enough time that the task proceeds into the trial's inter-trial interval, the model will detect this transition and select/fire a production which interrupts the pending retrieval request in an identical manner to the zero-back model. In most cases, however, this retrieval request will succeed before the inter-trial interval occurs, and the model will proceed to response. The model responds by comparing the identity of the retrieved stimulus to that of the current stimulus in the visual buffer, responding "target" for a match and "nontarget" for a nonmatch. If no attempt at retrieval was made, the model defaults to responding "nontarget" (as would occur on the first two trials in a two-back condition, where there is no stimulus that occurred "two" ago to compare against). These responses result in keypresses in an identical manner to that of the zero-back model.

Once the model has responded on a given trial, it proceeds to the updating of the stimulus-window. If no stimulus-window exists in the model's imaginal buffer (such as on the first trial of a block), the model will initialize this stimulus-window chunk within the imaginal

buffer by specifying a `one` slot with a value equal to the identity of the current stimulus, a `category` slot with a value equal to the stimulus category represented by the goal buffer, and `two` and `three` slots that are associated with "nothing". In parallel, it will add to the visual buffer's representation of the current stimulus a new slot named `seen` and associate it with a value of "one". This operation reflects the integration of top-down knowledge with bottom-up sensory processing, and labels the current stimulus as having been seen "once" ago.

If a stimulus-window chunk does exist in the imaginal buffer, the model will instead update the named number slots (`one`, `two`, and `three` and potentially `four`) of this chunk. To do so, the model begins by identifying the contents of the slot of the stimulus-window chunk specified by the `nback` slot of the goal buffer. It will make a retrieval request for this item, specifying the category of the stimulus, the identity of the stimulus, and a value of the `seen` slot equal to the value specified by the current `count` of the update (which, in the first slot update, is equal to the `nback` slot of the goal buffer). In layman's terms, the model is attempting to remember what it has "seen" that many occurrences ago. Two things are done with the chunk returned by this retrieval request; the identity of the retrieved item is placed into the *n+1* slot, and the `seen` slot of the chunk that has been retrieved is updated with the *n+1* value. So, if the model successfully retrieved the chunk named by the imaginal buffer's stimulus-window chunk's `two` slot (`seen` "two" ago), the identity of that chunk is placed into the `three` slot of the imaginal buffer's stimulus-window chunk and a new memory of having seen that stimulus "three" ago is created. This new memory exists in declarative memory (after harvesting of the retrieval buffer) alongside older memories of having seen the same stimulus "one" and "two" ago. Once the *n+1* slot of the stimulus-window chunk and the `seen` slot of the retrieved chunk has been updated, the model "counts down" from the *n* slot to move the contents of the *n-1* slot

to the *n* slot, in the same manner as has just been described. This process continues to update the named number slots of the stimulus-window chunk in reverse-order until only the `one` slot remains (that is, after the"old" contents of the `one` slot have been used to update the `two` slot), upon which it will fill the `one` slot with the visual buffer's information of the current stimulus, and instantiate the `seen` slot in this visual buffer chunk to "one". The model will recognize when a slot in the stimulus-window chunk has a value of "nothing", and will proceed directly to the previous slot without attempting to update the next slot with "nothing". If the updating loop ends during the inter-trial interval, the information from the current stimulus is no longer available to the model, and there are two consequences: the stimulus-window's `one` slot is not updated, and a memory of having seen that stimulus "one" ago is not created. If the updating loop has not finished by the time a new stimulus is presented, the loop is broken and the stimulus-window is left in an "unfinished" state as the model attends to the new stimulus.

In most cases, though, the updating of the stimulus-window finishes before the next stimulus is presented. In this case, the model begins rehearsing each item named in the stimulus-window in sequential (not reverse) order. This rehearsal procedure operates in a similar manner to the stimulus-window updating procedure, but does not edit the contents of the stimulus-window. Instead, the model counts through the slots of the stimulus-window, and for each slot, makes a retrieval request specifying the category listed by the `category` slot of the goal buffer and the identity of the stimulus named by the corresponding slot of the stimulus-window chunk. For these retrieval requests, the `:mp-value` parameter was set to "nil", ensuring that the specified item is retrieved. When the model has rehearse the last existing number-named slot in the stimulus-window chunk, it will detect that there is no "next" slot to rehearse and instead reset the rehearsal loop back to the `one` slot. This rehearsal proceeds until

one of two events occur: either the model attempts to rehearse a stimulus-window slot with "nothing" in it, or the task advances from an inter-trial interval to a new task phase. In either of those cases, the model will stop rehearsing and re-orient to being prepared to interact with a new stimulus. This rehearsal procedure has the effect of boosting the activation of the chunks in declarative memory that represent the stimuli listed in the stimulus-window; as it begins with the most recent item and loops after the oldest item, the model on average has more opportunities to rehearse more recently experienced stimuli when compared to older stimuli.

Four parameters were chosen to be estimated for the N-back model, due to their potential to affect the ability of the model to perform the task: the goal buffer spreading activation value W_g, the imaginal buffer spreading activation value W_i, the base-level activation decay parameter d, and the latency factor scale parameter F. These parameters are a subset of those estimated for the zero-back model, and details of the effect of each of these parameters on the general functioning of the model can be found in section 2.4.2. In the case of the N-back model, the cue-stimulus similarity value c was not estimated as there are no cued stimuli in the versions of the task that this model was designed to be capable of accomplishing; as such, the value of this parameter would have no effect on the functioning of the model. A visualization of the model's ability to retrieve the chunk representation of the block's cue can be seen in Figure 4.

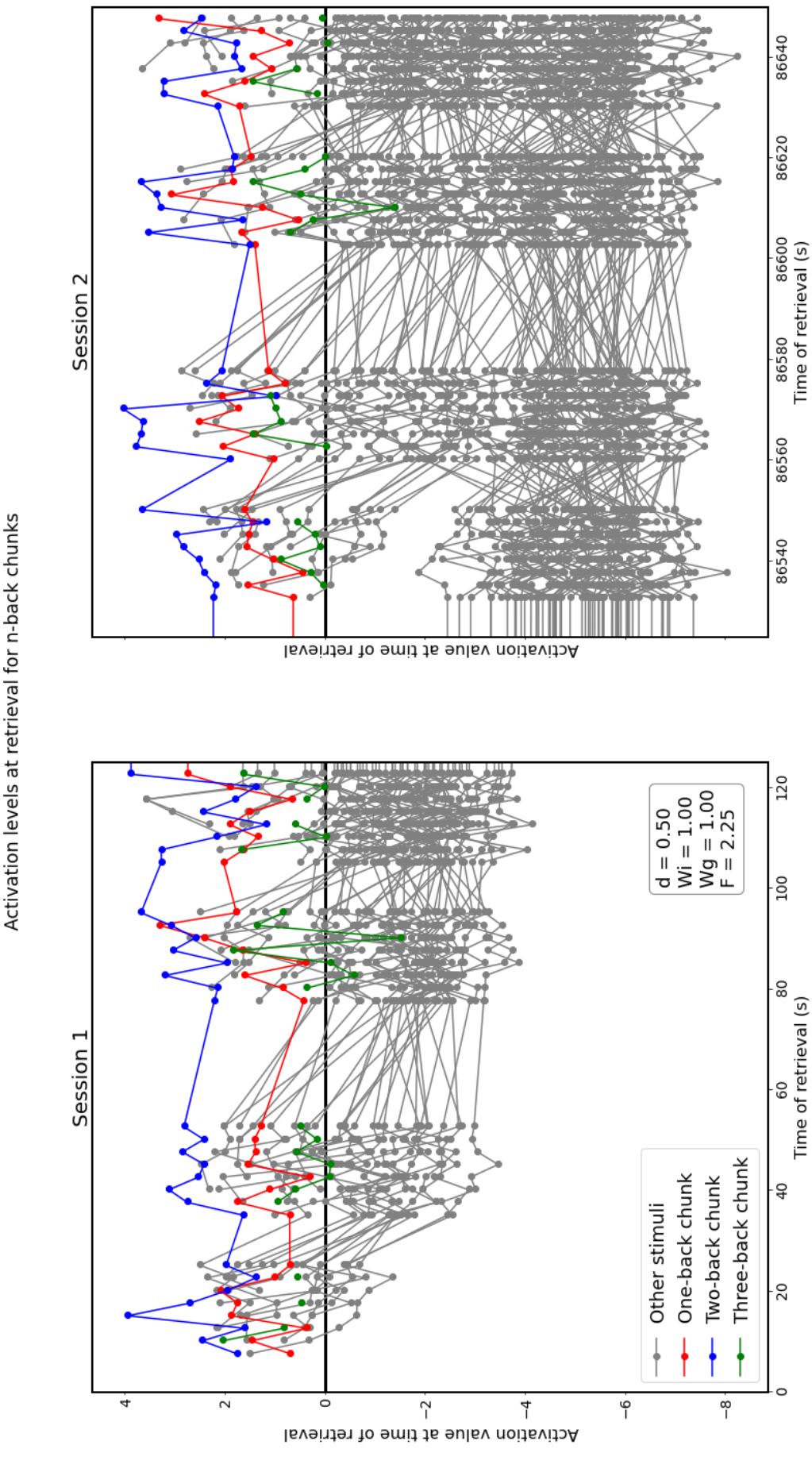
Activation levels at retrieval for n-back chunks
Session 1
Session 2
Activation value at time of retrieval
Time of retrieval (s)
Other stimuli
One-back chunk
Two-back chunk
Three-back chunk
d = 0.50
Wi = 1.00
Wg = 1.00
F = 2.25

Figure 4. Chunk activation at time of retrieval for the N-back model at default parameter settings. Left plot shows chunk activations in session 1 of the task; the plot on the right shows chunk activations during session 2. Plotted on the x-axis in each is the time of retrieval (relative to the beginning of the task), while the y-axis of each plot represents the activation value of a given chunk at the time of retrieval. Colored points denote the activation value of chunks that represent block cues, while grey points denote the activation value of stimulus chunks. Horizontal black line in each plot represents an activation value of zero. Inset in the left plot's bottom left is a legend with the mapping of each color to each chunk type; inset in the left plot's bottom right are the parameter values that were used to generate the presented values.

2.4.4 Individual-specific model fitting

Participants who did not achieve greater-than-chance performance (binomial test) on the target, lure, or nontarget conditions were not included in the parameter estimation process or any subsequent analysis. Participants were filtered separately on these criteria for the zero-back and N-back model fitting procedures. Additionally, a small subset of participants (11 out of 178) did not experience the same order of stimulus presentation as the remainder of the participants, and so this subset was also excluded from parameter estimation under both models as well as all subsequent analyses.

To determine the set of parameter values that resulted in the best model fit for each participant, bound-constrained minimization of the root mean square error (RMSE) between measures of the participant's responses and the model's predictions of those responses was applied through the Python SciPy library's optimize.minimize() function. The RMSE between a given participant's responses and the model's predictions was computed from a combination of the response times associated with each trial and two types of accuracy measures: 1. the target, lure, and nontarget accuracies (the mean accuracy of all trials in these conditions); and 2. the block-wise accuracies (the mean accuracy of each block of trials). For participants, the trial-by-trial response times were sourced directly from the HCP data, while the participant's responses were compared to a list of correct responses for each trial to first determine the trial-by-trial accuracies before subsetting these trials into the target/lure/nontarget and block-wise groups and calculating the mean accuracy in each. For the model, the trial-by-trial response times were computed by subtracting the time of the task device's presentation of a stimulus from the time of the model's response to that stimulus. The target/lure/nontarget and block-wise accuracies were calculated by separate methods for the zero-back and the two-back model, and

are described in sections 2.4.4.1 and 2.4.4.2. For both participant and model data, missing response times (on no-response trials) were replaced with the mean of all existing response times for that participant/model run, and the accuracy of no-response trials was coded as "incorrect". The participant and model response times were then divided by two so that they were on the same scale as the accuracy measures, with a minimum value of zero and a maximum value of one (as the maximum allowable response time in the task was 2 s). Once these measures had been determined for both the participant and the parameterized model output, they were organized so that two vectors were produced: one containing the participant's response times and accuracy measures, and the other containing the same for the model. The RMSE between the participant and the model was then calculated by first subtracting the participant's vector from the model's vector and then taking the square root of the mean of the squared elements of this difference vector.

This calculated RMSE value was minimized using Powell's optimization method (Powell, 1964) as implemented in the *optimize* packages of Python's *SciPy* library (Virtanen et al., 2020). In this procedure, the minimum of a scalar function that possesses one or more input variables is determined in an iterative manner. In this case, the input variables to this scalar function are the values of the parameters to be estimated, and the scalar output is the RMSE between the model's parameterized predictions and a specific participant's responses. The smaller the RMSE, the closer the model's predictions are to a given participant's responses, and so the set of parameter values that minimize this RMSE are considered to be the set of values that characterize that given participant.

Powell's method works by performing a sequential search along a set of search vectors, beginning at an initial point that is specified as an input (Powell, 1964). When the minima along

each of the search vectors has been determined, the displacement vector between the newly-found position (set of values) and the previous position is added to the search vector list as a new search direction, and the search vector that contributed most to the new position is removed from the search vector list. This "dropping" of a search vector contributes to the mitigation of linear dependence between search vectors, as the dropped vector is a major component of the newly added vector (Powell, 1964; Vetterling et al., 1992). Compared to other methods (e.g. the Nelder-Mead simplex method), Powell's method has the advantage of allowing the user to specify bounds on the parameter space, thus preventing the model from running under conditions where parameters are not meaningful (such as negative values of the memory decay parameter *d*).

2.4.4.1 Zero-back model fitting

As described in section 2.4.2, five parameters were fit through Powell's method for the zero-back model: the cue-stimulus similarity value *c*; the goal buffer spreading activation value W_g, the imaginal buffer spreading activation value W_i, the base-level activation decay parameter *d*, and the latency factor scale parameter *F*. The cue-stimulus similarity value *c* was initialized at a value of 0.0 with bounds (-1.0,1.0); while *c* is recommended to be bounded between (-1.0,0.0), it is not impossible to have positive values of *c*, and testing during implementation of the model-fitting procedure indicated that the optimize.minimize() procedure would become "stuck" at a *c* value of zero without being "allowed to explore" positive values and determine that they likely lead to a poor model fit. The goal and imaginal buffer spreading activation values W_g and W_i were each initialized at a value of 1.0 with bounds of (0.0,2.0); while the ACT-R default value for W_g is 0.0, the ACT-R default value for W_i is 1.0 and so this was chosen as the initial point for both parameters due to the usually competing influence they have during retrieval. The

base-level activation decay parameter *d* was initialized at 0.5, the ACT-R default value, with bounds of (0.2,0.8); these bounds are representative of previous findings on the range of possible values of decay rate that match individual performance in memory task (Sense et al., 2016; Zhou et al., 2021). The latency factor scale parameter *F* was initialized with a value of 2.25 (long enough that a failure to retrieve would eclipse into the intertrial interval) with bounds of (1.0,3.5).

Accuracy for the zero-back model was calculated by determining if the chunk in the retrieval buffer at the time of response on a given trial was the cue chunk for the block of that trial. As there were no other sources of noise in the model's production of a response, if the correct cue chunk had been retrieved by the model, it would make a correct response on the basis of comparison to the chunk in the imaginal buffer (representing the current stimulus). This manner of calculating accuracy was employed due to the disproportionate number of nontarget trials (approximately 60% of all trials). If accuracy according to the model's keypress response was used, there was the possibility that the model would retrieve incorrect chunks but still appear to be "correct" on nontarget trials, as the keypress response would indicate that there was a mismatch between the retrieved chunk and the chunk representing the current stimulus. By scoring the model's accuracy according to the retrieved chunk, it is ensured that trials marked "correct" are those where the model was able to determine whether or not the stimulus was a match to the blocks' cue, and trials marked "incorrect" are those where the model had retrieved incorrect information to make this determination on.

2.4.4.2 N-back model fitting

Four parameters were fit through the Powell procedure for the two-back model: the goal buffer spreading activation value W_g, the imaginal buffer spreading activation value W_i, the

base-level activation decay parameter *d*, and the latency factor scale parameter *F* (described in section 2.4.2, as these parameters were included in the zero-back model). The initial values and bounds input to the RMSE minimization procedure for each of these parameters were identical to those used for zero-back model fitting. Accuracy for the two-back model was calculated by combining keypress accuracy with an accuracy measure similar to that used for the zero-back model, in which the chunk that the model retrieved as the stimulus that occurred two presentations ago is compared to the chunk representing the stimulus that actually occurred two presentations ago. If a trial's keypress accuracy was incorrect, the trial was marked as incorrect; if the keypress accuracy was correct, then the trial was marked as correct only if the chunk that the model retrieved as the stimulus was actually the chunk of the stimulus that occurred two presentations ago.

2.4.5 Evaluation of parameter estimates

2.4.5.1 Evaluation of fit

To evaluate the fit of the parameterized models to the participant data, each model was run 100 times for each participant that had been characterized in that model's fitting process. These participant-specific 100 model runs used the set of parameter values estimated for the participant in question. These model runs were averaged together on a trial-by-trial basis, producing a set of model-predicted mean trial-by-trial response times and accuracies for each participant. Within each participant, the RMSE between these model-predicted mean responses and the participant's responses was computed for trial-by-trial target, lure, and nontarget response times and accuracies. Per-participant differences in the model's prediction of mean response times and accuracies across all trials as well as the target, lure, and nontarget conditions

were examined. In addition to these measures of fit, the correlation between the model's predictions and participant responses was calculated for average target, lure, and nontarget response times and accuracies as well as mean response times and accuracies across all trials.

2.4.5.2 Evaluation of predictive efficacy

To determine the contribution of each estimated parameter to a model's predictive efficacy, a "decremental leave-one-out" (dLOO) procedure was implemented. The dLOO can be thought of as a stepwise regression procedure, but applied in reverse. In this procedure, a set of models utilizing a subset of the estimated parameters are first produced from the full parameter set n by applying n choose k, where $k = n - 1$. For each participant and each model in this set, the k chosen parameters are set to the participant's estimated values, while the "left out" parameter is set to the mean of that parameter's estimates across participants for that model. Each of these models were run 100 times for each participant (using the participant's estimated values), and the model runs were averaged together on a trial-by-trial basis before the mean predicted response time and accuracy was calculated for each participant under each model. For each model, the R^2 between the model's predicted mean response times/accuracies and the participant's corresponding measures is first calculated, and then the response time and accuracy R^2 values are averaged. What results is a measure of how well each model predicts the group of participant's response times and accuracies. The model with the largest mean R^2 is determined as the "best-fitting" model, and the parameter that was "left out" of this model is "decremented" from the procedure, leaving $n - 1$ eligible parameters. The procedure then repeats, and a new subset of models are produced from the now-reduced parameter set $n - 1$ by applying n choose k again (here, the new n is now equal to $n - 1$, and the new k is equal to $n - 2$). For subsequent model runs, any "decremented" parameters are set of the mean of that parameter's estimates, just as the

"left out" parameters are. The procedure continues until only two parameters are eligible to take on the individually-estimated values. Of the two models that each include the participant-specific estimates of one of these two parameters (with all other parameters being set to the group's mean estimated value), the model (and therefore, the single parameter) with the larger mean R^2 value is determined to be the most essential to predicting differences between individuals. This procedure provides a rank-ordering of the importance of each parameter to the prediction of differences between individuals, and also allows for the visualization of the contribution of each of these parameters to the predictive efficacy of the model.

2.5 Resting-state fMRI prediction of Individual-specific parameter estimates

2.5.1 Brain parcellation and calculation of resting-state functional connectivity

To calculate the participant-specific resting-state functional connectivity matrices, each participant's resting-state fMRI data was first parcellated into a set of 264 regions on the basis of a pre-defined parcellation template (Power et al., 2011). This parcellation template was produced through a combination of two methods: 1. a meta-analysis of a large task-based fMRI dataset to identify a set of voxels that reliably demonstrate significant changes in activity that are associated with specific behaviors or task events; and 2. a functional-connectivity mapping technique (Cohen et al., 2008) that was applied to a secondary eyes-open fixation resting-state fMRI dataset. These two approaches each generated a set of ROI centroids which were merged together in a non-overlapping fashion (with preference given to the meta-analysis-derived ROIs) to produce the set of 264 ROI centroids. The volume of the ROIs was modeled as a 10 mm diameter sphere centered on each of the 264 centroids.

For each ROI in this parcellation, the set of voxels contained by a given ROI in an individual participant's data was first determined before taking the average of the timeseries of that set of voxels. This produced a mean timeseries for each ROI for each participant, which was subsequently used to compute the resting-state functional connectivity matrix.

To produce the resting-state functional connectivity matrix for each participant, the partial correlation between each pair of the 264 ROI timeseries was computed. While the standard method of computing functional connectivity uses the Pearson's *r* value between two timeseries (Biswal et al., 1995), use of the partial correlation controls for the potential influence of co-occurring third variables (such as the time-dependent activity of other brain ROIs) on the observed relationship between the two ROIs in question, providing a more accurate estimation of the direct connectivity between two regions (Cole et al., 2016; Smith, Vidaurre, et al., 2013).

2.5.2 Lasso regression

Once the resting-state functional connectivity matrix was computed for each participant, a series of Lasso regressions were performed on these matrices to identify the functional connectivity features that best predict individual values of the parameters estimated for a given ACT-R model. Lasso regression is a variant of standard linear regression that combines feature selection with parameter fitting (Tibshirani, 1996). In standard linear regression, the values of the β weights used to scale a set of regressors are determined by minimizing the quantity:

$$||y - \beta X||_2$$

where y is the value to be predicted, X is the regressor, β is the weight the regressor is scaled by, and the notation $||\mathbf{v}||_2$ represents the L^2 norm of a vector $\mathbf{v}$. Lasso regression similarly produces β weights which linearly scale a set of regressors; however, an additional penalty term ensures that

the complexity of the model is minimized by reducing the beta value of unnecessary regressors to zero:

$$\beta = argmin(||y - \beta X||_2 + \lambda||\beta||_1)$$

in which λ controls the balance between model simplicity (the reduction in the number of regressors through the penalty term $\lambda||\beta||_1$) and model accuracy (through the standard ordinary least squares term $||y - \beta X||_2$). In other words, the Lasso procedure seeks to minimize both the residual sum of squares (as in standard linear regression) and the sum of the absolute values of each regressor β weight, at a certain value of λ. At λ = 0, the Lasso regression reduces to the standard linear regression, and all regressors are "kept"; as λ increases, regressors are eliminated through setting their β weight to zero in order to satisfy the constraint. In this application, the Lasso regression serves to eliminate a number of functional connections that do not contribute to the model's ability to predict participant-specific parameter estimates.

To find the optimal value of λ for a given criterion and set of regressors, a cross-validation procedure can be applied. Here, a sequence of possible λ values are generated, and the ability of the Lasso model in predicting the criterion at each value of λ is determined using leave-one-out validation (LOOV). In LOOV, the model is fit *n*-1 times, where *n* is the number of observations on the criterion variable. In each round of model fitting, a different observation is left out while the model is fit to the remaining *n*-1 values. The model's mean error in predicting the criterion variable on the left-out observation is then determined for all tested values of λ, and the value of λ that produces the smallest cross-validation error is chosen.

Chapter 3. Results

3.1 Group-level Behavioral Performance

Mean response times and accuracies across all trials, as well as for trials belonging to the target, lure, and nontarget conditions separately, were calculated within the zero-back and two-back conditions for each participant. Response time and accuracy means and standard deviations for these sub-conditions within the zero-back and two-back conditions can be found in Table 1, while histograms of these participant measures in the zero-back condition can be seen in Figure 5, and equivalent histograms for the two-back condition can be found in Figure 6. On average, participants responded faster in the zero-back condition than the two-back condition, as well as showing greater accuracy in the zero-back condition.

The relationships of these measures between the zero-back and two-back conditions were also examined. In each sub-condition (all trials, target trials, lure trials, and nontarget trials), participant response times and accuracies were significantly correlated between the zero-back and two-back conditions. The Pearson's *r* values and associated *p* values can be found in Table 2, and scatterplots of these relationships can be seen in Figure 7.

3.2 Zero-back Model Results

3.2.1 Idiographic parameter estimation

Participants who did not achieve greater-than-chance performance on the target, lure, or nontarget conditions of the zero-back component of the task were not included in the parameter estimation procedure or any subsequent analysis of the zero-back model. For the zero-back

Table 1. Descriptive measures of participant performance in the zero-back and two-back conditions of the HCP N-back task. Mean response times are in seconds; accuracy calculated as the proportion of correct trials to total trials within the listed sub-conditions.

	Response Time		Accuracy	
	M	*SD*	*M*	*SD*
Zero-back				
All trials	0.82	0.14	0.91	0.10
Targets	0.84	0.15	0.81	0.21
Lures	0.80	0.16	0.90	0.16
Nontargets	0.82	0.14	0.97	0.05
Two-back				
All trials	1.03	0.12	0.85	0.09
Targets	0.99	0.14	0.79	0.17
Lures	1.10	0.15	0.73	0.19
Nontargets	1.02	0.13	0.91	0.08

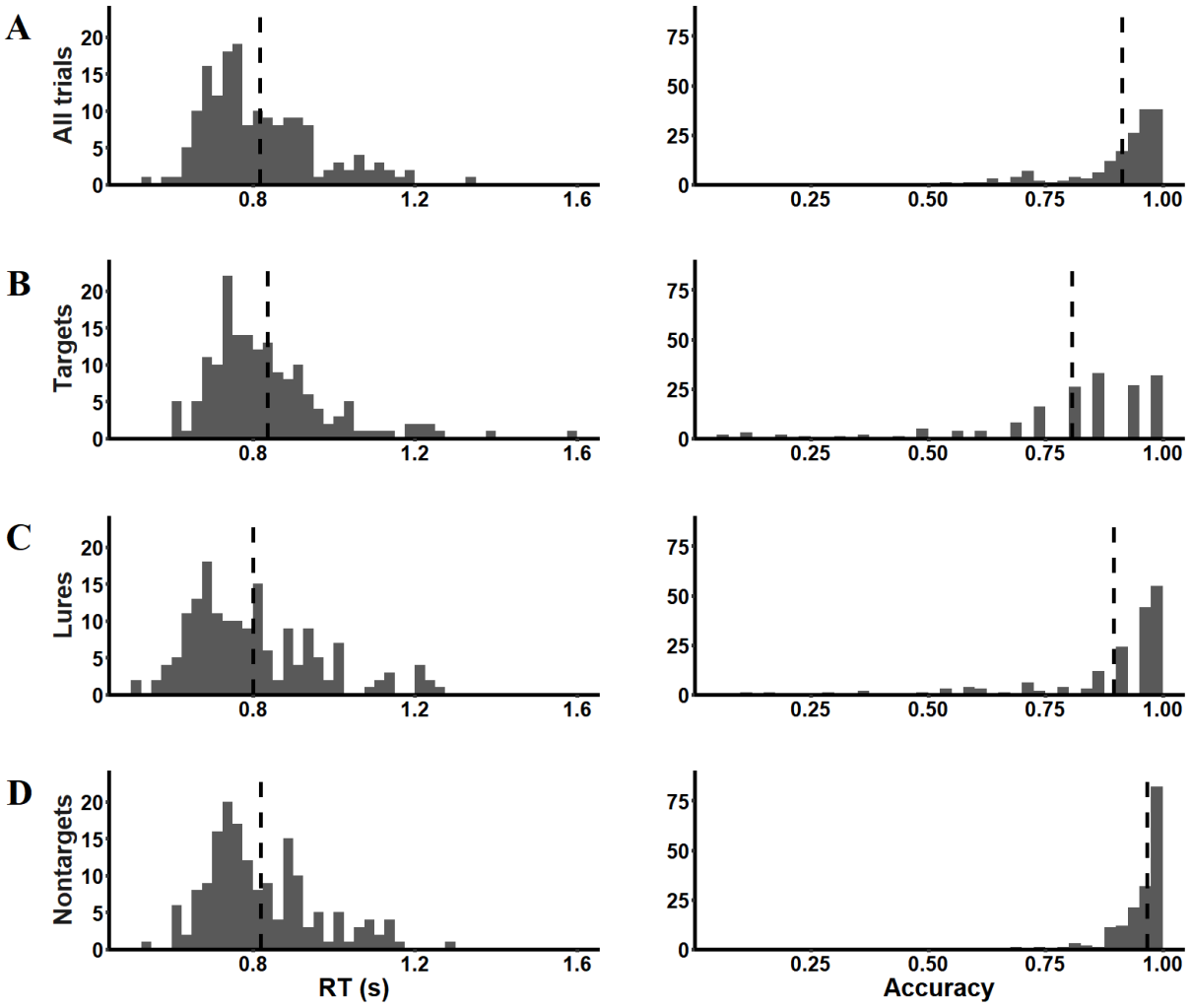

Figure 5. Histograms of participant mean response time and accuracy A. across all trials; B. across target trials; C. across lure trials, and D. across nontarget trials of the zero-back condition of the HCP N-back task. Plotted on the y-axis is the number of participants with a certain value of response time or accuracy. Vertical dashed bars denote the mean of the sample.

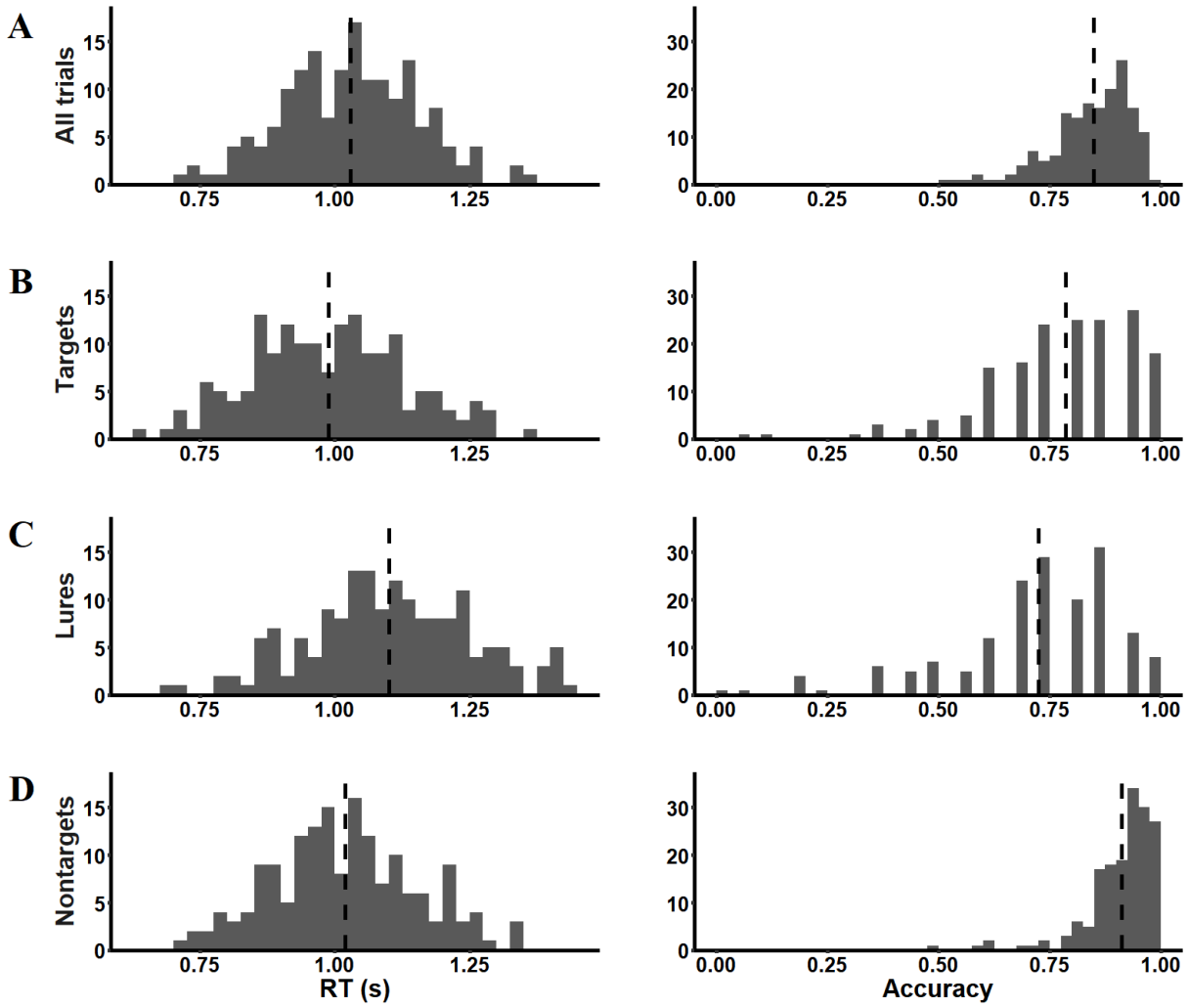

Figure 6. Histograms of participant mean response time and accuracy A. across all trials; B. across target trials; C. across lure trials, and D. across nontarget trials of the two-back condition of the HCP N-back task. Plotted on the y-axis is the number of participants with a certain value of response time or accuracy. Vertical dashed bars denote the mean of the sample.

Table 2. Correlations between participant zero-back and two-back response times and accuracies. Shown is the Pearson's *r* of the relationship, alongside the significance *p* of the relationship.

	Response Time		Accuracy	
	r	$p <$	r	$p <$
All trials	0.60	0.001	0.54	0.001
Targets	0.43	0.001	0.24	0.01
Lures	0.46	0.001	0.66	0.001
Nontargets	0.60	0.001	0.46	0.001

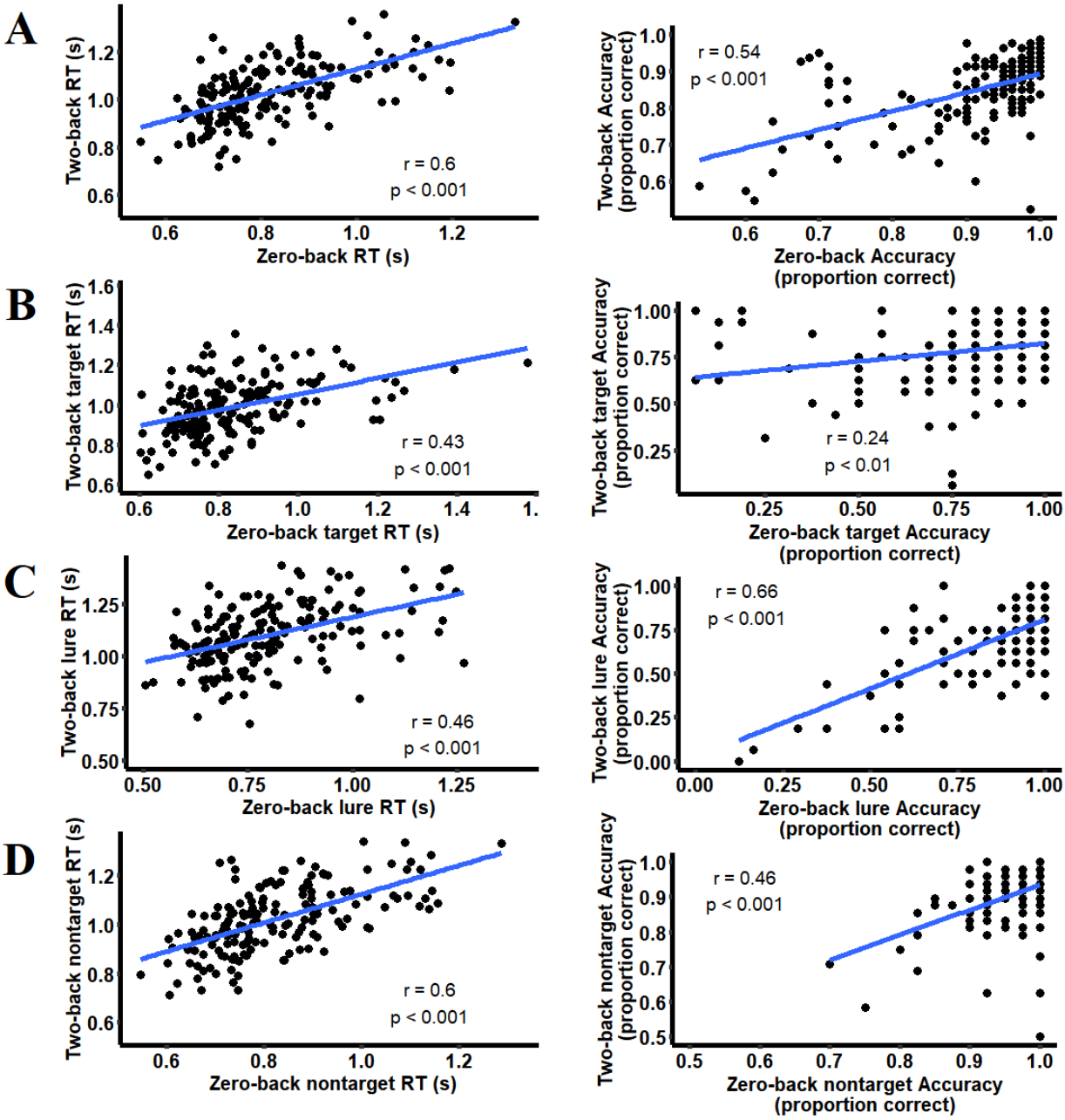

Figure 7. Scatterplots of the relationship between participant zero-back and two-back response time and accuracy A. across all trials; B. across target trials; C. across lure trials; and D. across nontarget trials. Annotated in each plot is the Pearson's *r* value of the relationship and the significance of the correlation. Embedded in each plot as a blue line is the linear model fit of the zero-back data to the two-back data.

condition, greater-than-chance performance was defined as 12 or more correct responses out of 16 possible attempts for target trials (one-sided binomial test, $N = 16$, $K = 12$, $p < 0.05$), 17 or more correct responses out of 24 possible attempts for lure trials (one-sided binomial test, $N = 24$, $K = 17$, $p < 0.05$), and 26 or more correct responses out of 40 possible attempts for nontarget trials (one-sided binomial test, $N = 40$, $K = 26$, $p < 0.05$). This resulted in the exclusion of 35 out of 167 participants.

A set of five parameters (the cue-stimulus similarity value c; the goal buffer spreading activation value W_g, the imaginal buffer spreading activation value W_i, the base-level activation decay parameter d, and the latency factor scale parameter F) were considered in the fitting of the zero-back model to the participant data from the zero-back component of the task. For each participant, the set of parameter values that minimized the RMSE between trial-by-trial RTs and block-wise/condition-wise (target, lure, and nontarget) accuracies were estimated. Means, standard deviations, and the range of estimated values across participants for each parameter can be found in Table 3; histograms of the estimates for each of these parameters across participants can be seen in Figure 8. As W_g and W_i provide a complementary but opposing influence on the retrieval process in this model (except in the case of target stimuli, for which they both promote the retrieval of the correct cue chunk), the difference between these two parameter estimates (W_g - W_i) within participants was calculated. The mean difference between the participant-specific estimates of these parameters was 0.24 ± 0.51 with a range of (-0.85, 1.38). This mean of differences was significantly different than 0 (one-sample t-test; $t(131) = 5.45$, $p < 0.001$; a two-sample paired t-test of the difference in means would reduce to this one-sample test of the differences), and almost 70% (92/132) of participants were estimated to have a W_g value greater than their W_i. A histogram of the within-participant differences in W_g and W_i estimates can also

Table 3. Means, standard deviations, and ranges of participant-specific parameter estimates for the zero-back model. W_g: goal buffer spreading activation value, W_i: imaginal buffer spreading activation value; d: base-level activation decay parameter, F: latency factor scale parameter; c: cue-stimulus similarity value.

Parameter	*M*	*SD*	*Range*
W_g	0.93	0.33	0.27-1.74
W_i	0.69	0.35	0.06-1.71
d	0.52	0.10	0.29-0.71
F	2.49	0.43	1.31-3.29
c	-0.43	0.22	-0.88-0.19

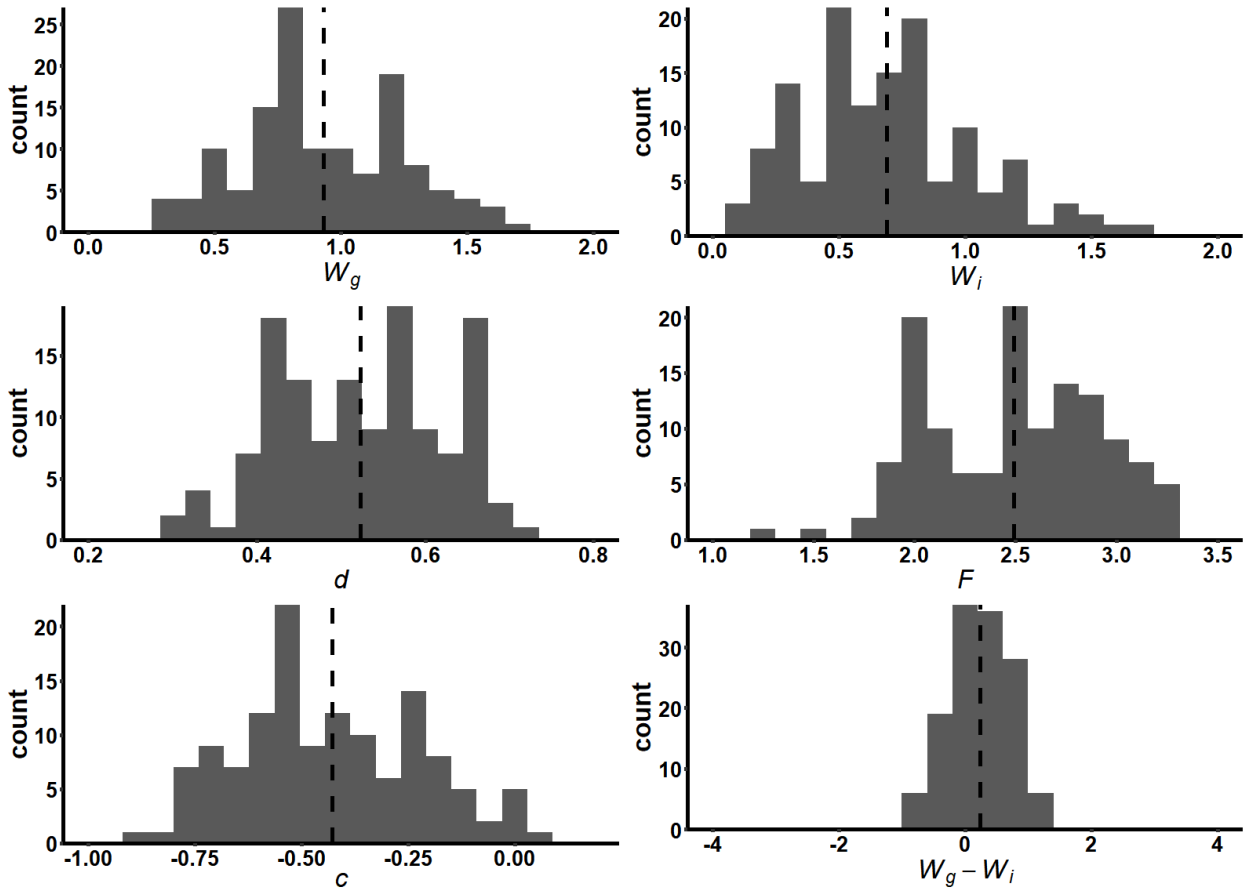

Figure 8. Histograms of the zero-back model participant-specific parameter estimates. Plotted on the y-axis is the number of participants with a certain value of the parameter estimate. The title of the x-axis denotes the parameter for which estimates are displayed (W_g: goal buffer spreading activation; W_i: imaginal buffer spreading activation; d: base-level decay rate; F: latency scale factor; c: cue-stimulus chunk similarity. W_g - W_i denotes the difference between goal and imaginal buffer spreading activation value estimates on a per participant basis). Vertical dashed bars denote the mean of the estimates of the specified parameter.

be found in Figure 8. Additionally, while the majority of participants were found to have a negative *c* value, the *c* value for 2% of participants was estimated as slightly positive. It is possible for ACT-R similarity values to take on a positive value, and the result in this case would be an increased likelihood in retrieving a stimulus chunk over a cue chunk of the same stimulus identity; this would not necessarily prevent the model (or participant) from performing appropriately.

3.2.2 Goodness of fit to participant measures

To determine the degree to which the participant-specific parameterized zero-back model predicts performance in the zero-back condition on an individual basis, the model's predicted mean response time and accuracy, when produced using each participant's set of estimated parameter values, was compared to the participant's measures of the same. These comparisons were made between model and participant aggregate measures over all trials, as well as those over the target, lure, and nontarget trials individually. Table 4 contains, in part, the mean and standard deviations of the model's predicted response times and accuracies within these sub-conditions. Histograms of these participant measures and model predictions can be found in Figure 9 (response times) and Figure 10 (accuracies).

The differences between the participant's measures and the model's predictions of these measures (model - participant) were calculated on a per participant basis. While it was observed that the model consistently underpredicted the duration of the participant's response times across the condition-wise measures (one-sample *t*-test results in Table 4), the model slightly over-predicted accuracy for target trials, while underpredicting accuracy for lure and nontarget trials.

Table 4. Mean and standard deviations of the zero-back model's predicted response times and accuracies, as well as the within-participant difference between these predicted measures and the participant's equivalents, across all trials as well as for target, lure, and nontarget trials. Shown alongside the mean difference between model and participant measures is the results of a one-sample t-test of whether the row's mean of differences was significantly different than zero. Listed at the right end of the table are the mean and standard deviation across participants of the RMSE between the model's trial-by-trial predictions and the participant's observed trial-by-trial behavior in the target, lure, and nontarget conditions.

	Model		Model - Participant Measure				RMSE	
	M	*SD*	*M*	*SD*	$t(131)$	$p <$	*M*	*SD*
Response Times								
All trials	0.64	0.09	-0.14	0.09	-18.4	0.001		
Targets	0.66	0.09	-0.14	0.09	-17.4	0.001	0.30	0.09
Lures	0.60	0.08	-0.16	0.11	-16.3	0.001	0.30	0.10
Nontargets	0.65	0.09	-0.13	0.09	-16.3	0.001	0.28	0.08
Accuracies								
All trials	0.89	0.08	-0.06	0.08	-8.56	0.001		
Targets	0.93	0.07	0.03	0.10	3.82	0.001	0.29	0.14
Lures	0.88	0.10	-0.07	0.11	-8.08	0.001	0.21	0.12
Nontargets	0.89	0.07	-0.09	0.08	-13.2	0.001	0.17	0.10

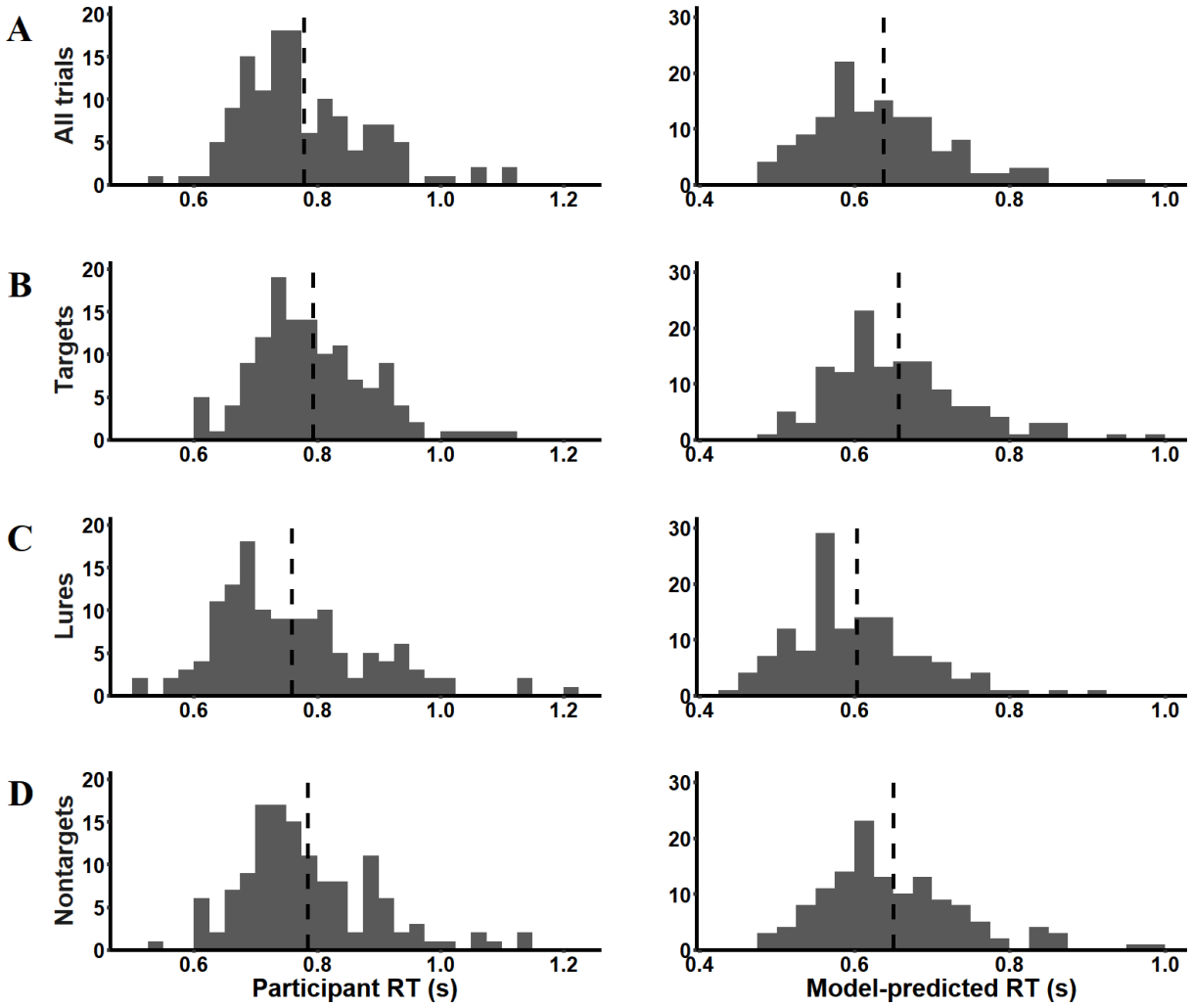

Figure 9. Histograms of zero-back participant and model-predicted response times. Participants include only those for whom zero-back model parameter estimates were produced. Model-predicted response times were produced by running the model 100 times with relevant parameters set to each participant's estimated values, then taking the average of the response times of those runs. Plotted on the y-axis is the number of participants or model predictions with a certain response time. Vertical dashed bars denote the mean of the sample.

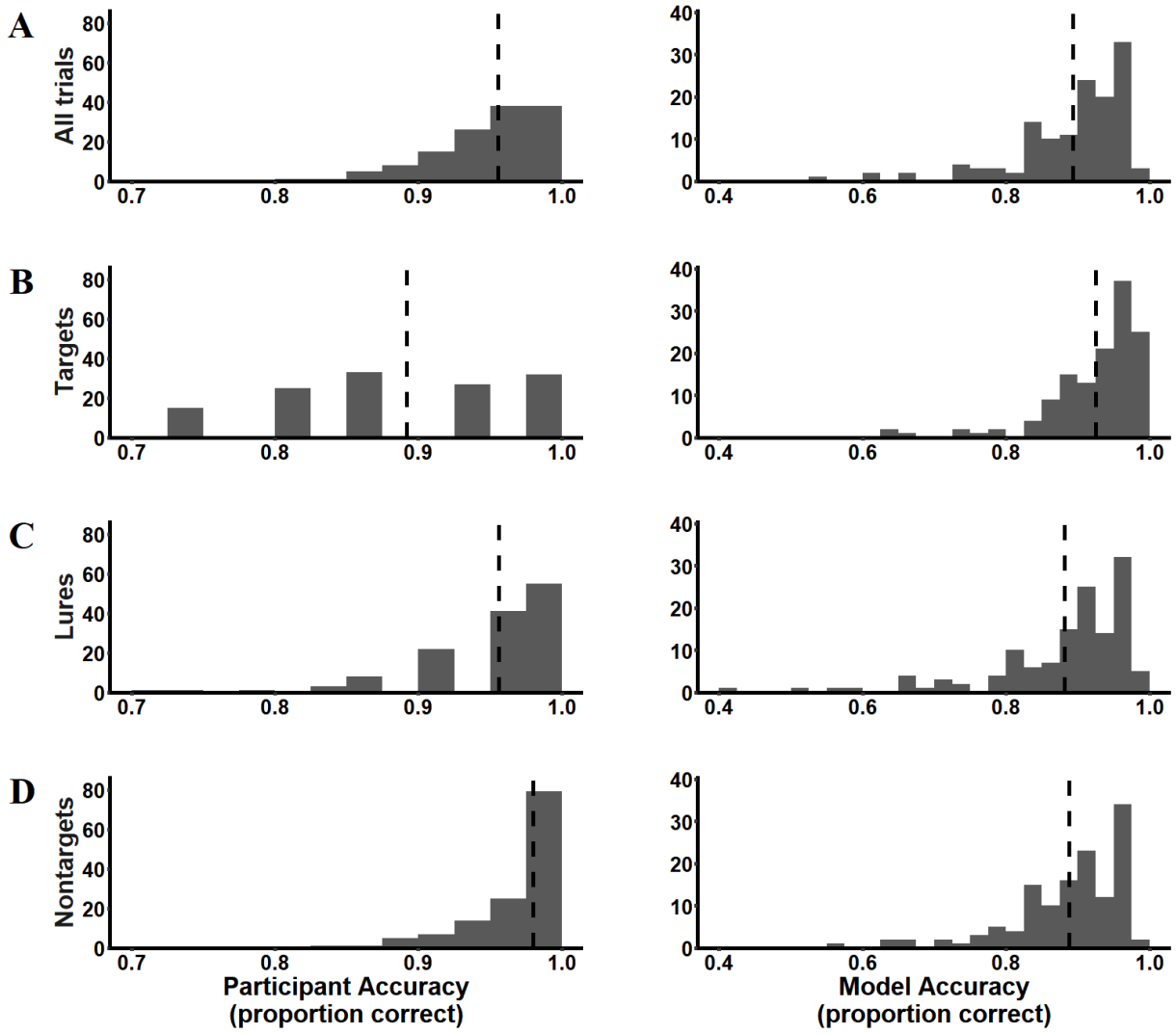

Figure 10. Histograms of zero-back participant and model-predicted accuracies. Participants include only those for whom zero-back model parameter estimates were produced. Model-predicted accuracies were produced by running the model 100 times with relevant parameters set to each participant's estimated values, then taking the average of the accuracies of those runs. Plotted on the y-axis is the number of participants or model predictions with a certain response time. Vertical dashed bars denote the mean of the sample.

Furthermore, the trial-by-trial RMSE between the model's predictions of target, lure, and nontarget response times and accuracies and the participant's measures of the same was calculated for each participant. The mean and standard deviation of these RMSEs across participants can be found in Table 4, while Figure 11 presents histograms of these values. In zero-back response times, the mean of the target RMSEs was not different than the mean of the lure RMSEs (one-sample t-test of the participant-wise differences; $t(131) = -0.18$, $p = 0.86$), but the means of target and lure RMSEs were each different than the nontarget RMSEs (target - nontarget one-sample t-test $t(131) = 2.10$, $p < 0.05$; lure - nontarget one-sample t-test $t(131) = 3.13$, $p < 0.01$). In zero-back accuracies, the means of the target, lure, and nontarget RMSEs were each different from one another (target - lure one-sample t-test $t(131) = 6.14$, $p < 0.001$; target - nontarget one-sample t-test $t(131) = 9.34$, $p < 0.001$; lure - nontarget one-sample t-test $t(131) = 5.12$, $p < 0.001$).

In addition to these RMSE measures, the correlation between the participant mean response time and mean accuracy (across all trials as well as in the target, lure, and nontarget conditions) and the model's predictions of these measures was calculated. The Pearson's r value and significance of these relationships are shown in Table 5, while scatterplots visualizing these relationships can be seen in Figure 12. Across all trials, the model's predictions of participant mean response times was strongly positively correlated with the participant's measures, and the model's predictions of participant mean accuracy was moderately correlated to the participant measures. In the target, lure, and nontarget stimulus conditions, the model significantly predicted the participant's mean target response times in each case, but only trended towards significance in the prediction of target, lure, and nontarget accuracies.

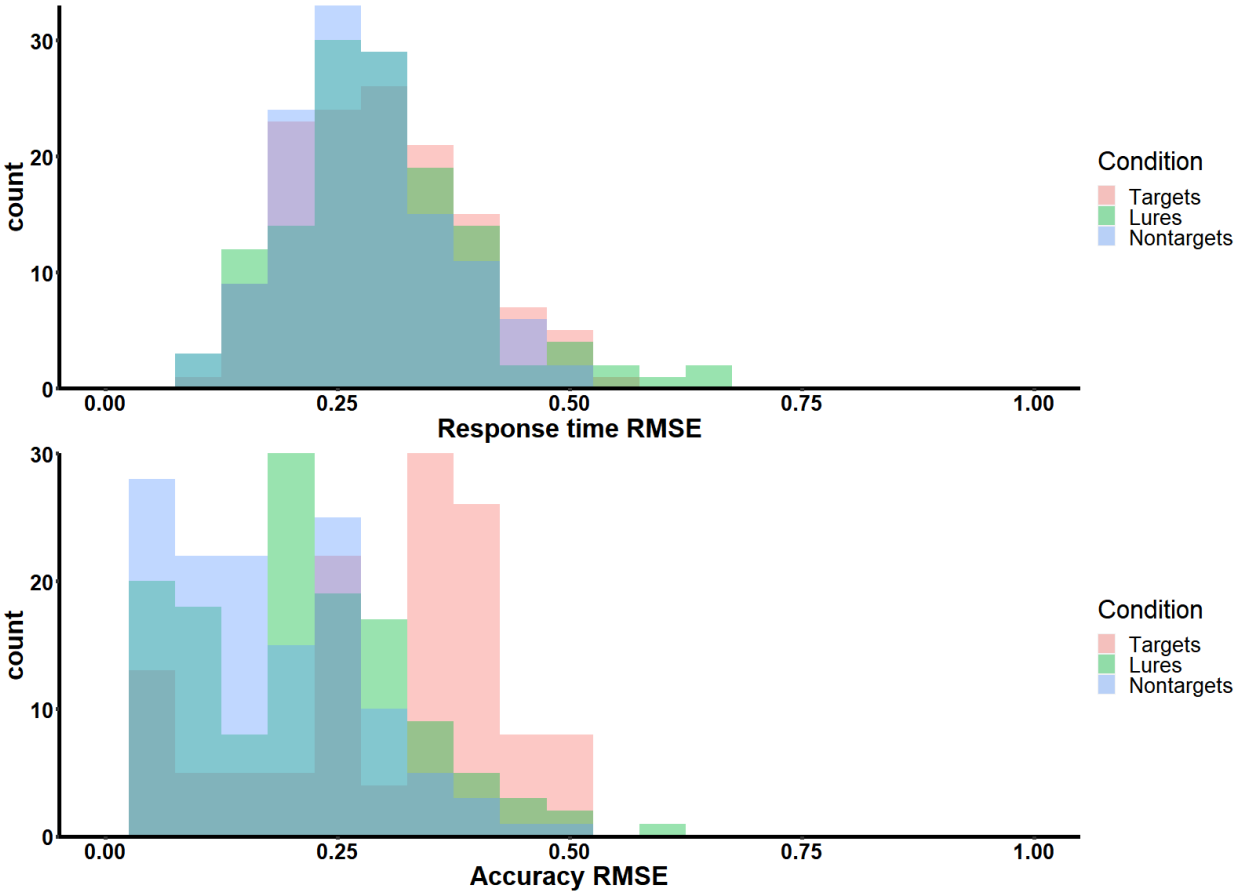

Figure 11. Histograms of RMSE values between the zero-back model's predictions of target, lure, and nontarget response times (top) and accuracies (bottom) and the participant's measures of the same. Plotted on the y-axis is the number of participants for whom the fitted model predictions produced a certain RMSE value.

Table 5. Correlations between the zero-back model's participant-specific-parameterized predictions of mean response times and accuracies and the observed participant measures of the same within the listed sub-conditions. Shown is the Pearson's *r* of the relationship, alongside the significance *p* of the relationship.

	Response Time		Accuracy	
	r	$p <$	r	$p <$
All trials	0.59	0.001	0.20	0.05
Targets	0.54	0.001	0.16	$= 0.06$
Lures	0.50	0.001	0.17	$= 0.05$
Nontargets	0.58	0.001	0.16	$= 0.06$

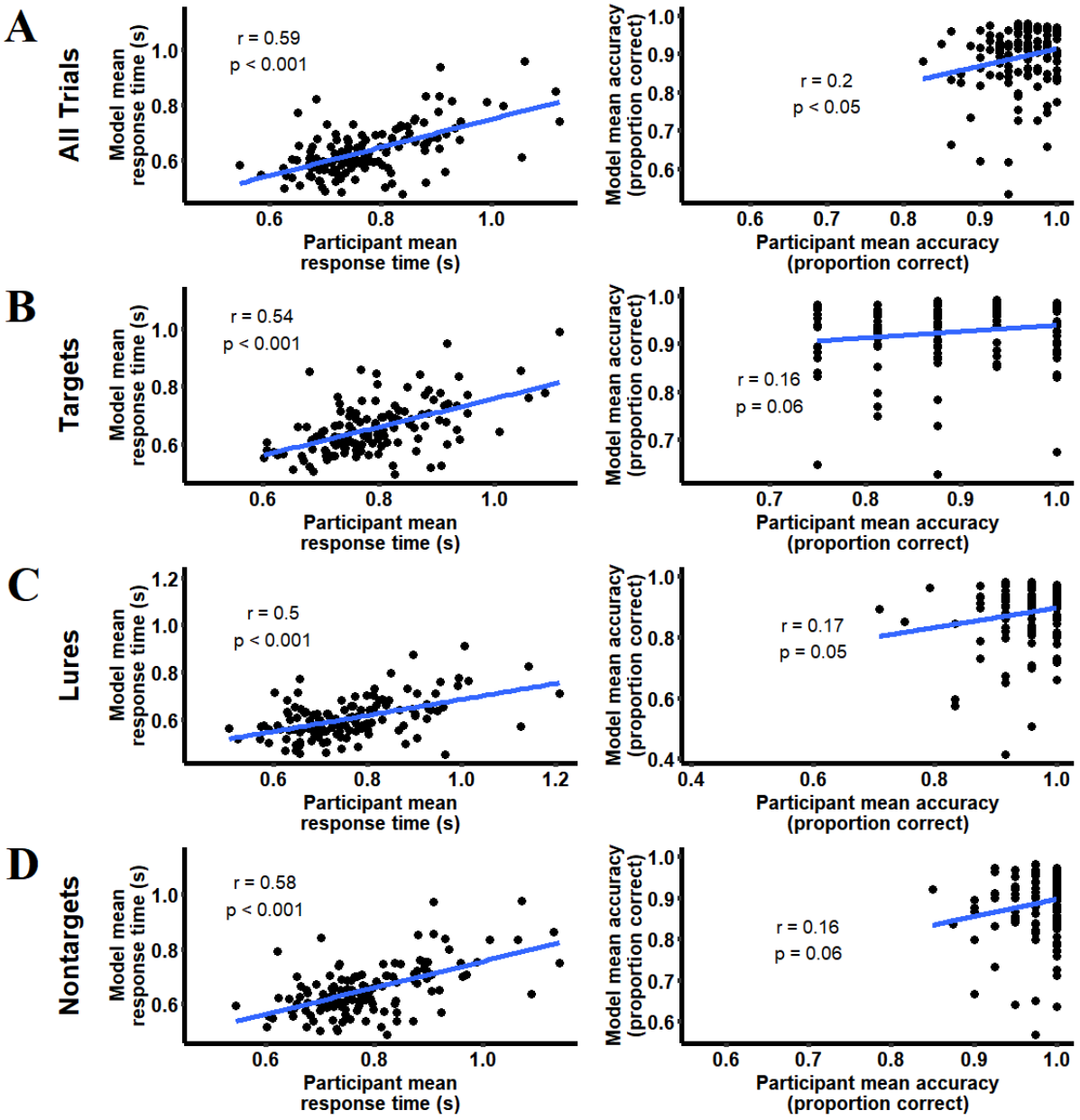

Figure 12. Scatterplots of the relationship between zero-back participant mean response times/accuracies and zero-back model-predicted mean response times/accuracies A. across all trials; B. across target trials; C. across lure trials; and D. across nontarget trials. Annotated in each plot is the Pearson's *r* value of the relationship and the significance of the correlation. Embedded in each plot as a blue line is the linear model fit of the model predictions to the participant data.

3.2.3 Decremental leave-one-out procedure

To further examine the individual contribution of each estimated parameter to the zero-back model's ability to predict participant behavior, the parameter set was subjected to the decremental leave-one-out procedure, described in section 2.4.5.2. Histograms displaying the relationship between model predictions and participant responses for all models in all iterations of the dLOO procedure can be seen in Figure 13 (response times) and Figure 14 (accuracies). In the initial iteration's set of five models, the model that utilized participant-specific estimates of the parameters W_g, W_i, d and F (with the value of c being set to the mean across estimates of c, or "left out") was the best fitting model across response time and accuracy measures (response time $R^2 = 0.36$, accuracy $R^2 = 0.053$), and so the cue-stimulus similarity parameter c was "decremented". In the second iteration, the model that included the participant-specific estimates of W_g, W_i, and F was observed to best-fit the participant measures (response time $R^2 = 0.36$, accuracy $R^2 = 0.10$), and so the decay parameter d was also "decremented". In the third iteration, the model that considered the participant-specific estimates of W_g and F "won out" (response time $R^2 = 0.29$, accuracy $R^2 = 0.13$), and W_i was "decremented". The final iteration of the dLOO procedure demonstrated that for the zero-back model's predictions of behavior under the zero-back task, the participant-specific estimates of the goal buffer spreading activation parameter W_g alone were sufficient to accurately predict both participant response time (Pearson's $r = 0.52$, $p < 0.001$) and accuracy (Pearson's $r = 0.35$, $p < 0.001$), while the participant-specific estimates of the latency factor F were unable to predict neither response time (Pearson's $r = 0.14$, $p = 0.11$) nor accuracy (Pearson's $r = 0.03$, $p = 0.72$).

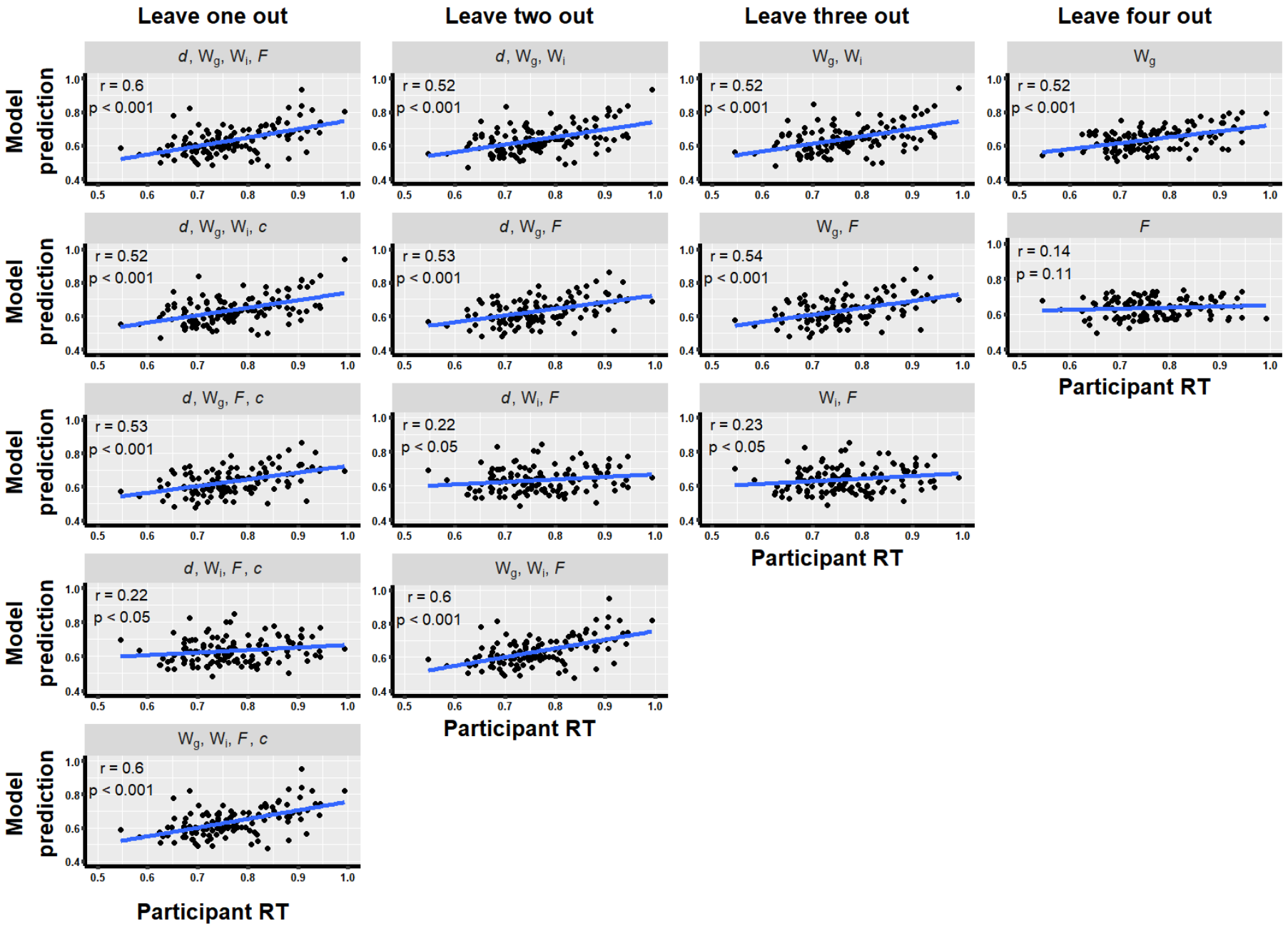
Leave one out
Leave two out
Leave three out
Leave four out
d, Wg, Wi, F
r = 0.6
p < 0.001
d, Wg, Wi
r = 0.52
p < 0.001
Wg, Wi
r = 0.52
p < 0.001
Wg
r = 0.52
p < 0.001
d, Wg, Wi, c
r = 0.52
p < 0.001
d, Wg, F
r = 0.53
p < 0.001
Wg, F
r = 0.54
p < 0.001
F
r = 0.14
p = 0.11
d, Wg, F, c
r = 0.53
p < 0.001
d, Wi, F
r = 0.22
p < 0.05
Wi, F
r = 0.23
p < 0.05
d, Wi, F, c
r = 0.22
p < 0.05
Wg, Wi, F
r = 0.6
p < 0.001
Wg, Wi, F, c
r = 0.6
p < 0.001
Model prediction
Participant RT

Figure 13. Scatterplots of the relationship between zero-back participant mean response times and model-predicted mean response times for each model considered in each iteration of the dLOO procedure (columns). The parameters for which the model used the participant-specific estimates are labeled at the top of each subplot (W_g: goal buffer spreading activation; W_i: imaginal buffer spreading activation; d: base-level decay rate; F: latency scale factor; c: cue-stimulus chunk similarity). Annotated in each plot is the Pearson's r value of the relationship and the significance of the correlation. Embedded in each plot as a blue line is the linear model fit of the model predictions to the participant data.

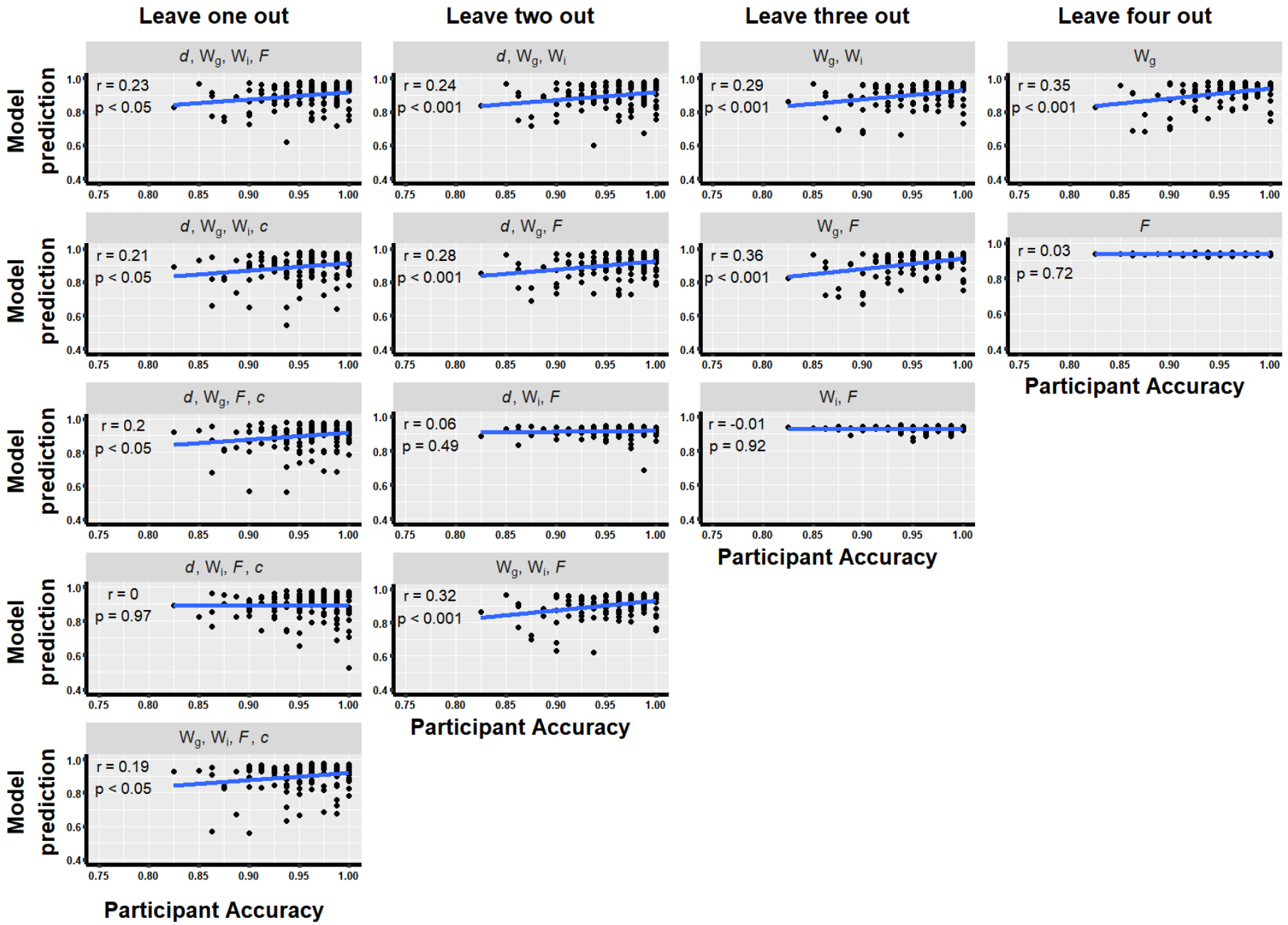
Leave one out
Leave two out
Leave three out
Leave four out
Model prediction
Participant Accuracy
d, Wg, Wi, F
r = 0.23
p < 0.05
d, Wg, Wi
r = 0.24
p < 0.001
Wg, Wi
r = 0.29
p < 0.001
Wg
r = 0.35
p < 0.001
d, Wg, Wi, c
r = 0.21
p < 0.05
d, Wg, F
r = 0.28
p < 0.001
Wg, F
r = 0.36
p < 0.001
F
r = 0.03
p = 0.72
d, Wg, F, c
r = 0.2
p < 0.05
d, Wi, F
r = 0.06
p = 0.49
Wi, F
r = -0.01
p = 0.92
d, Wi, F, c
r = 0
p = 0.97
Wg, Wi, F
r = 0.32
p < 0.001
Wg, Wi, F, c
r = 0.19
p < 0.05
0.4
0.6
0.8
1.0
0.75
0.80
0.85
0.90
0.95
1.00

Figure 14. Scatterplots of the relationship between zero-back participant mean accuracies and model-predicted mean accuracies for each model considered in each iteration of the dLOO procedure (columns). The parameters for which the model used the participant-specific estimates are labeled at the top of each subplot (W_g: goal buffer spreading activation; W_i: imaginal buffer spreading activation; d: base-level decay rate; F: latency scale factor; c: cue-stimulus chunk similarity). Annotated in each plot is the Pearson's r value of the relationship and the significance of the correlation. Embedded in each plot as a blue line is the linear model fit of the model predictions to the participant data.

3.2.4 Identification of predictive resting-state functional connectivity and prediction of individual-specific parameter values

As the participant-specific estimates of the goal buffer spreading activation parameter W_g were solely predictive of both mean response time and mean accuracy, identification of predictive resting-state functional connections and subsequent prediction of individual-specific parameter values will be reported for only W_g here.

Through the Lasso regression cross-validation procedure, it was determined that the W_g Lasso regression's optimal value of λ was 0.024 (λ = 0.024), resulting in the inclusion of 89 out of a possible 34,716 regressors (functional connections; 0.26% of possible predictors). Visualizations of the results of the cross-validation procedure can be seen in Figure 15. These 89 functional connections included a total of 129 brain regions from 14 different functional networks. Figure 16 shows location of each brain region and the β values of the 10 functional connections with the largest absolute beta values of the set of identified functional connections. See table 6 for the β values associated with these 10 functional connections; these values are visualized as bars between brain region nodes in Figure 16. Additionally, for each node included in the set of 89 identified functional connections, the node's "importance" in the prediction of the goal buffer spreading activation parameter W_g was computed by summing the absolute value of the β values of the connections that a given node participates in. See Figure 17 for a visualization of the importance of each node included in the 10 functional connections that are presented in Figure 16/Table 6; Table 7 presents importance values for each of these nodes.

The fitted Lasso regression equation was then used to predict participant-specific value of the W_g parameter, on the basis of the participant's functional connectivity values in the 89 functional connections identified through the Lasso regression fitting process. These

Lasso-predicted values of the W_g parameter were strongly correlated to the W_g parameter estimate produced by fitting the ACT-R zero-back model to the participant's behavior (Pearson's $r = 0.99$, $p < 0.001$). Figure 18 shows the scatterplot of Lasso-predicted W_g values against ACT-R model fitted W_g values.

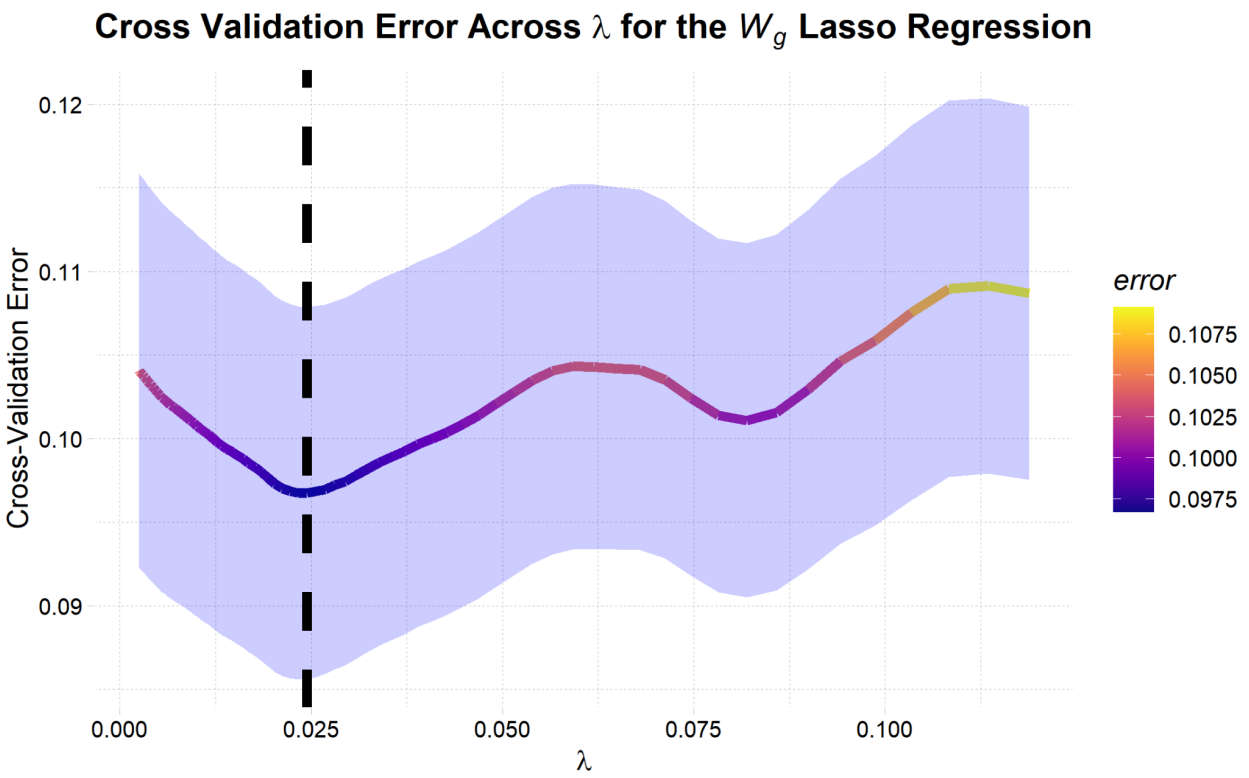

Figure 15. Results of the cross-validation procedure for the Lasso regression of the participant-specific functional connectome on the zero-back model estimated W_g parameter. Plotted on the y-axis is the cross-validation error of the regression model's prediction of participant-specific W_g at a given value of λ, which is also mapped to the color of the plotted line. Shaded blue region represents the standard deviation of the regression model's cross-validation error across held-out participants, while the vertical dashed black line denotes the value of λ that produced the minimum cross-validation error.

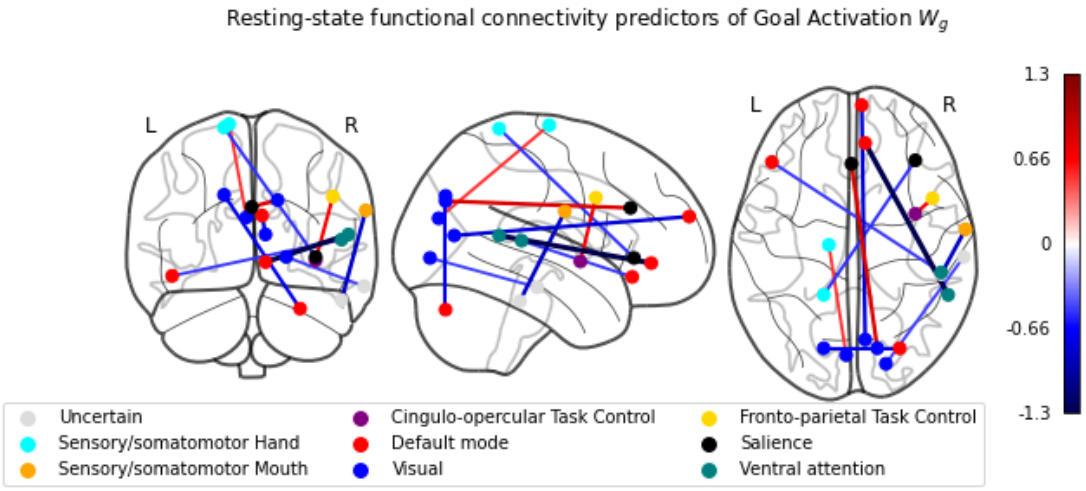

Figure 16. Functional connectome nodes and connections identified to be significantly predictive of participant-specific estimates of the goal buffer spreading activation parameter W_g. Shown here are the 10 functional connections with the largest absolute beta values of the set of identified functional connections (89 total). Nodes are color-coded by their assigned network according to (Power et al., 2011). Thickness and color of bars between nodes denotes the estimated β value for that node-pair's functional connectivity value.

Table 6. The 10 largest (absolute value) Lasso regression-estimated β values for the prediction of the zero-back model W_g parameter. Listed are the networks of each of the two nodes in the identified connection, the specific node numbers of the paired nodes (in the Power (2011) parcellation), and the Lasso regression-estimated β value for the connection.

Connected Networks	Specific Nodes	Beta Value
Default mode-Ventral attention	110-240	-1.320621
Uncertain-Sensory/somatomotor Mouth	10-46	-0.868669
Visual-Salience	159-215	0.840435
Default mode-Visual	106-145	-0.805250
Default mode-Visual	127-166	-0.725804
Cingulo-opercular Task Control-Fronto-parietal...	52-186	0.720632
Sensory/somatomotor Hand-Salience	38-210	-0.545266
Sensory/somatomotor Hand-Visual	35-167	0.532757
Default mode-Ventral attention	137-238	-0.523943
Uncertain-Visual	9-158	-0.498597

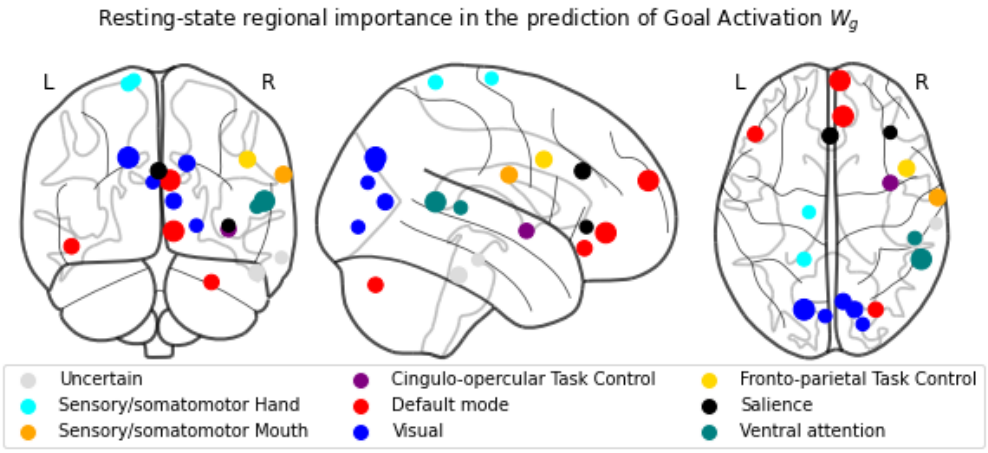

Figure 17. Importance of functional connectome nodes that participate in connections identified to be significantly predictive of participant-specific estimates of the goal buffer spreading activation parameter W_g. The size of each node is scaled by the node's importance value. Shown here are the nodes included in the 10 functional connections with the largest absolute beta values of the set of identified functional connections (89 total). Nodes are color-coded by their assigned network according to (Power et al., 2011).

Table 7. The importance value (sum of the absolute values of each β associated with the identified functional connections that a node participates in) for each node included in the set of functional connections with the 10 largest β values for the Lasso regression prediction of the zero-back model goal buffer spreading activation parameter W_g. Listed is the specific node number in the Power (2011) parcellation, the network that the node is a member of, the X, Y, and Z coordinates of the node in MNI space, and the computed importance value of the node.

ROI	Network	X	Y	Z	Importance Value
166	Visual	-16	-77	34	1.404774
110	Default mode	8	42	-5	1.335948
106	Default mode	6	64	22	1.329925
240	Ventral attention	56	-46	11	1.324744
10	Uncertain	52	-34	-27	0.868669
46	Sensory/somatomotor Mouth	66	-8	25	0.868669
186	Fronto-parietal Task Control	47	10	33	0.853969
215	Salience	0	30	27	0.840435
159	Visual	15	-77	31	0.840435
145	Visual	8	-72	11	0.805250
137	Default mode	-46	31	-13	0.746941
127	Default mode	28	-77	-32	0.725804
52	Cingulo-opercular Task Control	37	1	-4	0.720632
38	Sensory/somatomotor Hand	-16	-46	73	0.626086
238	Ventral attention	52	-33	8	0.574663
167	Visual	-3	-81	21	0.566681
158	Visual	20	-86	-2	0.564807
210	Salience	37	32	-2	0.545266
35	Sensory/somatomotor Hand	-13	-17	75	0.532757
9	Uncertain	65	-24	-19	0.498597

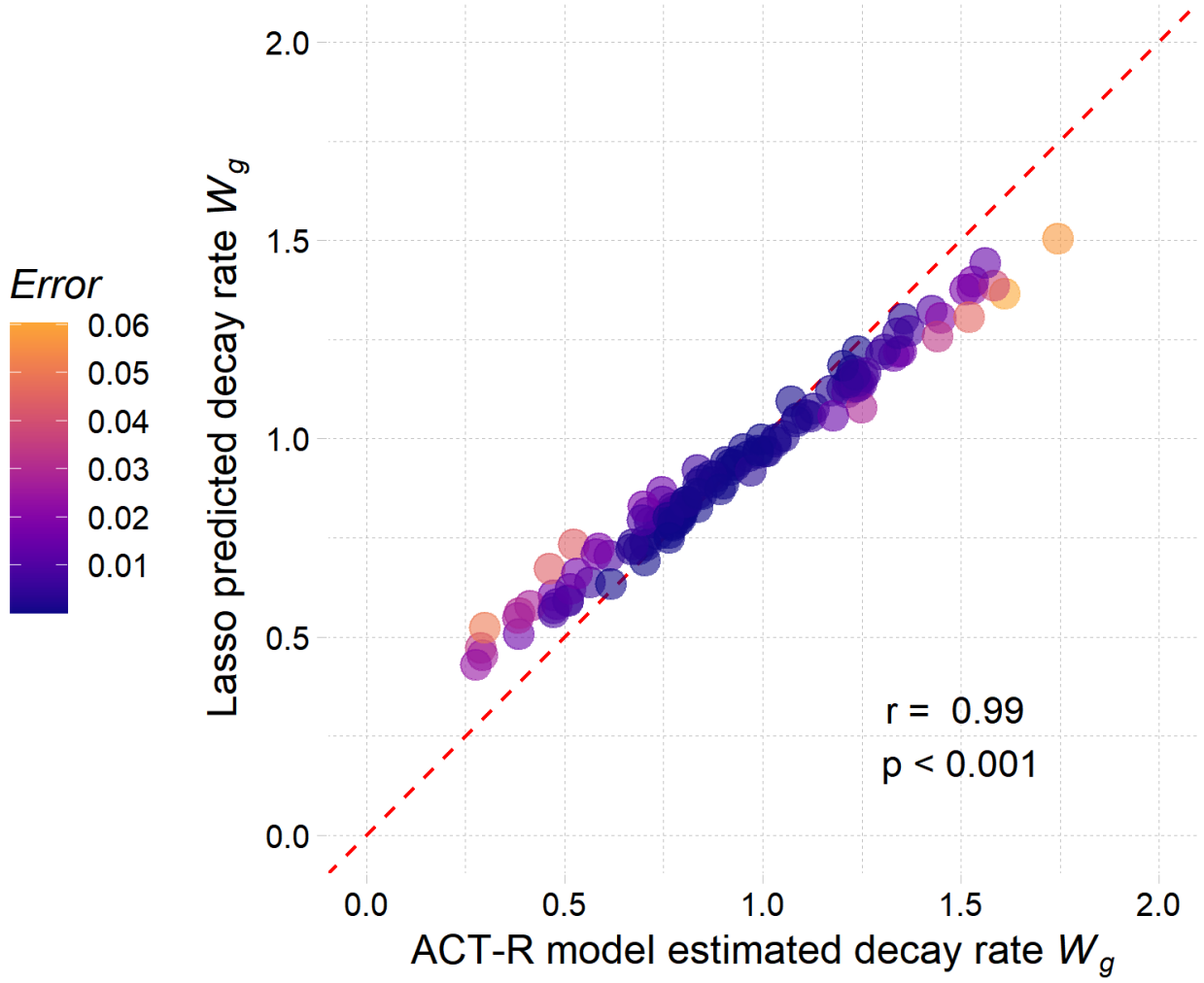

Figure 18. Scatterplot of Lasso-predicted W_g values against zero-back ACT-R model fitted W_g values. Points are colored according to the squared error between Lasso-predicted values and model-estimated values. Dashed red line denotes the identity line. The Pearson's *r* value of the relationship and the significance of the correlation is annotated in the bottom right.

3.3 N-back Model Results

3.3.1 Model performance across N-back conditions at default parameter settings

Before parameters were fit to the participant's behavior, the performance of the N-back model was simulated at the initialization parameter settings (see section 2.4.1) 100 times for each of the one-, two-, and three-back conditions. This was done to assess the overall reliability of the model's structure and the degree to which it can capture average human performance realistically. Table 8 presents the mean and standard deviation of the model's response time and accuracy predictions across the 100 runs (mean response time was only calculated across all trials for each of the N-back conditions, as it showed very little variability within-condition as well as across conditions, but mean accuracy was calculated over all trials as well as within target, lure, and nontarget trials), while Figure 19 presents histograms of these accuracy and response time predictions of the N-back model.

To examine changes in the model's performance across the one-, two-, and three-back conditions, a fixed-effects linear model with the N-back condition serving as a predictor for the overall, target, lure, and nontarget accuracies was established. This model was observed to significantly predict the accuracy measures ($F(1,1198) = 3081$, $p < 0.001$), and the N-back condition regressor was found to have a significant, negative β value, indicating that the model's accuracy decreased as the N-back value increased ($\beta = -0.20$, $p < 0.001$).

Table 8. N-back model means and standard deviations over 100 model runs at default parameter settings for response times across all trials as well as accuracies across all trials and within target, lure, and nontarget trials for one-, two-, and three-back conditions. Mean response times are in seconds; accuracy calculated as the proportion of correct trials to total trials within the listed sub-conditions.

	Response Time		Accuracy	
	M	*SD*	*M*	*SD*
One-back				
All trials	0.56	0.01	0.96	0.03
Targets			0.91	0.08
Lures			0.96	0.04
Nontargets			0.98	0.03
Two-back				
All trials	0.55	0.01	0.84	0.04
Targets			0.73	0.11
Lures			0.83	0.09
Nontargets			0.88	0.04
Three-back				
All trials	0.54	0.01	0.59	0.04
Targets			0.40	0.12
Lures			0.55	0.11
Nontargets			0.66	0.04

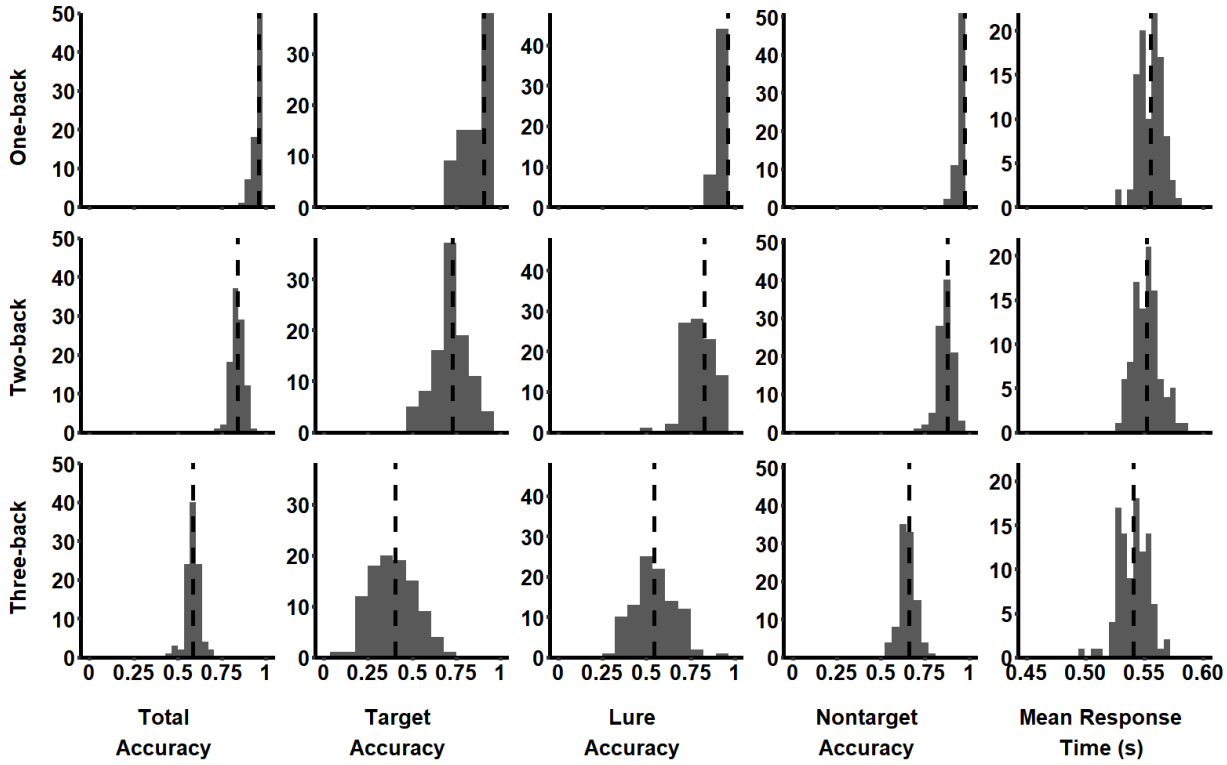

Figure 19. Histograms of the N-back model's mean accuracy and response time predictions for the one-, two-, and three-back conditions at the initialization parameter values for the estimated parameters W_g, W_i, d, and F. Rows show results from each condition; columns show the labeled accuracy or response time measure. Vertical dashed bars denote the mean of the sample.

3.3.2 Idiographic parameter estimation

Participants who did not achieve greater-than-chance performance on the target, lure, or nontarget conditions of the two-back component of the task were not included in the parameter estimation procedure or any subsequent analysis of the N-back model. Greater-than-chance performance was defined as 12 or more correct responses out of 16 possible attempts for target trials (one-sided binomial test, $N = 16$, $K = 12$, $p < 0.05$), 12 or more correct responses out of 16 possible attempts for lure trials (one-sided binomial test, $N = 16$, $K = 12$, $p < 0.05$), and 31 or more correct responses out of 48 possible attempts for nontarget trials (one-sided binomial test, $N = 48$, $K = 31$, $p < 0.05$). This resulted in the exclusion of 84 out of 167 participants.

A set of four parameters (the same as the zero-back model, minus the cue-stimulus similarity c) were considered in the fitting of the N-back model to the participant data from the two-back component of the task. For each participant, the set of parameter values that minimized the RMSE between trial-by-trial RTs and block-wise/condition-wise (target, lure, and nontarget) accuracies were estimated. Means, standard deviations, and the range of estimated values across participants for each parameter estimated under the N-back model can be found in Table 9; histograms of the estimates for each of these parameters, as well as the difference between W_g and W_i estimates, can be seen in Figure 20. As was done for the zero-back model, the difference between these two parameter estimates (W_g - W_i) within participants was examined. The mean difference between the participant-specific estimates of these parameters was 0.26 ± 0.54 with a range of (-1.18, 1.27). This mean of differences was significantly different than 0 ($t(82) = 4.40$, $p < 0.001$), and almost 70% (58/83) of participants were estimated to have a W_g value greater than their W_i. A histogram of the per-participant differences between W_g and W_i estimates can also be found in Figure 20.

Table 9. Means, standard deviations, and ranges of participant-specific parameter estimates for the N-back model. W_g: goal buffer spreading activation value, W_i: imaginal buffer spreading activation value; d: base-level activation decay parameter, F: latency factor scale parameter.

Parameter	*M*	*SD*	*Range*
W_g	0.95	0.39	0.06-1.75
W_i	0.70	0.29	0.18-1.65
d	0.51	0.10	0.29-0.67
F	2.33	0.45	1.20-3.14

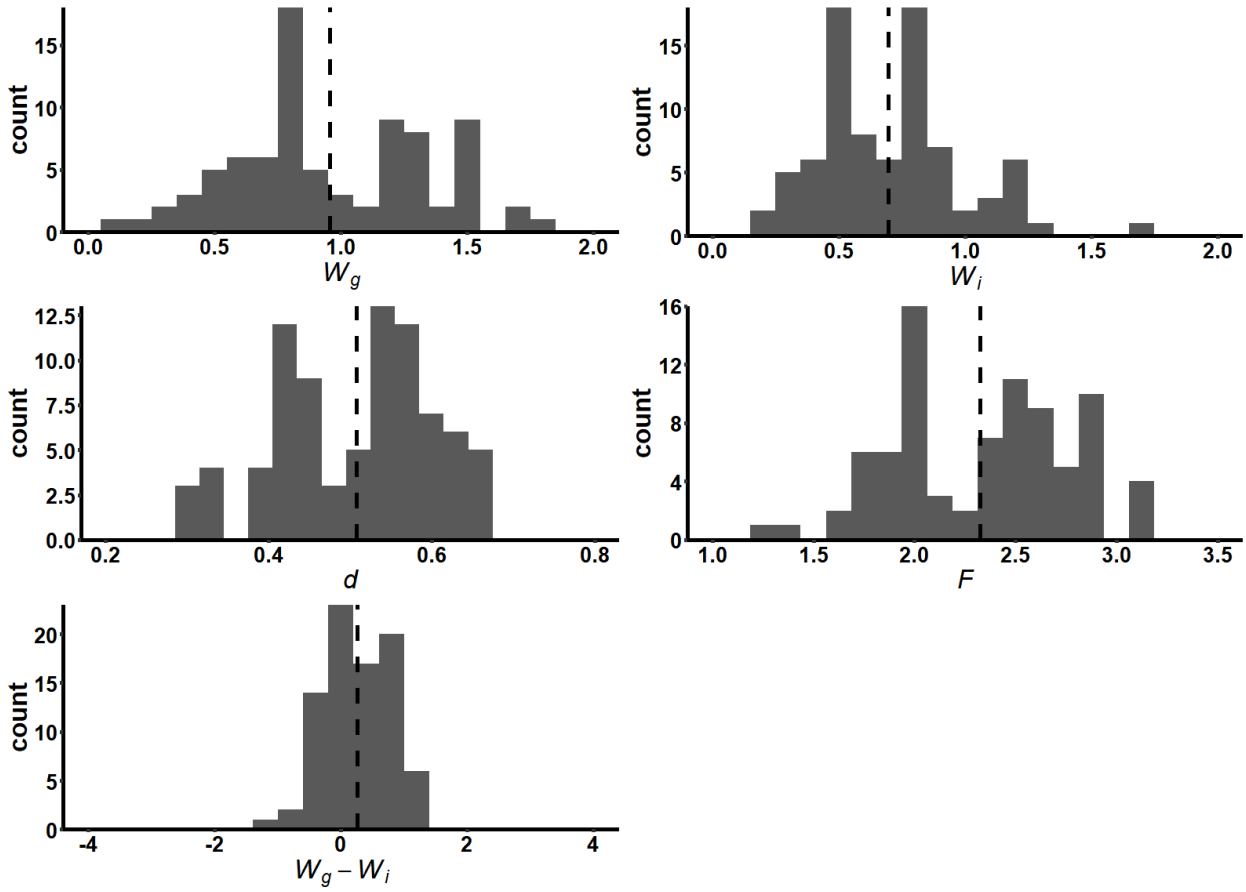

Figure 20. Histograms of the N-back model participant-specific parameter estimates. Plotted on the y-axis is the number of participants with a certain value of the parameter estimate. The title of the x-axis denotes the parameter for which estimates are displayed (W_g: goal buffer spreading activation; W_i: imaginal buffer spreading activation; d: base-level decay rate; F: latency scale factor. W_g - W_i denotes the difference between goal and imaginal buffer spreading activation value estimates on a per participant basis). Vertical dashed bars denote the mean of the estimates of the specified parameter.

3.3.3 Goodness of fit to participant measures

The measures that were applied to evaluate the goodness of fit of the zero-back model were also applied to the N-back model's parameterized predictions of two-back performance. Table 10 contains, in part, the mean and standard deviations of the model's predicted response times and accuracies within these sub-conditions. Histograms of these participant measures and model predictions can be found in Figure 21 (response times) and Figure 22 (accuracies). Akin to the zero-back model, it was observed that the N-back model consistently underpredicted the duration of the participant's response times across the condition-wise measures (one-sample *t*-test results in Table 10). The N-back model, however, also underpredicted accuracy across these measures.

Trial-by-trial RMSEs were also calculated in an identical manner to the same that were computed for the zero-back model. The mean and standard deviation of these RMSEs across participants can be found in Table 10, while Figure 23 presents histograms of these values. In two-back response times, the mean of the lure RMSEs was not different than the mean of the nontarget RMSEs (one-sample t-test of the participant-wise differences; $t(82) = -1.60$, $p = 0.11$), but the means of lure and nontarget RMSEs were each different than the target RMSEs (target - lure one-sample t-test $t(82) = -13.9$, $p < 0.001$; target - nontarget one-sample t-test $t(82) = -14.9$, $p < 0.001$). In two-back accuracies, the means of the target, lure, and nontarget RMSEs were each different from one another (target - lure one-sample t-test $t(82) = 2.91$, $p < 0.01$; target - nontarget one-sample t-test $t(82) = 11.8$, $p < 0.001$; lure - nontarget one-sample t-test $t(82) = 7.71$, $p < 0.001$).

In addition to these RMSE measures, the correlation between the participant mean response time and mean accuracy (across all trials as well as in the target, lure, and nontarget

conditions) and the model's predictions of these measures was calculated. The Pearson's *r* value and significance of these relationships are shown in Table 11, while scatterplots visualizing these relationships can be seen in Figure 24. It was observed that the N-back model's predictions of participant mean response time and accuracy across all trials were significantly correlated to the participant's measures. The N-back model also significantly predicted mean response time and accuracy in the target and nontarget conditions, but only predicted mean response time for the lure condition - the model did not significantly predict participant lure accuracies in the two-back condition.

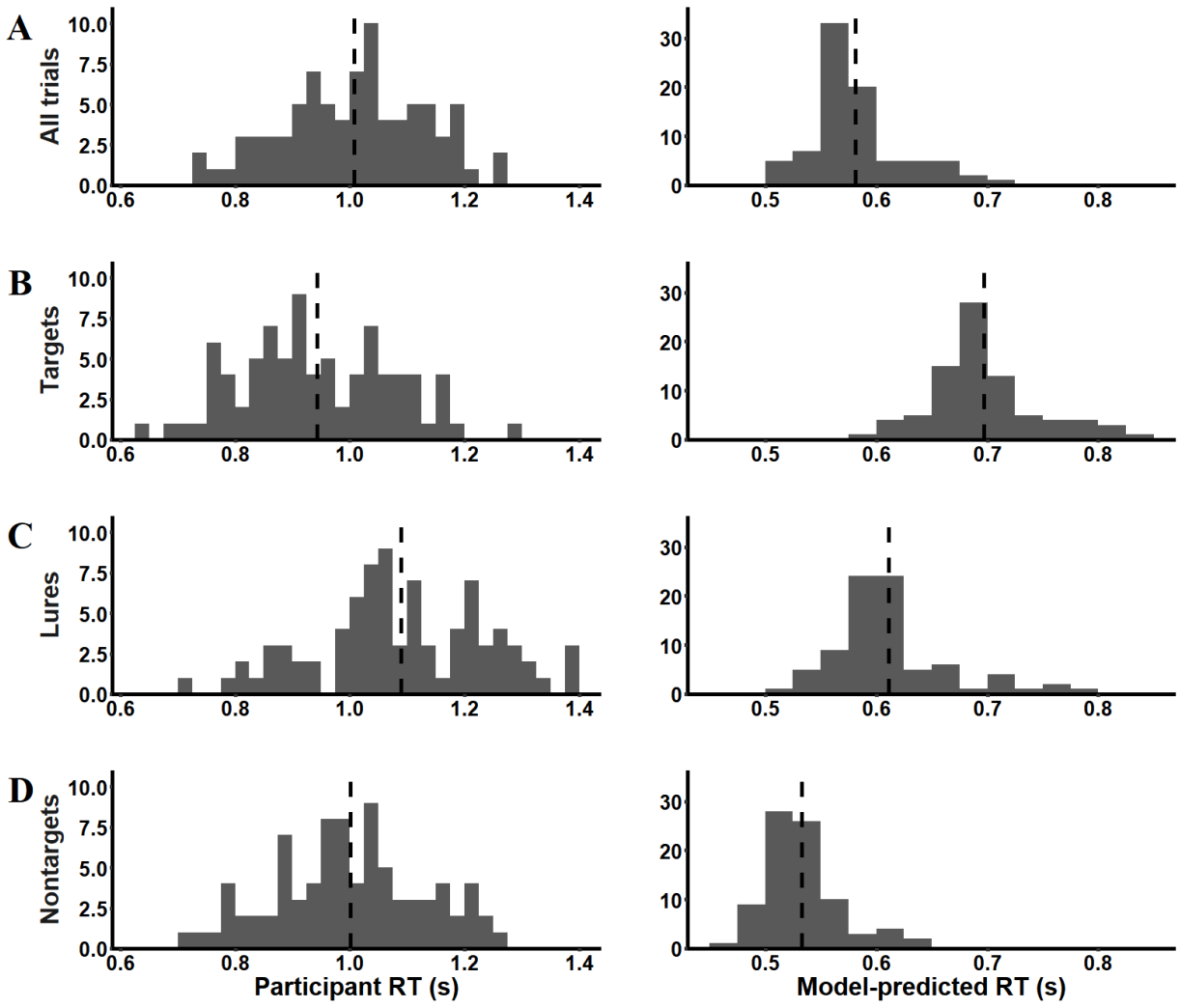

Figure 21. Histograms of two-back participant and N-back model-predicted response times. Participants include only those for whom N-back model parameter estimates were produced. Model-predicted response times were produced by running the model 100 times with relevant parameters set to each participant's estimated values, then taking the average of the response times of those runs. Plotted on the y-axis is the number of participants or model predictions with a certain response time. Vertical dashed bars denote the mean of the sample.

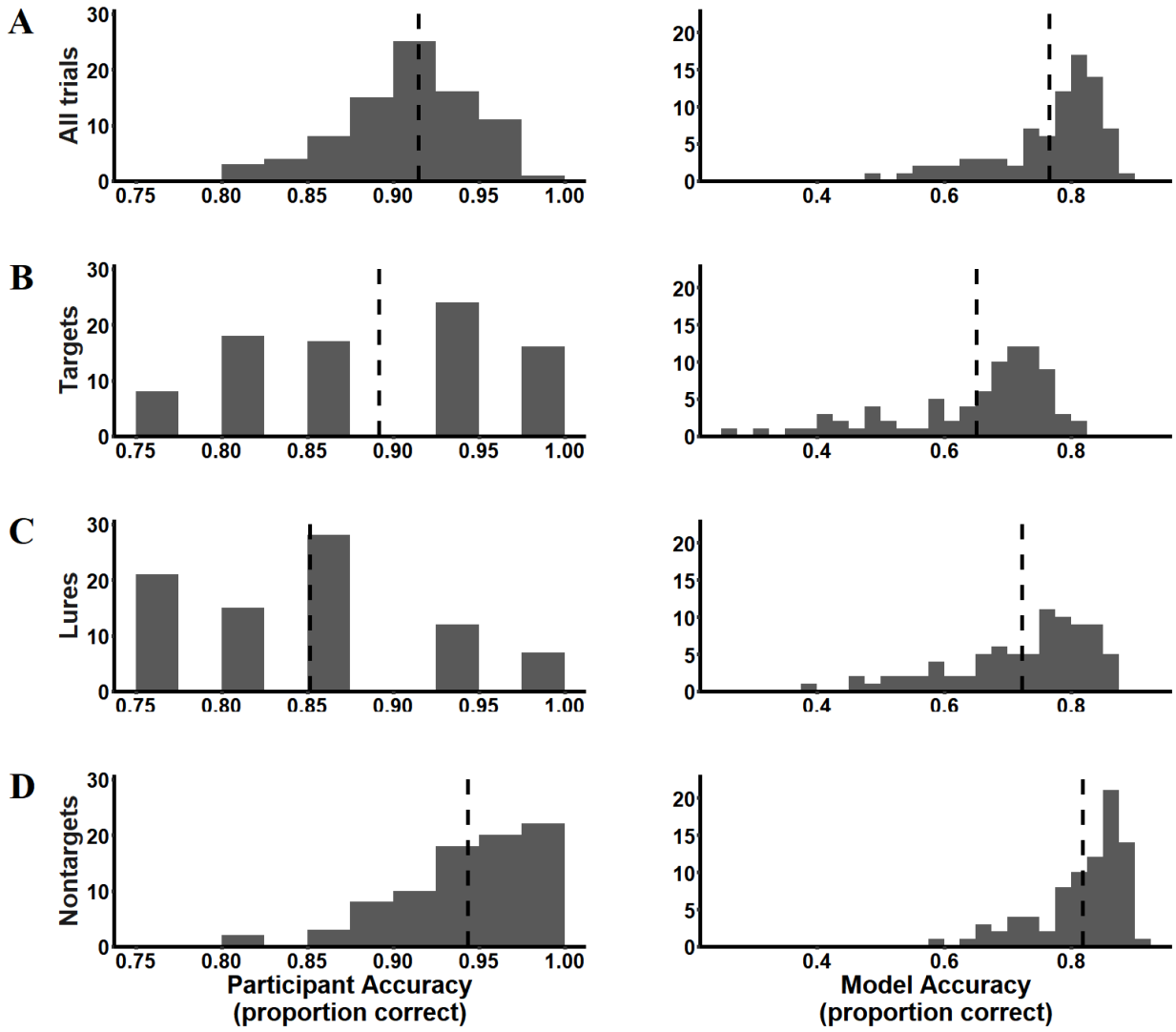

Figure 22. Histograms of two-back participant and N-back model-predicted accuracies. Participants include only those for whom N-back model parameter estimates were produced. Model-predicted accuracies were produced by running the model 100 times with relevant parameters set to each participant's estimated values, then taking the average of the accuracies of those runs. Plotted on the y-axis is the number of participants or model predictions with a certain response time. Vertical dashed bars denote the mean of the sample.

Table 10. Mean and standard deviations of the N-back model's predicted response times and accuracies, as well as the within-participant difference between these predicted measures and the participant's equivalents, across all trials as well as for target, lure, and nontarget trials. Shown alongside the mean difference between model and participant measures is the results of a one-sample t-test of whether the row's mean of differences was significantly different than zero. Listed at the right end of the table are the mean and standard deviation across participants of the RMSE between the model's trial-by-trial predictions and the participant's observed trial-by-trial behavior in the target, lure, and nontarget conditions.

	Model		Model - Participant Measure				RMSE	
	M	*SD*	*M*	*SD*	$t(82)$	$p <$	*M*	*SD*
Response Times								
All trials	0.58	0.04	-0.43	0.11	-36.4	0.001		
Targets	0.70	0.05	-0.25	0.12	-18.7	0.001	0.39	0.11
Lures	0.61	0.05	-0.48	0.13	-33.6	0.001	0.56	0.13
Nontargets	0.53	0.03	-0.47	0.12	-36.3	0.001	0.58	0.11
Accuracies								
All trials	0.77	0.09	-0.14	0.08	-16.2	0.001		
Targets	0.65	0.12	-0.24	0.13	-16.9	0.001	0.42	0.10
Lures	0.72	0.11	-0.13	0.12	-9.54	0.001	0.40	0.09
Nontargets	0.82	0.07	-0.13	0.07	-16.1	0.001	0.33	0.08

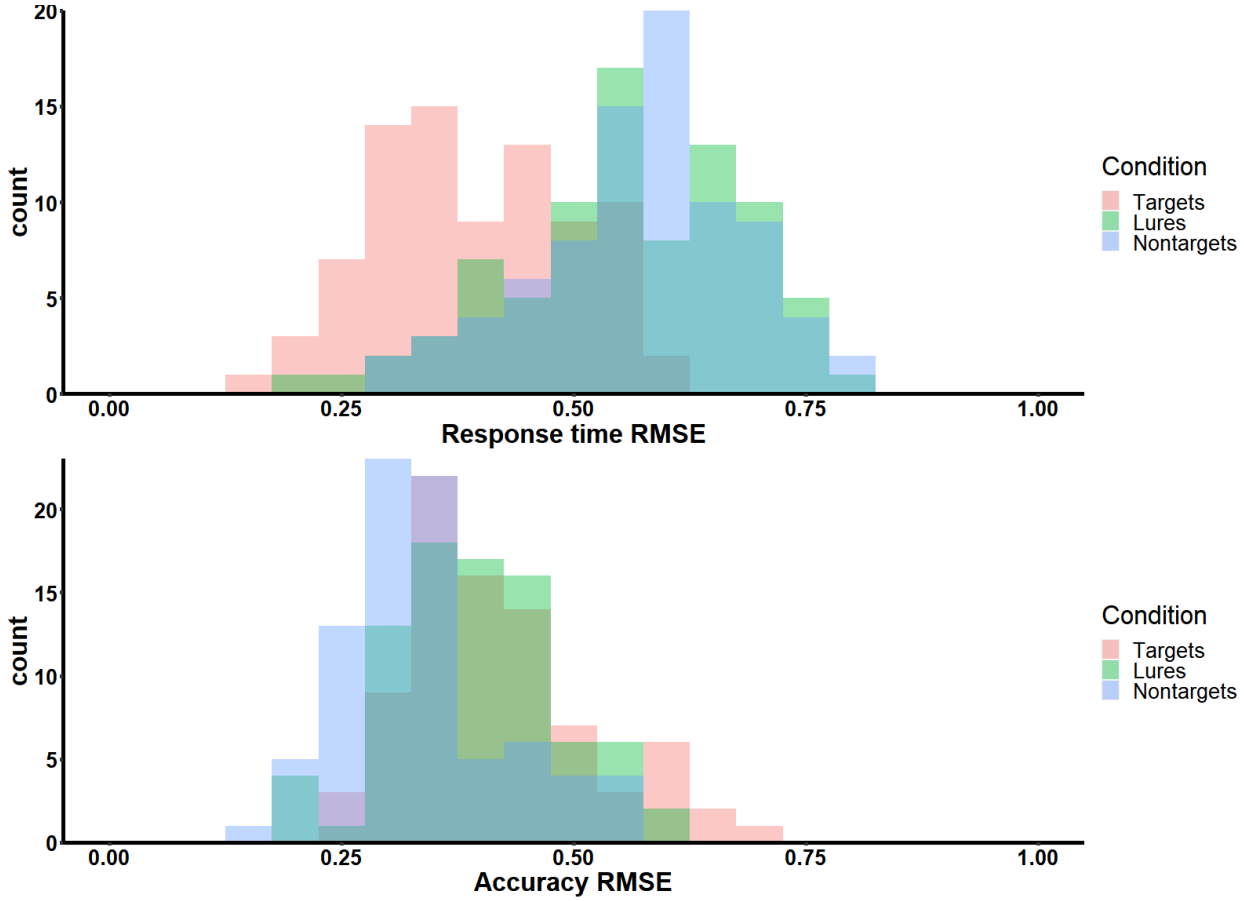

Figure 23. Histograms of RMSE values between the participant's measures of target, lure, and nontarget response times (top) and accuracies (bottom) in the two-back condition and the N-back model's predictions of the same. Plotted on the y-axis is the number of participants for whom the fitted model predictions produced a certain RMSE value.

Table 11. Correlations between the N-back model's participant-specific-parameterized predictions of mean response times and accuracies and the observed participant measures of the same within the listed sub-conditions. Shown is the Pearson's *r* of the relationship, alongside the significance *p* of the relationship.

	Response Time		Accuracy	
	r	$p <$	r	$p <$
All trials	0.53	0.001	0.32	0.01
Targets	0.43	0.001	0.25	0.05
Lures	0.52	0.001	0.17	$= 0.13$
Nontargets	0.48	0.001	0.25	0.05

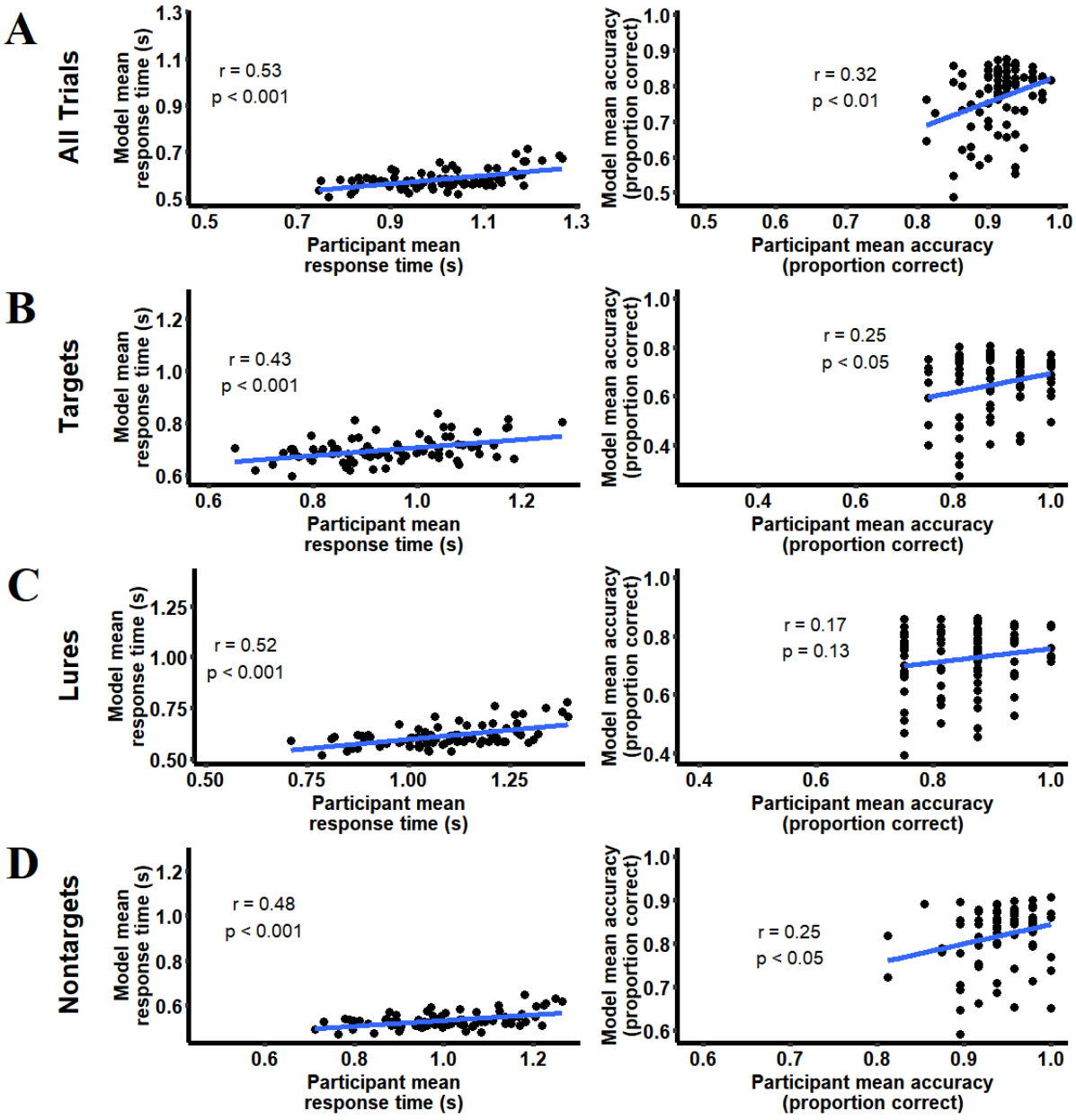

Figure 24. Scatterplots of the relationship between two-back participant mean response times/accuracies and N-back model-predicted mean response times/accuracies A. across all trials; B. across target trials; C. across lure trials; and D. across nontarget trials. Annotated in each plot is the Pearson's *r* value of the relationship and the significance of the correlation. Embedded in each plot as a blue line is the linear model fit of the model predictions to the participant data.

3.3.4 Comparison to zero-back parameter estimates

As four of the parameters estimated for individual participants under the N-back model were shared by the zero-back model (the W_g, W_i, d, and F parameters), these estimates were compared between the two models within participants. Only participants for whom parameters had been estimated under both the zero-back model and the N-back model were considered in this analysis, resulting in the inclusion of a total of 78 participants. For each participant and each of the four parameters, the N-back model's estimate of the parameter value was subtracted from the zero-back model's estimate, and then the distribution of these differences across participants was inspected. The distribution of differences in parameter estimates between the zero-back and N-back models for each of the four parameters can be seen in Figure 25. It was determined that the mean of the distribution of differences for both the W_g and W_i parameters was not significantly different than zero (W_g mean of differences = 0.02, $t(77) = 0.38$, $p = 0.71$; W_i mean of differences = 0.04, $t(77) = 0.68$, $p = 0.5$), while the mean of the distribution of differences for the d and F parameters was significantly different than zero (d mean of differences = 0.03, $t(77) = 2.1$, $p < 0.05$; F mean of differences = 0.19, $t(77) = 2.6$, $p < 0.05$).

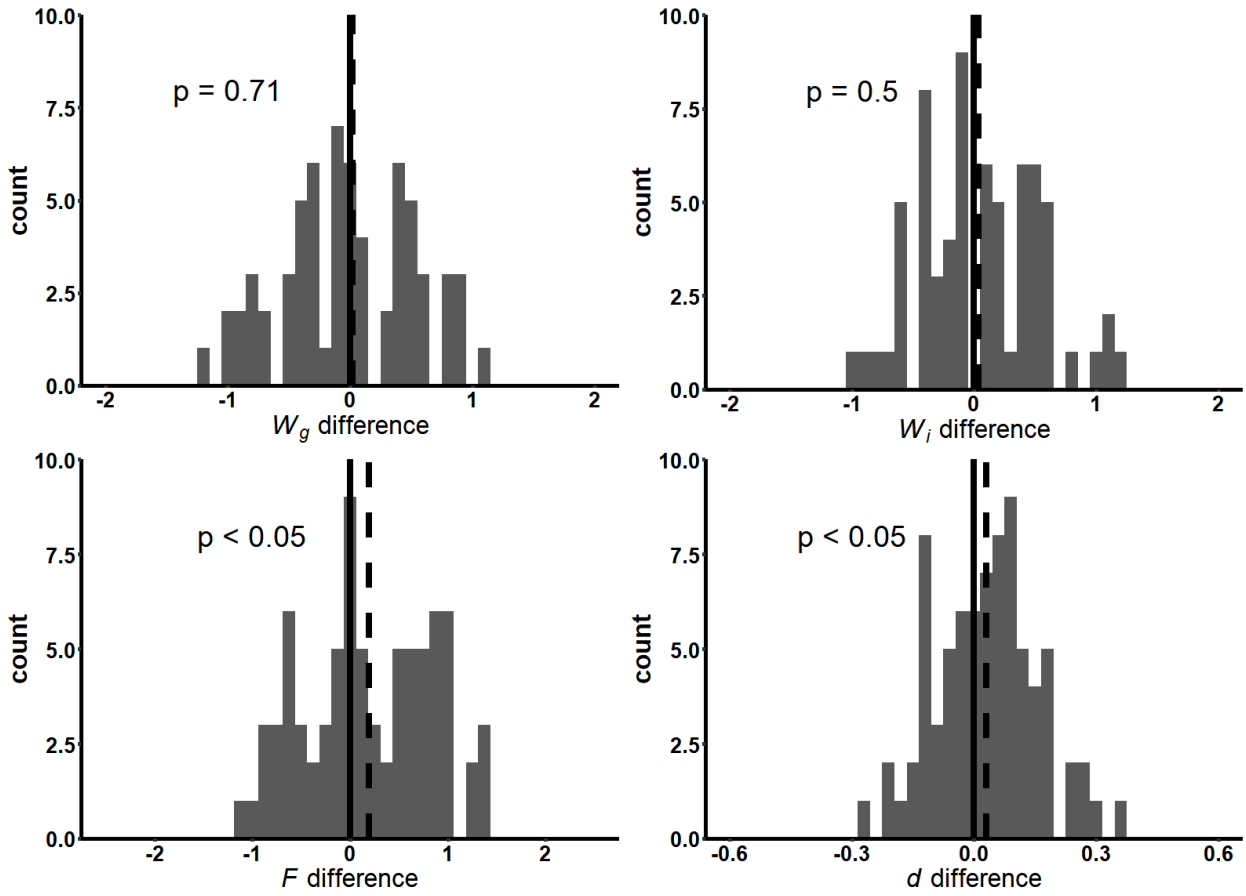

Figure 25. Participant-specific differences in parameter estimates between zero-back and N-back models. Plotted on the y-axis is the number of participants with a certain value of the parameter estimate difference. The title of the x-axis denotes the parameter for which estimates are displayed (W_g: goal buffer spreading activation; W_i: imaginal buffer spreading activation; d: base-level decay rate; F: latency scale factor). Vertical dashed bars denote the mean of the estimates of the specified parameter.

3.3.5 Decremental leave-one-out procedure

The dLOO procedure was also applied to the parameterized predictions of the N-back model. In the first iteration of the procedure, the model that included the participant-specific W_i, d, and F parameters, with W_g set to the mean of the parameter estimates ("left out"), best fit the participant response times and accuracies on average (response time $R^2 = 0.14$, accuracy $R^2 = 0.13$). In the second iteration, the model that utilized the participant-specific d and F parameters (with W_g and W_i "left out") won out (response time $R^2 = 0.05$, accuracy $R^2 = 0.12$). In the final iteration, the model that set only the d parameter to the individual participant's estimates was successful at predicting response times (Pearson's $r = 0.35$, $p < 0.001$) but not accuracies (Pearson's $r = 0.08$, $p = 0.47$), while the opposite was true of the model that set only the F parameter - it successfully predicted accuracies (Pearson's $r = 0.36$, $p < 0.001$) but not response times (Pearson's $r = 0.05$, $p = 0.68$). Histograms displaying the relationship between model predictions and participant responses for all models in all iterations of the dLOO procedure can be seen in Figure 26 (response times) and Figure 27 (accuracies).

3.3.6 Identification of predictive resting-state functional connectivity and prediction of individual-specific parameter values

Identification of predictive resting-state functional connections and subsequent prediction of individual-specific parameter values will be reported here for the N-back model parameters d and F, as the dLOO procedure revealed that these two parameters were critical for the N-back model's ability to predict participant-specific response times and accuracies, respectively.

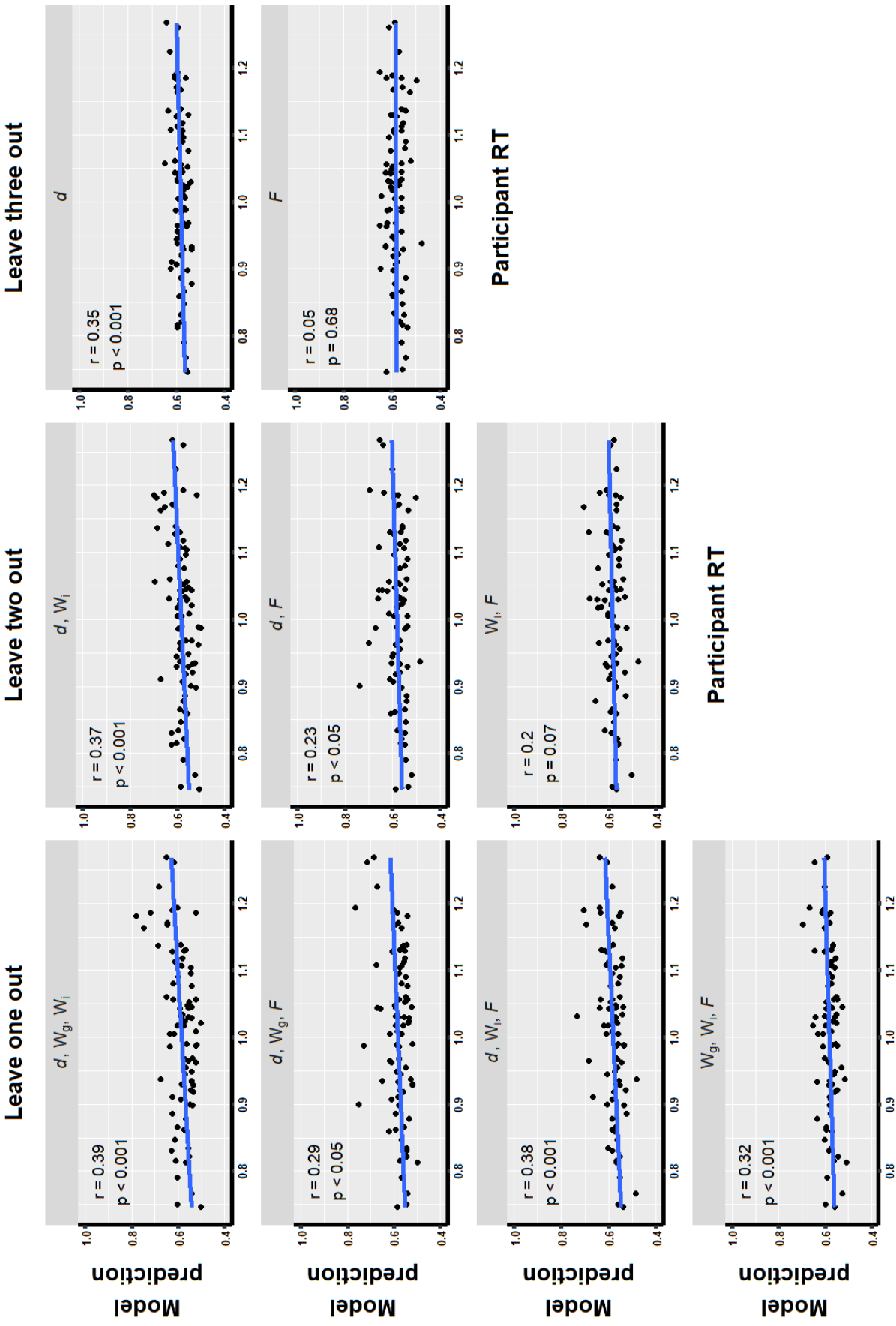
Leave one out
Leave two out
Leave three out
d, Wg, Wi
d, Wi
d
r = 0.39
p < 0.001
r = 0.37
p < 0.001
r = 0.35
p < 0.001
d, Wg, F
d, F
F
r = 0.29
p < 0.05
r = 0.23
p < 0.05
r = 0.05
p = 0.68
d, Wi, F
Wi, F
r = 0.38
p < 0.001
r = 0.2
p = 0.07
Wg, Wi, F
r = 0.32
p < 0.001
Model prediction
Participant RT
Participant RT
Participant RT

Figure 26. Scatterplots of the relationship between two-back participant mean response times and N-back model-predicted mean response times for each model considered in each iteration of the dLOO procedure (columns). The parameters for which the model used the participant-specific estimates are labeled at the top of each subplot (W_g: goal buffer spreading activation; W_i: imaginal buffer spreading activation; d: base-level decay rate; F: latency scale factor). Annotated in each plot is the Pearson's r value of the relationship and the significance of the correlation. Embedded in each plot as a blue line is the linear model fit of the model predictions to the participant data.

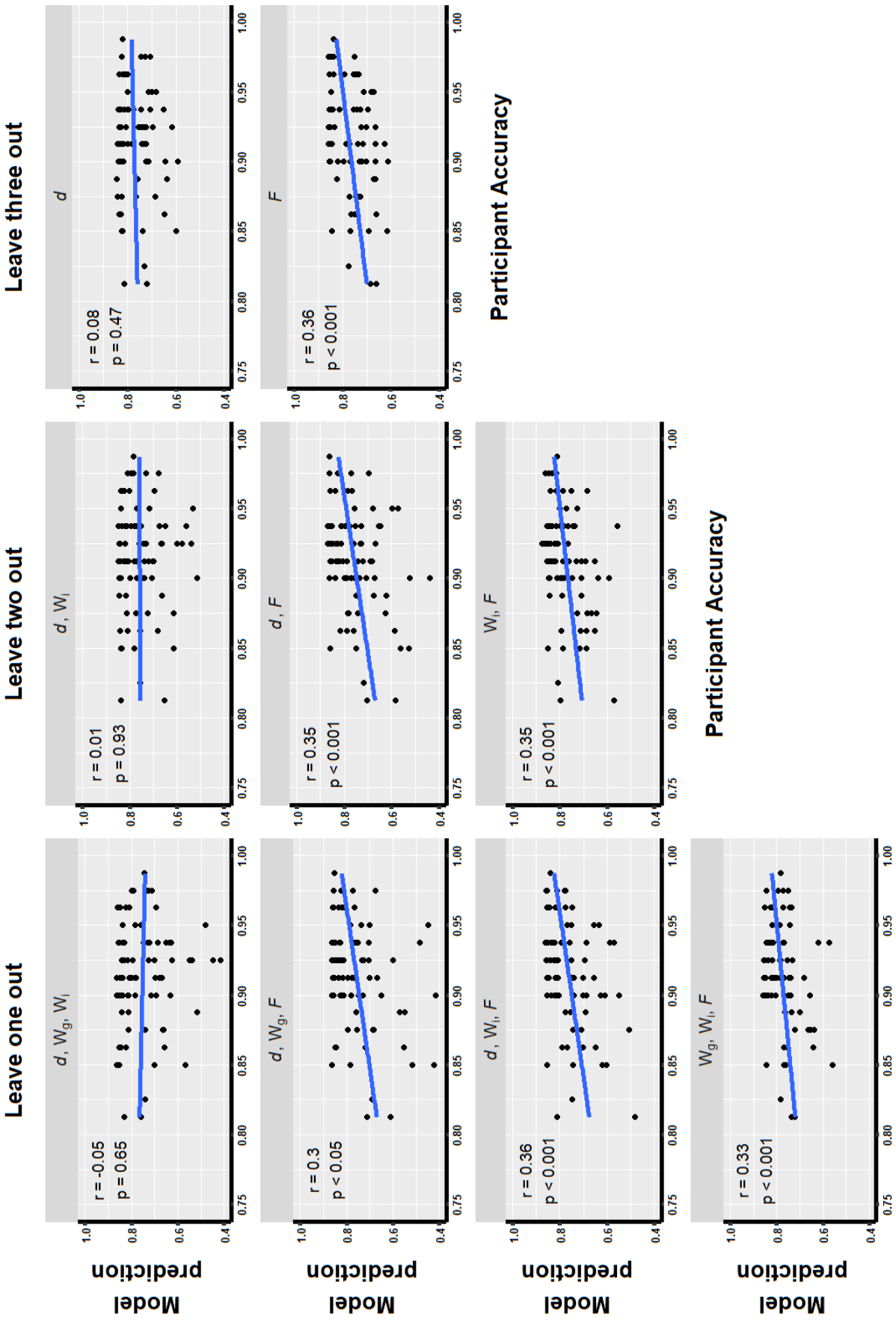
Leave one out
Leave two out
Leave three out
d, Wg, Wi
d, Wi
d
d, Wg, F
d, F
F
Wi, F
Wg, Wi, F
r = -0.05
p = 0.65
r = 0.01
p = 0.93
r = 0.08
p = 0.47
r = 0.3
p < 0.05
r = 0.35
p < 0.001
r = 0.36
p < 0.001
r = 0.36
p < 0.001
r = 0.35
p < 0.001
r = 0.33
p < 0.001
Model prediction
Participant Accuracy
Participant Accuracy

Figure 27. Scatterplots of the relationship between two-back participant mean accuracies and N-back model-predicted mean accuracies for each model considered in each iteration of the dLOO procedure (columns). The parameters for which the model used the participant-specific estimates are labeled at the top of each subplot (W_g: goal buffer spreading activation; W_i: imaginal buffer spreading activation; d: base-level decay rate; F: latency scale factor). Annotated in each plot is the Pearson's r value of the relationship and the significance of the correlation. Embedded in each plot as a blue line is the linear model fit of the model predictions to the participant data.

For the decay rate parameter *d*, the Lasso regression cross-validation procedure determined that the optimal λ value was 0.027 (λ = 0.027), resulting in the inclusion of 19 out of a possible 34,716 regressors (0.0005% of possible predictors). Visualizations of the results of the cross-validation procedure can be seen in Figure 28. These 19 functional connections included a total of 35 brain regions from 10 different functional networks; the location of each brain region and the β values of the 10 functional connections with the largest absolute beta values of the set of identified functional connections are visualized in Figure 29, while Table 12 contains the β values associated with these 10 functional connections. The "importance" of each node in the set of 19 identified functional connections was calculated in the same manner as before; Figure 30 presents a visualization of the importance of each node included in the 10 functional connections that are presented in Figure 29/Table 12, while Table 13 presents importance values for each of these nodes.

The fitted β values were then used in conjunction with the participant-specific functional connectivity values from the 19 identified connections to predict individual values of *d*. These regression-based predictions were strongly correlated to the *d* values obtained by fitting the ACT-R N-back model to the participant's behavior in the two-back condition (Pearson's $r = 0.87$, $p < 0.001$). Figure 31 shows the scatterplot of Lasso-predicted *d* values against ACT-R model fitted *d* values.

For the latency factor parameter *F*, the Lasso regression cross-validation procedure determined that the optimal λ value was 0.15 (λ = 0.15), resulting in the inclusion of 11 out of a possible 34,716 regressors (0.0003% of possible predictors). Visualizations of the results of the cross-validation procedure can be seen in Figure 32. These 11 functional connections included a total of 22 brain regions from 9 different functional networks; the location of each brain region

and the β values of the 10 functional connections with the largest absolute beta values of the set of identified functional connections are visualized in Figure 33, while Table 14 contains the β values associated with these 10 functional connections. Figure 34 presents a visualization of the importance of each node included in the 10 functional connections that are presented in Figure 33/Table 14, while Table 15 presents importance values for each of these nodes.

As was done for the decay rate parameter *d*, individual values of *F* were predicted using the β values fitted through Lasso regression and the participant-specific functional connectivity values from the 11 connections identified by the fitting process. The regression-predicted *F* values were again strongly correlated to the *F* values obtained by fitting the ACT-R N-back model to the participant's behavior (Pearson's $r = 0.82$, $p < 0.001$). Figure 35 shows the scatterplot of Lasso-predicted *F* values against ACT-R model fitted *F* values.

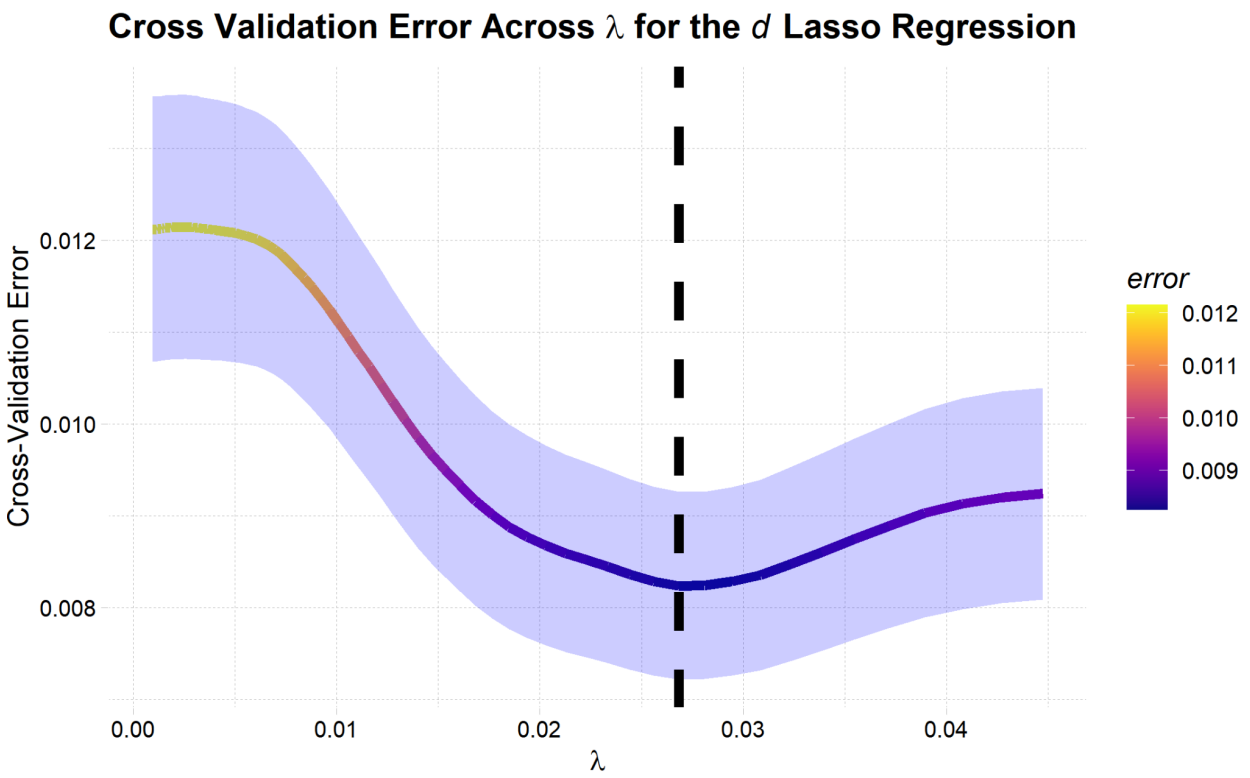

Figure 28. Results of the cross-validation procedure for the Lasso regression of the participant-specific functional connectome on the N-back model estimated *d* parameter. Plotted on the y-axis is the cross-validation error of the regression model's prediction of participant-specific *d* at a given value of λ, which is also mapped to the color of the plotted line. Shaded blue region represents the standard deviation of the regression model's cross-validation error across held-out participants, while the vertical dashed black line denotes the value of λ that produced the minimum cross-validation error.

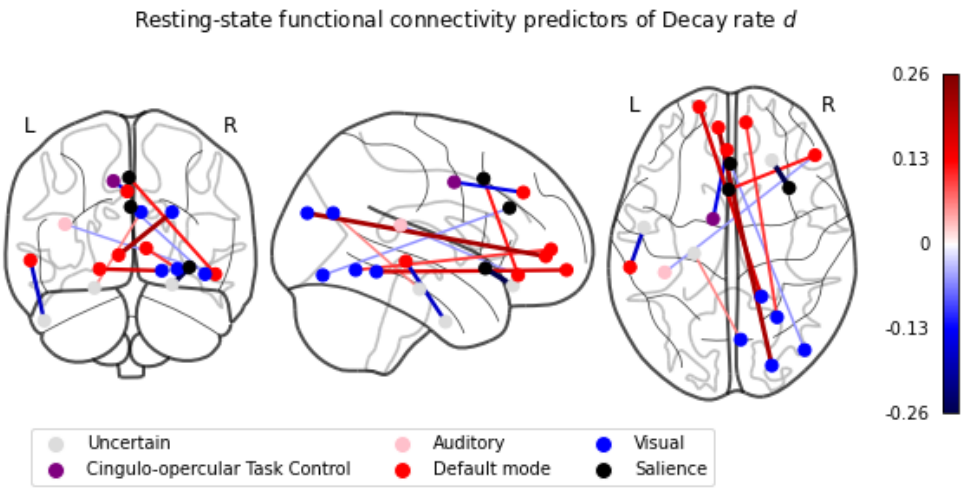

Figure 29. Functional connectome nodes and connections identified to be significantly predictive of participant-specific N-back model estimates of the decay rate parameter *d*. Shown here are the 10 functional connections with the largest absolute beta values of the set of identified functional connections (19 total). Nodes are color-coded by their assigned network according to (Power et al., 2011). Thickness and color of bars between nodes denotes the estimated β value for that node-pair's functional connectivity value.

Table 12. The 10 largest (absolute value) Lasso regression-estimated β values for the prediction of the N-back model *d* parameter. Listed are the networks of each of the two nodes in the identified connection, the specific node numbers of the paired nodes (in the Power (2011) parcellation), and the Lasso regression-estimated β value for the connection.

Connected Networks	Specific Nodes	Beta Value
Uncertain-Salience	3-211	-0.262971
Default mode-Visual	107-162	0.222448
Uncertain-Default mode	118-250	-0.172962
Default mode-Visual	78-143	0.172294
Cingulo-opercular Task Control-Default mode	51-112	-0.147302
Default mode-Salience	139-213	0.123106
Default mode-Visual	108-150	0.103079
Uncertain-Visual	6-163	0.062212
Visual-Salience	153-215	-0.056924
Auditory-Default mode	64-139	-0.051601

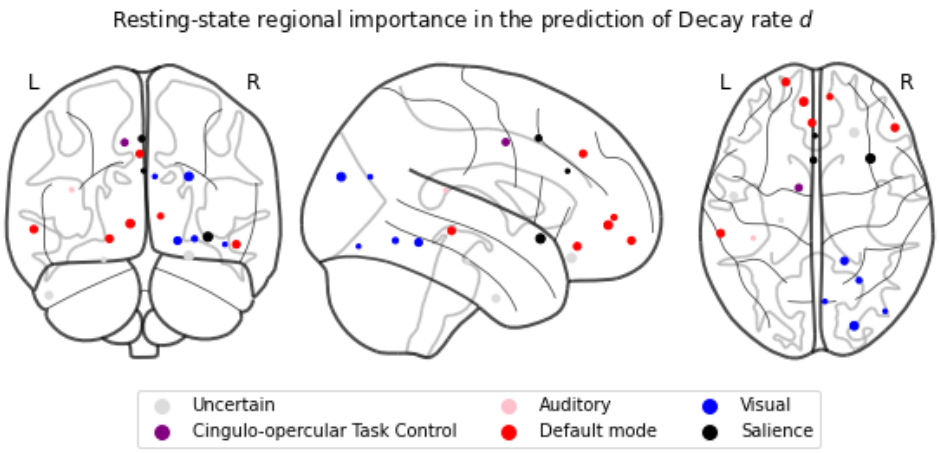

Figure 30. Importance of functional connectome nodes that participate in connections identified to be significantly predictive of participant-specific N-back model estimates of the decay rate parameter *d*. The size of each node is scaled by the node's importance value. Shown here are the nodes included in the 10 functional connections with the largest absolute beta values of the set of identified functional connections (19 total). Nodes are color-coded by their assigned network according to (Power et al., 2011).

Table 13. The importance value (sum of the absolute values of each β associated with the identified functional connections that a node participates in) for each node included in the set of functional connections with the 10 largest β values for the Lasso regression prediction of the N-back model goal buffer spreading activation parameter *d*. Listed is the specific node number in the Power (2011) parcellation, the network that the node is a member of, the X, Y, and Z coordinates of the node in MNI space, and the computed importance value of the node.

ROI	Network	X	Y	Z	Importance Value
3	Uncertain	24	32	-18	0.262971
211	Salience	34	16	-8	0.262971
107	Default mode	-7	51	-1	0.222448
162	Visual	24	-87	24	0.222448
139	Default mode	49	35	-12	0.174707
118	Default mode	-58	-30	-4	0.172962
250	Uncertain	-50	-7	-39	0.172962
78	Default mode	-18	63	-9	0.172294
143	Visual	18	-47	-10	0.172294
112	Default mode	-2	38	36	0.159889
51	Cingulo-opercular Task Control	-10	-2	42	0.147302
213	Salience	-1	15	44	0.123106
108	Default mode	9	54	3	0.103079
150	Visual	27	-59	-9	0.103079
6	Uncertain	-21	-22	-20	0.062212
163	Visual	6	-72	24	0.062212
153	Visual	43	-78	-12	0.056924
215	Salience	0	30	27	0.056924
64	Auditory	-38	-33	17	0.051601

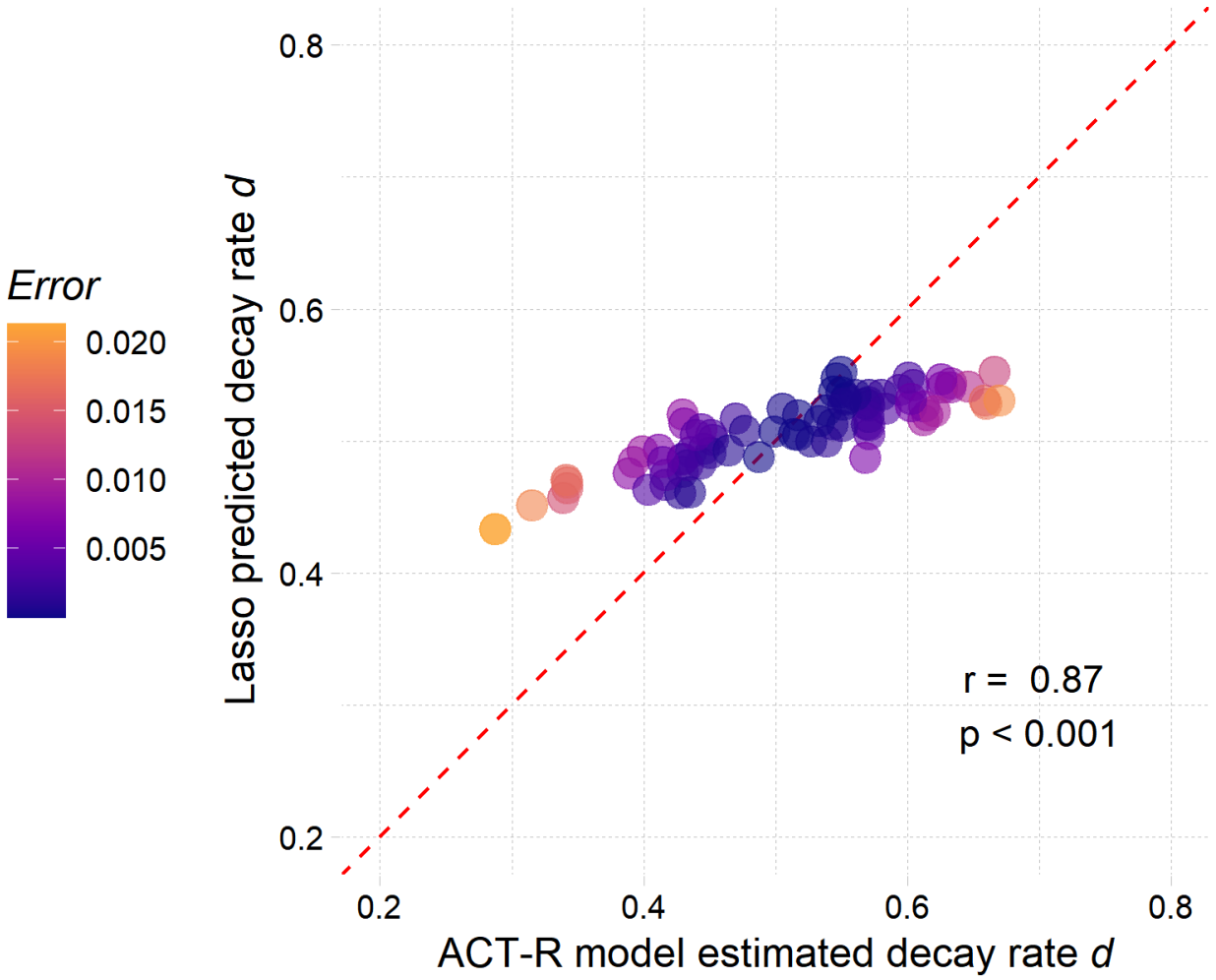

Figure 31. Scatterplot of Lasso-predicted *d* values against N-back ACT-R model fitted *d* values. Points are colored according to the squared error between Lasso-predicted values and model-estimated values. Dashed red line denotes the identity line. The Pearson's *r* value of the relationship and the significance of the correlation is annotated in the bottom right.

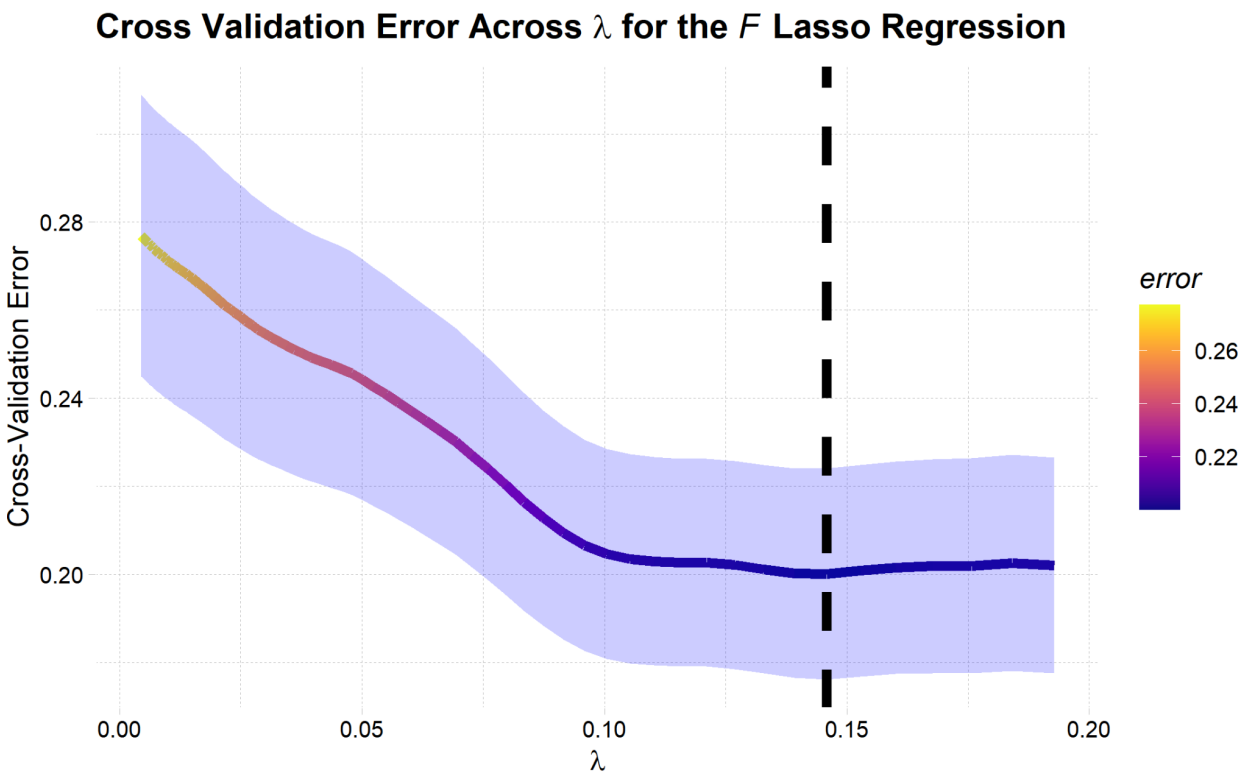

Figure 32. Results of the cross-validation procedure for the Lasso regression of the participant-specific functional connectome on the N-back model estimated *F* parameter. Plotted on the y-axis is the cross-validation error of the regression model's prediction of participant-specific *F* at a given value of λ, which is also mapped to the color of the plotted line. Shaded blue region represents the standard deviation of the regression model's cross-validation error across held-out participants, while the vertical dashed black line denotes the value of λ that produced the minimum cross-validation error.

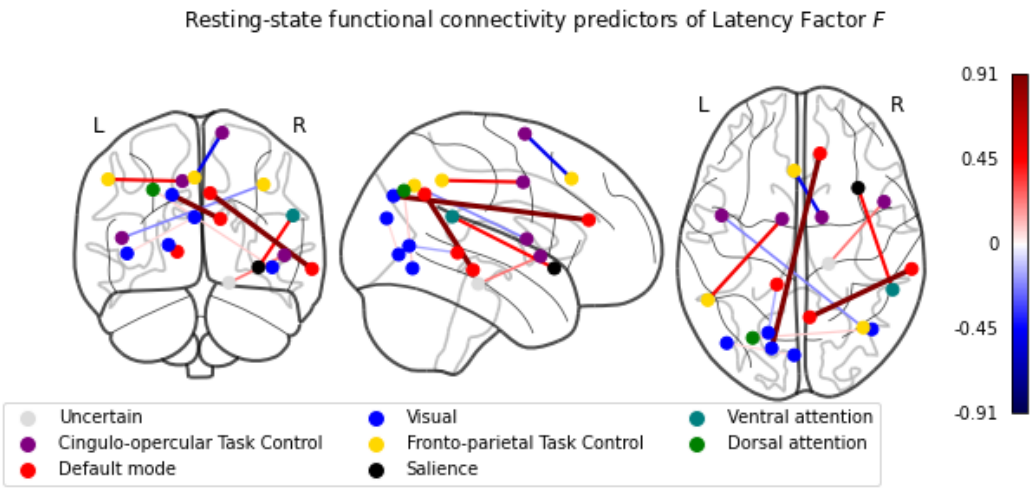

Figure 33. Functional connectome nodes and connections identified to be significantly predictive of participant-specific N-back model estimates of the decay rate parameter *F*. Shown here are the 10 functional connections with the largest absolute beta values of the set of identified functional connections (11 total). Nodes are color-coded by their assigned network according to (Power et al., 2011). Thickness and color of bars between nodes denotes the estimated β value for that node-pair's functional connectivity value.

Table 14. The 10 largest (absolute value) Lasso regression-estimated β values for the prediction of the N-back model *F* parameter. Listed are the networks of each of the two nodes in the identified connection, the specific node numbers of the paired nodes (in the Power (2011) parcellation), and the Lasso regression-estimated β value for the connection.

Connected Networks	Specific Nodes	Beta Value
Default mode-Visual	122-166	0.906182
Default mode-Default mode	89-119	0.898376
Cingulo-opercular Task Control-Fronto-parietal...	53-202	-0.488986
Cingulo-opercular Task Control-Fronto-parietal...	51-177	0.485709
Salience-Ventral attention	211-235	0.449267
Uncertain-Cingulo-opercular Task Control	7-56	0.246135
Cingulo-opercular Task Control-Fronto-parietal...	55-194	-0.217990
Default mode-Visual	77-152	-0.153264
Visual-Dorsal attention	161-260	0.089319
Visual-Visual	164-167	0.069286

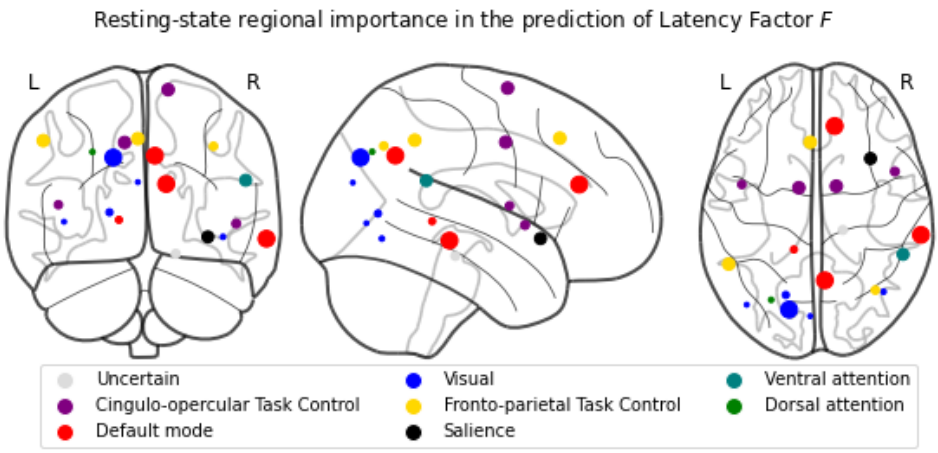

Figure 34. Importance of functional connectome nodes that participate in connections identified to be significantly predictive of participant-specific N-back model estimates of the decay rate parameter *F*. The size of each node is scaled by the node's importance value. Shown here are the nodes included in the 10 functional connections with the largest absolute beta values of the set of identified functional connections (11 total). Nodes are color-coded by their assigned network according to (Power et al., 2011).

Table 15. The importance value (sum of the absolute values of each β associated with the identified functional connections that a node participates in) for each node included in the set of functional connections with the 10 largest β values for the Lasso regression prediction of the N-back model goal buffer spreading activation parameter *F*. Listed is the specific node number in the Power (2011) parcellation, the network that the node is a member of, the X, Y, and Z coordinates of the node in MNI space, and the computed importance value of the node.

ROI	Network	X	Y	Z	Importance Value
122	Default mode	12	36	20	0.906182
166	Visual	-16	-77	34	0.906182
89	Default mode	6	-59	35	0.898376
119	Default mode	65	-31	-9	0.898376
202	Fronto-parietal Task Control	-3	26	44	0.488986
53	Cingulo-opercular Task Control	13	-1	70	0.488986
51	Cingulo-opercular Task Control	-10	-2	42	0.485709
177	Fronto-parietal Task Control	-53	-49	43	0.485709
235	Ventral attention	54	-43	22	0.449267
211	Salience	34	16	-8	0.449267
7	Uncertain	17	-28	-17	0.246135
56	Cingulo-opercular Task Control	49	8	-1	0.246135
55	Cingulo-opercular Task Control	-45	0	9	0.217990
194	Fronto-parietal Task Control	37	-65	40	0.217990
77	Default mode	-13	-40	1	0.153264
152	Visual	-18	-68	5	0.153264
161	Visual	42	-66	-8	0.089319
260	Dorsal attention	-27	-71	37	0.089319
167	Visual	-3	-81	21	0.069286
164	Visual	-42	-74	0	0.069286

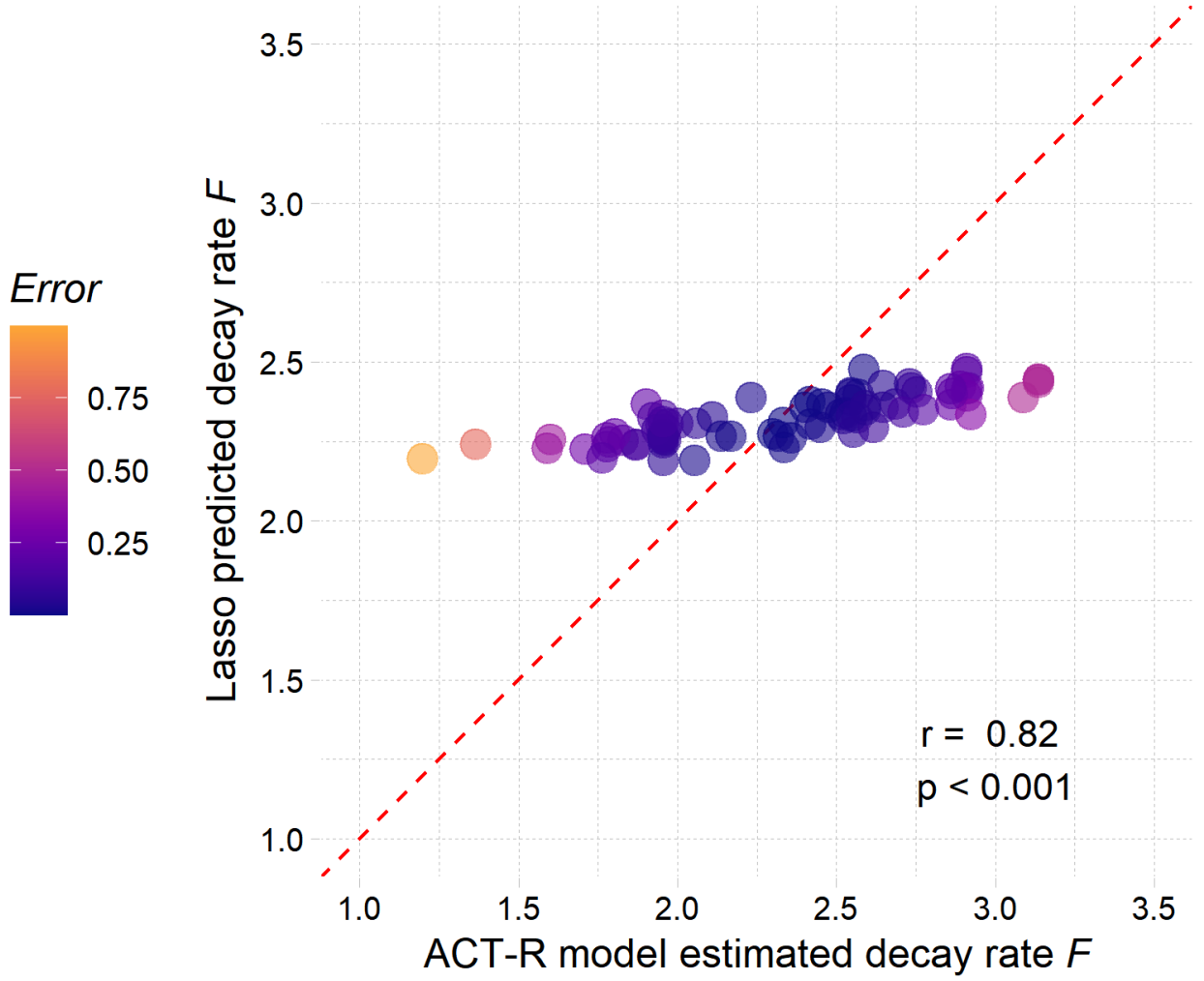

Figure 35. Scatterplot of Lasso-predicted *F* values against zero-back ACT-R model fitted *F* values. Points are colored according to the squared error between Lasso-predicted values and model-estimated values. Dashed red line denotes the identity line. The Pearson's *r* value of the relationship and the significance of the correlation is annotated in the bottom right.

Chapter 4. Discussion

The present work utilized a subset of the freely available Human Connectome Project Young Adult dataset (Van Essen et al., 2013) to investigate the relationship between model-based characterizations of behavior and resting-state functional connectivity in individuals. Specifically, the behavior of each individual in a group of participants under the zero-back and two-back condition of the HCP N-back task (a task that is thought to place demands on working memory ability (Jaeggi et al., 2010; Michael J. Kane et al., 2007; K. M. Miller et al., 2009)) was characterized through the fitting of two different models (one for the zero-back condition, and the other for the two-back condition) created using the ACT-R cognitive modeling architecture. Model fitting was performed by determining, for each participant, the values for a set of parameters under each model that resulted in the most accurate model prediction of the participant's response time and accuracy measures in the corresponding condition of the task. These parameters can be interpreted in terms of working memory constructs (such as attention and decay) and therefore the fit parameter values potentially reflect individual differences in a multicomponent view of working memory function. The predictive efficacy of these fit parameters was then examined by performing a unique "decremental leave-one-out" procedure, in which the weakest parameter was iteratively determined and removed from consideration until only those that adequately predicted individual differences in reaction time and accuracy remained for examination. This procedure identified the parameters that are most essential for the model's ability to predict individual differences in behavior. If these parameters are reflective of individual differences in working memory ability (observed through the model's characterization of differences in behavior), then these metrics of working memory function are

likely to be reflected in the individual's brain connectivity at a resting state, as has been previously demonstrated (Sala-Llonch et al., 2012). To determine whether this was the case, participant-specific resting state functional connectivity measures between a predefined parcellation of brain regions (Power et al., 2011) were regressed on the individual values of the identified essential parameters using the Lasso method (Tibshirani, 1996). This technique minimizes the number of regressors that are included in the predictive model - which, in this case, serves to identify connections between and within specific brain networks at rest that are most essential for prediction of the model parameters (and therefore, characterizations of working memory ability).

4.1 Correspondence of ACT-R Model Parameters to Working Memory Constructs

The five total ACT-R model parameters that were utilized between the two models to capture individual differences in behavior under the N-back task can be interpreted under common conceptualizations of the psychological constructs that compose working memory. This is important to examine as working memory is thought to be a multicomponent process (A. D. Baddeley & Hitch, 1974; Nelson Cowan, 1999), and so individuals may differ in the functioning of any number of these components, resulting in observable differences in behavior; application of the ACT-R model allows for the examination of how these working memory components contribute to these differences in behavior. Subsequently, these components can be related to individual differences in brain activity to more fully delineate the correspondences between neurophysiology, cognitive functioning, and behavior.

One parameter, the cue-stimulus similarity *c*, is only estimated for participants under the zero-back model. It is not included in the N-back model as there is no differentiation between stimuli presented as a cue and "normal" stimuli in conditions of the N-back task outside of

zero-back. In ACT-R, the similarity value contributes to the partial matching component of a chunk's activation value by "penalizing" a chunk when the value of a chunk's slot does not exactly match the value for that slot specified in an ACT-R retrieval request. The degree of this penalty is determined by the similarity between the value specified by the request and the existing value in the chunk's matching slot; the greater the similarity between representations, the smaller the penalty. This effect can be thought of as capturing the role of interference in working memory. While this may seem at odds with the standard conceptualization of interference as a form of destructive competition between shared features, the correspondence is clarified by the distinction that ACT-R makes between the slots of chunks, and the values associated with those slots. The slots themselves are the shared features that are being compared: between that of the retrieval request (which themselves are informed by the model's current representation of information throughout it's buffers) and potential chunk matches in memory. Shared features that possess common values have the potential to cause interference through the retrieval of an incorrect chunk - because the "incorrect" chunk is not penalized to the same degree as chunks that share features with diverging values, it is not as "distinct" in memory from the "correct" chunk as these other chunks, and is more likely to be retrieved. It is in this manner that the cue-stimulus similarity parameter *c* has the potential to capture interference between the representation of a block's cue and representations of previously-seen stimuli; a stimulus representation may be "mistaken" for a cue representation, interfering with the representation of the intended item.

Of the five parameters that were estimated under both the zero-back and N-back models, two should be discussed in conjunction - the goal and imaginal buffer spreading activation components, W_g and W_i. They are estimated for each participant for both the zero-back and

N-back models, and in ACT-R terms, these parameters define the amount of spreading activation that chunks in ACT-R's declarative memory module receive from chunks represented in the goal and imaginal buffers. A declarative memory chunk that shares a greater number of slot values with the chunks in the goal/imaginal buffers will receive more spreading activation than a chunk that shares a fewer amount of slot values. This spreading activation increases the chunk's overall activation and makes it more likely to be retrieved as a result. This effect can be interpreted in working memory terms as attention sourced from goal-directed representation and representation of current information for the ACT-R goal and imaginal buffers, respectively. From this perspective, the ACT-R concept of spreading activation is similar to Cowan's role of attention in the Embedded-Processes working memory model (Nelson Cowan, 1999) - attention, which can be sourced from multiple systems, serves to increase the availability of a subset of representations from a set that has already become partially active through other means. This attention is directed by the representation of the source - the goal buffer spreads activation to features that are shared by it's current chunk representation, and the imaginal buffer does the same. This role of attention is also reflected in the Baddeley and Hitch multicomponent model (A. D. Baddeley & Hitch, 1974). Baddeley and Hitch conceptualize the multicomponent model's central executive as an agglomeration of multiple executive functions that provide the capability to divide the focus of attention across multiple representations (the concepts of "control" and "scope" of attention), with the goal of performing some form of processing with the focused representations. While ACT-R does not specify a particular "central executive" module, the architecture captures the role of Baddeley and Hitch's central executive through the spreading activation mechanism and production system. From this perspective, "control" of attention in ACT-R is sourced from the active representation in the buffers that provide spreading activation

to chunks in declarative memory alongside the productions that govern the dynamically changing contents of these buffers. “Scope” of attention, meanwhile, is modeled by the division of a buffer’s spreading activation over the number of slots possessed by the buffer’s currently represented chunk. The separate contributions of the goal and imaginal buffers can be considered as state-based information that the central executive refers to in order to allocate attention effectively. In this manner, the spreading activation mechanism of ACT-R (in the case of the zero- and N-back models, the goal and imaginal buffer spreading activation values W_g and W_i) captures a role of attention in working memory function that is common across multiple models of working memory.

The base-level activation decay parameter *d* is also estimated for each participant for both the zero-back and N-back models, and in the ACT-R architecture, this parameter dictates the decrease in the base-level activation of a chunk that occurs as time passes. When ACT-R’s optimized learning equation is used, as was the case in this work, the decay parameter *d* has an effect on two components of the base-level activation: the increase in activation sourced in increasing familiarity with a chunk (captured through the number of experiences of that chunk), and the decrease in activation sourced in the passage of time. In this schema, a smaller decay parameter value results in a smaller contribution of repeated experience to a chunk’s activation, but a slower decay of this activation over time. Larger decay parameter values result in the opposite - a large contribution of repeated experience to a chunk’s activation, but quick decay of this activation over time. In working memory terms, the base-level decay parameter *d* has a fairly equivalent interpretation - it is the time-related decay in the availability, or fidelity, of a given representation. While the role of decay in working memory representation is still under debate in the literature (Lewandowsky & Oberauer, 2009; Portrat et al., 2008), models of working memory

(such as the multicomponent or Embedded-Processes models) commonly consider the effect of decay on working memory representation. Just as ACT-R balances the time-related decay of activation by considering the "familiarity", or number of experiences, with a chunk of information, the models that include decay commonly address the effect through a mechanism such as rehearsal, which serves to increase the number of experiences with the decaying information. Rehearsal can maintain the activation of a chunk by offsetting the decay associated with the time over which the rehearsal occurs. While a rehearsal mechanism itself is not inherent to the ACT-R architecture, the contribution of the number of experiences with a chunk to the base-level activation, and the interaction of this term with the decay rate, enables ACT-R to be sensitive to the *act* of rehearsal. The N-back model presented in this work implements updating and rehearsal mechanisms that capitalize on this capability - the manner in which the N-back model updates it's representation of the last N+1 stimuli results in increased activation of these chunks, which in turn would support accurate rehearsal of the newly-updated information. Subsequent rehearsal then reinforces these active representations and serves to maintain their availability until the task next demands their use.

The latency factor scale parameter *F* is the last parameter that is estimated for each participant under both the zero-back and N-back models. In ACT-R, this parameter scales the amount of time necessary to either retrieve a chunk from the declarative memory module or determine that the retrieval attempt has failed. The former duration is also reliant on the activation value of the chunk that is retrieved, while the latter is instead dependent on the activation threshold value, which determines which chunks are available for retrieval (i.e., any chunk with an activation value above the threshold). The latency factor scale parameter *F* therefore reflects the processing speed of performing a particular operation in working memory -

attempted retrieval/encoding of new information. Successful retrievals are sped by increased activation in the target of retrieval - the more "available" a chunk is, the more easily it is retrieved and encoded for use in current behavior. On the other hand, unsuccessful attempts at retrieval need to have some form of "stopping criterion", otherwise the attempt would last indefinitely until some information eventually crosses the threshold of activation - in a naturalistic setting, this persistence could potentially be less advantageous than "giving up" and allowing further experience to modulate activation values and availability of information.

As an inherently multicomponent architecture in itself, ACT-R is well-suited to characterize the multiple components thought to contribute to working memory. The five total parameters estimated between the two ACT-R models each have reasonable interpretations under working memory theory, and are therefore viable in capturing individual differences in working memory ability sourced in differences in these facets of WM. These parameters influence the behavior of the ACT-R models in a synergistic fashion, with each contributing in parallel to the model's ability to perform the task. Particularly, these five parameters represent some of the critical elements of working memory, including attention, interference, decay, and processing speed. While these correspondences have been defined between the working memory constructs and ACT-R elements that have the most in common, it should be stressed that they operate in a synergistic fashion. For example, attention (through the influence of spreading activation) increases the activation of potential targets of retrieval, but the sources of this spreading activation also contribute to the defined features of the retrieval request, presenting the potential for interference. The combined influence of attention and interference modulates both the likelihood of the retrieval of a particular representation as well as the speed at which it is retrieved; whether a particular representation is retrieved subsequently impacts the influence of

decay on that representation's activation. In this manner, ACT-R serves well in reflecting the integrated, multicomponent nature of working memory function.

4.2 Plausibility of Implemented ACT-R Models

While the ACT-R model parameters considered by this work feasibly capture a number of elements of working memory theory, it is also important to consider whether the ACT-R models themselves represent plausible cognitive strategies for completing a task that demands working memory function. This is because each ACT-R model presented here implements a singular strategy for completing the task; there is no guarantee that the strategy implemented in the model is the only strategy to complete the task, or the implemented strategy is used by an individual participant. However, the goal of this study was not to differentiate between different strategies in working memory function (which could be captured by the implementation of multiple models, and the subsequent determination of the best-fitting model for each participant), but instead to capture differences in the contribution of elements of working memory to behavior. If the strategy of the model is specific enough to perform the task accurately, yet general enough to capture overarching patterns of behavior between individuals, then differences in the parameter values between participants (through the fit of the model to different patterns of behavior) should reflect differences in these working memory components.

As has been discussed in section 4.1, both the zero-back and two-back models naturally leverage a major defining feature of working memory theory - the contribution of multiple cognitive components which each participate in cognitive processes aside from working memory. Individual differences in working memory ability were characterized through task-general elements of ACT-R (the parameters described in section 4.1) that contribute to almost all ACT-R models, rather than through custom-designed elements specific to completion of the N-back task.

As the ACT-R cognitive architecture is successful in modeling a wide array of cognitive functions, these task-general elements are known to play vital roles in ACT-R's ability to capture human behavior.

The utilization of a rehearsal mechanism in each model's strategy is another example of the application of a common aspect in working memory theory. Rehearsal as a form of maintenance was utilized by Ebbinghaus, is explicitly considered by Baddeley and Hitch's multicomponent model, and is commonly considered to be a viable strategy for the maintenance of information over short periods of time (A. D. Baddeley & Hitch, 1974; Ebbinghaus, 1913; Oberauer, 2019). Additionally, rehearsal begins to be used as a maintenance strategy in childhood (Tam et al., 2010), and this use continues into adulthood (Kaushanskaya & Yoo, 2011). The N-back model rehearses a "window" of recently presented items explicitly; when a response has been made on the current trial and the representation of the last N+1 presented stimuli has been updated (through a mechanism that is very similar to that of rehearsal), the N-back model will begin rehearsing the memories of the stimuli included in the updated "window" by retrieving the chunk representation of each stimulus in a serial manner. An assumption was made that rehearsal of successfully-updated information is "easy" by disabling ACT-R's partial matching mechanism for the rehearsal retrievals; the result is that the model is guaranteed to be successful in retrieving the intended chunk from the declarative memory module. The zero-back model does not rehearse as a standalone action, as the N-back model does; however, it does "rehearse" the relevant information (the identity of the cue stimulus) on each trial by retrieving this information to be used as a basis of response. In both cases, the model's strategy serves to enhance the activation value of the relevant chunks through repeated experience, effectively rehearsing the information.

The final point of consideration regarding plausibility is specific to the N-back model. While participants of the Human Connectome Project only completed the zero-back and two-back conditions of the N-back task, other studies have additionally characterized behavior in the one-back and three-back conditions of the task (Jaeggi et al., 2010; K. M. Miller et al., 2009). In general, as the value of N increases, accuracy decreases, with a corresponding increase in response time. If the N-back model implements a working memory strategy that reasonably resembles that of human participants, then it should not only be capable of completing the one-back and three-back conditions, but also show similar behavioral dynamics across these three conditions. In fact, the model (which was developed using only the two-back condition as a reference) does exactly that: the final N-back model that was capable of performing the two-back condition at default parameter settings was exposed to one-back and three-back conditions, and was able to complete the task successfully with no further modifications. Moreso, the performance of the model across the three conditions mimics that of humans; it demonstrates best performance in the one-back condition, with accuracy on target, lure, and nontarget trials decreasing as the value of N increases. While this default-parameter-value model does not capture every pattern in participant group's behavior in the two-back condition (for example, the participants in this study were, on average, more accurate on target trials than lure trials, while the model shows the opposite), the model's average accuracy is almost identical to the participant's, and the target, lure, and nontarget accuracies are fairly commensurate. Additionally, the model's predicted accuracies across the N-back conditions closely align with those reported in previous research (Jaeggi et al., 2010; K. M. Miller et al., 2009). This ability to reasonably predict accuracy values and patterns of accuracy across conditions suggests that the N-back

model implements a reasonable strategy for representation, updating, and rehearsal of information necessary to complete the N-back task.

Despite this success, the model consistently underpredicts response times, and moreso, makes a flat prediction across the three conditions. This is due to the model's procedure for determining a response; when a new stimulus is presented, the model simply makes a retrieval attempt for a chunk that represents the stimulus associated with the "two" slot of the stimulus window chunk in the imaginal buffer (which the model has been updating and rehearsing as the task proceeds), and then responds on the basis of whether the retrieved chunk is a match to the stimulus the model is currently viewing. There is no temporal delay associated with the model determining what is in the "two" slot of the chunk in the imaginal buffer. Humans, however, may have to engage in a time-consuming process of scanning or evaluating the contents of information stored during WM performance in a serial manner (Kessler & Oberauer, 2015). If items in WM must be scanned through and addressed one at a time to determine temporal ordering, then a human WM item representation likely includes some episodic quality. This is partially captured by the N-back model's "seen" slot specification, which is relied upon during updating to increase the likelihood that the stimulus that was seen (for example) two presentations ago is retrieved to be updated into the "three" slot, but this slot is not specified during the retrieval attempt ahead of response. This is because the model already "knows" what it believes occurred two presentations ago, on the basis of the stimulus window chunk in the imaginal buffer, and so this retrieval attempt can be interpreted as a representation of a potential working memory scanning process. The serial nature of a scanning process is also considered by the model's updating and rehearsal mechanisms, which sequentially address each slot of the stimulus window chunk in the imaginal buffer. However, it is limited in the extent to which it can

capture a scanning process in a number of related manners. The retrieval that occurs before response is not just "scanning" over elements in the model's representation of the contents of working memory (the stimulus window chunk in the imaginal buffer), but instead the contents of ACT-R's declarative memory system. Additionally, the duration of this retrieval is not dependent on the number of items that are currently represented, but rather the latency scale factor parameter *F* (in conjunction with the activation value of the retrieved chunk); the *F* parameter also influences the speed of retrievals that occur during updating and rehearsal. It is unreasonable to believe that these two different types of retrievals represent the same process, as the updating/rehearsal retrievals respect a potential "scanning process" by addressing the slots of the stimulus window chunk in a serial manner. As a result, the value of the model's *F* parameter is constrained by how quickly updating and rehearsal must occur in order to represent the task's information with enough fidelity to adequately predict participant accuracies. This interpretation is supported by the observation that the decremental leave-one-out procedure determined that *F* was critical to the N-back model's prediction of participant accuracies. The same procedure determined that the base-level activation decay parameter *d* was essential in the prediction of response times for the N-back model, suggesting a predominant role of the activation value of the chunk that is retrieved before response in the fitting of participant response times - the more quickly this activation has decayed since last rehearsed (i.e., the larger the value of *d*), the slower the model's response time will be (through the duration of the model's retrieval). Altogether, this pattern of model behavior suggests that the N-back model implements a reasonable strategy for representation, updating, and rehearsal of information necessary to complete the N-back task, but does not consider a time-consuming process that occurs between the presentation of a new stimulus and the generation of a response.

In conclusion, the two ACT-R models considered in this work each include a number of elements that promote their plausibility as models of cognitive strategy in the zero-back and N-back conditions of the N-back task paradigm. They both reflect the multicomponent nature of popular working memory theory through their consideration of a number of parameters that can be interpreted as individual components within this theoretical framework. Additionally, these parameters are themselves central to a multicomponent theory of general cognition, reflecting the notion that the cognitive processes that contribute to working memory function are independent functions that support other cognitive behaviors. These components are utilized to implement a common strategy of rehearsal, where each model repeatedly experiences the information necessary to complete the task on a time scale that is aligned with the relevancy of the information; for the zero-back model, over trials, and for the N-back model, in-between trials. For the N-back model in particular, this strategy results in model behavior that mirrors observed human accuracies across a range of N-back conditions, but does not reflect human response times in the same. This suggests that the N-back model does not include processes that humans may engage in before response, but likely captures maintenance processes that occur after responding and before the external environment changes in a manner that demands new interaction. Together, these factors make it likely that the models will be capable of fitting differences between the behavior of individuals in the N-back task, thereby providing a manner of observing individual differences in the contribution of multiple independent cognitive functions to working memory ability.

4.3 Individual Differences in Model-based Characterizations of Working Memory Ability

As the zero-back and N-back models provide eminently plausible accounts of a general working memory strategy in the zero-back and two-back conditions of the N-back task, there is potential for these models to be capable of characterizing differences in participants through differences in parameter values such as those discussed in section 4.1. The interpretation of these parameters as working memory constructs allows for the investigation of the contribution of multiple cognitive components to working memory function, while the decremental leave-one-out procedure allows for the determination of the parameters/cognitive components that were most essential to the model's ability to predict differences in individual performance. Furthermore, as these parameters were estimated in two conditions of the N-back task that place different demands on working memory, the reliability of these parameter estimates within individuals and between cognitive contexts can be examined.

The participant-specific parameter estimates that resulted from fitting of the zero-back model to individual behavior in the zero-back condition showed reasonable variability across participants, indicating the model was capable of capturing individual behavior through variation in these parameters. If one or more parameter estimates showed little variability, it would indicate that the model was reliant on that specific value to produce successful predictions across participants, reducing the validity of that parameter's working memory interpretation. It was observed that the goal buffer spreading activation parameter W_g was greater than the imaginal buffer spreading activation parameter W_i for a majority of participants, indicating a reliance on goal-related information for the zero-back's working memory strategy. The information held in the goal buffer that is relevant to the model's ability to perform the task successfully includes the representation of "cue stimuli" as a relevant factor in retrieval, and so the model predicts that

these participant engage in a strategy of determining what their memory of the cue is on each trial before comparing to the current stimulus. This observation was reinforced by the decremental leave-one-out procedure's result that individual-specific estimates of W_g alone were adequate in explaining individual differences in both response time and accuracy. This finding of the role of W_g is commensurate with previous research (Daily et al., 2001) and supports the role of goal-directed attention in working memory (Nelson Cowan, 1999; Engle & Kane, 2004; M. J. Kane et al., 2001). However, approximately 30% of participants were estimated to have a larger W_i value, indicating that for some individuals, representation of the current stimulus drove retrieval attempts — in other words, they may be searching for a memory of having seen the current stimulus as a cue in the past, rather than actively attempting to determine what their memory of the current cue is. This process would be driven by attention sourced in representation of the currently relevant information, rather than goal-directed attention. The decremental leave-one-out procedure again supported this observation by revealing W_i to be the second-most influential parameter in prediction of participant behavior, although it by itself was not predictive of differences across participants in response time or accuracy.

Similar to the zero-back model, the N-back model also demonstrated reasonable variability across participants in the parameter estimates. It was again observed that the majority of participants exhibited a relatively greater contribution of goal-directed attention over attention sourced from currently relevant information, although the amount of information driving these sources of attention was different between the two models. In the N-back model, the goal buffer provides spreading activation to chunks in declarative memory which possessed a "seen" slot that had a value equal to the condtion's N-back value; the model was allocating goal-directed attention to memories of stimuli that had been seen two presentations ago. The imaginal buffer

provides spreading activation to chunks in declarative memory which match those represented as the last N+1 stimuli presented. In the case of target stimuli, this allocates attention (sourced from task-relevant information represented in working memory) to the proper chunk representation in declarative memory (aided by goal-driven attention), but also does so for other recent stimuli. In the case of stimuli older than those that occurred N presentations ago, the model has a memory (chunk representation) of having seen the item N presentations ago, which has decayed more so than the chunk representation of the stimulus actually seen N presentations ago. These stimuli still compete for retrieval (aided by imaginal-driven attention), providing the possibility for errors on target trials through the retrieval of an incorrect stimulus chunk representation. For lure trials, the imaginal buffer's spreading activation allocates attention to stimuli seen both N-1 and N+1 presentations ago, increasing the probability of a false alarm on these trials that becomes related to the amount that those two chunks have decayed. In this manner, the goal and imaginal buffer spreading activation parameters W_g and W_i of the N-back model capture the influence of multiple sources of attention, and provide the opportunity for the model to capture individual differences in sensitivity to targets and lures in the N-back task.

However, the decremental leave-one-out procedure did not identify either W_g or W_i as able to adequately predict participant response times or accuracies in the two-back condition; instead, it was observed that participant-specific values of the decay rate *d* predicted participant response times (but not accuracies), while participant-specific values of the latency scale factor *F* allowed the model to be predictive of participant accuracies (but not response times). This combined influence of *d* and *F* is likely due to the model's reliance on retrievals to support updating and rehearsal procedures, although on first pass, it seems to be reversed in dependency; as *F* scales retrieval times, it should directly affect response times (as the model's response is

made on the basis of the match between the current stimulus and a retrieved chunk). However, only one retrieval occurs before response is made on a trial, while many retrievals occur afterwards to support the updating and rehearsal processes. In order to fully update the stimulus window and have adequate time to rehearse before the next trial, these rehearsals must occur relatively quickly - constraining the value of *F*, as discussed in section 4.2. The dependency of the updating and rehearsal mechanisms on *F* result in the parameter's prediction of individual differences in accuracy — the faster an individual can perform these updating and rehearsal processes, the more likely it is that information regarding the last N+1 stimuli is accurately represented (as slower "updaters" may not finish updating before a new trial begins), and the more time an individual will have for rehearsal before the next trial. In parallel, this constraint on *F* leads to a reliance on *d* to capture differences in reaction times - the faster a chunk's activation decays, the longer the duration of the retrieval that occurs before response. Including a "scanning" process over the stimulus window, which aims to determine whether the current stimulus is held in the stimulus window (and if so, in which positions), may better represent the processes required to extract information from WM for use in ongoing behavior, removing the arbitrary constraint on *F* and resulting relationship with *d*.

Differences in the participant-specific estimates of parameters that are shared between the zero-back and N-back models can also be examined to determine the reliability of these estimates within individuals across two conditions of a task which require relatively different cognitive strategies. The more consistent these estimates in different cognitive contexts, the more likely they are to capture task-general elements of cognitive function that may be more likely to be reflected in resting-state brain activity. It was observed that, on average, the estimates of the goal and imaginal buffer spreading activation parameters W_g and W_i were no different between

the two models and task conditions. This is consistent with the observation of a relatively stronger influence of W_g over W_i in a majority of participants across the two models; additionally, roughly the same proportion of participants demonstrated this relationship between the two models. This implies that the influence of attention between these two conditions is roughly the same; again, this could be impacted by differences in information that drive the attention, as well as the competitive relationship between F and d in the N-back model. This competitive relationship in one model may underlie the observation that the decay rate d and latency scale factor F parameters were each found to have significantly different estimates between the zero-back and N-back models (zero-back minus N-back); however, the magnitude of these differences was small (the mean of per-participant differences in d was 0.03, while the mean of differences in F was 0.19).

Overall, the parameter estimation results primarily provided evidence for the involvement of goal-directed attention in working memory processes, which in this model serves to activate representations in long-term memory related to goal-oriented concepts. These representations become more available, increasing the likelihood of their retrieval for use in ongoing cognitive processing supporting behavior. Additionally, attention driven by information currently in working memory may activate representation of related concepts in long-term memory, allowing for a form of "caching" in which commonly referenced items are persistently more available. Decay of representation as well as speed of retrieval from LTM into working memory also had support for a role particularly in updating and rehearsal processes, although this may have been driven at least partially by an unintended arbitrary relationship between the two parameters.

4.4 Brain Regions and Networks Identified to be Predictive of Individual Working Memory Characterization

The zero-back and N-back models were shown to characterize individual differences in working memory function through differences in parameters reflective of cognitive constructs thought to contribute to working memory. Further evidence supporting the model's validity in capturing these constructs can be gained by relating these differences in construct estimates to individual differences in resting-state brain activity, specifically the functional connectivity between a set of network-associated regions (Power et al., 2011). As this resting state functional connectivity has been shown to be predictive of both task activation (Cole et al., 2016; Tavor et al., 2016) and performance (including in working memory task paradigms (Esposito et al., 2009; Sala-Llonch et al., 2012)), it is likely to reflect to some degree the contribution of these components to task performance. In order to pursue this notion, individual resting-state functional connectomes were regressed on participant-specific values of the parameters determined to be essential to each model's predictive efficacy (through the decremental leave-one-out procedure). The Lasso procedure balances minimization of the number of predictors with ability to predict the criterion, which in this case serves to identify the set of functional connections most necessary to predict individual values of these parameters. A limitation of the Lasso procedure is the large number of regressors that can potentially be included in the final regression model; while the use of partial correlations in the calculation of the functional connectome attempts to account for this, some regressors are bound to be correlated. As a result, the Lasso regression will "sparsify" and exclude functional connections that may be predictive of the criterion variable. However, as will be discussed here, the functional connections that are identified by the Lasso regressions in this study agree with results

previously reported in the literature. The regions that are included in these functional connections, including the networks they are thought to participate in, can be examined for correspondences with working memory-related processes. This approach serves to identify neural correlates of specific cognitive components of working memory.

Analysis of the zero-back model determined that the goal buffer spreading activation parameter W_g, representative of the influence of goal-directed attention, was essential to the model's ability to predict individual differences in both response times and accuracies. The Lasso regression of individual functional connectomes on these parameter values revealed a set of resting-state functional connections (measured as the correlation between regional timeseries) that included a number of regions that have been associated with theorized cognitive networks. The importance value of these regions, a metric which combines the number and strength of the predictive functional connections that a given region is involved in, can identify the regions most critical to prediction of these specific parameter estimates. It was observed that regions belonging to the default mode, ventral attention, and visual networks were among these critical regions. The degree of anticorrelation between the default mode network (DMN) and task-positive networks has previously been implicated in resting-state correspondences to working memory task performance (Sala-Llonch et al., 2012), as has the extent of DMN inactivation (Esposito et al., 2009), implying that the degree of inactivation of DMN relative to the activation of task-positive networks is important in success under working memory tasks. The observations here support this notion, as each of the DMN regions identified by the Lasso regression participated in functional connections with either the ventral attention or visual networks that were estimated to have anticorrelated relationships. The ventral attention network is implicated in the reorientation of attention to task-relevant environmental stimuli (Vossel et al.,

2014) while the visual network is comprised of regions dedicated to visual processing, including occipital cortex and a portion of parietal cortex (Power et al., 2011). The predictability of the functional connections between these regions indicates that the zero-back model's goal buffer spreading activation parameter W_g may capture elements of attentional orienting driven by changes in the visual scenery; while the model's goal buffer does not explicitly represent these changes, they do result changes to the goal buffer representation, particularly when a new block's cue stimulus is presented.

The N-back model was successful in capturing an updating/rehearsal mechanism (discussed in 4.2). The cognitive processes that support such a mechanism are likely to occur absent of an explicit task, as they support cognitive operations that are thought to occur in the resting state and without external input (Mazoyer et al., 2001). Accordingly, relation of the N-back model parameter estimates to resting-state functional connectivity are likely to identify functional connections between brain regions that are involved in these processes. Analysis of the N-back model revealed two different parameters to be significantly predictive of participant response time or accuracy; the base-level decay parameter *d* and the latency scale factor parameter *F*, respectively. For the decay parameter *d*, the Lasso regression again revealed a number of predictive functional connections that linked task-negative DMN regions to a number of task-positive networks, including the visual, salience, and cingulo-opercular task control networks The most predictive functional connections included a set connecting visual regions to DMN regions in the prefrontal cortex (PFC). However, while the Power (2011) parcellation labels a number of these regions as "visual", they are clearly located in the ventral pathway of the visual processing stream, thought to underlie the recognition of visual objects (see Figure 29) (Kravitz et al., 2013). As PFC has been previously implicated in working memory performance,

including updating and maintenance of WM information (Barch et al., 2013; Lamichhane et al., 2020), the role of these functional connections in the prediction of *d* may suggest that information sourced from processes supporting the recognition of task stimuli influences the amount of decay in WM representations, through interactions with prefrontal regions. The DMN may "gate" this influence in the resting-state, reducing the interference that automatically-attended visual information may generate in conjunction with ongoing WM representations.

The N-back model's latency scale factor *F* was found to be predictive of individual differences in two-back accuracies, and the Lasso regression of individual functional connectomes on these parameter estimates once again resulted in the identification of relationships between task-negative DMN regions and task-positive networks as being predictive of these individual parameter values. The predictive task-positive networks functionally connected to the DMN mainly included the visual network, and in this case, these regions were largely confined to occipital cortex; since *F* is reflective of processing speed, this relationship may be due to the necessity of surface-level processing in recognizing the stimulus quickly enough to subsequently engage in a search for the stimulus in WM. As noted in sections 4.2 and 4.3, the N-back model's value of *F* was largely governed by the updating and rehearsal processes; this makes interpretation of these task-positive/task-negative network connections difficult. Another set of functional connections, in this case between two task-positive regions, were also identified as being predictive of individual values of *F*; these two networks included the fronto-parietal task control network and the cingulo-opercular task control network. The former of these two networks is thought to implement cognitive control by flexibly coupling with other cognitive networks (such as attentional or default mode networks) to modulate their

activity in the support of goal-directed behavior; the latter has been implicated in the engagement of tonic alertness (Coste & Kleinschmidt, 2016; Sadaghiani & D'Esposito, 2015; Spreng et al., 2010; Zanto & Gazzaley, 2013). Successful performance of the N-back task demands a certain level of engagement or alertness; if incoming stimuli are not adequately processed, the memory of recently presented stimuli will not be reflective of the truth, leading to erroneous behavior. Additionally, if the representation of recently presented stimuli is not updated or rehearsed with care, the representation can become degraded, further increasing the chance of error (perhaps particularly in the case of lures, where "adjacent" information becomes more likely to be mistaken for the "target" information). The observed relationship between functional connectivity of the frontoparietal/cingulo-opercular task control networks and a parameterization of processing speed which was predictive of accuracy supports this interpretation: those that are better able to maintain alertness (through the cingulo-opercular network, mediated by the fronto-parietal network) are able to perform processing on task information in a more efficient manner, improving behavioral accuracy.

4.5 Conclusions

The present study aimed to determine relationships between behavioral performance in a working memory task paradigm, model-based characterizations of working memory function, and resting-state connectivity between brain regions in order to further understanding of the cognitive contributors of a multicomponent working memory system. Two ACT-R models of cognitive strategy under zero-back and two-back conditions of the N-back task were capable of capturing individual differences in task behavior through fitting of participant-specific parameter estimates, thereby providing a manner of examining differences in the contribution of multiple cognitive processes to working memory function. These parameter estimates were related to

individual functional connectivity in the resting state, identifying brain regions and networks related to the cognitive components represented by the model parameters. The identified contributions of parameters, as well as the associated brain regions were largely consistent with previous reports on the neural correlates of multicomponent working memory. This work furthers support for a multicomponent view of working memory, and begins to provide evidence for differences in these cognitive components as driving differences in working memory ability. Individuals may differ in manners such as their ability to allocate attention to intended representations (or inhibit attention towards unintended targets), the decay of their working memory representations, or the speed at which they process task-relevant information in working memory. These differences in cognition are attributable to differences in brain physiology, specifically the interactions between brain networks thought to support internally- and externally-directed cognition. The approach of the present study also extends manners in which the ACT-R cognitive modeling framework can be related to brain activity. Previous methods related the aggregate activity of ACT-R modules to the activity of specific brain regions; however, this approach assumes a localist, specificity-of-function perspective and aims to identify cognitive architectural aspects that are consistent across individuals. The current study extends the approach by considering differences between individuals and identifies related brain regions in a data-driven manner. Future work in this direction could include a number of perspectives. The ACT-R models could certainly be improved; the HCP participants performed the zero-back and two-back conditions in alternating blocks, and so one large improvement would be the consolidation of the zero-back and N-back models into one that is capable of performing both conditions (as well as the putative one- and three-back conditions). As noted earlier, the N-back model does not model a time-consuming process that apparently occurs

before response (perhaps a process of "scanning" working memory contents?); inclusion of this process may free the *d* and *F* parameters from the noted arbitrary dependency and allow all parameters of the model to better capture cognitive functioning. Another potential improvement to the N-back model strategy is to more fully specify the rehearsal procedure; in this study, a simplified version that was guaranteed to accurately rehearse the contents of the stimulus window was used. Outside of the strategy implemented by the current models, alternative strategies could be implemented as models (as in (Juvina & Taatgen, 2007)), and interactions between strategy and parameter values in the ability to account for individual differences could be examined. Regarding the use of fMRI, only resting-state functional brain data was considered in this study, despite the association of this data to characterizations of task behavior; subsequent research could examine the relationship between these parameters and task-based functional brain data, which was also collected in the HCP paradigm. This effort could be doubled by examining ACT-R predictions of the fMRI BOLD signal under the task: for instance, by including these predictions in the model-fitting process, thereby requiring that the ACT-R models make predictions of co-occurring behavior and brain activity. These approaches to the characterization of brain function, cognitive ability, and behavior have the potential to contribute to a growing understanding of the neuropsychological factors that differentiate individuals and define cognition.

Acknowledgements

Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

References

Anderson, J. R. (1983). Retrieval of information from long-term memory. *Science*, *220*(4592), 25–30.

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*(3), 261–295.

Anderson, J. R., Anderson, J. F., Ferris, J. L., Fincham, J. M., & Jung, K.-J. (2009). Lateral inferior prefrontal cortex and anterior cingulate cortex are engaged at different stages in the solution of insight problems. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(26), 10799–10804.

Anderson, J. R., Fincham, J. M., Qin, Y., & Stocco, A. (2008). A central circuit of the mind. *Trends in Cognitive Sciences*, *12*(4), 136–143.

Anderson, J. R., & Lebiere, C. J. (2014). *The Atomic Components of Thought*. Psychology Press.

Anderson, J. R., Qin, Y., Jung, K.-J., & Carter, C. S. (2007). Information-processing modules and their relative modality specificity. *Cognitive Psychology*, *54*(3), 185–217.

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In *Psychology of learning and motivation* (Vol. 2, pp. 89–195). Elsevier.

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin & Review*, *22*(2), 366–377.

Baddeley, A. (1992). Working memory. *Science*, *255*(5044), 556–559.

Baddeley, A. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Sciences*, *4*(11), 417–423.

Baddeley, A. (2003). Working memory: looking back and looking forward. *Nature Reviews. Neuroscience*, *4*(10), 829–839.

Baddeley, A. (2007). *Working Memory, Thought, and Action*. OUP Oxford.

Baddeley, A. D. (2004). The psychology of memory. *The Essential Handbook of Memory Disorders for Clinicians*, 1–13.

Baddeley, A. D., & Hitch, G. (1974). Working Memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 8, pp. 47–89). Academic Press.

Baddeley, A., & Wilson, B. A. (2002). Prose recall and amnesia: implications for the structure of working memory. *Neuropsychologia*, *40*(10), 1737–1743.

Baldo, J. V., & Dronkers, N. F. (2006). The role of inferior parietal and inferior frontal cortex in working memory. *Neuropsychology*, *20*(5), 529–538.

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., Glasser, M. F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J. M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A. Z., … WU-Minn HCP Consortium. (2013). Function in the human connectome: task-fMRI and individual differences in behavior. *NeuroImage*, *80*, 169–189.

Barrett, L. F., Tugade, M. M., & Engle, R. W. (2004). Individual Differences in Working Memory Capacity and Dual-Process Theories of the Mind. *Psychological Bulletin*, *130*(4), 553–573.

Barrouillet, P., & Camos, V. (2020). The time-based resource-sharing model of working memory. *Working Memory: State of the Science*, 85–115.

Barrouillet, P., Portrat, S., Vergauwe, E., Diependaele, K., & Camos, V. (2011). Further evidence for temporal decay in working memory: reply to Lewandowsky and Oberauer (2009) [Review of *Further evidence for temporal decay in working memory: reply to Lewandowsky and Oberauer (2009)*]. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *37*(5), 1302–1317.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. In *Magnetic Resonance in Medicine* (Vol. 34, Issue 4, pp. 537–541). https://doi.org/10.1002/mrm.1910340409

Borst, J. P., & Anderson, J. R. (2013). Using model-based functional MRI to locate working memory updates and declarative memory retrievals in the fronto-parietal network. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(5), 1628–1633.

Borst, J. P., & Anderson, J. R. (2017). A step-by-step tutorial on using the cognitive architecture ACT-R in combination with fMRI data. *Journal of Mathematical Psychology*, *76*, 94–103.

Borst, J. P., Taatgen, N. A., & van Rijn, H. (2011). Using a symbolic process model as input for model-based fMRI analysis: locating the neural correlates of problem state replacements. *NeuroImage*, *58*(1), 137–147.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends in Cognitive Sciences*, *8*(12), 539–546.

Bromiley, P., & Curley, S. P. (1992). Individual differences in risk taking. *Risk-Taking Behavior.*, *345*, 87–132.

Buckner, R. L., Kelley, W. M., & Petersen, S. E. (1999). Frontal cortex contributes to human memory formation. *Nature Neuroscience*, *2*(4), 311–314.

Bush, G., Luu, P., & Posner, M. I. (2000). Cognitive and emotional influences in anterior cingulate cortex. *Trends in Cognitive Sciences*, *4*(6), 215–222.

Cabeza, R., Dolcos, F., Graham, R., & Nyberg, L. (2002). Similarities and differences in the neural correlates of episodic memory retrieval and working memory. *NeuroImage*, *16*(2), 317–330.

Caceres, A., Hall, D. L., Zelaya, F. O., Williams, S. C. R., & Mehta, M. A. (2009). Measuring fMRI reliability with the intra-class correlation coefficient. *NeuroImage*, *45*(3), 758–768.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*(1), 31.

Cassenti, D. N., Kerick, S. E., & McDowell, K. (2011). Observing and modeling cognitive events through event-related potentials and ACT-R. *Cognitive Systems Research*, *12*(1), 56–65.

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, *54*(1), 1.

Chai, X. J., Castañón, A. N., Öngür, D., & Whitfield-Gabrieli, S. (2012). Anticorrelations in resting state networks without global signal regression. *NeuroImage*, *59*(2), 1420–1428.

Chen, Z., & Calhoun, V. (2018). Effect of Spatial Smoothing on Task fMRI ICA and Functional Connectivity. *Frontiers in Neuroscience*, *12*, 15.

Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: evidence for domain-specific working memory systems. *Memory & Cognition*, *30*(7), 1086–1095.

Cohen, A. L., Fair, D. A., Dosenbach, N. U. F., Miezin, F. M., Dierker, D., Van Essen, D. C., Schlaggar, B. L., & Petersen, S. E. (2008). Defining functional areas in individual human brains using resting functional connectivity MRI. *NeuroImage*, *41*(1), 45–57.

Cole, M. W., Ito, T., Bassett, D. S., & Schultz, D. H. (2016). Activity flow over resting-state networks shapes cognitive task activations. *Nature Neuroscience*, *19*(12), 1718–1726.

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review*, *82*(6), 407–428.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Zach Hambrick, D., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. In *Psychonomic Bulletin & Review* (Vol. 12, Issue 5, pp. 769–786). https://doi.org/10.3758/bf03196772

Coste, C. P., & Kleinschmidt, A. (2016). Cingulo-opercular network activity maintains alertness. *NeuroImage*, *128*, 264–272.

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin*, *104*(2), 163–191.

Cowan, N. (1999). An Embedded-Processes Model of working memory. In A. Miyake (Ed.), *Models of working memory: Mechanisms of active maintenance and executive control , (pp* (Vol. 506, pp. 62–101). Cambridge University Press, xx.

Cowan, N. (2001). The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *The Behavioral and Brain Sciences*, *24*(1), 87–114; discussion 114–185.

Cowan, N. (2008). What are the differences between long-term, short-term, and working memory?

Progress in Brain Research, *169*, 323–338.

Cowan, N. (2017). The many faces of working memory and short-term storage. *Psychonomic Bulletin & Review*, *24*(4), 1158–1170.

Daily, L. Z., Lovett, M. C., & Reder, L. M. (2001). Modeling individual differences in working memory performance: a source activation account. In *Cognitive Science* (Vol. 25, Issue 3, pp. 315–353). https://doi.org/10.1207/s15516709cog2503_1

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, *19*(4), 450–466.

D'Esposito, M., Postle, B. R., Ballard, D., & Lease, J. (1999). Maintenance versus manipulation of information held in working memory: an event-related fMRI study. *Brain and Cognition*, *41*(1), 66–86.

D'Esposito, M., Postle, B. R., & Rypma, B. (2000). Prefrontal cortical contributions to working memory: evidence from event-related fMRI studies. *Experimental Brain Research. Experimentelle Hirnforschung. Experimentation Cerebrale*, *133*(1), 3–11.

Dimov, C., Khader, P. H., Marewski, J. N., & Pachur, T. (2020). How to model the neurocognitive dynamics of decision making: A methodological primer with ACT-R. *Behavior Research Methods*, *52*(2), 857–880.

Drobyshevsky, A., Baumann, S. B., & Schneider, W. (2006). A rapid fMRI task battery for mapping of visual, motor, cognitive, and emotional function. *NeuroImage*, *31*(2), 732–744.

Dubois, J., & Adolphs, R. (2016). Building a Science of Individual Differences from fMRI. *Trends in Cognitive Sciences*, *20*(6), 425–443.

Ebbinghaus, H. (1913). Memory: A Contribution to Experimental Psychology, 1885. *New York: Teachers College, Columbia University*.

Engle, R. W., & Kane, M. J. (2004). Executive Attention, Working Memory Capacity, and a Two-Factor Theory of Cognitive Control. In B. H. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory, Vol* (Vol. 44, pp. 145–199). Elsevier Science, x.

Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent-variable approach. *Journal of Experimental Psychology. General*, *128*(3), 309–331.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review*, *102*(2), 211–245.

Esposito, F., Aragri, A., Latorre, V., Popolizio, T., Scarabino, T., Cirillo, S., Marciano, E., Tedeschi, G., & Di Salle, F. (2009). Does the default-mode functional connectivity of the brain correlate with working-memory performances? *Archives Italiennes de Biologie*, *147*(1-2), 11–20.

Felez-Nobrega, M., Foster, J. L., Puig-Ribera, A., Draheim, C., & Hillman, C. H. (2018). Measuring working memory in the Spanish population: Validation of a multiple shortened complex span task. *Psychological Assessment*, *30*(2), 274–279.

Foster, J. L., Shipstead, Z., Harrison, T. L., Hicks, K. L., Redick, T. S., & Engle, R. W. (2015). Shortened complex span tasks can reliably measure working memory capacity. *Memory & Cognition*, *43*(2), 226–236.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., Van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(27), 9673–9678.

Fuster, J. M., & Alexander, G. E. (1971). Neuron activity related to short-term memory. *Science*, *173*(3997), 652–654.

Gazzaley, A., Rissman, J., & D'Esposito, M. (2004). Functional connectivity during working memory maintenance. *Cognitive, Affective & Behavioral Neuroscience*, *4*(4), 580–599.

Glasser, M. F., Smith, S. M., Marcus, D. S., Andersson, J. L. R., Auerbach, E. J., Behrens, T. E. J., Coalson, T. S., Harms, M. P., Jenkinson, M., Moeller, S., Robinson, E. C., Sotiropoulos, S. N., Xu, J., Yacoub, E., Ugurbil, K., & Van Essen, D. C. (2016). The Human Connectome Project's neuroimaging approach. *Nature Neuroscience*, *19*(9), 1175–1187.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., Xu, J.,

Jbabdi, S., Webster, M., Polimeni, J. R., Van Essen, D. C., & Jenkinson, M. (2013). The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage*, *80*, 105–124.

Goldman-Rakic, P. S. (1995). Cellular basis of working memory. *Neuron*, *14*(3), 477–485.

Gonzalez, C., Lerch, J. F., & Lebiere, C. (2003). Instance-based learning in dynamic decision making. *Cognitive Science*, *27*(4), 591–635.

Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *NeuroImage*, *48*(1), 63–72.

Gunzelmann, G., Richard Moore, L., Salvucci, D. D., & Gluck, K. A. (2011). Sleep loss and driver performance: Quantitative predictions with zero free parameters. *Cognitive Systems Research*, *12*(2), 154–163.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(19), 6829–6833.

Jaeggi, S. M., Buschkuehl, M., Perrig, W. J., & Meier, B. (2010). The concurrent validity of the N-back task as a working memory measure. *Memory* , *18*(4), 394–412.

James, W. (1890). *The principles of psychology, Vol I*. https://doi.org/10.1037/10538-000

Jarrold, C., & Towse, J. N. (2006). Individual differences in working memory. *Neuroscience*, *139*(1), 39–50.

Jonides, J., Schumacher, E. H., Smith, E. E., Koeppe, R. A., Awh, E., Reuter-Lorenz, P. A., Marshuetz, C., & Willis, C. R. (1998). The role of parietal cortex in verbal working memory. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *18*(13), 5026–5034.

Juntu, J., Sijbers, J., Van Dyck, D., & Gielen, J. (2005). Bias Field Correction for MRI Images. *Computer Recognition Systems*, 543–551.

Juvina, I., & Taatgen, N. A. (2007). Modeling control strategies in the n-back task. *Proceedings of the 8th International Conference on Cognitive Modeling*, 73–78.

Kane, M. J., Bleckley, M. K., Conway, A. R., & Engle, R. W. (2001). A controlled-attention view of

working-memory capacity. *Journal of Experimental Psychology. General*, *130*(2), 169–183.

Kane, M. J., Conway, A. R. A., Miura, T. K., & Colflesh, G. J. H. (2007). Working memory, attention control, and the N-back task: a question of construct validity. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *33*(3), 615–622.

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005) [Review of *Working memory capacity and fluid intelligence are strongly related constructs: comment on Ackerman, Beier, and Boyle (2005)*]. *Psychological Bulletin*, *131*(1), 66–71.

Kaushanskaya, M., & Yoo, J. (2011). Rehearsal effects in adult word learning. *Language and Cognitive Processes*, *26*(1), 121–148.

Kessler, Y., & Oberauer, K. (2015). Forward scanning in verbal working memory updating. *Psychonomic Bulletin & Review*, *22*(6), 1770–1776.

Klingberg, T. (1998). Concurrent performance of two working memory tasks: potential mechanisms of interference. *Cerebral Cortex* , *8*(7), 593–601.

Kravitz, D. J., Saleem, K. S., Baker, C. I., Ungerleider, L. G., & Mishkin, M. (2013). The ventral visual pathway: an expanded neural framework for the processing of object quality. In *Trends in Cognitive Sciences* (Vol. 17, Issue 1, pp. 26–49). https://doi.org/10.1016/j.tics.2012.10.011

Lamichhane, B., Westbrook, A., Cole, M. W., & Braver, T. S. (2020). Exploring brain-behavior relationships in the N-back task. *NeuroImage*, *212*, 116683.

Lebiere, C. (1999). The dynamics of cognition: An ACT-R model of cognitive arithmetic. *Kognitionswissenschaft*, *8*(1), 5–19.

Lebiere, C., Gray, R., Salvucci, D., & West, R. (2003). Choice and learning under uncertainty: A case study in baseball batting. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *25*. https://escholarship.org/content/qt5p25k0wg/qt5p25k0wg.pdf

Lee, J. H., Durand, R., Gradinaru, V., Zhang, F., Goshen, I., Kim, D.-S., Fenno, L. E., Ramakrishnan, C., & Deisseroth, K. (2010). Global and local fMRI signals driven by neurons defined optogenetically

by type and wiring. *Nature*, *465*(7299), 788–792.

Lee, M. H., Smyser, C. D., & Shimony, J. S. (2013). Resting-state fMRI: a review of methods and clinical applications. *AJNR. American Journal of Neuroradiology*, *34*(10), 1866–1872.

Lengenfelder, J., Bryant, D., Diamond, B. J., Kalmar, J. H., Moore, N. B., & DeLuca, J. (2006). Processing speed interacts with working memory efficiency in multiple sclerosis. *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *21*(3), 229–238.

Lewandowsky, S., & Oberauer, K. (2009). No evidence for temporal decay in working memory. *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *35*(6), 1545–1551.

Liakakis, G., Nickel, J., & Seitz, R. J. (2011). Diversity of the inferior frontal gyrus—A meta-analysis of neuroimaging studies. *Behavioural Brain Research*, *225*(1), 341–347.

Litvan, I., Grafman, J., Vendrell, P., & Martinez, J. M. (1988). Slowed information processing in multiple sclerosis. *Archives of Neurology*, *45*(3), 281–285.

Liu, P., Calhoun, V., & Chen, Z. (2017). Functional overestimation due to spatial smoothing of fMRI data. *Journal of Neuroscience Methods*, *291*, 1–12.

Marewski, J. N., & Mehlhorn, K. (2011). Using the ACT-R architecture to specify 39 quantitative process models of decision making. *Judgment and Decision Making*, *6*(6), 439–519.

Mathy, F., Chekaf, M., & Cowan, N. (2018). Simple and Complex Working Memory Tasks Allow Similar Benefits of Information Compression. In *Journal of Cognition* (Vol. 1, Issue 1). https://doi.org/10.5334/joc.31

Mazoyer, B., Zago, L., Mellet, E., Bricogne, S., Etard, O., Houdé, O., Crivello, F., Joliot, M., Petit, L., & Tzourio-Mazoyer, N. (2001). Cortical networks for working memory and executive functions sustain the conscious resting state in man. *Brain Research Bulletin*, *54*(3), 287–298.

Miller, E. K., Lundqvist, M., & Bastos, A. M. (2018). Working Memory 2.0. *Neuron*, *100*(2), 463–475.

Miller, G. A. (1956). The magical number seven plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63*(2), 81–97.

Miller, K. M., Price, C. C., Okun, M. S., Montijo, H., & Bowers, D. (2009). Is the n-back task a valid neuropsychological measure for assessing working memory? *Archives of Clinical Neuropsychology: The Official Journal of the National Academy of Neuropsychologists*, *24*(7), 711–717.

Milner, B., Squire, L. R., & Kandel, E. R. (1998). Cognitive neuroscience and the study of memory. *Neuron*, *20*(3), 445–468.

Nijboer, M., Borst, J., van Rijn, H., & Taatgen, N. (2016). Contrasting single and multi-component working-memory systems in dual tasking. *Cognitive Psychology*, *86*, 1–26.

Oberauer, K. (2019). Is Rehearsal an Effective Maintenance Strategy for Working Memory? *Trends in Cognitive Sciences*, *23*(9), 798–809.

Oberauer, K., & Lewandowsky, S. (2013). Evidence against decay in verbal working memory. In *Journal of Experimental Psychology: General* (Vol. 142, Issue 2, pp. 380–411). https://doi.org/10.1037/a0029588

Oberauer, K., & Lewandowsky, S. (2014). Further evidence against decay in working memory. In *Journal of Memory and Language* (Vol. 73, pp. 15–30). https://doi.org/10.1016/j.jml.2014.02.003

Oberauer, K., Süß, H.-M., Wilhelm, O., & Sander, N. (2007). Individual differences in working memory capacity and reasoning ability. *Variation in Working Memory.*, *321*, 49–75.

Osaka, N., Osaka, M., Kondo, H., Morishita, M., Fukuyama, H., & Shibasaki, H. (2004). The neural basis of executive function in working memory: an fMRI study based on individual differences. *NeuroImage*, *21*(2), 623–631.

Owen, A. M., McMillan, K. M., Laird, A. R., & Bullmore, E. (2005). N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. In *Human Brain Mapping* (Vol. 25, Issue 1, pp. 46–59). https://doi.org/10.1002/hbm.20131

Oztekin, I., McElree, B., Staresina, B. P., & Davachi, L. (2009). Working memory retrieval: contributions of the left prefrontal cortex, the left posterior parietal cortex, and the hippocampus. *Journal of Cognitive Neuroscience*, *21*(3), 581–593.

Pessoa, L., Gutierrez, E., Bandettini, P., & Ungerleider, L. (2002). Neural correlates of visual working

memory: fMRI amplitude predicts task performance. *Neuron*, *35*(5), 975–987.

Portrat, S., Barrouillet, P., & Camos, V. (2008). Time-related decay or interference-based forgetting in working memory? *Journal of Experimental Psychology. Learning, Memory, and Cognition*, *34*(6), 1561–1564.

Powell, M. J. D. (1964). An efficient method for finding the minimum of a function of several variables without calculating derivatives. *Computer Journal*, *7*(2), 155–162.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., Vogel, A. C., Laumann, T. O., Miezin, F. M., Schlaggar, B. L., & Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, *72*(4), 665–678.

Raichle, M. E., MacLeod, A. M., Snyder, A. Z., Powers, W. J., Gusnard, D. A., & Shulman, G. L. (2001). A default mode of brain function. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(2), 676–682.

Rao, S. M., St Aubin-Faubert, P., & Leo, G. J. (1989). Information processing speed in patients with multiple sclerosis. *Journal of Clinical and Experimental Neuropsychology*, *11*(4), 471–477.

Redick, T. S., & Lindsey, D. R. B. (2013). Complex span and n-back measures of working memory: a meta-analysis. *Psychonomic Bulletin & Review*, *20*(6), 1102–1113.

Rogers, B. P., Morgan, V. L., Newton, A. T., & Gore, J. C. (2007). Assessing functional connectivity in the human brain by fMRI. *Magnetic Resonance Imaging*, *25*(10), 1347–1357.

Sadaghiani, S., & D'Esposito, M. (2015). Functional Characterization of the Cingulo-Opercular Network in the Maintenance of Tonic Alertness. *Cerebral Cortex* , *25*(9), 2763–2773.

Sala-Llonch, R., Peña-Gómez, C., Arenaza-Urquijo, E. M., Vidal-Piñeiro, D., Bargalló, N., Junqué, C., & Bartrés-Faz, D. (2012). Brain connectivity during resting state and subsequent working memory task predicts behavioural performance. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *48*(9), 1187–1196.

Sense, F., Behrens, F., Meijer, R. R., & van Rijn, H. (2016). An Individual's Rate of Forgetting is Stable Over Time but Differs Across Materials. *Topics in Cognitive Science*, *8*(1), 305–321.

Shallice, T., & Warrington, E. K. (1970). Independent functioning of verbal memory stores: a neuropsychological study. *The Quarterly Journal of Experimental Psychology*, *22*(2), 261–273.

Siero, J. C. W., Hermes, D., Hoogduin, H., Luijten, P. R., Ramsey, N. F., & Petridou, N. (2014). BOLD matches neuronal activity at the mm scale: a combined 7T fMRI and ECoG study in human sensorimotor cortex. *NeuroImage*, *101*, 177–184.

Skehan, P. (1991). Individual Differences in Second Language Learning. *Studies in Second Language Acquisition*, *13*(2), 275–298.

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D. A., Griffanti, L., Harms, M. P., Kelly, M., Laumann, T., Miller, K. L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A. Z., Vu, A. T., … Glasser, M. F. (2013). Resting-state fMRI in the Human Connectome Project. *NeuroImage*, *80*, 144–168.

Smith, S. M., Vidaurre, D., Beckmann, C. F., Glasser, M. F., Jenkinson, M., Miller, K. L., Nichols, T. E., Robinson, E. C., Salimi-Khorshidi, G., Woolrich, M. W., Barch, D. M., Uğurbil, K., & Van Essen, D. C. (2013). Functional connectomics from resting-state fMRI. *Trends in Cognitive Sciences*, *17*(12), 666–682.

Spreng, R. N., Stevens, W. D., Chamberlain, J. P., Gilmore, A. W., & Schacter, D. L. (2010). Default network activity, coupled with the frontoparietal control network, supports goal-directed cognition. *NeuroImage*, *53*(1), 303–317.

Tam, H., Jarrold, C., Baddeley, A. D., & Sabatos-DeVito, M. (2010). The development of memory maintenance: Children's use of phonological rehearsal and attentional refreshment in working memory tasks. In *Journal of Experimental Child Psychology* (Vol. 107, Issue 3, pp. 306–324). https://doi.org/10.1016/j.jecp.2010.05.006

Tavor, I., Parker Jones, O., Mars, R. B., Smith, S. M., Behrens, T. E., & Jbabdi, S. (2016). Task-free MRI predicts individual differences in brain activity during task performance. *Science*, *352*(6282), 216–220.

Tenison, C., & Anderson, J. R. (2016). Modeling the distinct phases of skill acquisition. *Journal of*

Experimental Psychology. Learning, Memory, and Cognition, *42*(5), 749–767.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, *58*(1), 267–288.

Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*(2), 127–154.

Uğurbil, K., Xu, J., Auerbach, E. J., Moeller, S., Vu, A. T., Duarte-Carvajalino, J. M., Lenglet, C., Wu, X., Schmitter, S., Van de Moortele, P. F., Strupp, J., Sapiro, G., De Martino, F., Wang, D., Harel, N., Garwood, M., Chen, L., Feinberg, D. A., Smith, S. M., … WU-Minn HCP Consortium. (2013). Pushing spatial and temporal resolution for functional and diffusion MRI in the Human Connectome Project. *NeuroImage*, *80*, 80–104.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*(1), 104–132.

van den Heuvel, M. P., & Hulshoff Pol, H. E. (2010). Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology: The Journal of the European College of Neuropsychopharmacology*, *20*(8), 519–534.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K., & WU-Minn HCP Consortium. (2013). The WU-Minn Human Connectome Project: an overview. *NeuroImage*, *80*, 62–79.

van Vugt, M. K. (2012). Relating ACT-R buffer activation to EEG activity during an attentional blink task. *Nele Rußwinkel| Uwe Drewitz| Hedderik van Rijn (eds.)*, 218.

van Vugt, M. K. (2014). Cognitive architectures as a tool for investigating the role of oscillatory power and coherence in cognition. In *NeuroImage* (Vol. 85, pp. 685–693). https://doi.org/10.1016/j.neuroimage.2013.09.076

van Vugt, M. K., Taatgen, N. A., Sackur, J., Bastian, M., Borst, J., & Mehlhorn, K. (2015). Modeling mind-wandering: a tool to better understand distraction. *Proceedings of the 13th International*

Conference on Cognitive Modeling, 252–257.

van Vugt, M. K., van der Velde, M., & ESM-MERGE Investigators. (2018). How does rumination impact cognition? A first mechanistic model. *Topics in Cognitive Science*, *10*(1), 175–191.

Vecchi, T., Richardson, J., & Cavallini, E. (2005). Passive storage versus active processing in working memory: Evidence from age-related variations in performance. *The European Journal of Cognitive Psychology*, *17*(4), 521–539.

Vetterling, W. T., Press, W. H., Teukolsky, S. A., & Flannery, B. P. (1992). *Numerical recipes: example book C (The Art of Scientific Computing)*. Press Syndicate of the University of Cambridge.

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … SciPy 1.0 Contributors. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272.

Vossel, S., Geng, J. J., & Fink, G. R. (2014). Dorsal and ventral attention systems: distinct neural circuits but collaborative roles. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *20*(2), 150–159.

WU-Minn, H. C. P. (2017). 1200 subjects data release reference manual. *URL Https://www. Humanconnectome. Org*. http://www.humanconnectome.org/storage/app/media/documentation/s1200/HCP_S1200_Release_Reference_Manual.pdf

Zanto, T. P., & Gazzaley, A. (2013). Fronto-parietal network: flexible hub of cognitive control [Review of *Fronto-parietal network: flexible hub of cognitive control*]. *Trends in Cognitive Sciences*, *17*(12), 602–603.

Zhou, P., Sense, F., van Rijn, H., & Stocco, A. (2021). Reflections of idiographic long-term memory characteristics in resting-state neuroimaging data. *Cognition*, *212*, 104660