

FIs working memory domain-general or domain-specific?

Nazbanou Nozari^{1,2} and Randi C. Martin³

¹Department of Psychological and Brain Sciences, Indiana University, Bloomington, Indiana, USA

²Cognitive Science Program, Indiana University, Bloomington, Indiana, USA

³Department of Psychological Sciences, Rice University, Houston, Texas, USA

Corresponding author: Nazbanou Nozari

Address: 1101 E 10th St., Bloomington, IN 47405

Phone number: 812-855-5818

Email: bnozari@iu.edu

Lab website: <https://www.nozarilab.com/>

Twitter handle: @NozariL

Abstract

Given the fundamental role of working memory (WM) in all domains of cognition, a central question has been whether WM domain-general. However, the term “domain-general” has been used in different, and sometimes misleading, ways. By reviewing recent evidence and biologically plausible models of WM, we show that the level of domain-generality varies substantially between three facets of WM: in terms of computations, WM is largely domain-general. In terms of neural correlates, it contains both domain-general and domain-specific elements. Finally, in terms of application, it is mostly domain-specific. This variance encourages a shift of focus towards uncovering domain-general computational principles and away from domain-general approaches to the analysis of individual differences and WM training, favoring newer perspectives, such as training-as-skill-learning.

Keywords

Working memory, domain-generality, brain training, resource models, neural correlates

The question of domain-generalality of WM

Working memory (WM) is the ability to maintain information over short periods of time, usually in the service of an ongoing task. Since different kinds of information may need to be temporarily maintained before they are acted upon, it is not surprising that the question of domain-generalality has come up in various accounts of WM. For example, the classic Baddeley and Hitch (1974) [1] model comprised a mixture of domain-specific visual and verbal components and a domain-general “central executive”. Later, some researchers adopted a mostly domain-general account [2], whereas others proposed a mostly domain-specific account [3]. However many such accounts do not address neural findings and lack computational specificity, or the versions that have computational implementations focus on specific aspects of WM or tasks [4,5], leaving the larger debate on domain-generalality unsettled. Aside from theoretical models, the assumption of domain-generalality is behind many practices in current psychological studies and interventions. Two prominent examples are the analysis of individual differences and brain training. When one WM task is used to predict (analysis of individual differences) or enhance (brain training) performance on a range of tasks that are hypothesized to tap WM, domain-generality of WM is tacitly assumed. It is, thus, of both theoretical and applied interest to verify the assumption of domain-generality of WM.

The goal of this paper is not to give a comprehensive overview of the perspectives adopted on domain-generality/specificity of WM over the years, as such overviews exist elsewhere [6]. Rather, the paper aims to propose a more useful taxonomy for studying domain-generality, comprising (a) domain-generality in computations, (b) domain-generality in neural substrates, and (c) domain-generality in application. The tenet of this work is that logical relations such as $a \rightarrow b$ and $b \rightarrow c$ (and consequently, $\neg b \rightarrow \neg a$ and $\neg c \rightarrow \neg b$) do not necessarily hold between these three components. For example, the same computations could be carried out by different neural populations, or the same neural population could carry out different computations. Starting with this logic, the next three sections examine domain-generality or specificity of WM in each component. The first section examines a computational principle with high plausibility for domain-specificity and shows that it is indeed domain-general. The second section demonstrates that potentially domain-general neural regions in WM can have a mixture of domain-general and

domain-specific functions. Finally, the third section uses a state-of-the-art neural model to show that domain-generality in neural correlates does not translate into domain-generality of application.

Domain-generality in computations

Principles of resource division in WM

The key question here is: If a set of computations can be domain-general, are they indeed domain-general or not? This question naturally excludes operations that are tied to domain-specific representations. For example, “verbal rehearsal” is clearly tied to phonological representations, so even if it is used for memorizing visual objects, it is still mediated through verbal labels, and is thus not a suitable candidate for examining domain-generality. A good test case for this purpose is the principle of resource division, because (a) limited capacity is one of the least disputed properties of human WM [7–9], (b) detailed models have been proposed to explain resource division (Fig. 1a-d), and (c) at least some of these models have neurobiological plausibility and directly link resource division to attention, which is another widely accepted process relevant to WM [10,11].

Two major classes of theories have been proposed to explain capacity limitations of WM, discrete and continuous models [7,9,12]. Discrete models were based on a clever observation in early studies, namely that the number, and not the complexity, of stimuli determined WM capacity [9]. Correspondingly, it was proposed that WM has a fixed capacity, defined as k slots, within which items are stored perfectly and beyond which performance is at chance (Fig1a). However, the introduction of continuous reproduction paradigms [13,14], which allowed for measuring the *precision* of a memory, revealed a decline in precision even from 1 to 3 items, i.e., below k . To account for this finding, the slot-plus-averaging (SA) model was proposed [14] (Fig. 1b). Although SA maintains the notion of slots by assuming that WM resource is divided into discrete quanta, these slots can be combined and shared between items, allowing for higher precision for items below k . The idea of slots-plus-averaging was later adopted by more models [15], but also received criticism [16–18].

The success of the SA model, which flexed the notion of fixed slots, gave rise to continuous resource models that assume WM to be a continuous resource with no upper bound k [12,19]. As the number of items increases, the precision with which they are encoded decreases. Thus, these

models can account for the data using a single process of continuous resource division, although such division may be equal (Fig.1c) or unequal (Fig. 1d) among items, the latter representing variable-precision resource models [17,19–21]. In continuous resource models, guessing is simply the extreme end of the continuous process of resource division, where an item receives little to no resources. Opposing this view are dual-process mixed-state models, which view pure guessing as a process qualitatively distinct from recall [22] (see underlying distributions in Fig.1a-d, and [23] for a unifying view).

Although both SA and variable-precision continuous resource models capture the error distributions well, continuous resource models are biologically more plausible [24]. While the notion of k slots is difficult to define neurally, resource limitation in continuous resource models is simply defined as neural gain, i.e., the amplitude of activation at the level of a neural population that has encoded a probability distribution over a given feature [23–26]. Higher gain means higher precision. Moreover, defining resource limitation as neural gain provides a natural link between WM and attention, which corroborates the close correspondence between the two at the level of brain [11] and behavior [27,28].

We must also note that a large body of research has examined the limits of memory by focusing on interference [29,30] and time [4,31] as two critical factors, but these models are not direct competitors of resource models, and are hence not discussed here. To summarize, despite some of the ongoing debates, the bulk of the evidence supports continuous resource models. However, most of the evidence comes from visual WM paradigms. Can the same principle be applied to verbal WM?

Extending continuous resource division from vision to language

Historically, visual and verbal WM have been assumed to be separate [32]. Empirically too, performance on visual and verbal domains are dissociable [33,34]. This dissociation is understandable given the very different nature of representations in these two domains. For one thing, visual stimuli are spread in *space*, whereas verbal stimuli are spread in *time*. Moreover, while the influence of category (color or phoneme labels) on perception has been reported in both domains [35,36], cf., [37,38], this influence is much stronger in the verbal domain. This is called categorical perception [36]; when presented with a series of artificially created stimuli between two phonemes, such as /b/ and /d/, participants do not experience them as a smooth continuum,

but as either /d/ or /b/. Categorical perception is a desirable property of speech processing, as differentiating between minimal pairs such as /beer/ and /deer/ critically depends on the identification of the right phoneme and not its acoustic details.

The continuous nature of visual representations makes a continuous resource model an intuitive model of visual WM, whereas the categorical nature of speech representations makes slot-based models a prime candidate for verbal WM (see Box 1). However, if the same principles operate across domains, continuous resource models should be able to account for the data in verbal WM. This proposition is not unreasonable, given that despite categorical perception, listeners are sensitive to subtle changes in the distribution of statistical information in acoustic dimensions even when the information does not flip the category [39,40], and such sensitivity is powerful enough to change production [41].

[Figure 1 about here]

To test this, the continuous reproduction paradigm was adapted to the auditory domain [42]. Four continua (/ba/–/da/, /ka/–/ga/, /la/–/ra/, and /sa/–/ʃa/) were created, each manipulating one acoustic property of the pair to generate seven equally-spaced. In the baseline phase, participants heard a single syllable and moved a continuous slider to mark the position of the syllable (Fig. 1e). This allowed for the calculation of a “deviation score” between the actual and the indicated position of the syllable, to measure precision. The results (Fig. 1e) showed that while the effect of category was not eliminated, participants were able to hear intermediate syllables. The authors then played 1, 2, or 4 syllables, sequentially, from different continua on each trial and probed one continuum at the end of the trial. After accounting for positional differences, the pattern was similar to that uncovered in visual WM paradigms, with a gradual increase in the deviation score as a function of increased set size even from 1 to 2 (Fig. 1f; see also [42] but see [43], for the criticism of their methodology).

Two follow-up experiments investigated the interaction between attention and WM, by using probabilistic cues. In the *pre-cue* condition, a number was presented before the sequence was played, indicating the position in the sequence where the syllable had the highest chance of being

probed. In the *retro-cue* condition, the same cue was presented but only after the syllables had been heard (Fig. 1g). The results showed that, for both pre- and retro-cues, deviation scores for the uncued items (i.e., those with a lower chance of being probed on a cued trial) were significantly higher than a baseline with no cueing. On the other hand, only pre-cues were successful at significantly decreasing deviation scores of the cued items (Fig.1h). Collectively, these experiments showed that the cued-item advantage was due to more precise encoding, whereas the uncued-item disadvantage was likely due to the release of the uncued items during maintenance. Importantly, this pattern was very similar to that reported for visual WM [29].

To summarize, when WM in speech perception is probed using a continuous reproduction paradigm, the results look strikingly similar to visual WM. This pattern is compatible with a continuous resource model, but, in theory, also with a mixed-state model that combines a discrete slot model with a pure guessing process. To minimize the chance of guessing, in another study, the framework was applied to speech production [44] (see Box 1, and Fig. 1i). Using a case-series approach, phonological errors were analyzed from four individuals with aphasia [44]. The precision of an error was defined as the distance between the target and error phoneme in the articulatory-phonetic space, weighted more heavily for features more diagnostic of minimal pairs, using ALINE ([45]; Fig. 1g). WM load was simply word length (3-9 letters; see Box 1). If phonological WM stores phonemes as discrete units, one would expect an increased probability of committing a phonological error with increasing word length, but there would be no reason to expect that such errors show a gradual increase in deviation from the target. Conversely, if phonemes in the phonological buffer have a graded representation, one would expect a continuous increase in deviation scores as a function of word length, as predicted by continuous resource models. This was indeed what was found (Fig. 1k). These findings match those reported for verbal WM in perception, as well as visual WM. Moreover, since phoneme selection within a word is unlikely to be driven by pure guessing (see Box 1), this pattern of results endorses the applicability of continuous resource models to verbal WM, despite the much greater utility of discrete categorical representations in speech vs. vision, pointing to domain-generalty in a core principle of processing.

[Box 1 about here]

Domain-generalities in neural substrates

A general framework has emerged in which WM is argued to be implemented in a network comprising sensory and fronto-parietal regions, with contributions from cerebellum, as well as subcortical regions including hippocampus, thalamus and basal ganglia [46–55]. Comprehensive reviews of the role of these regions exist elsewhere [49,56]. Rather, our goal here is to examine the existing evidence as it relates to domain-generalities or specificity of the neural correlates of WM. The target of such investigation will, obviously, not be the sensory cortex, as domain-specificity is undisputed in those areas. Rather, the focus will be on regions that are not clearly specialized to process a certain type of sensory information, such as the frontoparietal regions.

Although parietal regions have been implicated in both visual and verbal WM, different parts of the parietal lobe have been identified in these two domains. Specifically, the bulk of evidence in visual WM points to posterior parietal and intraparietal sulcus (IPS), whereas verbal WM studies often implicate the supramarginal gyrus (SMG) [57–65]. Even within the domain of language processing, researchers have identified separable neural correlates for semantic and phonological WM [66]. For example, [64] used representational similarity analysis (Fig. 3c) on data from participants judging either the phonological or semantic similarity of a memory item to a probe item (Fig. 3a, b). For the phonological task, they were able to decode phonological representations in superior temporal gyrus (STG, the speech perception region [67]) during encoding but not during maintenance across a delay period (Fig. 3d), whereas in the SMG, the opposite was the case (Fig. 3e; see also [63,65] for similar evidence using multi-voxel pattern analysis (MVPA) and transcranial magnetic stimulation (TMS)). On the other hand, semantic representations could not be decoded in either STG or SMG during either encoding or maintenance, showing the specificity of these regions for phonological WM. In contrast, both types of information could be decoded from angular gyrus (AG; Fig. 3f), although in a task-specific manner, e.g., phonological information could only be decoded in the phonological task but not in the semantic task.

Also informative about the neural correlates of WM are studies using lesion-symptom mapping (LSM) techniques, which examine the relationship between lesions and behavioral deficits in individuals with brain damage. Although several studies have examined the neural correlates of

WM using this method [68–70], very few studies have compared the neural regions involved in semantic and phonological processing within the same participants in tasks with comparable demands. In one such study [71], 94 participants were studied at the acute stage (within one week) of stroke, ruling out reorganization of function. To tap phonological WM, a digit matching task was used in which participants judged whether two lists were the same or different, and to tap semantic WM, a category probe task was used in which participants heard a word list and judged whether a probe word was in the same semantic category as any list word. For both tasks, WM load was manipulated by presenting lists of increasing length. Appropriate control measures were employed to identify regions specific to either semantic or phonological WM. In keeping with the neuroimaging results, the largest number of voxels relevant to phonological WM were in SMG (see also [72]). Several frontal and subcortical regions were also identified, all of which could plausibly be involved in articulatory rehearsal. In contrast, the largest regions related to semantic WM were in AG and a region spanning the opercular portion of the inferior frontal gyrus and the insula [73,74]. The next largest region was in the posterior superior temporal sulcus – a region often thought to link phonological with semantic representations [75]. These results show a clear double dissociation in neural correlates of semantic and phonological WM.

To summarize, studies of WM across visual and verbal domains, and even subdomains of language processing have implicated largely separable neural correlates for different types of information, pointing to domain-specificity in neural correlates. That said, there is also evidence that certain regions may be involved in aspects of processing that have a more domain-general flavor. For example, transcranial magnetic stimulation (TMS) of SMG affected the maintenance of a sequence regardless of the nature of items (verbal, spatial or motor)[76], raising the possibility that the sequencing demand of certain WM tasks may also critically involve this region, although this finding is not uncontested [64,74]. Similarly, as discussed earlier, both phonological and semantic representations could be decoded from AG, although each in a task-specific manner [64]. This finding fits well with the role of AG in episodic memory [77], which would predict its involvement in processing various kinds of information but each related to a specific context. Finally, there is ample evidence for the involvement of prefrontal regions, thalamus and basal ganglia in balancing maintenance and updating functions of WM (Box 2). It is reasonable to assume that, in so far as balancing these demands is required in WM tasks, the neural correlates involved in these gating

processes could be shared between tasks. However, these loops are triggered by sensory-motor mapping and task goals, which are, by definition, domain and task-specific.

[Figure 2 about here]

[Box 2 + Fig. 1 about here]

Domain-generalty in application

The assumption of domain-generalty has had two important outcomes: the emergence of WM training programs, and certain practices for the analysis of individual differences. Given the broader impact of the former, we focus on training in this section. However, the latter can also impact science in important ways, as discussed in Box 4. The general idea behind training programs is that extensive (and usually adaptive) training of WM using a small number of (computer-based) tasks will improve WM capacity across a wide array of tasks in real life (i.e., far transfer; [78–80]). The efficacy of brain-training methods in general, and WM training in specific, has been one of the most hotly debated topics in the past two decades. The reason is that WM is considered to underlie a very wide range of cognitive tasks, act as a building block of general intelligence [81,82], and be a critical locus of impairment in many neuropsychiatric conditions such as schizophrenia and attention-deficit hyperactivity disorder [83,84]. Therefore, it is natural that an easy tool for training WM would be highly appealing to both neurotypical and impaired populations. In fact, sales in the digital-brain-health market already well exceed \$1 billion, with more than one case of charges brought up by the Federal Trade Commission against brain-training companies for deceptive advertising practices [85].

There are numerous empirical studies, meta-analyses, and review papers arguing for or against the efficacy of WM training [86–90]. Similar disagreements exist about aging effects on WM training, with studies ranging from no effect to differential effects of training on some but not other tasks [91–93]. As with neural correlates of WM, our goal is not to review this literature but rather to examine whether far transfer is predicted by current, biologically plausible, theories of WM. Currently, three main theoretical accounts exist for transfer: neuroplasticity [94,95], representational overlap [e.g., 94,120], and skill learning accounts [97,98].

Proponents of far transfer often cite neuroplasticity as the foundation of their theoretical perspective. It is well-established that training WM induces plastic changes in certain brain regions (e.g., dorsolateral prefrontal cortex, DLPFC) by changing the local field potentials and recruiting a greater proportion of neurons with increased firing rates [99], and that such plasticity improves performance on trained tasks [94,100]. It is also well-known, as reviewed in the earlier sections, that there is some overlap in neural regions involved in WM processes across various tasks and domains [95]. Combining these two uncontroversial findings, it has been reasoned that, if there is some domain-generalty in neural correlates of WM and if such neural regions show plasticity, then inducing plastic changes through one task should lead to improved performance in other tasks [94,101]. Indeed, single-cell recordings from the DLPFC of rhesus monkeys show that changes induced during training are sufficient to benefit near-transfer, i.e., changes to the task parameters such as delay period or stimulus location in visual WM paradigms [99], and similar effects have been observed in humans [102]. The critical question, however, is whether far-transfer can be expected, when there is little or no overlap between stimuli and/or task goals, except for the necessity of holding on to information over short periods of time. The answer, according to recent neurobiologically plausible computational models of WM is no. Box 3 discusses an example model [103]. Critically, the model implements the uncontested assumptions of neuroplasticity and the undifferentiated neural space that overlaps across tasks. Nevertheless, while training does improve performance on the same task, as well as new stimuli within task [104], far transfer is not predicted by these models (see Box 3 for details).

[Box 3 + Fig. I about here]

The representational overlap view claims that the extent of transfer of WM training between two tasks depends on the degree of overlap between their features/processes. Although less general than the neuroplasticity account, the representational overlap account still predicts automatic transfer, and attributes the absence of transfer to the lack of overlap in the appropriate representations [e.g., 94,120]. It is, however, unclear which features and processes constitute the critical representational mediating transfer. A large-scale and well-controlled study [98] demonstrated that overlap in stimulus, response, or task, alone was not sufficient to reliably predict transfer, casting doubt on the utility of the representational account. The authors of this study, in turn, proposed the skill learning account.

According to the skill learning account, transfer is only expected when a new cognitive routine is learned during training that can later be applied to a new task [97,98]. Importantly, and in opposition to the other two accounts, transfer is not viewed as automatic and may very well comprise strategies that lie outside of core WM processes. No significant transfer is expected if the tasks are either too well-learned to allow room for new strategies to develop during training, or if they are too different to benefit from the application of the same newly developed cognitive routine. For example, N-back tasks, often used for WM training, entail routines that are not applicable to many other WM tasks, which explains the absence of transfer between them [105]. Similarly, the authors reported substantial transfer between complex span tasks within the same stimulus domain, but not across domains. This result indicates that cognitive routines may depend strongly on task and domain, making the skill learning view the most restricted of the three perspectives in predicting generalization in WM training, but also the most compatible with both theoretical models of WM and empirical evidence. Critically, the success of the account hinges on the fact that it does not posit an automatic mechanism for far transfer solely based on shared representations or neural substrates. Rather, it emphasizes the development of new cognitive pipelines through learning, some of which are outside of primary WM operations.

In summary, while the assumption of domain-generality in neural implementation is very much part of the cutting-edge computational models of WM, this assumption does not lead to the prediction of domain-generality in application. Quite the opposite, these models predict task-specific patterns of neural activity, and at best, generalization to new items within the same task

(see also Box 4 for complementary evidence from the analysis of individual differences). At the same time, a new and more restrictive perspective on transfer, namely transfer by learning new cognitive routines, shows much promise, although there are still open questions regarding what constitutes a new cognitive routine and what factors promote the learning of such routines in individuals.

Concluding remarks

The goal of this paper was to show that while the question of domain-generality of WM is an important one, to be truly informative, it must be posed separately for computations, neural substrates, and application, as the answer differs for each. We showed that, in terms of computations that are potentially applicable to different representations, WM is indeed domain-general. In terms of neural correlates, it contains both domain-general and domain-specific elements. Finally, in terms of application, it is mostly domain-specific. Thus, if progress is to be made, researchers must be clear about which aspect of domain-generality is being addressed, and mind the limits of logical inference when extending claims of domain-generality across these three components. The current review encourages a greater focus on cross-domain investigation of theoretical principles involved in WM operations, such as sequencing, and their potentially domain-general neural correlates, while at the same time implying a need for understanding why some non-sensory regions behave in a domain-specific manner in WM operations (see Outstanding Questions). This overview also encourages a major shift in perspectives regarding WM training, from an expectation of transfer as a natural and automatic consequence of domain-generality in neural correlates to a view of transfer as skill learning.

Acknowledgments

The authors would like to thank Rich Shiffrin and Rob Nosofsky for helpful discussions, Valeri Camos for her thoughtful comments on the manuscript, and Flora Bouchacourt and Timothy Buschman for their permission to reuse a figure. This work was supported by grants NSF-BCS-2346989 to N.N., NIH R01DC014976 to the Baylor College of Medicine, and T.L.L. Temple Foundation Neuroplasticity Laboratory Award to R.C.M.

References

1. Baddeley, A.D. and Hitch, G. (1974) Working Memory. In *Psychology of Learning and Motivation* 8 (Bower, G. H., ed), pp. 47–89, Academic Press
2. Cowan, N. *et al.* (2020) An embedded-processes approach to working memory. *Working Memory: The state of the science* 44
3. Logie, R.H. *et al.* (2021) Integrating theories of working memory. *Working memory: State of the science*
4. Burgess, N. and Hitch, G.J. (1999) Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review* 106, 551–581
5. Oberauer, K. and Lewandowsky, S. (2011) Modeling working memory: a computational implementation of the Time-Based Resource-Sharing theory. *Psychon Bull Rev* 18, 10–45
6. Logie, R. *et al.* (2020) *Working Memory: The state of the science*, Oxford University Press
7. Cowan, N. (2001) The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24, 87–114
8. Luck, S.J. and Vogel, E.K. (1997) The capacity of visual working memory for features and conjunctions. *Nature* 390, 279–281
9. Miller, G.A. (1956) The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63, 81–97
10. Awh, E. and Jonides, J. (2001) Overlapping mechanisms of attention and spatial working memory. *Trends in Cognitive Sciences* 5, 119–126
11. Panichello, M.F. and Buschman, T.J. (2021) Shared mechanisms underlie the control of working memory and attention. *Nature* 592, 601–605
12. Ma, W.J. *et al.* (2014) Changing concepts of working memory. *Nature neuroscience* 17, 347
13. Wilken, P. and Ma, W.J. (2004) A detection theory account of change detection. *Journal of Vision* 4, 11
14. Zhang, W. and Luck, S.J. (2008) Discrete fixed-resolution representations in visual working memory. *Nature* 453, 233–235
15. Cowan, N. and Rouder, J.N. (2009) Comment on “Dynamic Shifts of Limited Working Memory Resources in Human Vision.” *Science* 323, 877–877
16. Bays, P.M. (2018) Failure of self-consistency in the discrete resource model of visual working memory. *Cognitive Psychology* 105, 1–8
17. van den Berg, R. *et al.* (2012) Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences* 109, 8780–8785
18. Williams, J.R. *et al.* (2022) You cannot “count” how many items people remember in visual working memory: The importance of signal detection–based measures for understanding change detection performance. *Journal of Experimental Psychology: Human Perception and Performance* 48, 1390–1409
19. van den Berg, R. *et al.* (2014) Factorial comparison of working memory models. *Psychological Review* 121, 124–149
20. Fougny, D. *et al.* (2012) Variability in the quality of visual working memory. *Nat Commun* 3, 1229
21. Keshvari, S. *et al.* (2012) Probabilistic Computation in Human Perception under Variability in Encoding Precision. *PLOS ONE* 7, e40216

22. Nosofsky, R.M. and Donkin, C. (2016) Response-time evidence for mixed memory states in a sequential-presentation change-detection task. *Cognitive Psychology* 84, 31–62
23. Schneegans, S. *et al.* (2020) Stochastic sampling provides a unifying account of visual working memory limits. *Proceedings of the National Academy of Sciences* 117, 20959–20968
24. Bays, P.M. (2015) Spikes not slots: noise in neural populations limits working memory. *Trends in Cognitive Sciences* 19, 431–438
25. Ma, W.J. *et al.* (2006) Bayesian inference with probabilistic population codes. *Nat Neurosci* 9, 1432–1438
26. van den Berg, R. and Ma, W.J. (2018) A resource-rational theory of set size effects in human visual working memory. *eLife* 7, e34963
27. Kane, M.J. *et al.* (2001) A controlled-attention view of working-memory capacity. *Journal of Experimental Psychology: General* 130, 169–183
28. Rosselet-Jordan, F.L. *et al.* (2022) Role of attention in the associative relatedness effect in verbal working memory: Behavioral and chronometric perspectives. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 48, 1571
29. Oberauer, K. and Lin, H.-Y. (2017) An interference model of visual working memory. *Psychological Review* 124, 21–59
30. Nairne, J.S. (1990) A feature model of immediate memory. *Memory & Cognition* 18, 251–269
31. Barrouillet, P. *et al.* (2004) Time constraints and resource sharing in adults' working memory spans. *Journal of experimental psychology: General* 133, 83
32. Baddeley, A. (1992) Working Memory. *Science* 255, 556–559
33. Fougny, D. *et al.* (2015) Working memory storage is intrinsically domain specific. *Journal of Experimental Psychology: General* 144, 30–47
34. Logie, R.H. *et al.* (2014) When cognitive performance does not decline across the lifespan. In *Working Memory and Ageing*, Psychology Press
35. Holmes, K.J. and Regier, T. (2017) Categorical Perception Beyond the Basic Level: The Case of Warm and Cool Colors. *Cognitive Science* 41, 1135–1147
36. Liberman, A.M. *et al.* (1957) The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54, 358–368
37. Brown, A.M. *et al.* (2011) Color names, color categories, and color-cued visual search: Sometimes, color perception is not categorical. *Journal of Vision* 11, 2
38. McMurray, B. (2022) The myth of categorical perception. *The Journal of the Acoustical Society of America* 152, 3819–3842
39. Hodson, A.J. *et al.* (2023) Statistical learning across passive listening adjusts perceptual weights of speech input dimensions. *Cognition* 238, 105473
40. Idemaru, K. and Holt, L.L. (2020) Generalization of dimension-based statistical learning. *Attention, Perception, & Psychophysics* 82, 1744–1762
41. Murphy, T.K. *et al.* (2023) Transfer of statistical learning from passive speech perception to speech production. *Psychon Bull Rev* DOI: 10.3758/s13423-023-02399-8
42. Joseph, S. *et al.* (2015) Precision of working memory for speech sounds. *The Quarterly Journal of Experimental Psychology* 68, 2022–2040
43. Hepner, C.R. and Nozari, N. (2019) Resource allocation in phonological working memory: Same or different principles from vision? *Journal of Memory and Language* 106, 172–188

44. Black, J. and Nozari, N. (2023) Precision of phonological errors in aphasia supports resource models of phonological working memory in language production. *Cognitive Neuropsychology* 40, 1–24
45. Kondrak, G. (2003) Phonetic Alignment and Similarity. *Computers and the Humanities* 37, 273–291
46. Ashida, R. *et al.* (2019) Sensorimotor, language, and working memory representation within the human cerebellum. *Human Brain Mapping* 40, 4732–4747
47. Borders, A.A. *et al.* (2022) The hippocampus supports high-precision binding in visual working memory. *Hippocampus* 32, 217–230
48. Brissenden, J.A. *et al.* (2021) Stimulus-Specific Visual Working Memory Representations in Human Cerebellar Lobule VIIb/VIIIa. *J. Neurosci.* 41, 1033–1045
49. Christophel, T.B. *et al.* (2017) The Distributed Nature of Working Memory. *Trends in Cognitive Sciences* 21, 111–124
50. Devereett, B. *et al.* (2019) Cerebellar disruption impairs working memory during evidence accumulation. *Nat Commun* 10, 3128
51. Dimakopoulos, V. *et al.* (2022) Information flows from hippocampus to auditory cortex during replay of verbal working memory items. *eLife* 11, e78677
52. Liu, W. *et al.* (2021) Substantia Nigra Integrity Correlates with Sequential Working Memory in Parkinson’s Disease. *J. Neurosci.* 41, 6304–6313
53. Song, D. *et al.* (2020) The lateralization of left hippocampal CA3 during the retrieval of spatial working memory. *Nat Commun* 11, 2901
54. Wang, Y. *et al.* (2021) A cortico-basal ganglia-thalamo-cortical channel underlying short-term memory. *Neuron* 109, 3486–3499
55. Zhang, Y. *et al.* (2021) Cholinergic suppression of hippocampal sharp-wave ripples impairs working memory. *Proceedings of the National Academy of Sciences* 118, e2016432118
56. Emch, M. *et al.* (2019) Neural Correlates of Verbal Working Memory: An fMRI Meta-Analysis. *Front. Hum. Neurosci.* 13
57. Crottaz-Herbette, S. *et al.* (2004) Modality effects in verbal working memory: differential prefrontal and parietal responses to auditory and visual stimuli. *NeuroImage* 21, 340–351
58. Duma, G.M. *et al.* (2019) Functional dissociation of anterior cingulate cortex and intraparietal sulcus in visual working memory. *Cortex* 121, 277–291
59. Lefco, R.W. *et al.* (2020) Gradients of functional organization in posterior parietal cortex revealed by visual attention, visual short-term memory, and intrinsic functional connectivity. *NeuroImage* 219, 117029
60. Li, X. *et al.* (2022) Delay activity during visual working memory: A meta-analysis of 30 fMRI experiments. *NeuroImage* 255, 119204
61. Praß, M. and de Haan, B. (2019) Multi-target attention and visual short-term memory capacity are closely linked in the intraparietal sulcus. *Human Brain Mapping* 40, 3589–3605
62. Xu, Y. (2023) Parietal-driven visual working memory representation in occipito-temporal cortex. *Current Biology* 33, 4516–4523.e5
63. Yue, Q. *et al.* (2019) Non-perceptual Regions in the Left Inferior Parietal Lobe Support Phonological Short-term Memory: Evidence for a Buffer Account? *Cereb Cortex* 29, 1398–1413
64. Yue, Q. and Martin, R.C. (2021) Maintaining verbal short-term memory representations in non-perceptual parietal regions. *Cortex* 138, 72–89

65. Yue, Q. and Martin, R.C. (2022) Phonological Working Memory Representations in the Left Inferior Parietal Lobe in the Face of Distraction and Neural Stimulation. *Front. Hum. Neurosci.* 16
66. Martin, R.C. *et al.* (2021) Domain-specific working memory: Perspectives from cognitive neuropsychology. In *Working memory: State of the science*, pp. 235–281, Oxford University Press
67. Nozari, N. (2022) The neural basis of word production. In *The Oxford Handbook of the Mental Lexicon* (Papafragou, A. *et al.*, eds), pp. 0, Oxford University Press
68. Ghaleh, M. *et al.* (2020) Dissociable Mechanisms of Verbal Working Memory Revealed through Multivariate Lesion Mapping. *Cerebral Cortex* 30, 2542–2554
69. Ivanova, M.V. *et al.* (2018) Neural mechanisms of two different verbal working memory tasks: A VLSM study. *Neuropsychologia* 115, 25–41
70. Pisoni, A. *et al.* (2019) The neural correlates of auditory-verbal short-term memory: a voxel-based lesion-symptom mapping study on 103 patients after glioma removal. *Brain Struct Funct* 224, 2199–2211
71. Ding, J. *et al.* (2020) Dissociation between frontal and temporal-parietal contributions to connected speech in acute stroke. *Brain* 143, 862–876
72. Pillay, S.B. *et al.* (2014) Cerebral localization of impaired phonological retrieval during rhyme judgment. *Annals of Neurology* 76, 738–746
73. Barde, L.H. *et al.* (2006) The role of left inferior frontal gyrus (LIFG) in semantic short-term memory: A comparison of two case studies. *Brain and Language* 99, 82–83
74. Hamilton, A.C. *et al.* (2009) Converging functional magnetic resonance imaging evidence for a role of the left inferior frontal lobe in semantic retention during language comprehension. *Cognitive Neuropsychology* 26, 685–704
75. Hickok, G. *et al.* (2009) Area Spt in the Human Planum Temporale Supports Sensory-Motor Integration for Speech Processing. *Journal of Neurophysiology* 101, 2725–2732
76. Guidali, G. *et al.* (2019) Keeping order in the brain: The supramarginal gyrus and serial order in short-term memory. *Cortex* 119, 89–99
77. Humphreys, G.F. *et al.* (2021) A Unifying Account of Angular Gyrus Contributions to Episodic and Semantic Cognition. *Trends in Neurosciences* 44, 452–463
78. Jaeggi, S.M. *et al.* (2008) Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences* 105, 6829–6833
79. Klingberg, T. *et al.* (2002) Training of Working Memory in Children With ADHD. *Journal of Clinical and Experimental Neuropsychology* 24, 781–791
80. Morrison, A.B. and Chein, J.M. (2011) Does working memory training work? The promise and challenges of enhancing cognition by training working memory. *Psychon Bull Rev* 18, 46–60
81. Conway, A.R.A. *et al.* (2003) Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences* 7, 547–552
82. Shipstead, Z. *et al.* (2012) Is working memory training effective? *Psychological Bulletin* 138, 628–654
83. Braun, U. *et al.* (2021) Brain network dynamics during working memory are modulated by dopamine and diminished in schizophrenia. *Nat Commun* 12, 3478
84. Ramos, A.A. *et al.* (2020) A meta-analysis on verbal working memory in children and adolescents with ADHD. *The Clinical Neuropsychologist* 34, 873–898

85. Simons, D.J. *et al.* (2016) Do “Brain-Training” Programs Work? *Psychol Sci Public Interest* 17, 103–186
86. Au, J. *et al.* (2016) There is no convincing evidence that working memory training is NOT effective: A reply to Melby-Lervåg and Hulme (2015). *Psychon Bull Rev* 23, 331–337
87. Karbach, J. and Verhaeghen, P. (2014) Making Working Memory Work: A Meta-Analysis of Executive-Control and Working Memory Training in Older Adults. *Psychol Sci* 25, 2027–2037
88. Melby-Lervåg, M. and Hulme, C. (2016) There is no convincing evidence that working memory training is effective: A reply to Au *et al.* (2014) and Karbach and Verhaeghen (2014). *Psychon Bull Rev* 23, 324–330
89. Pergher, V. *et al.* (2020) Divergent Research Methods Limit Understanding of Working Memory Training. *J Cogn Enhanc* 4, 100–120
90. Sala, G. and Gobet, F. (2020) Working memory training in typically developing children: A multilevel meta-analysis. *Psychon Bull Rev* 27, 423–434
91. Zinke, K. *et al.* (2014) Working memory training and transfer in older adults: Effects of age, baseline performance, and training gains. *Developmental Psychology* 50, 304–315
92. Brehmer, Y. *et al.* (2012) Working-memory training in younger and older adults: training gains, transfer, and maintenance. *Front. Hum. Neurosci.* 6
93. Guye, S. and von Bastian, C.C. (2017) Working memory training in older adults: Bayesian evidence supporting the absence of transfer. *Psychology and Aging* 32, 732–746
94. Klingberg, T. (2010) Training and plasticity of working memory. *Trends in Cognitive Sciences* 14, 317–324
95. McNab, F. and Klingberg, T. (2008) Prefrontal cortex and basal ganglia control access to working memory. *Nat Neurosci* 11, 103–107
96. von Bastian, C.C. *et al.* (2019) How strong is the evidence for the effectiveness of working memory training. *Cognitive and working memory training: Perspectives from psychology, neuroscience, and human development* 58, 1–23
97. Fellman, D. *et al.* (2020) The role of strategy use in working memory training outcomes. *Journal of Memory and Language* 110, 104064
98. Gathercole, S.E. *et al.* (2019) Working memory training involves learning new skills. *Journal of Memory and Language* 105, 19–42
99. Tang, H. *et al.* (2022) Prefrontal cortical plasticity during learning of cognitive tasks. *Nat Commun* 13, 90
100. von Bastian, C.C. and Oberauer, K. (2014) Effects and mechanisms of working memory training: a review. *Psychological Research* 78, 803–820
101. Iordan, A.D. *et al.* (2020) Neural correlates of working memory training: Evidence for plasticity in older adults. *NeuroImage* 217, 116887
102. Holmes, J. *et al.* (2019) Are Working Memory Training Effects Paradigm-Specific? *Frontiers in Psychology* 10
103. Bouchacourt, F. and Buschman, T.J. (2019) A flexible model of working memory. *Neuron* 103, 147–160
104. Lundqvist, M. *et al.* (2023) Working memory control dynamics follow principles of spatial computing. *Nat Commun* 14, 1429
105. Soveri, A. *et al.* (2017) Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychon Bull Rev* 24, 1077–1096

106. Dell, G.S. (1986) A spreading-activation theory of retrieval in sentence production. *Psychological Review* 93, 283–321
107. Nozari, N. and Pinet, S. (2020) A critical review of the behavioral, neuroimaging, and electrophysiological studies of co-activation of representations during word production. *Journal of Neurolinguistics* 53, 100875
108. Pinet, S. and Nozari, N. (2018) “Twisting fingers”: The case for interactivity in typed language production. *Psychon Bull Rev* 25, 1449–1457
109. Nozari, N. and Dell, G.S. (2009) More on lexical bias: How efficient can a “lexical editor” be? *Journal of Memory and Language* 60, 291–307
110. Caramazza, A. *et al.* (1986) The role of the (output) phonological buffer in reading, writing, and repetition. *Cognitive Neuropsychology* 3, 37–76
111. Fischer-Baum, S. (2018) Chapter Two - A Common Representation of Serial Position in Language and Memory. In *Psychology of Learning and Motivation* 68 (Federmeier, K. D. and Watson, D. G., eds), pp. 31–54, Academic Press
112. Nozari, N. and Dell, G.S. (2013) How damaged brains repeat words: A computational approach. *Brain and Language* 126, 327–337
113. O'Reilly, R.C. and Frank, M.J. (2006) Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. *Neural Computation* 18, 283–328
114. Lin, H.-Y. and Oberauer, K. (2022) An interference model for visual working memory: Applications to the change detection task. *Cognitive Psychology* 133, 101463
115. Boran, E. *et al.* (2019) Persistent hippocampal neural firing and hippocampal-cortical coupling predict verbal working memory load. *Science Advances* 5, eaav3687
116. Murray, J.D. *et al.* (2017) Stable population coding for working memory coexists with heterogeneous neural dynamics in prefrontal cortex. *Proceedings of the National Academy of Sciences* 114, 394–399
117. Bocincova, A. *et al.* (2022) Neural signature of flexible coding in prefrontal cortex. *Proceedings of the National Academy of Sciences* 119, e2200400119
118. Rigotti, M. *et al.* (2013) The importance of mixed selectivity in complex cognitive tasks. *Nature* 497, 585–590
119. Daneman, M. and Carpenter, P.A. (1980) Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior* 19, 450–466
120. Conway, A.R.A. (1996) Individual Differences in Working Memory Capacity: More Evidence for a General Capacity Theory. *Memory* 4, 577–590
121. Oberauer, K. *et al.* (2007) Individual differences in working memory capacity and reasoning ability. *Variation in working memory*
122. Cokely, E.T. *et al.* (2006) Sources of individual differences in working memory: Contributions of strategy to capacity. *Psychon Bull Rev* 13, 991–997
123. Arslan, B. *et al.* (2023) Trust my gesture or my word: How do listeners choose the information channel during communication? *Journal of Experimental Psychology: Learning, Memory, and Cognition* DOI: 10.1037/xlm0001253
124. De Renzi, E. and Nichelli, P. (1975) Verbal and Non-Verbal Short-Term Memory Impairment Following Hemispheric Damage. *Cortex* 11, 341–354
125. Shah, P. and Miyake, A. (1996) The separability of working memory resources for spatial thinking and language processing: An individual differences approach. *Journal of Experimental Psychology: General* 125, 4–27

126. Martin, R.C. *et al.* (1994) Language Processing and Working Memory: Neuropsychological Evidence for Separate Phonological and Semantic Capacities. *Journal of Memory and Language* 33, 83–111
127. Martin, R.C. and He, T. (2004) Semantic short-term memory and its role in sentence processing: A replication. *Brain and Language* 89, 76–82
128. Martin, R.C. (2021) The Critical Role of Semantic Working Memory in Language Comprehension and Production. *Curr Dir Psychol Sci* 30, 283–291
129. Martin, R.C. and Schnur, T.T. (2019) Independent contributions of semantic and phonological working memory to spontaneous speech in acute stroke. *Cortex* 112, 58–68
130. Tan, Y. and Martin, R.C. (2018) Verbal short-term memory capacities and executive function in semantic and syntactic interference resolution during sentence comprehension: Evidence from aphasia. *Neuropsychologia* 113, 111–125

Figure captions and Boxes

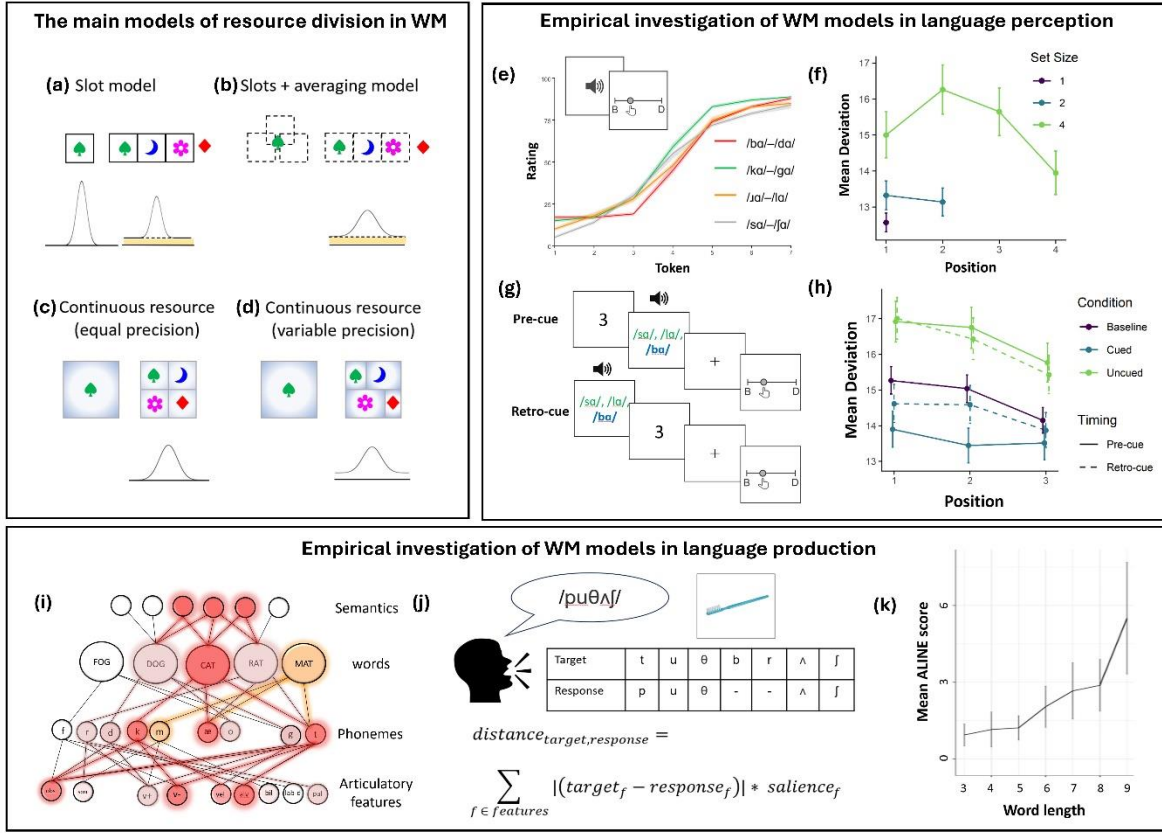


Figure 1. Theoretical and empirical aspects of resource division in WM. (a-d) Schematics of four accounts of resource division when one item (left) or four items (right) are to be stored, adapted from [12]. (a) Slot model ($k = 3$). (b) Slots + averaging (SA) model. (c) Continuous resource (equal precision) model. (d) Continuous resource (variable precision) model. Distributions under each figure show error around the true feature value of the probed item. This distribution is similar across models when one item is to be stored (hence shown once). When the number of items increases beyond capacity, all models predict increased error. In slot models, this distribution is a mixture of high-precision responses (white) and random guesses (yellow); in SA model, a mixture of low-resolution recall (white) and random guesses (yellow). In equal-precision continuous resource model, error increases continuously with the number of items, and in variable-precision, the distribution is an infinite mixture of distributions with varying error rates. (a) and (b) are mixed-state models. (e) Ratings of tokens in the perceptual task. (f) Changes to mean

deviation scores in the perceptual task as a function of number of items and position. **(g)** The cueing paradigm. Uncued = non-cued items on a cued trial. Baseline (not shown) had no cues. **(h)** Mean deviations scores in (g). **(i)** Schematic of the activation in the language production system for target “cat” (Box 1). Orange indicates activation through feedback. **(j)** Example trial and ALINE coding in the production task. **(k)** Mean ALINE distance as a function of word length in (j).

Box 1. Phonological errors as a tool for studying WM

Origin of phonological errors. The word production system is a hierarchical system comprising several layers of representations, from semantics to articulatory-phonetic features that guide motor production [67,106,107] (Fig. 1i). To say a word like “cat”, speakers must first activate the concept of cat (shown in Fig. 1i as distributed semantic features). Activation then spreads to lexical items, phonemes, and articulatory phonetic features (shown only partially for clarity). The system has three key properties that create potential for errors: (a) Spreading activation activates all the representations that are connected to the currently activated nodes, regardless of their target/non-target status. (b) The system is cascaded, meaning that even items that are not selected in one layer still send some activation down to subsequent layers. (c) The system is interactive [108,109], meaning that activation not only flows from upper to lower layers, but also backwards (shown only in one part of the system for visual clarity). This feature is important for producing phonological errors (e.g., “mat” instead of “cat”). When the lexical item “cat” is activated, it activates phonemes /k/, /æ/ and /t/. Through feedback between phonemes and lexical items, /æ/ and /t/ activate “mat”, which in turn activates its onset /m/, the misselection of which leads to the phonological error “mat”. As such, activation reverberates between phonemes and lexical items that support them, making the retrieval of other phonemes in the same words easier.

Relationship to WM. Several findings suggest that sequencing phonemes within a word taps into similar memory processes as memorizing items within a list, albeit less explicitly. The evidence includes the length effect, i.e., an increase in the probability of phonological errors in reading, writing, and repetition, as the word length increases, above and beyond chance [110], as well as strikingly similar positional effects in phonological errors in picture naming and reading and memory errors in serial recall tasks [111]. Furthermore, the inability to maintain information about

phonology in WM (e.g., whether a probe word rhymed with another word in a list) is predictive of phonological errors [112]. However, while items in a typical serial recall task are usually independent of one another, hence giving rise to responses that are simply guesses, phonemes within a word are not. Even if a phoneme is hard to retrieve, it still receives support from the lexical item, as well as other correctly retrieved phonemes. This dynamic makes selection by pure guessing a much less likely scenario in word production compared to a typical memory experiment.

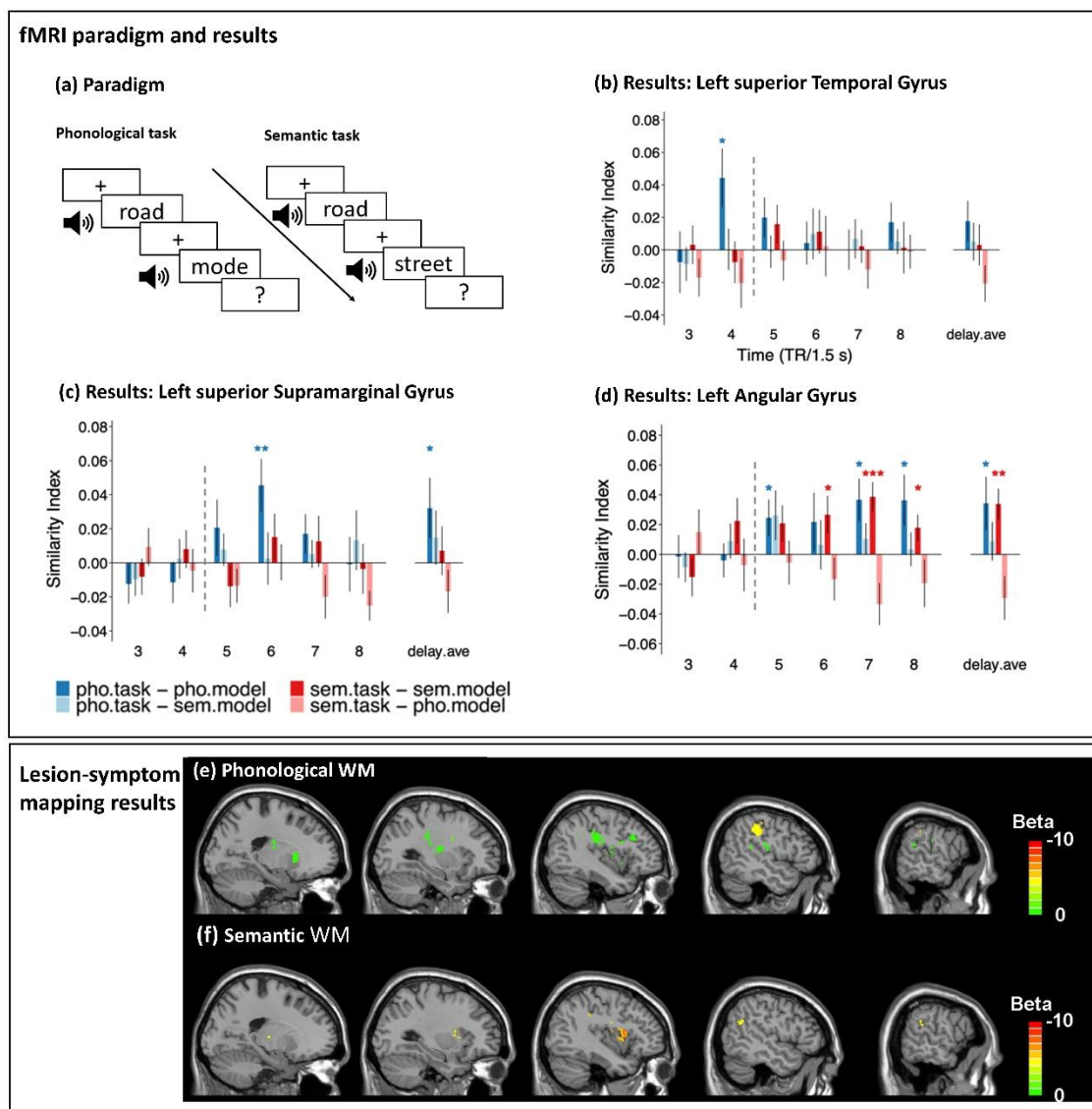


Figure 2. Neural correlates of phonological and semantic WM. (a) Paradigm for testing phonological and semantic WM. (b-d) ROI-based results of representational similarity analysis in (b) Left superior temporal gyrus (STG), (c) left supramarginal gyrus (SMG), and (d) left angular gyrus (AG). Error bars represent the standard error of the mean. Dashed lines indicate the typical boundaries between the encoding period and the delay period. pho: phonological; sem: semantic. Asterisks indicate the significance of one-sample t-test: * $p < .05$, ** $p < .01$, *** $p < .001$. Adapted from [64]. (e) Beta values of the regions significantly associated with decreased performance in phonological and (f) semantic WM after accounting for lesion volume, input processing, and the respective opposing WM task (p values < 0.05); adapted from [71].

Box 2. What might domain-general neural regions do in WM tasks?

Aside from regions in the parietal cortex discussed in the text such as AG, several other regions, including prefrontal cortex (PFC), thalamus, and basal ganglia have been implicated in domain-general WM processes. One of the most detailed accounts of the role of these regions is [113]’s gating model, a schematic of which is shown in Figure I. The model explains the trade-off between maintaining and updating memories in PFC over short periods of time. Sensory input is mapped onto motor outputs via posterior cortical regions. PFC’s role is to modulate this mapping based on prior information and task goals. PFC’s flexibility in updating is adjusted through gating via baso-cortical loops. Thalamus always excites the PFC, however this excitation is inhibited by substantia nigra (SN). Two pathways via dorsal striatum modulate this interaction. The direct pathway is activated when the Go neurons with excitatory D1 receptors in dorsal striatum are activated. They directly inhibit SN, which leads to the disinhibition of thalamus and updating in PFC. The indirect pathway is activated when the NoGo neurons with inhibitory D2 receptors in dorsal striatum are activated. This pathway counteracts the direct pathway by inhibiting globus pallidus. Since globus pallidus itself has an inhibitory effect on SN, its inhibition leads to the activation of SN and further inhibition of thalamus, which prevents updating in PFC.

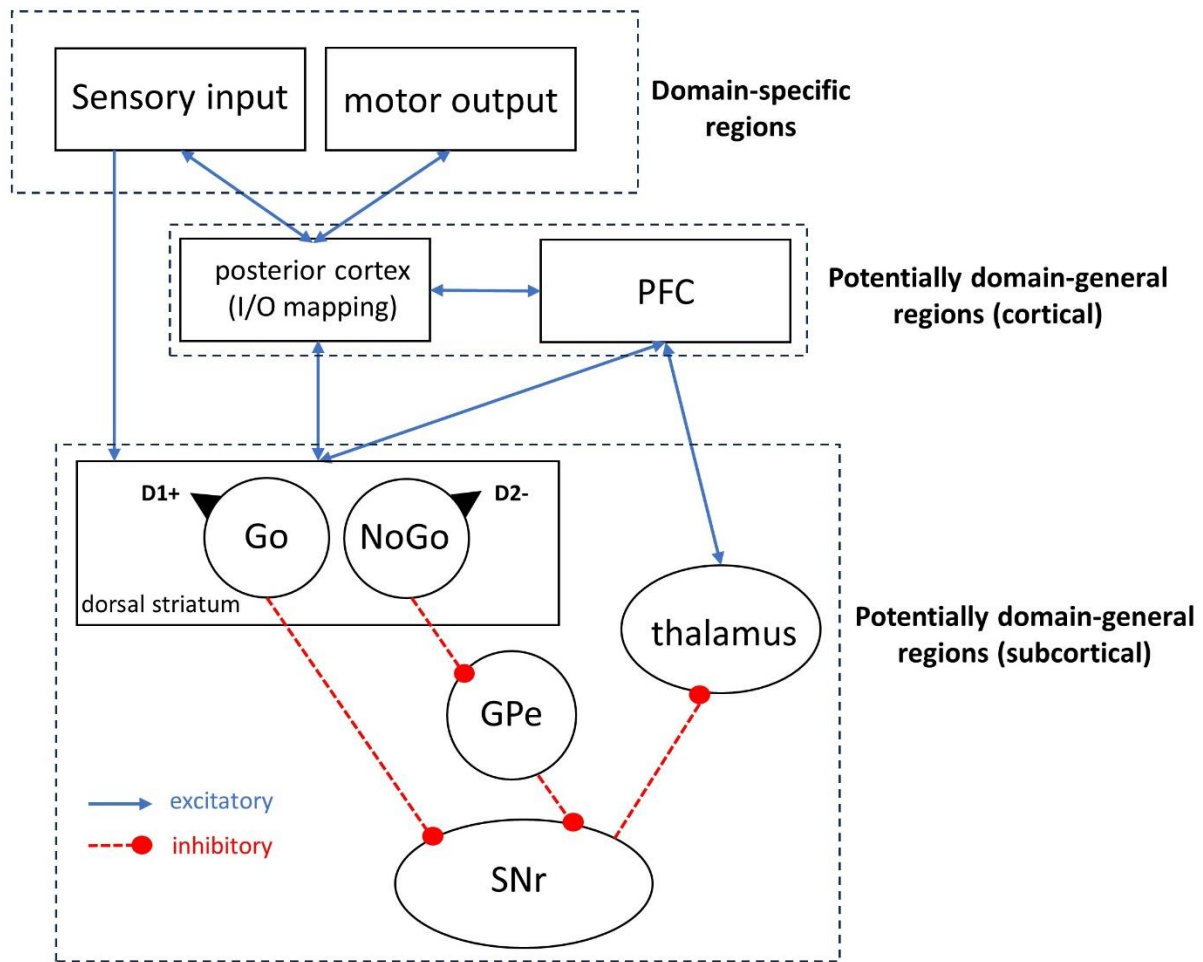


Figure I. Schematic of potentially domain-general and domain-specific regions in WM. GPe = globus pallidus, I = input, O = output, PFC = prefrontal cortex, SNr = substantia nigra. Adapted from [113].

Box 3. Do biologically-plausible models of WM predict far transfer?

Models of WM must explain many properties, such as its flexible nature for temporarily storing various kinds of information, its capacity limitation, and sensitivity to interference [114], as well as neural findings, such as the increase in neural activity with WM load [115], involvement of a distributed network comprising sensory cortex and frontoparietal and hippocampal regions [49], and the dual static-dynamic nature of neural representations in WM [116]. One recent model that successfully captures all of this is [103] (Fig. I), which models WM as random and recurrent connectivity between a structured network (representing sensory regions) and an unstructured random network (representing regions such as PFC and hippocampus; see also newer versions that also implement fast Hebbian synaptic plasticity [117]).

The idea of WM as a workspace is instantiated by the random connections to the random network, which generate a high-dimensional space within which various types of information can be temporarily stored, such as those reported in PFC or the hippocampus [118]. As such, the model very much embraces the idea of undifferentiated neural regions involved in WM operations across various tasks and domains, and could be considered a great candidate for demonstrating domain-generalization in application. Yet, even within such a model, while learning optimizes performance on trained items, benefits do not generalize to untrained items. The reason is that training increases the correlation between the input in sensory regions and the random network, and is therefore specific to those connections. Most recent advances, such as spatial computing models, allow generalization to new items within the same task, but even in these models, different tasks generate unique patterns of neural activity [104].

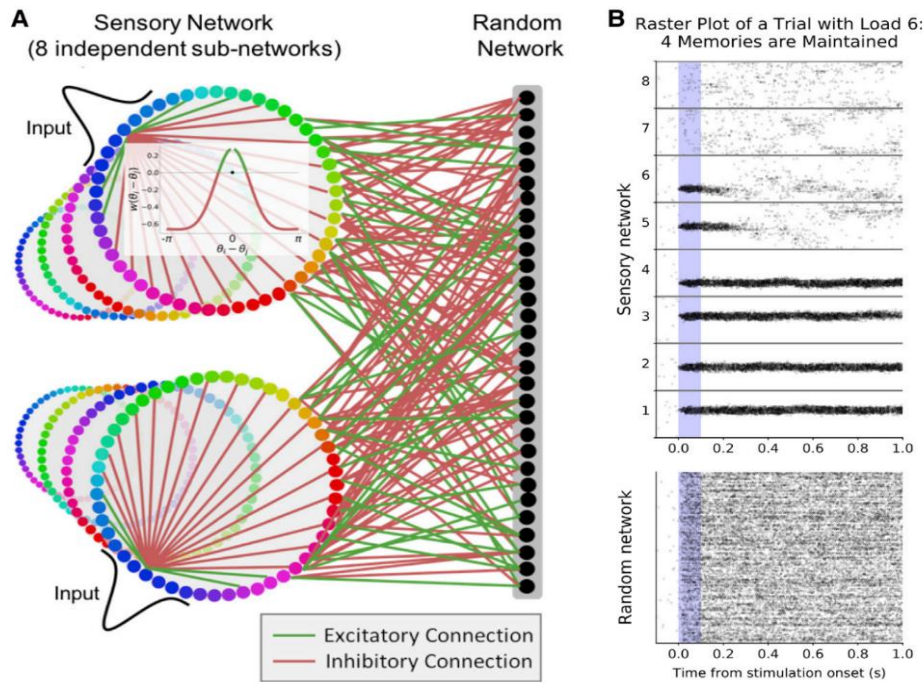


Figure 1. Flexible WM model [103]. (A) Model layout. The sensory network is composed of 8 ring-like sub-networks. The inset shows center-surround connectivity within a sensory sub-network. The connections to the random network are randomly assigned and balanced. (B) Raster plot of an example trial with 8 sensory sub-networks (512 neurons each) randomly connected to the same random network (1,024 neurons). Six sensory sub-networks (1–6) receive a Gaussian input for 0.1 s during the “stimulus presentation” period (shaded blue region). Representations are maintained (without external drive) for four of the inputs. Reproduced from [103].

Box 4. Analysis of individual differences and domain-generalty of WM

WM tasks are some of the most popular tasks used in the analysis of individual differences, as they often predict performance in tasks such as language comprehension, production, and reasoning [119–121]. If scores on the same WM task predict performance across an array of other tasks, it is concluded that a domain-general WM ability underlies these tasks. To see the problems with this interpretation, recall the concepts discussed in the text in the context of training and transfer. For example, in line with [98], complex span tasks have been shown to measure strategic processing

in interference control, and it is individual differences in this trainable strategy, and not general WM capacity, that predicts variation in performance in other tasks [122].

But analysis of individual differences, if conducted properly, can be a source of evidence for testing domain-generalty/specificity of WM. For example, many studies have shown that while WM tasks can be predictive of performance on other tasks, such correlation is specific to domain or task characteristics [123–125](see [66] for a review). Most convincing in this vein are reports of double dissociations within one domain, language processing, in individuals with brain damage. For example, double dissociations have been reported between semantic and phonological WM abilities in individuals with aphasia [126,127]; see [66,128] for reviews. Importantly, these two types of impairments have different consequences for language processing: phonological WM deficits lead to impaired verbatim repetition, whereas semantic WM deficits lead to greater disruption of comprehension and elaboration of content during language production [129,130]. In summary, both neurotypical and neuropsychological data on individual differences in WM point to domain-specificity of WM in application, complementing the data from training studies.