Wiley Interdisciplinary Reviews: Cognitive Science

WILEY

WIREs COGNITIVE SCIENCE

PERSPECTIVE OPEN ACCESS

Working Memory Is as Working Memory Does: A Pluralist Take on the Center of the Mind

Javier Gomez-Lavin

Philosophy, Purdue University, West Lafayette, Indiana, USA

Correspondence: Javier Gomez-Lavin (jglavin@purdue.edu)

Received: 7 May 2024 | **Revised:** 13 July 2024 | **Accepted:** 18 August 2024

Edited by: Mohamad El Haj, Editor

Funding: The author received no specific funding for this work.

Keywords: central cognition | maintenance | manipulation | multiple realizability | working memory

ABSTRACT

Working memory is thought to be the psychological capacity that enables us to maintain or manipulate information no longer in our environment for goal-directed action. Recent work argues that working memory is not a so-called natural kind and in turn cannot explain the cognitive processes attributed to it. This paper first clarifies the scope of this earlier critique and argues for a pluralist account of working memory. Under this account, working memory is variously realized by many mechanisms that contribute to the maintenance and manipulation of information across tasks. This view in effect updates one of the earliest pluralist formulations of working memory. Juxtaposing this view against deflationary descriptions allows us to delineate two gradients that help us chart various accounts of working memory and identify their respective theoretical commitments. In turn, we can isolate those accounts that fail to accord with the evidence supporting a pluralist view, and we can begin to rehabilitate working memory as a pluralist, and ultimately more informative, construct.

1 | Introduction

In an earlier paper (Gomez-Lavin 2021), I argued that working memory, or the capacity described by cognitive psychology that enables us to maintain or manipulate information that is no longer present in the environment in pursuit of goal-directed behaviors, isn't particularly useful. More specifically, I've said at various points that working memory isn't a natural kind, that it's an explanatorily empty term, that it obfuscates our search for the processes that underlie cognition, that it merely redescribes our intuitive commitments about cognition, and—in other venues—I've even claimed that it "doesn't exist" (Gomez-Lavin 2023). That's a long litany of pernicious roles mapped onto a construct that I've suggested isn't very useful. While I largely stand by my earlier criticisms, I want to take the opportunity to clarify the scope and impact of this critique to help delineate just what's at stake and which research programs are in need of revision if something like my view is correct.

There is a temptation to treat working memory—this ability we have to keep information at the forefront of our mind—as a monolithic construct and when we do so we risk ignoring much of its inherent complexity. Instead, I'd argue that working memory is and has, since its inception as a psychological term in 1960, been a plural construct (Miller, Galanter, and Pribram 1960). Furthermore, I think that many scientists and research programs focused on uncovering how working memory works are aware of this complexity. These projects are often aimed at uncovering exactly which systems in the brain are maintaining or manipulating specific pieces of information for successful task performance. Rather, it's when we export the construct from these contexts into other fields, like operational

or clinical psychology or philosophy, that we most risk succumbing to this latent pressure and deflating its polysemous nature.

But there's more; namely, the key that allows us to appreciate, or even rehabilitate working memory, is in valuing precisely this complexity. In this sense, I see this project as fundamentally conciliatory: Yes, there is no univocal system that realizes all the functions we've attributed to working memory, and, yes, working memory can't explain how cognitive processes arise, and, yes, the label risks obfuscating its own inherent complexity, but we might yet be able to excise and mold a useful and explanatory iteration of working memory. Namely, one that treats the construct as a plural, mosaic-like entity that charts the many ways that brains and creatures can maintain and manipulate information. To get there, I'll review in broad strokes the critical argument from the earlier paper and introduce a number of what I'll term "explanatory gradients" that can help us locate exactly which views or projects are threatened by this critical take. Finally, I'll gesture at some ways that we can excise and harness a plural, mosaic-like version of working memory.

2 | A Dilemma

Suppose you've set up two factor authentication on a website that requires you to enter a one-time login code that you receive via an app. The message arrives with the news that your code is "649195" at which point you have to somehow keep those six digits in your mind long enough for you to return to the browser and punch them in. How do you do it? If you're like most adults, you'll rehearse the digits using inner speech in a loop until you can type them in. If someone calls you in the middle of this process, it's very likely that you will have completely forgotten the code by the time you return to the task.[1] This ability that allows us to keep information in a "live" state at the forefront of our minds poised to guide future behavior, to be forgotten at a moment's notice, or in some cases to become encoded into long-term memory, characterizes working memory.[2]

Though a feature of our daily cognitive experience, there's no single accepted model for this capacity. Cowan 2017 summarizes nine distinct views ranging from the so-called "standard" or multicomponent model (consult Figure 1 below) crafted by Baddeley in the 1970s, to newer "emergent" and state-based views derived from the latest neuroscience. Despite the lack of consensus, we can distill a functional profile for working memory that accords with nearly every extant view in the psychological literature, and indeed, a similar formulation will be present at the outset of most scientific papers on the subject. I call this a "generic" account of working memory (Gomez-Lavin 2021, 211).

> **The generic account of working memory:** it is our capacity to maintain or manipulate some information, no longer immediately accessible to us (e.g., by perception), for limited durations in the service of goal-directed behaviour.

In their own formulations authors will typically spell out a precise amount of information to be held over specific time-frame, or they will privilege the manipulation of information over its maintenance, but this view nicely captures a more general functional profile that we're after.

In the earlier paper, I appealed to the apparatus of "natural kinds" to help set up a dilemma for this generic account of

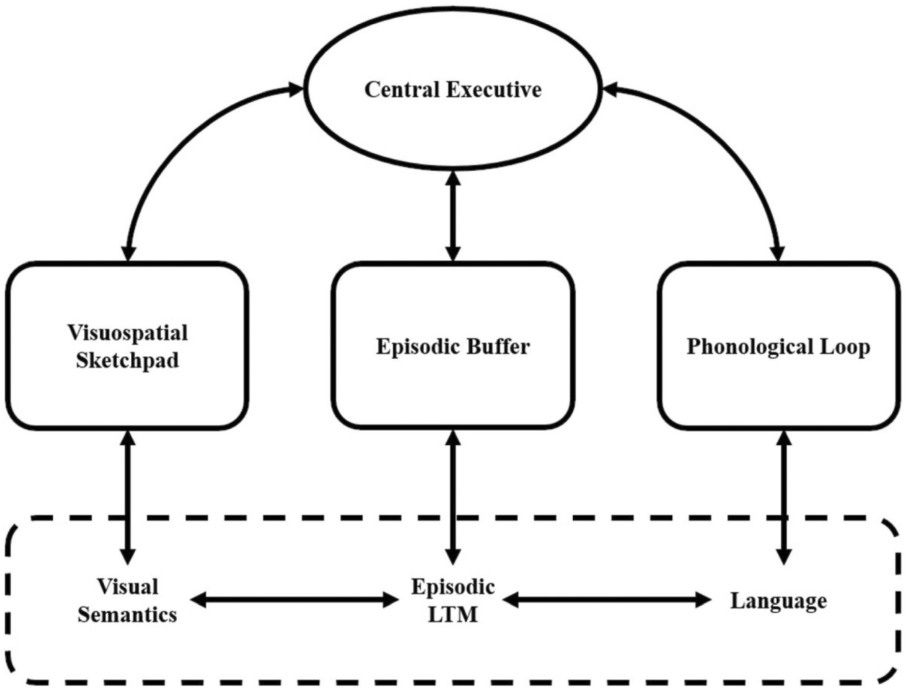

FIGURE 1 | A version of Baddeley's multicomponent model of working memory, a capacity comprised of the Central Executive, Visuospatial Sketchpad, Phonological Loop, and the newly added Episodic Buffer. Stored knowledge that is accessed by the relevant subcomponents of working memory is represented in the lower, dotted box (adapted from Gomez-Lavin 2021).

working memory, in effect implementing a tried and true strategy to investigate the status of categories in our scientific practice (cf. Griffiths 1997 for the case of emotions in psychology and Machery 2009 for the more general case of concepts tout court). I chose to frame the paper around this technical term from the philosophy of science as it provides a ready link between one's metaphysical and epistemological commitments; namely, that one's search for the "furniture of the universe," should be answerable to our best epistemic (i.e., scientific) practices (Khalidi 2023, 5). Ultimately, though I review the broad moves reliant on "natural kinds" from the earlier paper, not much rests on the status of these kinds (e.g., whether they're "real" or their exact metaphysics).[3] Rather, I see our use of this term as a temporary aid that helps us clearly place our implicit commitments about the structure and realizers of working memory side by side with what we hope it can explain.

Proponents of natural kindhood, then, expect that one of the aims of scientific practice is to identify real structures that make up our world; in other words, science should cut nature at its "joints." We can then explain phenomena of interest by appealing to the (micro)structure that gives rise to a given natural kind's properties along with their relevant causal interactions (see, for example, Khalidi's 2013 view which has a relatively low-barrier of entry for natural kind terms). By suggesting that working memory is a natural kind we're in effect saying that it might be a real entity in our (cognitive) ontology and that we can explain cognitive phenomena involving it by appealing to its structure and properties. Around the same time that I began this critical project, philosophical treatments of working memory appeared that explicitly identified working memory as a natural kind term (Carruthers 2015, 180).

How does this generate a dilemma? Natural kinds aren't thought to be plural entities. Gold has the atomic number 79 and if someone tried to sell you a chunk of metal that looked like gold, felt like gold, but whose elemental composition was off by a single proton, you wouldn't have gold.[4] Hence, if we treat working memory as a natural kind, it should have a single nature, ideally with a single mechanism that underwrites the properties we attribute to it.

Decades of neuroscientific research did, in fact, yield a plausible mechanism thought to underwrite working memory; namely, increased stimulus-specific neural firing during delay-periods, specifically that which occurs in prefrontal regions.[5] It's this neural firing that was thought to encode stimulus-specific features (e.g., its position or orientation in the visual field) while the stimulus disappears from the subject's environment, and hence this firing was thought to be the neural realizer of working memory. However, it turns out that we can find evidence of this kind of neural firing across the cortex (see, e.g., Christophel et al. 2017) and across many tasks that we would, from an a priori perspective, assume to be distinct (e.g., Zhou et al. 2022; Ikkai and Curtis 2011; and Jerde et al. 2012), so it's not clear that said mechanism is sufficient for working memory. Worse yet, it appears that there may be other neural mechanisms beyond simple increased rates of neural firing that are involved in the maintenance of information across delay periods, and as such it's not strictly necessary.[6] As such, it looks like we can't capture an account of working memory that isolates a single mechanism specific to its activity.

But we may still be justified in treating working memory as a natural kind or part of our (cognitive) ontology based on its unique functional profile; recall, that we specified earlier that it's the capacity that enables us to maintain or manipulate information for goal-directed behavior. However, this yields a further problem: it turns out that the maintenance of information is a common function that's realized across the cortex by many systems across many tasks (Christophel et al. 2017). This makes intuitive sense, as information producing and consuming systems that exist over time are going to find various ways to hold onto information for later periods. We can think of maintenance as a common currency that occurs in virtually every cognitive task; after all, most experiments require participants to at the very least learn and maintain the task instructions. More recently, cognitive neuroscientists have reintroduced and reinforced a stricter divide between short-term and working memory, restricting *bona fide* working memory tasks to those that require not only the maintenance but also the manipulation of information (Postle 2016). So, for instance, tasks that require that participants do something with the information—such as alphabetizing a string of characters or using a stimulus to help guide a future decision in a subsequent task—would count as genuine working memory tasks.

While the move to manipulation might provide firmer footing if we're hunting for a univocal, functional account of working memory it does so at a significant cost. Suppose we identify a manipulation task, such as re-alphabetizing a string of letters that are briefly presented on a screen, and we further find that successful performance on this task also requires the recruitment of a stable pattern of neural activity across several cortical regions—disruption of which (e.g., by TMS) results in significant degradation of task performance. Have we found the neural realizer of a robust, functional construct of working memory? No, we've found a stable pattern of neural activity (and perhaps even a neural representation) that is correlated with and likely causally linked to the successful alphabetization of characters. Whether that pattern of neural activity and whatever computations it realizes generalize to other cognitive tasks is an important empirical question. However, it's unlikely that the neural mechanisms underlying our ability to re-alphabetize a string of letters will be all and only those that also underwrite the vast swaths of cognitive labor with which we've burdened the construct of working memory for the last half-century, from future planning and problem solving, to even consciousness itself (Prinz 2012; Carruthers 2015). This will be more relevant in the next section, but as a heuristic the more cognitive processes one ties to working memory performance, the less likely it is that version of working memory will be subserved by a single or even small set of mechanisms.[7]

The problem is exacerbated as there's no consensus for what makes a task a manipulation task per se. Intuitively, manipulation tasks require additional cognitive work than mere maintenance of information, and difference is borne out in the kinds of tasks that are routinely corralled under the label of manipulation.[8] But there are no arguments addressing why those (and only those) additional operations count as "manipulating" information. Indeed, one of the first psychological models of working memory offered by Atkinson and Shiffrin 1971 equated the "maintenance" of task-relevant

information with active rehearsal, or the *refreshing* (i.e., manipulation) of values stored in simple mnemonic registers. This isn't merely a pedantic observation. Clearly, arithmetic or alphabetization are in some way harder than keeping a two-factor code at the fore of your mind, but that's because we're requiring participants to perform additional operations. Again, it may be that we've localized a stable neural representation or system involved in alphabetization, or arithmetic, or n-back performance, but that won't necessarily generalize to other cognitive achievements associated with working memory, like recalling and thinking through a recipe.

So where does this leave us? Recall that there are no mechanisms that are jointly necessary and sufficient for performance across working memory tasks. Neural mechanisms like increased neural firing occur more generally, and other mechanisms—including sub-threshold dynamics, calcium kinetics or glial action, may also play a role in successful task performance. Treating working memory as an entity unified by its functional profile—a functional "kind," if you will—isn't very helpful either.[9] Maintenance of information is subserved by a number of regions and is implicated in almost any robust cognitive process. The move to informational manipulation might help, but only to the extent that we can make a case for what counts as manipulation. That is, we might find stable neural representations or patterns recruited by a task aimed at measuring informational manipulation, such as in a re-alphabetization task. However, it's likely that what we've found is not a correlate of manipulation per se, but instead a signature of how brains successfully manage that additional cognitive task (e.g., re-alphabetization). This isn't a devastating finding, in fact, we may have even discovered an important fact about how the nervous system gives rise to a specific cognitive process. But at the same time we aren't warranted in generalizing this correlation to all or most of the cognitive achievements that we've classically associated with working memory, from reading comprehension to future planning. Indeed, this perspective might help us understand why the transfer effects of so-called "working memory training," where practice with one working memory task is thought to–or at least is hoped to–improve proficiency on other working memory tasks, is muted at best (see Redick 2019; and Redick, Wiemers, and Engle 2020). If there is no latent construct behind the scenes that's being employed in both, say, alphabetization and re-numeration, then it's not surprising that training on one task won't necessarily transfer proficiency to the other.

Of course, the muted evidence from working memory training needn't stop us from doing the hard work of coming up with a better account of manipulation. For instance, we might focus less on task-demands that seem intuitively linked to manipulation and instead appeal to operationalizable attributes across cognitive processes that could be tied to a latent construct of manipulation, like resistance to distractors (i.e., inhibitory strength) or degrading with cognitive load and so forth.[10] Eventually we might even settle on a functional description of WORKING MEMORY[11] as just being the combination of the right kind of informational maintenance with a properly parameterized kind of informational manipulation. I think this fine-tuning avenue is an open possibility and may even be a stepping stone to help explain a version of WORKING MEMORY that has deep philosophical appeal. That is, as a site of control and agency over our cognitive lives. As a source for this sense that, sometimes, we willfully marshal our cognitive resources to get something hard done.[12] This feeling of control associated with, and perhaps essential to, this iteration of WORKING MEMORY might, then, be captured by the right blend of informational maintenance and manipulation.

There are at least two reasons to think that this fine-tuning approach is not exhaustive. First is the issue of empirical tractability: it's not obvious that we could reliably test this sense of control or agency and correlate it with a proprietary blend of maintenance and manipulation. It's more likely that there is a messy, many-to-many and contextually dependent mapping between our feelings of cognitive control and the maintenance and manipulation of information (cf. Anderson 2014 on contextualism and Khalidi 2023 on many-to-many mappings in cognitive science). But grant that we could settle on a specific blend of maintenance and manipulation, what then would we do with those tasks that demand more or less informational maintenance or which require manipulation in a way that falls outside our properly parameterized bounds? Discarding them as useful windows onto *some* cognitive process seems wasteful. To reconsider re-alphabetization, whatever demands it places on maintenance and manipulation may not be the same as our hypothetical paragon exemplar of WORKING MEMORY, but it may still reveal something important about the computational characteristics of, and the relevant neural patterns supporting, cognitive processes like alphabetization. And that might be important, if for instance, you're interested in educational psychology or child development. I think the search for *the right* combination of maintenance and manipulation has us missing the forest for the trees. The upshot of a pluralist view of working memory—one that sees it as variously realized across the dimensions of informational maintenance and manipulation—is that we can begin to describe the many cognitive processes and achievements that we've long tied to working memory in terms of their specific demands on these dimensions. In turn, we may find distinct neural patterns or processes that support successful performance on tasks that require differing combinations of maintenance and manipulation. The hope, then, is that we can detail a more informative mapping between cognitive achievements, their informational demands, and their neural realizers than we have at present, where much of this story is distorted through the opaque bottleneck of working memory.

At this point we can kick away the ladder of natural kindhood, which we only touched at the outset of the project. Working memory's purview over our cognitive lives is—and has always been—large. The larger one grants, the more processes and cognitive achievements one corrals under the term, the less likely it is that any single mechanism, system, or functional profile will be isolated whose operation can explain how all these processes come about. In contrast, by homing in on very specific processes—like alphabetization—we might identify a reliable pattern of activity associated with proficient performance, but this is unlikely to generalize to other cognitive tasks, even within the working memory "family." This is the core motivating the dilemma that framed the earlier paper: a broad concept of working memory will have many realizers. Working memory tied to a narrow task description might isolate a single mechanism or system, but its operation will not generalize to a useful

scale. Now we can turn to the question of what we should do about this situation.

3 | Just What's at Risk and How to Move On?

There are two points to consider if something like the following view is correct: That working memory under any interesting level of description is plurally realized by many neural mechanisms, representations, and systems across a diversity of tasks. First, such a gloss would not be out of place in the early days of cognitive psychology when the term was first appropriated by Miller, Galanter, and Pribram 1960 from computer science (Newell and Simon 1956, specifically). In their own words,

> ... we need some special place to store [Plans]. The special place may be on a sheet of paper. Or (who knows?) it may be somewhere in the frontal lobes of the brain. Without committing ourselves to any specific machinery, therefore, we should like to speak of the memory we use for the executing of our Plans as kind of quick access, "working memory" (64).

And again,

> The kind of working memory that people prefer to use when they are executing a Plan seems to represent a characteristic difference. One person will insist on writing things down, running his life from a calendar pad, whereas another person will keep in his own head everything he intends to do (120).

Since the earliest days, then, working memory is as working memory does. It's what enables us to craft, store, evaluate, and manifest our Plans for future action. As it turns out, that functional flexibility abets a pluralist perspective.

The second point, then, builds on this inherent and longstanding pluralism to chart a pair of what I'll term "explanatory gradients." These axes help us visualize the explanatory demands that the construct is subject to, and in turn this landscape helps us describe the relevant commitments of research projects aimed at understanding working memory. Finally, it's by placing research projects on this map that we can begin to identify those research projects most at risk from this back-to-basics move toward a plural construct of working memory.

On the left side of the horizontal axis (see Figure 2 below) we have projects committed to a univocal description of working memory, while on the opposite side we have pluralist accounts. So, for instance, we can group those focused on determining the neural correlate of working memory task performance (i.e., the "prefrontal dogma" of the 1990's championed by Goldman-Rakic) with research projects that posit a latent construct of working memory that can manifest across cognitive task domains (as is the case with many proponents of working memory training). Contrast this cluster with pluralist views that

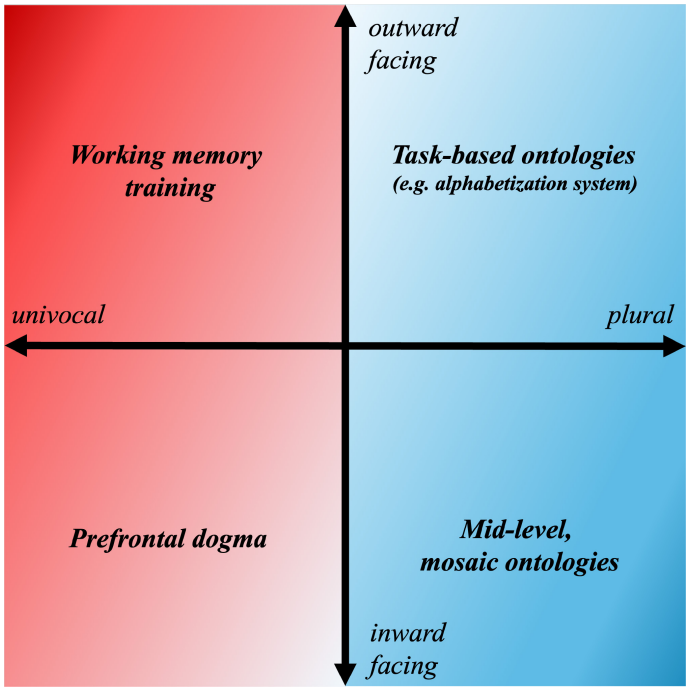

FIGURE 2 | This chart depicts two explanatory gradients that help us locate various accounts or applications of working memory. On the horizontal axis, we can discriminate between accounts of working memory that focus on a single system ('univocal' views) as opposed to those that treat it as a pluralist construct (e.g., Miller, Galanter, and Pribram 1960's own early version). Orthogonal to this, we can differentiate views that focus on the "outward-facing" role working memory can play, where it is placed as a proper part of the explanation of other cognitive achievements. This contrasts with what I've termed "inward-facing" accounts that attempt to describe the subpersonal, neural, and computational processes that give rise to working memory-like functions. The colors denote which projects are at greatest risk if something like the aforementioned account is correct, with red signaling greater concern and blue greater promise.

I've only begun to describe towards the tail end of the last section. Specifically, what we might term "task-based" ontologies, (adapted from Burnston 2024) in which a cognitive type serves as a heuristic for a series of behavioral contrasts measured by tasks that each may have distinct neural underpinnings, and my own favored approach of constructing what I've termed "mid-level" or mosaic ontologies, where we detail the many pathways brains can take to maintain and manipulate information in the service of behavior.

Orthogonal to this first axis, we can envision a second, anchored by what I've come to term the inward and outward aspects of working memory. Broadly, we can think of inward-facing projects as those whose explanatory domain and consequent methods and questions are constrained by the construct of working memory itself. These might be technically nuanced, meticulous attempts to identify and characterize the neural and computational dynamics instantiated in brains when performing typical working memory tasks. They are the kinds of papers that we've seen peppered in throughout this review. In other words, these are projects concerned with how working memory works. Opposite these lie ambitious projects that tend to the more familiar (to philosophers at least) landscape of functional and cognitive processes. It's here that working memory is placed facing *outwards*, posited as a proper part or a scientifically-vetted explanation of a litany person-level cognitive achievements from consciousness (Prinz 2012), reflection (Carruthers 2015), imagination (Reuland 2010), problem-solving (Baddeley 2007, 2010), intelligence (Curry 2021), to decision making (Evans and Stanovich 2013; Augusto 2024), and it's here where we see working memory positioned as an entry point for interventions aimed at improving human cognition (e.g., Atkinson, Allen, and Waterman 2021).

This axis intertwines at least two aspects that merit further discussion.[13] The first is this inward and outward quality, which depends on whether we treat working memory as an element that plays a role in explanations of further phenomena, like reflection or attention-deficit disorder, or as setting the bounds of an explanatory arena by being itself a target of explanation. Notice that while this seems to mirror treating working memory as an *explanans* and an *explanandum* respectively, that may be too simplistic. After all, working memory could be a thing to be explained (i.e., an explanandum) despite being positioned as an outward-facing part of another phenomenon. For instance, Susie may have failed her math test because she had working memory deficiencies, but we can ask what caused these deficiencies (she may have stayed up too late cramming for the exam!). Admittedly it is harder to see how it might work in the inward-facing direction, but if we were to fine-tune a proprietary blend of maintenance and manipulation that captures a philosophically rich sense of WORKING MEMORY, as I discussed earlier, then we can imagine cases where it might serve as an explanans (e.g., upon TMS application the subject failed to report a sense of control over his actions, because of the disrupted recruitment of systems involved in WORKING MEMORY manipulation).

Seen from this axis working memory also straddles at least two identities, both serving as a hypothesized functional system or set of systems that is realized in creatures like us, and as an entity whose existence helps explain a plethora of cognitive activities and that, in the best case, might serve as point of access and intervention to improve our cognitive lives. In this sense, working memory shares a similar explanatory silhouette to other bridge concepts in the history of science, including, for instance, the Mendelian Gene in biology, the concept of Force in physics, and the notion of inflammation in medicine.[14]

I'd argue that it's those projects premised on working memory having a univocal nature, and who position working memory outward to explain vast swaths of our cognitive lives, that are on the wrong path if our aim is to uncover how the activities of brains and people result in our cognitive achievements. Why? In one sense, because such a project would be simply too good to be true: to have a simple set of rules or mechanisms that underwrite much of cognition would be too easy. But more importantly, we see that such a view doesn't isn't supported: the effects of so-called working memory training are ambiguous at best and the mechanisms supporting working memory operate across tasks and the cortex.

At the same time, this realization, and the explanatory gradients that we've mapped above help us chart a possible pathway to rehabilitate working memory into something that contrary to my earlier claims *is* more explanatory. Namely, by focusing on inward-facing, plural accounts that begin to describe the many ways that brains, and perhaps even people and groups, maintain and manipulate information no longer present in the environment in the service of goal-directed behavior. A further upshot of this "mosaic" approach is that it isn't as revisionary as it might seem. Whereas earlier we struggled navigating the explanatory bottleneck formed by both tying working memory to a litany of cognitive achievements, from problem-solving to consciousness, and hunting for its proprietary functional profile or neural realizer, under a pluralist account we can take advantage of much of the extant work. As I mentioned at the outset, research projects clustered towards the inward-facing aspect are likely aware of complex array of mechanisms involved in supporting working memory task performance. Hence, many of these potential routes that begin from a cognitive process and trace a path through a functional description of maintenance and manipulation all the way down to a candidate neural or computational realizer are already implicitly outlined in the literature. Of course, these "routes" aren't presently cashed out in these terms, but by redescribing cognitive achievements in terms of their demands on the maintenance and manipulation of information, and by associating these patterns of maintenance and manipulation with specific computational or neural descriptions, we can bypass the earlier explanatory hurdle. Mapping these patterns and pathways generates a "mosaic ontology" of cognition, where task performance is tied to stable patterns of maintenance and manipulation of information realized by an array of neural mechanisms across the cortex. Ultimately we should expect this mosaic to yield a finer-level of description, explanation, and intervention than our present construct of working memory.

In effect, we're cracking the construct of working memory open, and by embracing its complexity we can begin to glimpse a more informative and explanatory landscape of cognition. In the end, perhaps the best guide going forward and piecing together our cognitive mosaic comes from listening to and taking seriously the intuition driving earlier takes, where working memory *is* as working memory *does*.

Author Contributions

Javier Gomez-Lavin: conceptualization (lead), visualization (lead), writing – original draft (lead), writing – review and editing (lead).

Acknowledgments

I'd like to thank audiences at the Southern Society for Philosophy and Psychology, Midwest Memory Mayhem, the Summer Seminars in Neuroscience and Philosophy (SSNAP), the Philosophy of Biology at Dolphin Beach (PBDB) Workshop, the School of Philosophy, Psychology and Language Sciences Speakers Series at the University of Edinburgh, the University of Cincinnati, the Bochum-Grenoble Memory Colloquium Series and the International Memory Reading Group for their helpful feedback over numerous earlier drafts of this project. I'd especially like to thank Tony Chemero, David Colaço, Peter Godfrey-Smith, Sarah Robins, Tillmann Vierkant, and Wayne Wu for their specific comments on presentations of this work and Edouard Machery along with the fellows and members of the Center for Philosophy of Science for their detailed feedback and support of this project during my residency at the University of Pittsburgh.

Conflicts of Interest

The author declares no conflicts of interest.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Related Wires Articles

Visual working memory

Working memory

Endnotes

[1] This is an updated version of the classic example of remembering a phone number long enough to write it down. However, astute readers will note that most of the time one doesn't even need to memorize a two-factor code as it will automatically populate if you're on a smartphone. As an important caveat: Younger people or those with mild cognitive impairment (e.g., in early stages of Alzheimer's disease) might have to repeat them aloud in order not to lose the information.

[2] In this sense it is similar to James' concept of *primary memory* or the capacity we have to bridge the recently experienced past to the active present (James 1890). Consult also Aristotle's faculty of *phantasia* (Gomez-Lavin and Humphreys 2022).

[3] Khalidi 2023 does a nice job reviewing the history of this term and notes that the use of "natural" as opposed to "real" kinds may have been a historical accident (p. 7). Hence, while I may alternate between "real" and "natural" kind terms, but I put little weight on any difference between them. Thanks to an anonymous reviewer for prompting me to clarify the introduction of these terms.

[4] Of course, there's the classic example by Kim 1992 of jade, a semi-precious stone that is comprised of two distinct minerals, jadeite and nephrite. In this sense, jade, while a conventional kind is not a natural kind as it lacks a univocal micro-structure.

[5] This became known as the "prefrontal dogma," wherein increased firing in dorsolateral prefrontal areas was thought to encode the contents of working memory. Patricia Goldman-Rakic and others crystallized this view in the 1990s (Funahashi, Bruce, and Goldman-Rakic 1989; Goldman-Rakic 1995). See Postle 2006 for a critical take on this dogma.

[6] For instance, consider the recent discussion over the possibility of so-called "activity-silent" states wherein novel mechanisms independent from activity-induced changes in action potential firing (e.g., calcium kinetics and sub-threshold neural-population level dynamics) where posited to play a role and encode stimulus-specific information during tasks (see Lewis-Peacock et al. 2012; LaRocque, Lewis-Peacock, and Postle 2014; LaRocque et al. 2015; Lewis-Peacock, Drysdale, and Postle 2015; Stokes 2015; and Sprague, Ester, and Serences 2016 as some examples and compare Barbosa, Lozano-Soldevilla, and Compte 2021 who argue that much of the evidence for "activity-silent" states is an experimental or statistical artifact).

[7] This is a general feature, or bug under certain perspectives, of faculty psychology more generally (e.g., the problems faced by even earlier faculty theories, including Aristotle's initial attempt, as Gomez-Lavin and Humphreys 2022 argues).

[8] For instance, consult tab. 2 in Barbey, Koenigs, and Grafman 2013 which lists an array of tasks that are seemingly arbitrarily assigned to either the "function" of maintenance, monitoring, manipulation, or reasoning.

[9] Whether this would count as a "real" or "natural" kind will depend on one's metaphysical commitments. The important point here is whether we might want to keep this functional iteration of working memory around to help in our epistemic practices (thanks to an anonymous reviewer for pushing me to clarify this point).

[10] Thanks to an anonymous reviewer for prompting me to think through this fruitful objection.

[11] Here I'm using small caps to designate an idealized version of the concept.

[12] I'm indebted to Peter Godfrey-Smith and Tillmann Vierkant for pushing me to consider this facet of working memory.

[13] Thanks to an anonymous reviewer for prompting me to think through this.

[14] For the Mendelian Gene analogy, consult Griffiths and Stotz 2013 (23). I'm indebted to Arnon Levy for that connection and to Zvi Biener and Guy Dove for the other two analogies, respectively. My sense is that these bridge concepts do significant work in the history of science by linking lower-level aspects of implementation and higher-level functional or computational abstractions; however, in doing so they may relax our epistemic constraints and may promote some explanatory slippage as phenomena are traced from lower to higher levels and causes inferred in the opposite direction.

References

Anderson, M. L. December 2014. *After Phrenology: Neural Reuse and the Interactive Brain*. Cambridge: MIT Press. https://doi.org/10.7551/mitpress/10111.001.0001.

Atkinson, A. L., R. J. Allen, and A. H. Waterman. 2021. "Exploring the Understanding and Experience of Working Memory in Teaching Professionals: A Large-Sample Questionnaire Study." *Teaching and Teacher Education* 103: 103343.

Atkinson, R. C., and R. M. Shiffrin. 1971. "The Control Processes of Short Term Memory (Report No. 173)." In *Institute for Mathematical Studies in the Social Sciences*. Stanford, CA: Mathematical Studies in the Social Sciences.

Augusto, R. 2024. "Two Kinds of Process or Two Kinds of Processing? Disambiguating Dual-Process Theories." *Review of Philosophy and Psychology* 15, no. 1: 277–298. https://doi.org/10.1007/s13164-023-00673-1.

Baddeley, A. 2007. *Working Memory, Thought, and Action*. New York: Oxford University Press.

Baddeley, A. 2010. "Working Memory." *Current Biology* 20: R136–R140.

Barbey, A. K., M. Koenigs, and J. Grafman. 2013. "Dorsolateral Prefrontal Contributions to Human Working Memory." *Cortex* 49, no. 5: 1195–1205. https://doi.org/10.1016/j.cortex.2012.05.022.

Barbosa, J., D. Lozano-Soldevilla, and A. Compte. 2021. "Pinging the Brain With Visual Impulses Reveals Electrically Active, Not Activity-Silent, Working Memories." *PLoS Biology* 19, no. 10: e3001436. https://doi.org/10.1371/journal.pbio.3001436.

Burnston, D. 2024. "Cognitive Ontologies, Task Ontologies and Explanation in Cognitive Neuroscience." In *The Tools of Neuroscience Experiment: Philosophical and Scientific Perspectives*, edited by J. Bickle, C. F. Craver, and A.-S. Barwich. New York: Routledge.

Carruthers, P. 2015. *The Centered Mind: What the Science of Working Memory Shows Us About the Nature of Human Thought*. New York: Oxford University Press.

Christophel, T. B., P. C. Klink, B. Spitzer, P. R. Roelfsema, and J. D. Haynes. 2017. "The Distributed Nature of Working Memory." *Trends in Cognitive Science* 21, no. 2: 111–124.

Cowan, N. 2017. "The Many Faces of Working Memory and Short-Term Storage." *Psychonomic Bulletin and Review* 24: 1158–1170.

Curry, D. S. 2021. "g as bridge model." *Philosophy of Science* 88, no. 5: 1067–1078.

Evans, J. S. B., and K. E. Stanovich. 2013. "Dual Process Theories of Higher Cognition: Advancing the Debate." *Perspectives on Psychological Science* 8: 223–241.

Funahashi, S., C. J. Bruce, and P. Goldman-Rakic. 1989. "Mnemonic Coding of Visual Space in the monkey's Dorsolateral Prefrontal Cortex." *Journal of Neurophysiology* 61: 331–349.

Goldman-Rakic, P. S. 1995. "Cellular Basis of Working Memory." *Neuron* 14, no. 3: 477–485. https://doi.org/10.1016/0896-6273(95)90304-6.

Gomez-Lavin, J. 2021. "Working Memory Is Not a Natural Kind and Cannot Explain Central Cognition." *Review of Philosophy and Psychology* 12: 199–225.

Gomez-Lavin, J. 2023. "Twitter profile [@jgomezlavin]." https://twitter.com/jgomezlavin.

Gomez-Lavin, J., and J. Humphreys. 2022. "Striking at the Heart of Cognition: Aristotelian Phantasia, Working Memory, and Psychological Explanation." *Medicina nei Secoli* 34.2: 13–38. https://doi.org/10.13133/2531-7288/2647.

Griffiths, P. E. 1997. *What Emotions Really Are: The Problem of Psychological Categories*. Chicago: University of Chicago Press.

Griffiths, P., and K. Stotz. 2013. *Genetics and Philosophy: An Introduction*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511744082.

Ikkai, A., and C. E. Curtis. 2011. "Common Neural Mechanisms Supporting Spatial Working Memory, Attention and Motor Intention." *Neuropsychologia* 49, no. 6: 1428–1434. https://doi.org/10.1016/j.neuropsychologia.2010.12.020.

James, W. 1890. *The Principles of Psychology*. Vol. 1. New York: Holt.

Jerde, T. A., E. P. Merriam, A. C. Riggall, J. H. Hedges, and C. E. Curtis. 2012. "Prioritized Maps of Space in Human Frontoparietal Cortex." *Journal of Neuroscience* 32, no. 48: 17382–17390. https://doi.org/10.1523/JNEUROSCI.3810-12.2012.

Khalidi, M. A. 2013. *Natural Categories and Human Kinds: Classification in the Natural and Social Sciences*. Cambridge: Cambridge University Press.

Khalidi, M. A. 2023. *Cognitive Ontology: Taxonomic Practices in the Mind-Brain Sciences*. Cambridge: Cambridge University Press.

Kim, J. 1992. "Multiple Realization and the Metaphysics of Reduction." *Philosophy and Phenomenological Research* 52: 1–26.

LaRocque, J. J., J. A. Lewis-Peacock, and B. R. Postle. 2014. "Multiple Neural States of Representation in Short-Term Memory? It's a Matter of Attention." *Frontiers in Human Neuroscience* 8, no. 5: 1–14.

LaRocque, J. J., A. S. Eichenbaum, M. J. Starrett, et al. 2015. "The Short- and Long-Term Fate of Memory Items Retained Outside the Focus of Attention." *Memory & Cognition* 43, no. 3: 453–468. (Special Issue on Working Memory).

Lewis-Peacock, J. A., A. T. Drysdale, K. Oberauer, and B. R. Postle. 2012. "Neural Evidence for a Distinction Between Short-Term Memory and the Focus of Attention." *Journal of Cognitive Neuroscience* 24, no. 1: 61–79.

Lewis-Peacock, J. A., A. T. Drysdale, and B. R. Postle. 2015. "Neural Evidence for the Flexible Control of Mental Representations." *Cerebral Cortex* 25, no. 10: 3303–3313.

Machery, E. 2009. *Doing Without Concepts*. New York: Oxford University Press.

Miller, G., E. Galanter, and K. H. Pribram. 1960. *Plans and the Structure of Behavior*. New York: Henry Holt and Company.

Newell, A., and H. A. Simon. 1956. "The Logic Theory Machine—A Complex Information Processing System." In *IRE Transactions on Information Theory*, vol. 2, 61–79. Santa Monica, CA: RAND Corporation.

Postle, B. R. 2006. "Working Memory as an Emergent Property of the Mind and Brain." *Neuroscience* 139: 23–28.

Postle, B. R. 2016. "Neural Bases of the Short-Term Retention of Visual Information." In *Mechanisms of Sensory Working Memory: Attention and Performance XXV*, edited by P. Jolicoeur, C. Lefebvre, J. MartinezTrujillo, and J. MartinezTrujillo, 43–58. London: Academic Press.

Prinz, J. 2012. *The Conscious Brain*. New York: Oxford University Press.

Redick, T. S. 2019. "The Hype Cycle of Working Memory Training." *Current Directions in Psychological Science* 28, no. 5: 423–429. https://doi.org/10.1177/0963721419848668.

Redick, T. S., E. A. Wiemers, and R. W. Engle. 2020. "The Role of Proactive Interference in Working Memory Training and Transfer." *Psychological Research* 84: 1635–1654. https://doi.org/10.1007/s00426-019-01172-8.

Reuland, E. 2010. "Imagination, Planning and Working Memory: The Emergence of Language." *Current Anthropology* 51: S99–S110.

Sprague, T. C., E. F. Ester, and J. T. Serences. 2016. "Restoring Latent Visual Working Memory Representations in Human Cortex." *Neuron* 91: 694–707.

Stokes, M. G. 2015. "'Activity-Silent' Working Memory in Prefrontal Cortex: A Dynamic Coding Framework." *Trends in Cognitive Sciences* 19, no. 7: 394–405.

Zhou, Y., C. E. Curtis, K. K. Sreenivasan, and D. Fougnie. 2022. "Common Neural Mechanisms Control Attention and Working Memory." *Journal of Neuroscience* 42, no. 37: 7110–7120. https://doi.org/10.1523/JNEUROSCI.0443-22.2022.