SAPIENZA
UNIVERSITÀ DI ROMA

E-ISSN 2531-7288
ISSN 0394/9001

MEDICINA NEI SECOLI
Journal of History of Medicine and Medical Humanities
34/2 (2022) 13-38

Received: 31.10.2021
Accepted: 24.03.2022
DOI: 10.13133/2531-7288/2647

Corresponding author:
Javier Gomez-Lavin
jgomezlavin@gmail.com

Striking at the Heart of Cognition: Aristotelian Phantasia, Working Memory, and Psychological Explanation

Javier Gomez-Lavin
Purdue University, West Lafayette, Indiana, USA

Justin Humphreys
Villanova University, Villanova, Pennsylvania, USA

ABSTRACT

Striking at the Heart of Cognition

This paper examines a parallel between Aristotle's account of *phantasia* and contemporary psychological models of working memory, a capacity that enables the temporary maintenance and manipulation of information used in many behaviors. These two capacities, though developed within two distinct scientific paradigms, share a common strategy of psychological explanation, Aristotelian Faculty Psychology. This strategy individuates psychological components by their target-domains and functional roles. Working memory and *phantasia* result from an attempt to individuate the psychological components responsible for flexible thought and are thus implicated in most of our robust cognitive processes, from reading comprehension to problem solving. We then present two novel objections which suggest that these capacities cannot explain our ability to engage in flexible thought. To escape the resultant impasse, we survey alternatives and argue that most promising strategies depend on identifying the behaviors attributed to intelligent thought and action.

Keywords: *Phantasia* - Working Memory - Faculty Psychology - Psychological Explanation

1. Introduction: faculty psychology and general cognition

Ancient theories of the soul and modern scientific psychology are significantly different endeavors. Yet the investigation of parallels between them can help resolve tensions in ancient accounts and illuminate the lineage of concepts used in contemporary psychology[1]. We ar-

gue that an important parallel obtains between φαντασία (*phantasia*), a controversial concept in Aristotle's philosophy of mind, and working memory, a capacity that is fundamental to much of contemporary psychology. This parallel obtains because Aristotelian psychology and contemporary cognitive science both aim to explain our mental lives by means of postulated faculties, which are individuated by their activities and objects. This method of fragmentation is fruitful, describing how disparate sensory, motoric, motivational, and evaluative components organize much of our mental lives. *Phantasia* and working memory result from attempts to understand the enormous domain of flexible thought by this method. However, we argue that they fail to be genuinely explanatory because they are situated within the framework of what we term Aristotelian Faculty Psychology, which describes our mental lives by positing functional, sub-personal divisions[2]. Thus, we are faced with a dilemma: in order to gain satisfactory explanations of flexible cognition, we must either jettison such general capacities from our cognitive ontology or abandon the framework of Aristotelian faculty psychology altogether.

We make two main claims in this paper. First, we claim that Aristotle's concept of *phantasia* is significantly analogous to the modern concept of working memory. On the face of it, it is surprising that psychological capacities rooted in computational and information-theoretic methods should bear any similarity to those employed in an ancient theory of the psyche. Yet each capacity stems from a theory that posits faculties to explain and predict cognition by their constituent functions. Consequently, many of the functional relationships established in Aristotle's account of *phantasia* have significant analogues in models of working memory.

Second, we claim that both constructs are too broad to explain their target psychological phenomena. Though *phantasia* and working memory are supposed to operate within explanatory frameworks that require the individuation of psychological faculties, each is taken to serve as an arena for the realization of flexible, domain-general thought. As such, they are implicated in almost every robust cognitive process. It's precisely because of their breadth that they cannot be functionally individuated within the framework of Aristotelian Faculty Psychology.

The plan for the paper is as follows. We begin (sect. 2) with an overview of Aristotle's concept of *phantasia*, stressing its function of mediating between perceptual inputs and behavioral outputs. We claim that its mediating role establishes certain functional relationships within Aristotle's psychology. We then consider the development of working memory in modern cognitive psychology (sect. 3). Despite their divergent features, models of working memory consistently understand it to be responsible for flexible thought. Next (sect. 4) we make explicit the analogies and divergences between *phantasia* and working memory via a side-by-side comparison. We then (sect. 5) lay out our central negative argument, that *phantasia* and working memory, for analogous reasons, are conceived too broadly to be explanatory of any particular cognition. Indeed, we

argue that because they are posited as general faculties required for flexible thought, *phantasia* and working memory merely serve as new terms for those more fundamental processes thought to lie at the heart of cognition. Finally, (sect. 6) we canvass suggestions for resolving this dilemma about faculty psychology and flexible thought.

2. *Phantasia*: The aristotelian account

The soul is the form of a living animal, for Aristotle, while the body is its material. Isolating psychological faculties explains why members of a species—individuals that share a form—engage in their characteristic activities. For instance, it is a characteristic activity of honeybees to perform the waggle dance to indicate a source of food. But the waggle dance can only be detected at a distance by sight. Thus, the possession of the faculty of vision is explanatory of *why* bees dance: they dance *in order* to communicate information about a source of food.

A central question of Aristotle's psychology is, consequently, how to distinguish one faculty, δύναμις (*dunamis*), from another in order to achieve explanations of characteristic activities. Plato's *Republic* (477c-d) suggests two criteria for individuating a faculty: that which it produces, and that which it is about. In his *De Anima*, Aristotle employs the term "*antikeimenon*"—that which is "opposed" to or "corresponds" to the faculty—to capture both criteria (DA 415a14-22)[3]. The Aristotelian ἀντικείμενον (*antikeimenon*) is usually translated as "object," which preserves the ambiguity of these distinguishing criteria. The object of the visual faculty, for example, comprises both what it produces, the visual episode produced by the faculty of sight, and the proper domain over which the faculty is set, color[4]. These criteria for individuation are fruitful in demarcating the senses, though they become problematic when applied to the general faculty of mental representations, which Aristotle calls *phantasia*.

Aristotle's canonical account of *phantasia* in *De Anima* III.3 consists largely of a series of arguments meant to show that *phantasia* is neither perception, nor propositional belief, nor belief mixed with perception, as suggested by Plato (*Sophist* 264a-b). But demonstrating that it has a unique place in the division of the soul does not amount to a clear job description for *phantasia*. Scholars are forced to look to his numerous but somewhat diffuse uses of the term in other works in order to reconstruct Aristotle's positive account of *phantasia*, which is consequently a topic of perennial interest and debate. As Johansen puts it, a reconstructed theory that can unify these disparate and at times seemingly contradictory features of *phantasia* is the "holy grail" of Aristotelian scholarship[5].

Despite the difficulty of fitting together his many uses of the term, there is widespread agreement on Aristotle's definition of *phantasia*. After offering a battery of arguments that divide it from perception and belief, and arguing that active perception is required for *phantasia*, Aristotle defines *phantasia* in terms of psychological change.

> *So, if it involves nothing other than we have reported, and is as we have described it, then* phantasia *would be* a movement that comes about in virtue of the activity of perception. *Indeed, since sight is perception most of all, the name* phantasia *is derived from light* (phos)*, because without light there is no seeing. (DA III.3, 428b30-429a4)*[6].

Aristotle conceives of *phantasia* as a type of change or "motion," specifically a change brought about by active perception[7]. The definition of *phantasia* as a motion caused by actualized perception suggests a functional model of the soul (Fig 1).

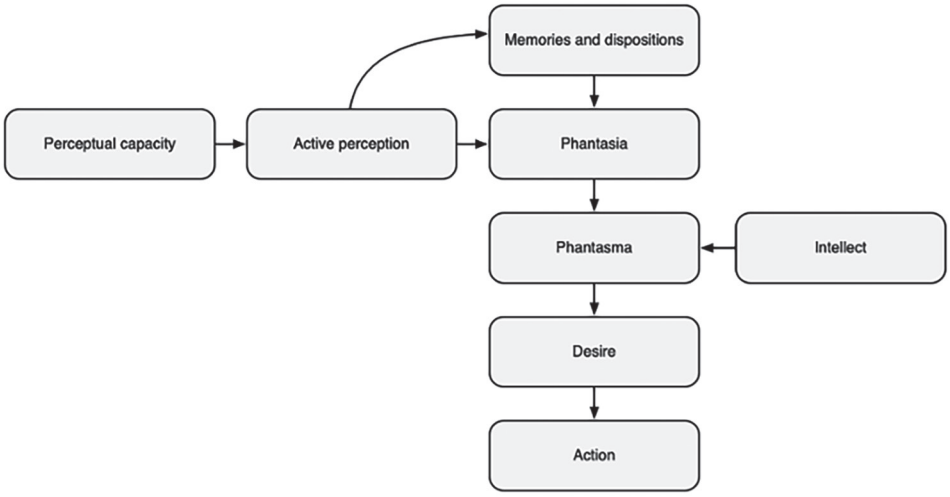

Fig. 1. Functional relationships attributed by Aristotle to the parts of the soul.

According to this model, activated perception sets off *phantasia*, which, guided by one's dispositions and memories, produces a φαντάσμα (*phantasma*). This *phantasma* gives occasion for intellectual cognition to occur, since it is in the *phantasma* that one thinks the object of intellect. At least in some cases, this cognitive activity can trigger desire, and thus animal locomotion. Though our argument depends only on the correctness of this outline of the functional relationships, considering each relation of functional dependence brings out correspondences between Aristotle's model and modern models of working memory.

The fundamental functional relation in Aristotle's model is the dependence of *phantasia* on active perception[8]. Aristotle conceives perception as the reception of the form of the object of perception without its matter, comparing it to the way wax receives the seal from a gold ring (DA 424a17-28). A debate in the scholarship concerns whether Aristotle thinks of this change as material or spiritual. For example, is Aristotle's "form without matter" formulation meant to describe part of the eye becoming red when one looks at a sunset, or is it meant to capture the immaterial, "spiritual" change

of becoming aware of redness[9]? Without taking sides on this interpretive issue, we can observe that Aristotle is committed to the idea that information about the objects is transferred in perception. Thus, within our reconstructed model the transfer of perceptual information initiates an episode of *phantasia*. Aristotle's stock example of this functional dependence of *phantasia* on perception is the "illusion" of large distant objects, such as the sun, appearing to be small. The *phantasma* or appearance that the sun's diameter is one is not the product of veridical perception, nor of the powers of judgment, but of *phantasia* (*Insom*. 460b16-18, cf. 458b25-29, DA 428b2-4). Here, active perception triggers *phantasia*, but *phantasia* itself is responsible for one's experience of the illusion.

Another set of functional dependencies obtain between *phantasia* and one's dispositions and memories. A coward seems to see his enemy when he is looking at something that bears a minimal similarity to the enemy (*Insom*.460b3-8). As in the case of illusion, *phantasia* presents something not given in the content of the perceptual episode. But in this case, the appearance of the enemy appears in part because the subject's emotions tend to weaken the discerning power of perception, so that φαντάσματα (*phantasmata*) crowd out the objects correctly discerned by perception. The emotion is the product of the coward's habituated response to his environment, so that the disposition has a top-down effect on the coward's judgment[10]. Indeed, Aristotle goes so far as to say that *phantasia* and memory belong to the same part of the soul (*De Mem* 450a22-25). But while *phantasia* and memory both operate on perceptual contents, they are not identical: *phantasia* is necessarily involved in explicit memory of past perceived objects, though *phantasia* as such has no particular tense[11].

The functional relationship between *phantasia* and *phantasmata* is perhaps the most straightforward: *phantasia* produces the phantasma (DA 427b17-20). More interestingly, while many, perhaps all, animals possess *phantasia*, in human beings a phantasma gives occasion for the intellect to present one with a universal. Hence, there is no thinking without an accompanying phantasma (DA 431a14-18), or as Aristotle more pointedly puts it, "One thinks the intelligible forms in *phantasmata*" (DA 431b2). In other words, in active thinking, *phantasia* provides the representational vehicle necessary for abstract thought[12].

The last set of functional relationships are between *phantasia* and the motive powers. In *De Motu Animalium*, Aristotle argues that cognitive powers—perception, *phantasia*, and intellect—could not on their own "be movers," but must determine desire, which acts as the mover of the animal (*De Motu* 701a33-36). The primary function of *phantasia* in action-contexts is presentational, since it makes some end seem good. Thus, while the cognitive powers alone are insufficient for action, they have the ability to affect the power of desire, which in turn acts as the proximate cause of an action[13]. Thus, *phantasia* confers the ability to envisage prospects, which when engaged with desire make *phantasia* indirectly responsible for locomotion[14].

Given this functional model of the psyche, what explanatory resources does *phantasia* offer Aristotle? While his predecessors often assumed a naïve empiricist view that attributes both perceptual success and perceptual error to the same capacity, on Aristotle's alternative account, false appearances need not be attributed to perception. The appearance of an oar bent in water, for example, casts no doubt on the veracity of perception, since *phantasia* is responsible for the appearance of the bending. Aristotle explains the appearance of a bent the oar by attributing the error to *phantasia* rather than to the veridical perceptual power.

Phantasia also connects the particulars afforded by perception to the intellectual grasp of a general fact. For example, geometers read off features of perceptually given diagrams of particular triangles as licensing universal inferences about triangles in general. Though geometry is not about material particulars with a definite quantity, in geometrical cognition "a quantity is nevertheless placed before the eyes, even though one does not think the quantity" (*Mem.*449b30-450a5). In this case, *phantasia* formulates a specific representation in which the universal, geometrical concept is thought[15]. Aristotle's *phantasia* model of the soul predicts novel facts, for example, that a geometer will be able to grasp a proposition such as "the internal angles of a triangle are equal to two right angles," on the basis of an observed phenomenon, that geometers look at triangles drawn on paper. The *phantasia* model is also explanatory of certain phenomena, for example, that an oar in water looks bent, in the light of certain pre-theoretical commitments a reasonable person should assume, such as that an oar might look bent even when it's not bent and one's eyes are working perfectly well. On the face of it, the *phantasia* model looks like a progressive research program, in the Lakatosian sense.

Nevertheless, we think that something has gone wrong with Aristotle's explanatory strategy, a mistake repeated in contemporary models of working memory[16]. In short, our objection is that *phantasia* attempts to explain too much and will consequently be implicated in every act of cognition that follows upon a sensory perception. The tremendous breadth of *phantasia* arises from intrinsic difficulty of explaining flexible cognition by means faculty psychology.

Consider "Aristotle's illusion," in which when touching a single object with one's fingers are crossed and eyes shut, "one object appears [φαίνεται] to be two" (*Insom.* 460b20-25). In this case, one cannot discern by touch a single object from two objects with similar textures. This imprecision of tactile discernment is corrected when one opens one's eyes. Vision is "more authoritative" than touch, forcing the judgment that one is touching a single object. Aristotle argues that whenever one perceives something, one acquires information about reality. But for touch, the information is not enough to determine number with accuracy. Since the information from touch is incomplete, and *phantasia* completes it, not into a correct representation, but into an incorrect model of the sensed object. Touching the object triggers *phantasia*, which produces the *phantasma* of two objects in contact with one's fingers. But once one's eyes

are open, one doesn't need *phantasia* to add that information, and one's judgment is more likely to be correct. Sight overrides not touch itself, but the *phantasia*-supplied aspects of the touch experience. Aristotle invokes *phantasia* to explain the erroneous representation of the object resulting from the imprecision of tactile discernment.
But what does this add to the mere description of the illusion? The phenomenon of interest is, after all, the inability of touch to discriminate correctly when one's fingers are crossed, that is, the false representation of a single object as two objects. Insofar as it is a general faculty of representation, invoking *phantasia* adds nothing to the mere description of the phenomenon. Aristotle's "explanation" of the illusion thus amounts to the vacuous statement that the cause of the illusory representation is the general faculty of representation. In what follows, we argue that explanations reliant on working memory involve the same risk of vacuous explanation.

3. Working memory: cognition in a computational paradigm

Suppose you're trying to make sense of a string of words grouped on a page. As the your eyes pass over each word, you must have some capacity to retain and combine them into a coherent whole, a meaningful sentence. This mental capacity to assemble our recently experienced past into a cohesive present, further allowing us to prepare for future action, is a fixture of our daily experience. The search for the source of this ability has propelled a longstanding research project in psychology and cognitive science, summarized by the deceptively simple phrase, *working memory*. This dyadic formulation is intuitive: one attempts to keep information in mind in order to *do* something with that information. For instance, to solve the twentieth-century problem of keeping a phone number in mind long enough to find a pen and paper to write it down, it is natural to describe the phenomenon as a *memory* that you've engaged to do some *work*.
What, then, is working memory? Despite having a clear intuitive grasp on what it's like to keep something at the front of one's mind, there is—unlike Aristotle's *phantasia*—no single privileged theory from an authoritative source that could be taken to adequately account for this phenomenon. For the purposes of this paper, we can characterize working memory as *a capacity that allows us to maintain and manipulate limited amounts of information, no longer in the environment, for limited durations, in the service of goal-directed behavior.* Commitment to this functional characterization comes as close to a consensus view in a field as diverse as cognitive psychology, with its practitioners divided by their choice of tools, methods, paradigms and even organisms of study[17]. As we shall see, even when contemporary philosophers and cognitive scientists express doubt about the ontological status of working memory, or deny that it is a natural kind, they begin with a similar sketch that situates working memory as the central arena for flexible, synthetic thought; c.f. Carruthers' recent treatment: "there is, indeed, a central workspace in the mind whose contents are always con-

scious. This is so-called 'working memory'"[18]. This implicit assumption that places working memory as a central bottleneck whose features train and shape most of our thoughts is not accidental to contemporary accounts of mind, but rather has a *history;* one which in this case we can trace to the conceptual scheme of the mid-twentieth century computational theory of mind from which working memory emerged.

The earliest mention of "working memory" of which we are aware comes from Newell and Simon's description of their Logical Theory Machine, which possessed multiple "working memories" that allowed the machine to store temporarily values for its ongoing operations—an analogue to the mathematician's pile of scrap paper[19]. George Miller and colleagues were likely the first to repurpose the term within a psychological theory, describing it as a capacity which allowed us to plan for future behavior[20]. By applying an information theoretic framework to human psychology, these authors noted that while working memory shared a similar capacity limit—around seven items that one could retain in mind—with other cognitive processes, such as absolute judgement, one could augment the amount of *information* carried by each item held by deploying heuristics. For instance, while trying to keep a ten-digit phone number in mind, one might "chunk" digits together to minimize the cognitive burden[21].

Atkinson and Shiffrin use the term working memory to describe the "short-term store," a central component of their model of information processing (Fig. 2)[22]. According to this model, the first to delineate working memory's functional role, stimuli from the environment are initially encoded in sensory specific registers, with some of this information continuing onto working memory. Within working memory, several control processes may subsequently act upon that information, yielding a behavioral response, or they can shunt that information along to a passive and permanent long-term memory store. Atkinson and Shiffrin thus attribute a wide swath of cognitive behavior to working memory:

> *Because* consciousness *is equated with [the short-term store], and because control processes are centered in and act through the [short-term store], this store is considered a 'working memory': a store in which* decisions are made, problems are solved, and information flow is directed[23].

The immense scope and role conceived for working memory within human information processing is apparent. It serves as a central hub for our cognitive lives, exhausting much of the cognitive space between perception and behavior.

With this model in view, we can generalize six properties inherent in Atkinson and Shiffrin's foundational functional account of working memory below:

1. *Wide Scope*: Working memory plays a far-reaching role in cognition.
2. *Maintenance and Manipulation:* Working memory operates by holding information in an active state and allowing further, derivative processes to access and operate on that information.

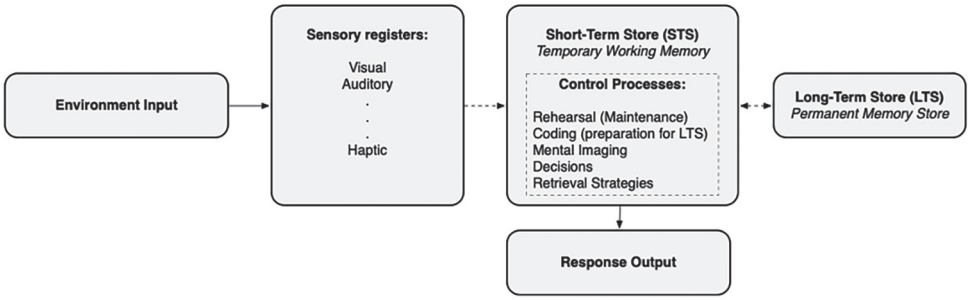

Fig. 2. A diagram, adapted from the original, of the functional relationships posited by Atkinson & Shiffrin's model of "human information processing." Bolded arrows indicate mandatory operations, while dotted arrows indicate contingent relations.

3. *Conscious*: Working memory is tied to phenomenal consciousness, and in this early model it is identified as the source of our conscious inner life[24].
4. *Voluntary:* Deploying working memory and its concomitant processes is an effortful procedure under our control[25].
5. *Capacity Limited:* Working memory can only maintain and manipulate a limited amount of information, perhaps around six "items," although strategies can increase the informational density of the items retained[26].
6. *Computational & Information Theoretic:* Working memory is assumed to decompose into a series of identifiable, computationally tractable, processes that can be modeled with the tools of mathematical psychology and information theory[27].

Like Aristotle's *phantasia*, working memory serves as a free agent in the arena of cognition, straddling perception and deliberative thought, and trading in the items of recent experience. Akin to a mathematician's pile of scrap paper, working memory functions primarily as a workspace for any number of thoughts from across the mind. But since working memory emerges from a computational and information-theoretic framework, we can ask a question that would be foreign to Aristotle: How does one begin to test and control such a flexible capacity?

Answering this question brought Atkinson and Shiffrin's model of working memory under criticism with the reintroduction of dual-task experiments in the 1970s[28]. In Baddeley and Hitch's pioneering experiments, participants were given a set of reasoning tasks of increasing complexity (for example, they were shown a stimulus set of "AB" and were asked a series of true or false questions, "Does A precede B?" and so on), while simultaneously asked to repeat a single word, ordered numbers, or random numbers[29]. Participants who repeated the words or ordered numbers performed close to control conditions, while those who repeated random numbers suffered performance deficits. These results should not be expected from a single domain gen-

eral working memory store, which instead would predict similar performance deficits across tasks and modalities. These findings, coupled with Shallice and Warrington's research on brain lesions, led Baddeley and Hitch to propose their "multicomponent" model of working memory, as reproduced in figure 3 below[30].

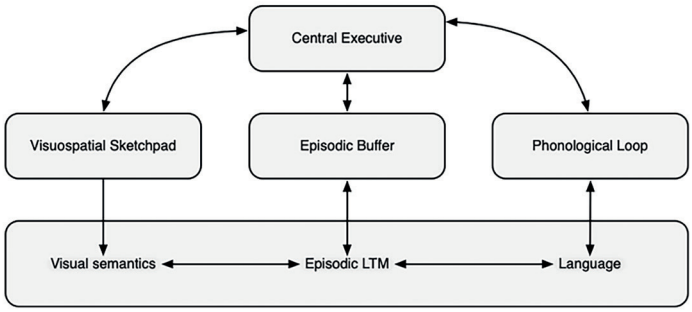

Fig. 3. A version of Baddeley's multicomponent model of working memory, a term which now describes the entire four-part system at top. Functions described in the lower box represent stored knowledge that is accessed by the relevant subcomponents of working memory (adapted from Baddeley)[31].

According to Baddeley and Hitch, both the reasoning task and the maintenance of a jumbled set of numbers compete for the same set of cognitive resources, while other dedicated systems can handle habitual phonological operations, such as repeating a single word or well-practiced string of numbers. Consequently, Baddeley *fragmented* Atkinson and Shiffrin's unitary "short-term store" into two sensory-specific subsidiary systems, the visuo-spatial sketchpad and the phonologoical loop, directed by a flexible controller, or the "central executive" and its episodic buffer. These subsystems, when working in tandem, were "assumed to be necessary in order to keep things in mind while performing complex tasks such as reasoning, comprehension and learning"[32]. Their model explains a range of findings and generates testable predictions, for example, the non-interference of visual stimuli with auditory rehearsal has become one of the most cited results of working memory research[33].

Baddeley's multicomponent model dominated the landscape of working memory for forty-years until the rise of more neuroscientifically tractable alternatives, at least in part because of its unabashed incorporation of perceptual faculties *within* working memory via its visual sketchpad and auditory loop[34]. As opposed to Atkinson and Shiffrin's separation of perception and working memory, Baddeley's model was able to leverage the already century-long use of psycho-physical measures of sensory stimuli to approach working memory, yielding an adaptable model that was already at home with psychologists' favored methods and stimuli[35].

Eventually, Baddeley introduced the episodic buffer to explain how working memory could maintain and make use of multi-modal cognitions. For example, when you en-

visage the scene of your having coffee this morning, with due effort, the smells, tastes, sights, and sounds of that experience might come into (albeit incomplete and "rough") focus. But this generative capacity, which Baddeley associated with consciousness, could not easily be explained by his original division of working memory into two sensory systems and a central executive, since there was no clear way to explain how working memory could access even episodic contents of long-term or autobiographical memory. So, like a *deus ex machina*, the episodic buffer was born[36].

The central executive itself also presents an epistemic conundrum. Its position within working memory begets a worry of homuncularity, since it "decides" what working memory will do, and by extension, what you do. Baddeley was aware of this concern, deeming the central executive a "conceptual ragbag" and a placeholder for the complex and little understood goings-on of cognition. Consequently, he took it that a central aim of future cognitive psychology should be to "sack" the central executive[37]. Indeed, some researchers, such as Logie, have attempted to do so, developing executiveless models of working memory[38]. However, these accounts run into similar quagmire that Fodor attributes to central cognitive processes generally, namely, how is it possible to give a reductive account of something as flexible as decision making and problem solving[39]? It seems that the fragmentation of working memory in the multicomponent model, and the resulting shuffling of general cognitive functions into a postulated central executive, does not so much resolve the problem of flexible cognition as reproduce it at a lower, perhaps sub-personal level, within the proposed model. To pose Fodor's problem slightly differently: how does one design an experiment in cognitive psychology that does *not* require working memory? After all, most psychological experiments need the subject to keep a set of instructions in mind, respond to a cue, and engage in some maintenance and manipulation of information that affects behavior[40]. If this functional characterization is accurate, then most of waking life is *suffused* with demands placed on working memory; from reading this paper to planning dinner and navigating the transit system, one is perpetually maintaining and manipulating information in the service of one's goals.

Turning to neuroscience to constrain working memory is an appealing strategy, and has become a central focus in working memory research in the last decade. But this move may not be that useful, either, as it may simply recapitulate Fodor's challenge at a different level. That is, the move from psychophysical to neurological methods will be unlikely to rectify what it, at its heart, a *conceptual problem;* i.e., working memory's seemingly necessary contribution to most interesting cognitive processes. For instance, Rottschy and colleagues performed a meta-analysis of 189 fMRI experiments on working memory, attention, and intention tasks and found a similar brain-wide pattern of activation common across them[41]. These results were echoed by Jerde and colleagues who used multivariate pattern analyses and determined that the brain seems to represent these *prima facie* distinct tasks similarly[42]. Finally, Christophel and colleagues' recent review sharpened these concerns by finding that many regions of the brain can

maintain sensory information throughout a delay-period common to working memory tasks[43]. They end their by arguing for a sea-change in the field:

> *perhaps the field of working memory should shift its focus from asking where in the brain working memories are stored to unraveling how a range of highly specialized brain areas together transform a sensory stimulus into an appropriate response and how this process is sustained as a working memory across delays*[44].

Of course, understanding what bridges perception to behavior is tantamount to understanding *cognition*. It seems that in half a century, we haven't moved far beyond Atkinson and Shiffrin's original model of human information processing and are certainly no closer to giving an adequate explanation of it. Though this might sound pessimistic, when we evaluate the risks posed by giving working memory such a sweeping role in our cognitive lives, we can begin to piece together a more explanatory picture of cognition. We think that this requires reconceiving of the maintenance and manipulation of information not as a single set of tools wielded by an immensely powerful central controller, but as generic features multiply realized at many levels not only in the brain, but by information-consuming systems generally.

Before proceeding it may be prudent to clarify an exegetical choice that we've made, which may in turn hedge off a few objections to our characterization of working memory. As we've already highlighted, there are a diverse array of working memory models on offer, and we've only reviewed a few—admittedly historically influential views—in the course of our paper. Likely it wouldn't be a difficult project for a clever scholar to propose a variant which sufficiently constrains working memory to a manageable domain; perhaps by explicitly tying it to a given cognitive task or stipulating its occurrence only under a litany of ecological conditions (e.g., holding exactly *n* amount of information over precisely timescale *t* etc.). Such a restricted working memory might escape the force of our negative argument; however, it would also jettison the conceptual appeal that has brought generations of scholars and scientists to study working memory in the first place[45]. Working memory *is supposed to be* "a temporary storage system under attentional control *that underpins our capacity for complex thought*"[46]. Our critique applies to any and all models of a similar ambition, as decomposing such a flexible, generative capacity yielding *complex thought* by individuating its constituent faculties and objects—as one might do under an Aristotelian strategy of Faculty Psychology—approaches an incoherent project. Hence our critique only requires that most of the oft-cited and used accounts of working memory are of a similar ambitious scope, and fortunately this is the case[47].

4. Analogies between *Phantasia* and working memory

We have been arguing that *phantasia* and working memory are both conceived as general faculties of mental representation, which mediate between perceptual inputs and behavioral outputs. But beyond this surface similarity, is working memory really

analogous to *phantasia*? Consider the features, discussed in the previous two sections (Table 1).

		Phantasia	***Working Memory***
Features associated with Phantasia			
1	*Derived from perception*	**Yes**	**Yes**
2	*Responsible for perceptual illusions and dreams*	**Yes**	**No**
3	*Related to memory*	**Yes**	**Yes**
4	*Envisaging*	**Yes**	**Yes**
5	*Responsible for locomotion*	**Yes**	**Unclear**
6	*Abstract thought*	**Yes**	**Yes**
Features associated with Working Memory			
7	*Closely related to consciousness*	**Unclear**	**Yes**
8	*Responsible for most thought*	**Yes**	**Yes**
9	*Maintain and manipulate representations*	**Yes**	**Yes**
10	*Under voluntary control*	**Unclear**	**Yes**
11	*Capacity limited*	**Unclear**	**Yes**
12	*Computationally described*	**No**	**Yes**

Table 1. Comparison of Phantasia and Working Memory.

Working memory and *phantasia* share many traits. They both involve representations derived from, but not simultaneous with perception. The contents of working memory, like the *phantasmata* of *phantasia,* are similar to the contents of their perceptual counterparts. Working memory is thought of as *a kind of memory,* and it is through working memory (in Baddeley's model, the episodic buffer) that we can access contents of our more robust short- and long-term memories[48]. Working memory and *phantasia* are taken to "underpin" complex thought, though many psychologists would allow that we might have propositional contents that are not themselves given in, or prompted by, working memory. For both *phantasia* and working memory, whether they are responsible for all thought will depend on how we describe thought. Working memory is, especially under a global-workplace model of consciousness, conceived of as a place where various modality-specific sensory contents are combined (or synthesized) and abstracted into a "higher-level" representation[49]. Furthermore, it is through working memory that our sensory knowledge is manipulated in the pursuit of some goal or task. This is why it enables us to *envision* possible outcomes and situations[50]. Similarly, Aristotle conceives of *phantasia* as what allows one to *envision* possible outcomes and situations.

It may seem less clear that working memory is necessary for locomotion. But if we understand locomotion as *goal-directed action*, then clearly working memory has a

role to play. Say you're trying to make a margarita. The goal of having a margarita in hand is translated into a series of steps that you must keep in mind in order to carry out the specific sub-actions that constitute margarita-making. Likewise, *phantasia* is supposed to be what connects perceptually given means, for example, the margarita glass and the salt, to the end you have deliberated upon, for example, that you should salt the rim of the glass in order to produce an optimal margarita. *Phantasia* and working memory alike are implicated in the action, since they are required to link your immediate bodily sensations and performances to the goal of the action those activities comprise.

Though *phantasia* and working memory are analogous, they are not identical. Working memory operates on conscious information, and is sometimes conceived as the faculty underlying consciousness[51]. Aristotle had no term for consciousness, so the relationship between *phantasia* and consciousness is unclear in his works. Nevertheless, in all the examples of which we are aware, it is plausible that one would be conscious of an Aristotelian *phantasma*. For instance, when a coward imagines his enemy to be lying in wait for him, he is conscious of his enemy, even though he is apt to be in error. Consequently, we think that the requirement that working memory operate on conscious information does not suggest that it plays a different *explanatory* role from *phantasia*.

We have been arguing that *phantasia* and working memory are implicated in many of the same psychological processes and have similar properties. Crucial to our considerations, however, is that when they are invoked in psychological explanations, they play the same explanatory roles. This is remarkable because it suggests that despite the computational turn that was supposed to transform our conception of mind, we are still working with the same basic faculty of maintenance and manipulation conceived by Aristotle. Though this might be taken as an unexpected approbation of Aristotle's foresight, we think this result is deeply problematic. Rather than suggesting that we are on the right track, the analogy implies that there is something wrong with both concepts, and leads to crucial objections against the explanatory strategy of *phantasia* and working memory.

5. Polemical objections to working memory and phantasia

We have been arguing that working memory in cognitive psychology and *phantasia* in Aristotelian psychology play the same explanatory role. Both are conceived as faculties of central cognition, meant to explain how perceptual content is used in non-perceptual thought. Moreover, both are well-described by the metaphor of the "workspace": it is within working memory or *phantasia* that a psychological subject is supposed to maintain and manipulate perceptual information, deploying it for a given task at hand[52].

Our critique of the explanatory effectiveness of working memory and *phantasia* has two related strands, which we've distilled into two objections. The first, the *cognitive suffusion objection* holds that invoking working memory or *phantasia* as an explanation of any specific cognitive behavior amounts to a bait and switch in which the explanans is just as mysterious as the explanandum. This results from the wide breadth and shallow functional description attributed to these capacities. Insofar as they serve as the workspace for cognitive activity, invoking either of them as genuinely explanatory of some cognitive behavior mistakes a change of terms for an illuminating description of the cognitive phenomenon under investigation. That is, the domain of working memory, or *phantasia*, is largely coextensive with domain of higher cognition—our imaginative ability to flexibly envision, plan and think through our present and future problems and goals. Working memory suffuses cognition. Thus, invoking "working memory" as an answer to why you are able to complete come cognitive task—say, keeping the steps of the recipe in mind as you make the margarita—amounts to little more than invoking "cognition" as an answer to how you manage cognitive work[53].

The computational framework within which working memory is ensconced does little to mitigate this conundrum. It's certainly easy to see how working memory's principal functions, of the maintenance and manipulation of information, could be operationalized in the compositional framework of an algorithm. As a caricature, consider working memory's role in our margarita case in the following computational sketch.

manipulate (MOVE: long term memory, "margarita recipe," [into] working memory);
maintain (REHEARSE: working memory, "margarita recipe," step one);
manipulate (EXECUTE: working memory, "margarita recipe," step one);
… and so on.

Could such a computational description safeguard working memory's explanatory role? We don't think so. First, while this caricature may meet a minimal description of computation, relying on the shallow functional description of working memory as a system that maintains and manipulates information yields a cascade of secondary functions that must be fleshed out: MOVE, REHEARSE, EXECUTE, and so on. The problem here becomes obvious when one realizes that in many models, working memory mediates most robust cognitive processes, including decision making, reading comprehension, flexible problem solving and the like[54]. Successfully enumerating and characterizing the secondary functions that instantiate the maintenance and manipulation of information for flexible thought is equivalent to giving a description for how we perform these cognitive behaviors in the first place. Providing a litany of these secondary functions would, once again, render working memory explanatorily empty.

Second, maintenance and manipulation are ubiquitous in any information consuming system. Indeed, it's hard to conceive of a productive information system that exists over time that does not retain and do something with information. Even a hypotheti-

cal state-based vending machine, without anything approximating a detailed memory, concepts, or rules, could be said to maintain and manipulate information when it moves between predefined states based on the amount and value of coinage deposited[55]. As Christophel and colleagues note, most areas of the brain that are responsible for producing sensory representations can also maintain those representations during the "delay periods" common across working memory tasks[56]. Indeed, one can detect maintenance and manipulation throughout the nervous system. *Every* neuron, insofar as it processes and propagates information, could be said to maintain and manipulate it via its membrane potential. Even the retina, a *prima facie* non-cognitive part of our nervous system, can be understood as maintaining information. As Aristotle himself noted, when you stare at a bright enough picture and then close your eyes, those light detecting cells continue to fire resulting in an after-image (Post. An. 99b36-100a1). Related cells are also responsible for laterally inhibiting—manipulating—one another, for example, in producing the Mach band illusion (Fig. 4).

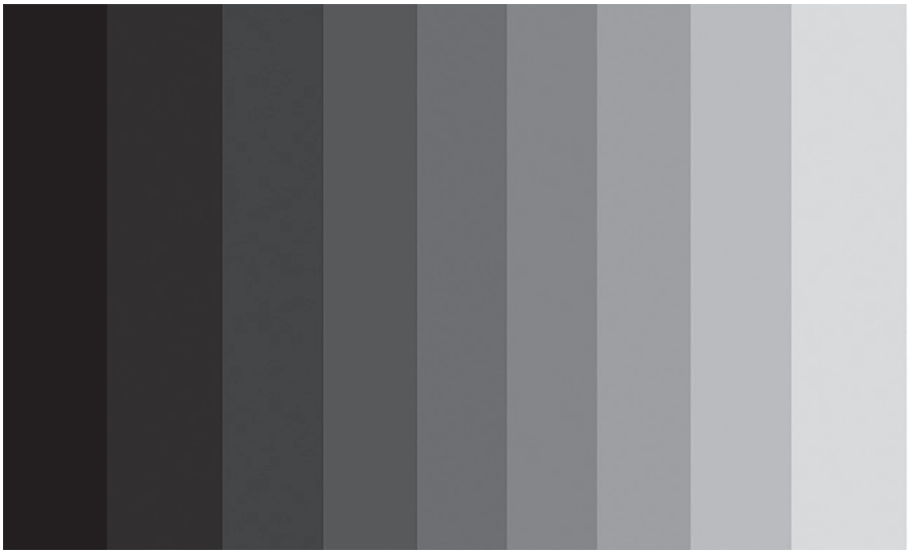

Fig. 4. Mach Band Illusion. As you view the increasingly brighter bands of grey, they each seem to possess a gradient where starting from the left side of each band they appear lighter and then become darker as they approach the next, lighter, band. This is an illusion produced by lateral inhibition amongst neurons in the retina.

The point of this survey is straightforward: maintenance and manipulation are generic to information consuming systems, like the nervous system. We shouldn't be surprised to find mechanisms and pathways at multiple levels that play a role in keeping information in mind and doing something with it.

As a reply to this suffusion objection, one might appeal to working memory models that fragment such a far-reaching capacity into sub-capacities, and thus render the

general capacity tractable to empirical intervention. For instance, Baddeley posits a mechanism that is supposed to explain how we can keep a phone number in mind while searching for a pen and paper: the *phonological loop*. In this model, once phonetic representations of digits are placed in the loop, a simple control structure is responsible for the *sotto voce* continuous rehearsal of those digits, while other cognitive resources are freed up, allowing us to do the work of finding the pen and paper and writing the number down[57]. Could our characterization of working memory as an overly broad capacity of maintenance and manipulation be little more than a straw man?

While it's true that the phonological "loop," and to a lesser extent the visuospatial "sketchpad" offer more tractable, perceptually grounded hypothetical mechanisms for the maintenance and manipulation of information, they cannot be considered solitary or independent systems. Indeed, Baddeley and Hitch refer to them as "slave" systems, which are governed exclusively by their central executive, which determines the what, when, and why of working memory[58]. The loop might retain phonetic contents, but only so you can use them. In the case prefaced above, this amounts to marshalling the cognitive, motivational, and motoric resources necessary to *write the phone number down* while you keep it in mind. This seemingly simple act requires the coordination of a good chunk of the brain. Appealing to these "mechanisms" of a loop and sketchpad do little to free us from the specter of homuncularity, since in fact, they seem to presuppose it.

This discussion of the fragmentation of working memory into subsidiary or perhaps sub-personal components already suggests our second objection: the *bottom-up mereological objection* is the notion that it is fallacious to assume that criteria that can individuate peripheral faculties can be successfully used to individuate those general faculties lying at the heart of cognition[59]. This objection echoes Fodor's concerns about any attempt to empirically characterize what he termed "central processors," those domain-general, informationally promiscuous, effortful, voluntary, and neurally distributed capacities that are marshalled in response to our higher cognitive needs, such as analogical reasoning and imaginative problem solving[60]. Reading Fodor's thesis of *modularity* primarily as an epistemic project, as opposed to a purely descriptive or metaphysical accounting of the modules taken to divvy up the tasks of perception, we can further understand those core, twin criteria Fodor posits of modules—their domain-specificity and their informational encapsulation—as outlining a strategy of *explanation* that surprisingly squares with its earlier Aristotelian counterpart. To be domain "specific" and informationally "encapsulated" is to be restricted in the kinds of questions a system can answer and the kinds of information it can consult in answering those questions[61]. In turn, these restrictions on the problems that a system can solve and limits on how they can go about solving them, render such a system amenable to empirical investigation; e.g., we can begin an examination of the visual system by controlling for the properties of visual stimuli. Central processors, whether dressed

in the jargon of *phantasia* or working memory, are by their nature *unbounded* in the kinds of questions they can entertain and are utterly agnostic to the kinds of information they can query. This in turn clarifies the challenge of explaining and predicting the operations of such domain-general, flexible processes—where does one *even begin* when trying to tame and steady such a boundless and shapeless construct? This is not local to theories of working memory, but is a problem for any faculty psychology. If one adopts Aristotle's idea that faculties are individuated by what they produce and their object-domains, as in the case of vision, it does not follow that cognition in general consists in a single activity or ranges over one set of objects.

6. Towards a post-aristotelian faculty psychology

The parallel functions attributed to *phantasia* and working memory and their purported role as arenas for flexible thought offer an opportunity to reflect on how we understand the science of thinking. What we've called Aristotelian Faculty Psychology aims to explain our mental life and its connection to behavior by individuating psychological capacities via their target-domains, attributes, or functional roles. But the explanatory effectiveness of this strategy decreases with the generality of the explanandum, reaching an asymptote when we attempt to characterize something as catholic as flexible thought. That much, we think, is shown by our *cognitive suffusion objection*, that working memory and *phantasia* re-describe the phenomenon they're tasked with explaining, and by our *bottom-up mereological objection*, that the subsidiary systems of working memory or *phantasia* could not exhaust their extensive purview. In effect, we have outlined the contours of an overarching dilemma: one must either accept the need for expansive realizers of flexible thought, in which case one loses the explanatory resources of faculty psychology, or one must hang onto faculty psychology while dismissing the need for expansive realizers, in which case one is left with much to explain—higher cognition—and little instruction on how to do so.

How do we move beyond this impasse brought about by a commitment to Aristotelian Faculty Psychology? We'll gesture at three possibilities worth exploring, which in turn suggest what a *post*-Aristotelian Faculty Psychology might look like. The first is a *mechanistic* description of cognitive processes, the second is a *generative* account, and the last is a doubling down on the individuation explicit in Aristotelian Faculty Psychology.

Mechanistic thinking in the philosophy of science has undergone a resurgence in the last thirty years [62]. While there are several accounts on offer, Glennan's complex systems approach to mechanisms is helpful for isolating how mechanisms can explain complex phenomena, like cognition. According to Glennan, "a mechanism for a behavior is a complex system that produces that behavior by the interaction of a number of parts, where the interactions between parts can be characterized by direct, invariant, change-relating generalizations[63]." How could isolating a specific psychological

mechanism provide an advantage over the explanatory strategies we have criticized? According to this definition, a mechanism is a mechanism for a *behavior*. Moreover, it is an *object*, since it consists of a number of parts, but it can also be described as *operating* by the interaction of those parts, where interaction is understood causally, as requiring the truth of certain counterfactuals. While the question of whether these complex systems understanding of mechanisms requires modularity is open, it at least some mechanisms could be embedded within one another. Thus, mechanisms can contribute to causal explanations due to their mereological relationships. For example, mechanism A might be part of mechanism B, so that A serves to produce a behavior that makes possible B's operation. Indeed, discovering the functional relationships among mechanisms is a primary goal of scientific research[64]. Though this understanding of mechanisms is quite general, and needs to be developed in greater detail in the psychological case, we think it could inform the design of experiments in cognitive psychology, as well as the interpretation of psychological results by philosophers.

Nevertheless, it may be that constitutively causal explanations and related interventionist models are underequipped to handle dynamic, noisy, and complex entities like the brain[65]. Supposing that the brain plays a major role in realizing cognition and its derivative behaviors, it may be better to consider strategies that allow for more metaphysical flexibility in exposing what are likely multiply realized and non-isomorphic relations between neural matter, cognitive processes, and behavior. Miracchi's *Generative Account* serves as an exemplar for how these descriptions might play out: rather than search for direct casual connections between a "basis model," in this case, neural dynamics, and an "agent model," in this case, cognitive behavior, we can instead posit and explore the space of *in virtue of* relations that a "generative model" requires[66]. By taking seriously and delineating different metaphysical levels of description and explanation, generative models could allow us to separate mere re-descriptions of phenomena from genuine, decomposable explanations. Moreover, moving away from strict and individual causal relations between basis and agent model descriptions allows us to appreciate the many ways that agent level behavior can arise[67]. Of course, this project requires an adequate description of the agent model, that is, of the particular *behaviors* that we want to explain. But neuroscientists have begun to argue that detailed descriptions of behavior could play an important role in neuroscientific theorizing and explanation[68].

Finally, we might try to salvage an important aspect of Aristotelian Faculty Psychology by reinterpreting its principle of individuation. Rather than conceive of faculties in terms of their target-domains or functions, we might instead individuate psychological components by *behavior*. Since faculties work in concert to produce our psychological life, this approach involved no claim that psychological faculties are *de re* separable from one another. Rather, the thought is that making the correct kinds of *de dicto* distinctions—in this case, in terms of behavior—will be explanatory of some family of psychological

functions. This approach borrows from the two prior accounts, although it requires the abandonment of any attempt to attribute our cognitive abilities to broad capacities. For instance, we might have a description—psychological or neural—of how we're able to maintain and manipulate auditory information in a classical Baddeley-styled dual task paradigm, perhaps even with something like the phonological loop, but *in no way* will this description generalize to cover all the cases we want to lump together as working memory. At the same time, this does not prevent us from identifying this description as the start of an explanation of *that particular task-behavior*. In fact, cognitive scientists are beginning to realize the explanatory virtues a tasked-based of a functional decomposition of psychological capacities[69]. This post-Aristotelian Faculty Psychology, which we term *Aristotelian Neo-Behaviorism*, takes behavior to offer the most stable epistemic window onto the workings of the mind. It's by holding behavior fixed that we can start to glimpse and individuate the processes that generate it.

The main point of this paper has been that the phenomenon of flexible thought—the genuine capacity to maintain and manipulate perceptual information, typically by consciously envisioning what one is thinking about—presents an impasse to the divide and conquer strategy that has driven the study of mental phenomena since Aristotle. It does so because any faculty postulated within an ontology of mind meant to realize this capacity will fail to be explanatory of the cognitions in which it is exercised. Likewise, though of little consolation for scholars of Aristotle, it's likely that there is no account of *phantasia* that makes its explanatory role coherent, because it cannot have one within it's own framework, that of Aristotelian Faculty Psychology. Yet in our view, this impasse is not insurmountable: we can hang on to the advantages of Aristotelian faculty psychology, principally its methodology of individuating faculties in order to explain cognitions, as long as we do not attribute an explanatory role to capacities for flexible thought, like *phantasia* and working memory. Thus, our negative view ultimately supports an optimism about psychological explanation. Far from being inherently mysterious, or depending on as yet undiscovered special methods of investigation, cognition is open to investigation, if only we isolate the correct mechanisms, make clear our ontological assumptions, and make behavior the ultimate tribunal by which we judge any proposed psychological theory.

Bibliography and notes

Acknowledgements

The authors would like to thank the audience at the 2016 Classics and Cognitive Theory conference hosted by NYU, Rosemary Twomey, Matthew Rachar, Iakovos Vasiliou, and Lisa Miracchi Titus and the MIRA Group at the University of Pennsylvania for their feedback and many helpful comments throughout the drafting of this paper. In addition we would like to Maria Cuellar for her help creating our figures.

1. For two prominent examples, consult: Singpurwalla R, Plato and the tripartition of the soul. In: Sisko JE (ed.), Philosophy of mind in antiquity. London: Routledge; 2019. pp. 101-119; Gendler TS, The third horse: On unendorsed association and human behaviour. Proc Aristot Soc. 2014;88:185-218.
2. The personal/subpersonal distinction is a thorny one, as discussed in: Drayson Z, The personal/subpersonal distinction. Philos. Compass 2014;9:338–346.We make no firm commitments regarding the specific and alternative readings of "sub-personal," but use it throughout for its connotations of conscious inaccessibility and decompositionality. Some faculties decompose into subsidiary systems whose workings may not be available to introspection. We would like to thank an anonymous reviewer for prompting us to clarify this point.
3. Cooper JM, Hutchinson DS (trans.), Plato: Complete Works. New York: Hackett; 1997; Barnes J (ed.), The complete works of Aristotle: the revised Oxford translation. Vols. 1-2. Princeton, NJ: Princeton University Press; 1984.
4. Johansen TK, The powers of Aristotle's soul. New York: Oxford University Press; 2012. p. 94.
5. Ibid. pp. 202-203.
6. Unless otherwise indicated, the translations are our own.
7. This has suggested to many commentators that Aristotle has in mind something like the British Empiricist conception, according to which imagination reproduces and is parasitic upon the perception that brings it about. Humphreys J, Aristotelian Imagination and Decaying Sense. Soc Imagin. 2019;5(1):37-55 argues that this reading is too strong: the definition does not commit Aristotle to a purely reproductive conception of phantasia, but to the weaker view that phantasia can't be active on its own, requiring a perception to trigger it.
8. Aristotle conceives of phantasia as being so much "like" perception (DA 429a5) that he sometimes describes them as sharing the same faculty (Insom. 459a15-20).
9. Consult the sources of this debate, Sorabji R, Body and soul in Aristotle. Philosophy 1974;49:63-89; Burnyeat MF, Is an Aristotelian philosophy of mind still credible? In: Nussbaum MC, Rorty AO (eds), Essays on Aristotle's de anima. Oxford: Oxford University Press; 1995. pp. 15-26.
10. For an analysis of Aristotle's discussion of habituation in the Nicomachean Ethics, consult Jimenez M, Aristotle on 'steering the young by pleasure and pain.' J Specul Philos. 2015;29(5):137-164.
11. We owe this point to an anonymous reviewer.
12. For discussion of the view that φαντάσματα (phantasmata) are not the objects of thought, but rather the vehicles for intellect to grasp its proper object, consult Jimenez ER. Aristotle's concept of mind. Cambridge: Cambridge University Press; 2017. pp. 63-67.
13. Though it is beyond the scope of our argument here, for discussion of the action-determining capacity of phantasia see DA 433b28-29, 434a2-7.
14. Scholars of Aristotle have offered various reconstructions of how Aristotle thinks cognitive powers can determine desire, and thus motivate action. Polansky (Polansky R, Aristotle's de anima. Pittsburgh. PA: Duquesne University Press; 2007. pp. 528-29.) argues that appetite, the most basic form of desire possessed by all animals, presupposes phantasia. In animal action, phantasia has the function of presenting something as pleasant or painful for the animal, and thereby sets off a desire for that thing. Moss (Moss J, Aristotle on the apparent good. New York: Oxford University Press; 2012. p. xii.) argues that

phantasia is the capacity responsible for the appearance of something as good. Though details of these accounts could be disputed (for example, one might ask how something can appear as pleasant without an agent acting on that appearance, or whether something can appear as good, without an agent thereby pursuing it), it is widely agreed that for Aristotle phantasmata that present something as good, and hence can determine what the animal desires.

15. For account of how this works in the geometrical case, see Humphreys J, Abstraction and diagrammatic reasoning Aristotle's philosophy of geometry. Apeiron 2017:50(2):197-224.
16. We thank an anonymous reviewer for pointing out that phantasia may be genuinely explanatory by Aristotle's lights. For it appears that Aristotle defines phantasia, gives it a substantial causal role in the behavior of animals, and, at least in the human case, says for the sake of what phantasia exists. However, if the phantasia model is genuinely explanatory according Aristotle, our objection will cut not only against the deployment of Aristotelian Faculty Psychology but also against the Aristotelian account of explanation itself.
17. Consult: Cowan N, The many faces of working memory and short-term storage. Psychon Bull Rev. 2017;24:1158-1170.
18. Carruthers P, On central cognition. Philos Stud. 2014;170:143-162. For an account arguing the negative claim that working memory is not a natural kind, consult: Gomez-Lavin J, Working memory is not a natural kind and cannot explain central cognition. Rev Philos Psychol. 2021;12:199-225.
19. Newell A, Simon HA, A logic theory machine and a complex information processing system. Santa Monica, California: The RAND Corporation; 1956.
20. Miller G, Galanter E, Pribram KH, Plans and the structure of behavior. New York: Henry Holt and Company; 1960. However, we should note that roots of working memory are anchored in a traditional beginning with 19th century psycho-physicists, from Wilhelm Wundt to William James, who theorized a distinction between primary and secondary memory (Consult: Atkinson RC, Shiffrin RM, The control processes of short term memory report number 173. Stanford: Institute for Mathematical Studies in the Social Sciences; 1971. p.1.). Secondary memory roughly maps onto what we might call long-term memory, wherein episodes from one's past—such as what you had for lunch three days ago—can be brought back to the fore of your mind. Primary memory, on the other hand, is closely tied to one's recent experience. Though this division between primary and secondary memory, which we would now call short- and long-term memory, was "discarded" with the turn to behaviorism in the early 20th century, it returned with the rise of mathematical and cognitive psychology in the midcentury.
21. Miller G, The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol Rev. 1956;63:81-97, for his discussion of information density. Aristotle also seems to countenance the possibility of compressing information in memory, by associating certain facts with places, for example, in a building. See e.g. De Memoria 452a12-16 and Insom. 458b20-22.
22. Atkinson RC, Shiffrin RM, ref. 20.
23. Ibid., p. 5. Emphases ours.
24. Ibid., pp. 4-5. Importantly, an explicit identification of working memory as the substrate or psychological root of consciousness drops out after the Atkinson and Shiffrin model. Still, many contemporary theories attempt to accommodate either the conscious experience of holding information in mind (e.g., like Baddeley's episodic buffer discussed later) or identify consciousness with some kind of "workspace" that shares many (if not

most) functional traits with these models of working memory (e.g, Global Workspace or Broadcast theories; for more, consult: Carruthers P, The centered mind: what the science of working memory shows us about the nature of human thought. New York: Oxford University Press; 2015.).

25. Ibid., p. 2.
26. Ibid., p. 4.
27. Ibid., p. 9.
28. James notes the existence of similar tasks in the 19th century; e.g. Paulhan's introspective poetry task. James W, The principles of psychology. Vol 1. New York: Holt; 1890. p. 408.
29. Baddeley A, Hitch G, Working memory. Psychol. Learn. Motiv. 1974;8:47-89.
30. Shallice T, Warrington EK. Independent functioning of verbal memory stores: a neuropsychological study. Q J Exp Psychol. 1970;22(2):261-273.
31. Baddeley A, Working memory. Curr Biol. 2010;20:R136-140.
32. Ibid., p. R136.
33. This model frames most empirical papers' introduction to the topic of working memory (consult, for example: Repovš G, Bresjanac M, Cognitive neuroscience of working memory: a prologue. Neuroscience 2006;139:1-3; Rypma B, Factors controlling neural activity during delayed-response task performance: testing a memory organization hypothesis of prefrontal function. Neuroscience 2006;139:223-23; Marchuetz C, Smith EE, Working memory for order information: multiple cognitive and neural mechanisms. Neuroscience 2006;139:195-200.). Furthermore, while explicit discussion of the 'voluntary control' of working drops out of their model, every other feature inherent in Atkinson and Shiffrin's model is carried over to Baddeley's iteration.
34. These alternatives include Cowan's "Long-term working memory" model, and "generic" models of working memory. Consult Cowan N, ref. 17.
35. Baddeley himself—as with many other early cognitive psychologists—came from a linguistics background and this is evident from most of his early work on the phonological buffer and its computational implementation (consult Baddeley A, Working memory, thought, and action. New York: Oxford University Press; 2007 for an overview). One suspects that the ready adaptation of preexisting linguistic and psychophysical paradigms and tasks encouraged by Baddeley's sensory approach to working memory facilitated its rapid adoption as the de facto model of working memory.
36. Though, initially it was proposed as a proper part of the central executive. Why this was a problem should be clear shortly.
37. This is recounted: Logie RH, Retiring the central executive. Q J Exp Psychol. 2012;69:2093-2109.
38. Ibid.
39. Fodor J. Modularity of Mind. Cambridge, MA: MIT Press; 1983. p. 107.
40. There are even some paradigms which claim to find results that support a parallel system of unconscious working memory (consult: Soto D, Mäntylä T, Silvanto J, Working memory without consciousness. Curr Bio. 2011;21:R912-R913), and that isn't so surprising, given its wide scope and role in cognition.
41. Rottschy C, Langer R, Dogan I, Reetz K, Laird AR, Schulz JB, Fox PT, Eickhoff SB, Modelling neural correlates of working memory: a coordinate-based metaanalysis. NeuroImage 2012;60(1):830-846.
42. Jerde TA, Merriam EP, Riggall AC, Hedges JH, Curtis CE, Prioritized maps of space in human frontoparietal cortex. J Neurosci. 2021;32(48):17382-17390.

43. Christophel TB, Klink PC, Spitzer B, Roelfsema PR, Haynes J, The distributed nature of working memory. Trends Cogn Sci. 2017;21(2):111-124.
44. Ibid., p. 120.
45. This largely simplifies and parrots a similar dilemma posed to proponents of working memory by Gomez-Lavin J, ref. 18. On the one hand, ambitious broadly-construed models will fail to adequately isolate the relevant mechanisms that should support the cognitive activities philosophers and psychologists are interested in studying, and, on the other, a narrower model hitched to a single mechanism will likely fail to generalize (and hence explain) that same diverse array of cognitive activities.
46. Baddeley A, ref 35, p. 1. Emphasis ours.
47. Besides Baddeley's multicomponent model still finding itself cited as a "standard view," (cf. Beukers AO, Buschman TJ, Cohen JD, Norman KD, Is activity silent working memory simply episodic memory?. Trends Cogn Sci. 2021;25:284-293.) a cursory review of hot-off-the-presses article recapitulates a preference for a broad and ambitious account of working memory "Higher cognition depends on working memory (wm), the ability to maintain and perform operations on internal representations of information no longer available in the environment" (cf. Hallenbeck GE, Sprague TC, Rahmati M, Sreenivasan KK, Curtis CE, Working memory representations in visual cortex mediate distraction effects. Nat Commun. 2021;12:4714.). As an exercise, take any modern cognitive neuroscience textbook and flip to the "working memory" subsection and you'll just as likely find a similar characterization, often with a Baddeley citation to boot (e.g., Gazzaniga MS, Ivry RB, Mangun GR, Cognitive neuroscience: the biology of the mind. 3rd ed. New York: W.W. Norton & Co.; 2009.).
48. Baddeley A, ref. 31.
49. Prinz J, The conscious Brain. New York: Oxford University Press; 2012. p. 320.
50. For example, consider how the visuospatial sketchpad is thought to play a crucial role in mental rotation tasks, where a participant must mentally translate and move a two-dimensional depiction of a three-dimensional object; e.g., Shepard RN, Metzler J, Mental rotation of three-dimensional objects. Science 1971;171:701-703.
51. Cf. Carruthers P, ref. 24.
52. As highlighted above, Carruthers P, ref. 18 makes this metaphor explicit.
53. One might instead argue that phantasia, as it is set out in the present literature, attempts to provide a satisfactory explanation along Aristotelian lines, and that we are being uncharitable by judging his project along the lines of a contemporary account of explanation borrowed from the philosophy of science; one that privileges a functional decomposition that ties psychological states to stimuli and behaviors. Under this admittedly anachronistic criterion phantasia falters as it does not provide a satisfactory decomposition of one of its primary functional roles: namely, helping to bridge the recently experienced past to the present in the service of (most if not all) thought. We argue that this is due, in part, to its wide sweeping purview. After all, what are the proper targets, objects or processes that exhaustively characterize thought? If one could adequately characterize the targets, objects, and processes of thought, then it's arguable that any need for functional middlemen would simply drop out of the account. At the same time, because these middlemen are wedded to most instances of thought, beginning to characterize them in terms of their targets, objects or processes (or to use the more contemporary language of stimuli, tasks and behaviors) is tantamount to the prior project of characterizing thought simpliciter. It's in this sense that we hold that one can move between these terms, from phantasia or

working memory to cognition (or vice versa), without a serious impact to the semantics of a statement set out in the terms of a functionalist, decompositional explanation. The normative force of our claim that a "mistake" has been made cuts principally against those that would accept a tokening of one of these purportedly subsidiary faculties, like working memory, as itself an explanation of some further cognitive event or process, for instance consciousness (i.e. Carruthers P, ref. 24.) and as such is aimed less directly at the Aristotelian project. We'd like to thank an anonymous reviewer for encouraging us to clarify this point.

54. Jonides J, Nee DE. Brain mechanisms of proactive interference in working memory. Curr Bio. 2006;21:R912.
55. This example is depicted in Fodor J. The mind-body problem. Sci Am. 1981;244:114-123.
56. Christophel TB, ref. 43.
57. Baddeley A, Logie R, Working memory: the multiple-component model. In: Miyake A & Shah P (eds), Models of Working Memory. New York: Cambridge University Press; 1999.
58. Cf. Baddeley A, ref 29.
59. This objection is adapted from Bennett and Hacker (cf. Bennett MR, Hacker PMS, History of cognitive neuroscience. Chichester: Wiley; 2008.), who take a more radical "top-down" approach wherein they generalize the objection to most mental terms. One commits a top-down mereological error when one attributes person-level psychological states exclusively to a sub-personal, psychological or neural capacity. While we're sympathetic to aspects of this top-down approach, it's not quite clear that working memory or phantasia are in the first place thought of as person-level states, like intelligence or judgement. They're meant to give an account of how these person-level states arise, and to do so they depend on the functional individuation of psychological capacities that have been—as we've discussed—so productive in helping us understand perceptual processes. Though Bennett and Hacker reintroduced this objection, versions of it have shadowed the development of the neural sciences since the time of Gall and his early phrenology. Gall, for instance, argued that there could be no "faculty" for judgement, as judgment is individuated by its target domains (e.g., judgement for arithmetic, for rhetoric, for music etc.), and so would proliferate throughout the mind and its substance, the brain (cf. Hollander B, In search of the soul, and the mechanism of thought, emotion, and conduct. London: Kegan Paul, Trench, Trubner & Company, Ltd.; 1920. p. 240). A similar mereological line can be found in Ladd's critique of early psychophysics (Ladd GT, Elements of physiological psychology. London: Longmans, Green & Co.; 1887. p. 545).
60. Cf. Fodor J, ref. 39.
61. Ibid., p. 102.
62. Consult: Craver C, Explaining the brain: mechanisms and the mosaic unity of neuroscience. New York: Oxford University Press; 2007. However, mechanisms have been invoked by philosophers as essential to at least some sorts of scientific explanation since the time of Aristotle. The Mechanica, a work attributed to Aristotle, describes a number of moving, physical systems geometrically, setting off the science of mechanics that described idealized or abstracted physical systems in mathematical terms. It is striking, then, that despite thinking of psychology as a systematic science, Aristotle made no such attempt to hypothesize mechanisms in that domain! Though mechanics continued to develop in the ancient works, for example, in the work of Archimedes and Heron, it was not until modern times that the concept of a mechanism again became a topic of

philosophical reflection, in the work of philosophers like Hobbes and Descartes. Yet even here, mechanical science was often conceived in opposition to psychology, perhaps most famously in Descartes' division of the unextended soul from the completely mechanical material world. But with the development of abstract laws of Newtonian physics, and later relativity and quantum mechanics, mechanisms seemed to pass out of philosophers' accounts of scientific method. Of course, scientists still employed mechanistic ways of thinking, for example, depicting specific processes with diagrams or physical models that portrayed causal and functional relationships. But the syntactical conception of science dominated the scene: whatever cognitive aids scientists might employ in practice, the real subject matter of science was taken to be universal laws.

63. Glennan S, Rethinking mechanistic explanation. Philos Sci. 2002;69:S344.
64. Consult: Woodward J, What is a mechanism? a counterfactual account. Philos Sci. 2002;69:S366-S377.
65. Consult Gomez-Lavin J, Why expect causation at all? A pessimistic parallel with neuroscience. Biol Philos. 2019;34:1-6 and Jonas E, Kording KP, Could a neuroscientist understand a microprocessor?. PloS Comput Biol. 2017;13:1-24.
66. Miracchi L, Generative explanation in cognitive science and the hard problem of consciousness. Philos Perspect. 2017;31:267-291; Miracchi L, A competence framework for artificial intelligence research. Philos Psychol. 2019; 5:588-633. Quotations are derived from page 283 of the 2017 paper and page 607 of the 2019 paper.
67. A way we like to characterize this strategy as applying to cognitive neuroscience is as follows: Rather than attempt to describe how all the possible states of a given neural population, say neurons in the prefrontal cortex, correlate with a given behavior—for example, holding items in mind in a working memory task—we can generalize across the data we have that suggest that the prefrontal cortex in all its messy neural glory is a difference maker to working memory task performance. While this doesn't exhaust the prefrontal cortex's role, it highlights its relevance given the parameters set by a certain kind of behavior; namely, task performance.
68. Krakauer JW, Ghazanfar AA, Gomez-Martin A, MacIver MA, Poeppel D, Neuroscience needs behavior: correcting a reductionist bias. Neuron 2017;93:480-490.
69. Consult: Burnston D, Getting over atomism: functional decomposition in complex neural systems. Br J Philos Sci. 2019;72(3). DOI: 10.1093/Bjps/Axz039