Revisiting the Mechanism Sketch Argument: A Task-Based Account of Functional Analysis

Aliya Rumana

Department of Philosophy, University of Texas at El Paso

Word Count: 8965 (all inclusive)

29-10-2025

Functional analysis is often thought to be distinct from mechanistic explanation because it abstracts away from the structural dependencies of organism behaviour. Piccinini & Craver (2011) argue this isn't sufficient to split functional analysis and mechanistic explanation into separate kinds of explanation. I'll argue that they are right, despite various objections. However, functional analysis involves task structure in a way that mechanistic explanation does not. I'll argue that if tasks are constituted independently of mechanisms, then functional analysis is distinct from mechanistic explanation. Thus, disagreement about the mechanism sketch argument may boil down to disagreement about the nature of tasks.

Keywords: functional analysis; mechanistic explanation; task analysis; constraint-based explanation; mechanism-task fit

The received (if disputed) view in philosophy of neuroscience is that explanation in cognitive and behavioural neuroscience is *mechanistic explanation*: we explain organism behaviour by representing the causally and spatiotemporally organised set of working parts on which organism behaviour depends (e.g., Machamer et al., 2000; Craver, 2007; Bechtel, 2008; Piccinini, 2021). Working parts are generally thought to be compound entities, which include both occurrences (activities or the instantiation of functional properties) and the objects that participate in the occurrences (entities or the instantiation of structural properties). Since there is an ongoing debate between activists (who prefer activity-talk) and passivists (who prefer property-talk), Kaiser & Krickel (2017) helpfully suggest that working parts be described in neutral terms as *object-involving occurrences*.

The received view in philosophy of psychology is that explanation in cognitive psychology is *functional analysis*. Cummins (1983) initially proposed that functional analysis explains a complex capacity possessed by a system by representing the organised set (called a "program") of simpler sub-capacities on which the system's capacity depends. Capacities are dispositions, but they are *manifested* by occurrences. As a result, most philosophers of psychology now characterise functional analysis as an analysis of occurrences: it explains an occurrence (which manifests a complex capacity possessed by a system) by representing the causal network of occurrences (which manifest the simpler sub-capacities), on which the explanandum occurrence causally depends. I'll employ the latter characterisation in this paper.

There are deep similarities between mechanistic explanation and functional analysis. Both are *componential*: both explain something about a system by representing something about its parts. Both involve *occurrence*: both explain certain occurrences, which involve a system (e.g., organism behaviour), by representing an organised set of other occurrences, which involve the system's parts (*inter alia*). The key difference, many think, is that mechanistic explanation explicitly represents the organism who is involved in behaviour and the organism parts that are involved in the explaining occurrences, whereas functional analysis abstracts away from the objects involved. This has led Piccinini & Craver (2011) (henceforth, "P&C") to argue that functional analysis is an incomplete form ("sketch") of mechanistic explanation: it is just mechanistic explanation that explicitly represents occurrences but abstracts away from the objects involved in them. This is the "mechanism sketch argument" (henceforth, "MSA").

One response to the MSA is to maintain that whether to include structural (i.e., objectual) dependencies is a "big enough" difference to split functional analysis and mechanistic explanation apart (e.g., Weiskopf, 2011; Barrett, 2014; Shapiro, 2017; Rosenberg, 2018). I'm sceptical that this works, for reasons we'll see below. There is another avenue for responding to the MSA, though. After all, there is a neglected sense in which functional analysis is *task-dependent*: psychologists incorporate empirical information about participant behaviour into a task analysis to develop a functional analysis of participant behaviour (Rumana, 2022). Although tasks are often used to find evidence for assembling mechanistic explanations, they don't seem to be constitutively involved in mechanistic explanations (Rumana, 2025). So, task-dependence may be a more promising way of splitting functional analysis and mechanistic explanation apart.

Whether this objection to the MSA succeeds depends on our conception of task. Suppose we take a *mechanism-dependent* conception of task, such that tasks are just what mechanisms (or their parts) do, or perhaps, what they have the function to do. Then it would be unclear how different functional analysis really is from mechanistic explanation. For more success, we'd need to take a *mechanism-independent* conception of task, such as that tasks involve concrete situations that ground the normative status of possible responses to them—independent of the normative commitments of either experimenters or participants. Then task-dependence would introduce a novel element to functional analysis that can't be reduced to mechanistic explanation. This should be enough to split functional analysis and mechanistic explanation into separate kinds.

In this paper, I'll argue that whether functional analysis is genuinely distinct from mechanistic explanation depends on whether we endorse a conception of task that is constituted independently of the mechanisms that perform the task. In §1, I'll develop a version of the MSA that is neutral to controversies about the nature of explanation. In §2, I'll develop a version of the MSA that addresses objections to P&C's version of the MSA, which justifies the inseparability of structural and functional dependencies. In §3, I'll develop a mechanism-independent conception of task. In §4, I'll show that if tasks are mechanism-independent in this way, they play a constitutive role in functional analysis. In §5, I'll show that if we endorse a mechanism-independent conception of task, we're ultimately committed to rejecting the reconstructed MSA. In §6, I'll conclude that disagreement about the MSA ultimately boils down to disagreement about the correct way to characterise the tasks that mechanisms perform.

§1. Explanatory Kinds

Piccinini & Craver's (P&C's) (2011) mechanism sketch argument (MSA) starts with a plausible intuition: if adding information about the structural dependencies of organism behaviour is sufficient to turn functional analysis into mechanistic explanation, then the two aren't genuinely distinct kinds of explanation. However, it's difficult to cash out this intuition without taking controversial positions in the philosophy of explanation. I suspect this is part of the reason why the critical response to the MSA has been so harsh: P&C seem to be committed to the ontic account and it's unclear whether critics recognise this. In this section, my goal is to reconstruct the MSA in a way that maximises neutrality to accounts of explanation. In doing so, I hope to start recovering the plausibility of P&C's insight.

I'll reconstruct the MSA in two separate steps, one in this section and another in the next. Also, I'll develop it mostly independently of P&C's presentation, to streamline our reconstruction. To start, consider a rudimentary argument, which forms the skeleton of the MSA:

> **Main premise:** functional analysis and mechanistic explanation represent the same kind of ontological dependencies.[1]
>
> **Conclusion:** therefore, functional analysis and mechanistic explanation belong to the same kind of explanation.

This argument is clearly invalid unless we introduce a bridging premise, which addresses the relation between the ontological dependencies that explanations represent and the kinds of explanation to which they belong. However, we have at least two important options for the bridge premise: a stronger, more controversial premise that P&C endorse and a weaker, less controversial one that I'll endorse in this paper. I've chosen to start with the rudimentary form of the P&C's argument to make this choice explicit.

While this rudimentary argument is incomplete (e.g., invalid), it's already controversial to some extent. After all, its premise takes for granted that functional analysis and mechanistic explanations are representations. One way to disagree with this is to say that explanations (of phenomena) are the ontological dependencies (of the phenomena) themselves. This claim is often associated with the ontic account of explanation (*a la* Salmon, 1989; Craver, 2007, 2014). More recently, though, many have suggested that the core insights of the ontic account can be retained by taking explanations to be representations of ontological dependencies (Illari, 2011; Illari & Williamson, 2011; Craver, 2019). Consistent with this, I'll reconstruct P&C's argument under the assumption that explanations are representations. However, I think my reconstruction could be rephrased back into "ontic terminology" without loss.

Another way to disagree with our rudimentary argument is to say that explanations are fictional or non-representational models. This is an increasingly popular view, especially for capturing the way in which ideal models explain phenomena (e.g., Godfrey-Smith, 2006; Kennedy, 2012; Batterman & Rice, 2014; Frigg, 2010; Frigg & Nguyen, 2016). There's a moderate version of

[1] By the 'dependencies of a phenomenon', we could be referring to the conditions on which a phenomenon depends or the relations by which it depends on those conditions. I don't think anything important hangs on this distinction, so I'll refer to both jointly as "dependencies".

this view that is compatible with the premise of our rudimentary argument. We could say that *some* explanations (e.g., ideal and fictional ones) are non-representational, other explanations are representational, and both functional analysis and mechanistic explanation belong to the latter kind (c.f., Bokulich, 2018). This concession won't satisfy those who deny *any* explanatory models are representational, of course, but I doubt any version of the MSA will impress them.

Let's turn to the business of finding a bridging premise, to convert our rudimentary argument into a valid one. The easiest way to bridge the gap between the premise and conclusion is:

> **Conditional bridge premise:** functional analysis and mechanistic explanation belong to the same kind of explanation if they represent the same kind of ontological dependencies.

This bridge premise is a conditional claim, not a biconditional one (unlike the alternative bridge premise below). Thus, it claims that one criterion (among potentially many) for lumping explanations into the same kind is that the explanations represent the same kind of ontological dependencies. To flag this point, I'll refer to the kind of explanation picked out by this criterion as "target-based kinds", since the ontological dependencies are the representational targets of explanation.

Thus, this conditional bridge premise fits comfortably with a permissive pluralism about explanatory kinds. For example, we could split explanatory models into different kinds if they use different formalisms, like when Chemero & Silberstein (2008) distinguish dynamical models from mechanistic explanations because they use differential equations and mechanistic explanations generally do not (c.f., Kaplan & Craver, 2011). Or we could split them into different kinds if they are guided by different metaphors, like when Ross (2021) distinguishes between explanations that invoke pathways vs. mechanisms. We could call these "vehicle-based kinds", since models are vehicles for explanatory representation. Our bridge premise could make plausible claims about target-based kinds that wouldn't be plausible for vehicle-based kinds.

However, P&C close the inferential gap in our rudimentary MSA with a stronger bridge premise. In particular, they seem to endorse a biconditional claim, rather than a conditional one:

> **Biconditional bridge premise:** functional analysis and mechanistic explanation belong to the same kind of explanation if *and only if* they represent the same kind of ontological dependencies.

Prima facie, this bridge premise is committed to the narrow monist claim that all explanatory kinds are target-based kinds. But there's a more pluralist way of interpreting this premise. We could grant that there are plenty of explanatory kinds, but only target-based kinds are "objective" or "natural" or "privileged". Consistent with this proposal, the biconditional bridge premise could be understood as a claim about privileged explanatory kinds: they (but not unprivileged explanatory kinds) are target-based kinds.

Why might we think target-based kinds are privileged explanatory kinds? One reason we might think this is the ontic account of explanation, which claims that models derive their status (and power) as explanations from their targets, the ontological dependencies that they represent (Craver, 2019; c.f., Wright & van Eck, 2018). This gives the representational targets of explanation a privileged role: demarcating which models are explanatory (the ones that represent

ontological dependencies) and which are not (the ones that don't) (c.f., d'Alessandro, 2020). To be clear, the ontic account isn't an account of what distinguishes types of explanation per se. My point, though, is just that if representational targets can confer explanatory status (and power) on models, as the ontic account suggests, then they plausibly have the power to individuate privileged kinds of explanation too.[2]

I happen to be sympathetic to this stronger premise, but it is very controversial (e.g., Lycan, 2005; Bechtel, 2008; Batterman & Rice, 2014; Bokulich, 2016, 2018; Wright & van Eck, 2018). However, I won't dispute whether this stronger premise is true. After all, arguments should avoid stronger premises when their conclusions can be deduced from weaker premises. We can broaden the appeal of the MSA by selecting the conditional (vs. biconditional) bridge premise. Moreover, this will help us divert disagreement about the MSA from tangential issues in the philosophy of explanation to issues in the philosophy of psychology and neuroscience: i.e., whether functional analysis and mechanistic explanation represent the same or different kinds of ontological dependencies.

Finally, I should clarify that while P&C set out to challenge the assumption that functional analysis and mechanistic explanation are distinct, they don't stop there. After all, they reach an asymmetrical conclusion: functional analyses are incomplete "sketches" of mechanistic explanations, but not vice versa. The reason, they say, is that the set of ontological dependencies that functional analysis and mechanistic explanation represent form something called a *mechanism*. It is in this (target-based) sense that functional analysis and mechanistic explanation are both mechanistic explanations (i.e., they are explanations of mechanisms). While this further claim warrants evaluation, I'm more interested in evaluating their symmetrical claims. Thus, we'll limit our consideration to the "core" of P&C's MSA, as represented by the symmetrical premises and conclusion we've considered here.

§2. Object Involvement

Prima facie, there is an obvious objection to the MSA: functional analysis represents only the functional dependencies of organism behaviour whereas mechanistic explanation represents both its functional and structural dependencies, so they must belong to different target-based kinds of explanation. P&C reject this objection, for reasons that critics have complained are unclear at best or question-begging at worst (Shapiro, 2017). In this section, I'll develop a rejoinder to this objection that defends the MSA for a particular variant of explanation—singular non-contrastive explanations of individual organism behaviours. However, I'll concede that this objection succeeds for other variants of explanation. Thus, I'll propose a scope restriction for the MSA.

Let's start with P&C's argument for why explanations of organism behaviour ought to represent both structural and functional dependencies (i.e. dependencies on object and occurrence, respectively). Their key premise is that object and occurrence are ontologically co-dependent. Critics don't disagree with this (Shapiro, 2017). From this, however, P&C argue that a representation of an active system's dependencies will be incomplete unless it represents *both* their functional and structural dependencies. To reach this conclusion, it seems that P&C must

[2] The latent assumption here is that special status (or "privilege") transfers from properties to the types individuated by those properties. I'm sure one could reasonably disagree with this assumption, but I'll see aside that doubt here.

implicitly endorse a bridge premise: the co-dependency of object and occurrence make them *inseparable*. It is this bridge premise that most critics wish to reject: the distinction between objects and occurrences strikes most of us as sufficiently robust to make it possible to represent occurrences separately from the objects involved in them (Weiskopf, 2011; Rosenberg, 2018).

To press this point, critics of the MSA have pointed to well-known reasons for representing functional dependencies separately from structural dependencies. They claim that functional dependencies are more *stable* or *projectable* (more supportive of generalisation) than structural dependencies, due to multiple realisation (Weiskopf, 2011; Rosenberg, 2018; c.f., Fodor, 1974; Pylyshyn 1984). At the same time, many neuroscientists and philosophers of neuroscience would argue that there is a sense in which objects are more stable than occurrences, as in *neural reuse* (Anderson, 2015; Burnston, 2018). To be clear, I don't think these claims are inconsistent: object and occurrence may be more stable with respect to different phenomena. My point is just that there are reasons from both psychology and neuroscience for separately representing object and occurrence.

One rejoinder would be to double down on the view that the distinction between objects and occurrences isn't sufficiently robust to make it possible to represent occurrences without representing the objects involved in them. For example, Kim (1992, p. 326) argues that functional properties aren't *projectible* (they can't support lawlike generalisation) when they are multiply realised, because they are realised by objects with different causal powers. His classic example is that even though jadeite and nephrite share similar functional properties (e.g., resist plastic deformation, scatter similar bands of green light), functional generalisations over jadeite aren't projectible to nephrite because jadeite and nephrite are different kinds of objects with different kinds of causal powers. If that's right, functional generalisations over occurrences must be implicitly indexed to the objects involved, because they wouldn't be projectible otherwise.[3]

By comparison, I prefer a second rejoinder, which concedes that MSA fails for certain types of explanation but insists it succeeds for others. First, we could concede that the MSA fails for the generic explanation of *types* of organism behaviour. *Contra* Kim (1992) and the first rejoinder, we could allow that functional dependencies are projectible even when they are multiply realisable (i.e., when multiple different objects could be involved). In other words, general types of organism behaviour could have stable functional dependencies without having stable structural dependencies. This would be a good reason for functional analysis to represent the (stable, projectable) functional dependencies of organism behavioural types without representing their (unstable, unprojectable) structural dependencies.

Considerations of stability, projectability, generalisability, etc. are relevant for generic explanations, but they aren't relevant for specific explanations of *tokens* of organism behaviour, i.e., *individual* organism behaviours. Specific explanations should represent the dependencies of individual organism behaviours regardless of whether those dependencies are found in other individuals. Thus, restricting the scope of MSA to singular explanations screens offs the key advantage touted for representing functional dependencies without representing structural ones.

[3] I thank Jeremy Pober for raising this point.

Moreover, it is quite compatible with the fact that New Mechanists like P&C generally focus on singular explanation, whereas philosophers of psychology prefer to focus on generic explanation.

However, the MSA isn't out of the woods yet. Consider the distinction between contrastive and noncontrastive singular explanations. Dretske (1988) gives a classic example of the former: the explanation for why Socrates died rather than lived. Few things are relevant to contrastive explanations: e.g., that Socrates consumed a beverage with vs. without poison. By comparison, an explanation for the event wherein Socrates died would be an example of a non-contrastive explanation. What's relevant to it is much more encompassing: *everything* that could make *any* difference to the event of Socrates' death, including but not limited to the fact that he consumed a beverage with vs. without poison (Craver & Kaplan, 2020, pp. 296–297).

Shapiro (2017) cites a study by Sternberg (1969) as an example. Sternberg presented participants with a list of numbers, subjected them to a delay, presented them another (test) number, and then asked them whether the test number appeared on the list. He found that participants exhibited equal response times, when test numbers were vs. weren't present on the list. He explained that this suggests that participants must have (unconsciously) checked the test number against every entry on the list before they responded. Otherwise, if they responded as soon as they found a match, they would have responded faster on average when there was a match (since they would usually find a match *before* they reached the end of the list) than when there wasn't a match (since they would first have to reach the end of the list).

We can interpret Sternberg as taking up a contrastive singular explanandum: if t is the response time for a given participant when the test number appears on the list, why would t, rather than $t^*>t$, have been the response time for that participant if the test number hadn't appeared on the list? Since this explanandum is contrastive, he can address it with a relatively narrow explanans: the participant checked the test number against every entry on the list before they responded, rather than responding as soon as they found a match. This explanans is so narrow, in fact, that it only represents a functional dependency: that checking the test number against every entry on the list occurred before participants responded.[4]

As Shapiro (2017) notes, representing this functional dependency seems sufficient to address the contrastive explanandum: there is no need to further represent the objects involved when participants check the test number against every entry on the list before they responded. Perhaps, P&C could find some lack that can only be resolved by representing structural dependencies, but I struggle to see it. I think it would be better to concede that the MSA fails for contrastive singular explanations—i.e., the explanation for why individual organism behaviours have certain properties rather than others.

Again, though, we can screen off this advantage just by restricting the scope of the MSA—this time to non-contrastive singular explanations of individual organism behaviours. After all, recall that non-contrastive explanations are "totalistic" in the sense that they must represent all difference-makers to every facet of that phenomenon, not just one difference-maker. In the Sternberg example, this would include every difference-maker to the event wherein a participant responds to the task. Then there would be no basis for separating different kinds of dependencies

[4] I thank Andrew Rubner for drawing out this interpretation of Shapiro's (2017) argument.

for the individual organism behaviour, such as between structural and functional dependencies. All difference makers, both object and occurrence, are relevant to non-contrastive singular explanations of organism behaviour.[5]

This scope restriction is consistent with a focus on non-contrastive singular explanation in work by Craver and colleagues. Craver & Kaplan (2020, pp. 296–7) acknowledge this focus: "Mechanists often characterize phenomena such as 'working memory' or 'the action potential' as multifaceted. These construct-terms are shorthand for a host of features, each of which must be explained to explain the multifaceted phenomenon in its entirety." This quote suggests that P&C might have been focusing on non-contrastive singular explanations when they articulated their MSA, even if they didn't explicitly restrict the scope of their argument to non-contrastive singular explanations. I propose that we make this scope restriction explicit.

Overall, then, I propose that the following is the most serious, plausible formulation of the MSA:

1. **Main premise:** singular (non-contrastive) functional analysis and singular (non-contrastive) mechanistic explanation represent the dependencies of individual organism behaviour on occurrences involving objects that are parts of the individual organism.
2. **Conditional bridge premise:** singular (non-contrastive) functional analysis and singular (non-contrastive) mechanistic explanation belong to the same kind of explanation if they represent the same kind of ontological dependencies.
3. **Conclusion:** therefore, singular (non-contrastive) functional analysis and singular (non-contrastive) mechanistic explanation belong to the same target-based kind of explanation.

We might worry whether these two scope restrictions water down the MSA too far. However, there's plenty of room for reasonable disagreement about this weaker conclusion. I'll demonstrate this when I raise my objection to it in §5.

§3. Task Structure

The lesson from the MSA is that singular functional analysis will end up non-distinct from singular mechanistic explanation unless it represents the dependencies of organism behaviour on something *outside* any object or occurrence in the mechanism for organism behaviour. A good candidate for this is the *task*, which the organism performs by generating behaviour. Of course, though, appealing to task will only help us respond to the MSA if we can characterise tasks such that they are *independent* from the mechanisms that perform them. To this end, I'll provide a mechanism-independent conception of tasks in this section. This conception of task will prove controversial, but I'm ultimately interested in defending a conditional claim: disagreement about the MSA could turn on a deeper disagreement about how to conceptualise tasks.

To develop a mechanism-independent account of tasks, it will help to consider a real task from a psychological experiment in some depth. I elect to consider the recall task designed by Sternberg (1969) and reviewed by Shapiro (2017) in his response to the MSA. Sternberg gives an *abstract* description of this task situation: he gave participants a list of numbers, subjected them to a delay period, presented them with a number (which may or may not have featured on the original list),

[5] As Craver & Kaplan (2020) note, this standard of relevance is too encompassing for scientists with limited resources, but we could plausibly insist that it is appropriately encompassing for individuating explanatory kinds.

and then asked them to indicate whether the number featured on the original list by pulling one of two levers (one for indicating "yes" and another for indicating "no"). Sternberg evaluated participant responses as "correct" if they followed his instructions and "incorrect" if they didn't follow his instructions.

Both Sternberg and Shapiro end their consideration of task structure here, but there's much more to say about it (at least, for a mechanism-independent account of task). When Sternberg reports that he provided participants with a list of numbers in the first phase and then a number in the second stage, what he provided them were papers with ink markings. He ensured that these ink markings *realised* symbols, which *denoted* numbers. He also ensured that in each trial, the numbers denoted by the symbols realised by ink markings in the initial prompt either *contained* or *didn't contain* the number denoted by the symbol realised by ink markings in the test prompt.

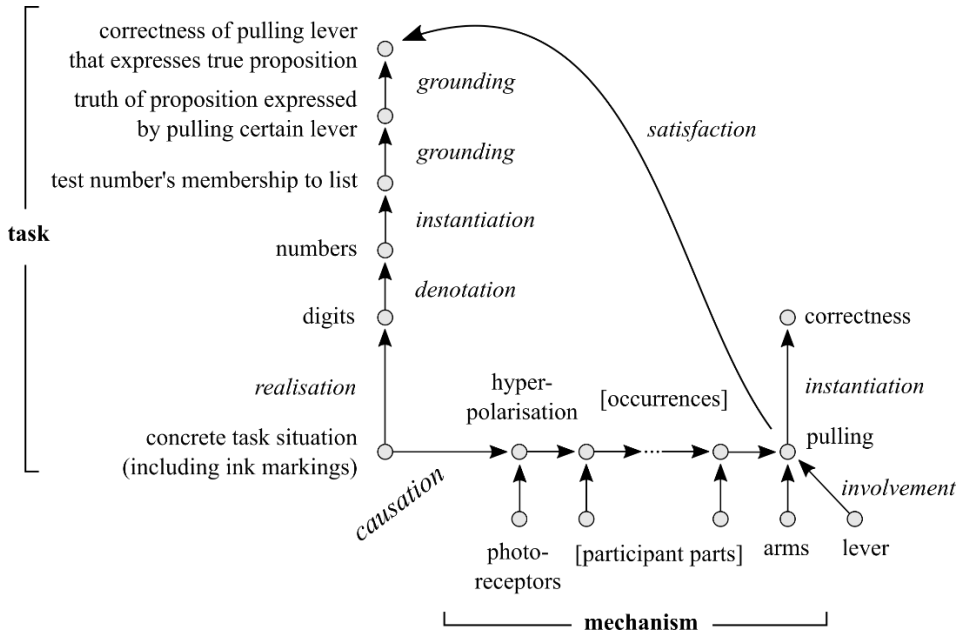

Figure 1. A schematic metaphysical model depicting the two basic determinants of Sternberg's (1969) concrete task situation on two axes: (a) the task on the vertical axis and (b) the mechanism for participant behaviour on the horizontal axis. Nodes depict relata named in solid font; arrows depict relations named in italic font. In §4, I'll argue that the explanandum for functional analysis is the (non-accidental) satisfaction of the correctness conditions for the task by the participant response, which non-causally determines the correctness of the participant response.

Sternberg doesn't say whether he gave written or verbal instructions to participants but suppose he gave them verbal instructions. What he would have provided them, then, were sound waves. He'd ensured that these sound waves *realised* phonemes, which *composed* words. In turn, these words would have *composed* sentences, which would have *expressed* instructions. I propose (not uncontroversially) that these instructions (whichever way Sternberg provided them) *grounded* the conditions under which possible responses to the task would instantiate the properties of correctness and incorrectness. Perhaps, we might say, they *generated* a convention (within the

task situation) whereby pulling one lever is correct iff the list of numbers contains the test number and pulling the other lever is correct otherwise.[6]

Or perhaps, it would be better to say that Sternberg's instructions not only *grounded* a convention (which decided which lever expresses which proposition) but also *referred* to objective norms of accuracy (which decided that correct responses are lever-pullings that express true propositions). How exactly to characterise these relations (realisation, denotation, containment, composition, expression, grounding) and their relata (symbols, phonemes, words, numbers, instructions, conventions, true propositions, accuracy norms) is an open question for several major areas in philosophy: metaphysics, philosophy of language, philosophy of mathematics, normative theory, and metanormative theory.

It would be imprudent for us to take any particular view of how these relations or their relata should be characterised. My response to the MSA only requires three commitments here. First, Sternberg created a concrete situation that gave rise to various abstracta: symbols, phonemes, words, numbers, instructions, conventions, true propositions, and possibly even objective norms of accuracy. Second, this concrete situation "gave rise" to these abstracta via non-causal, non-compositional relations: realisation, denotation, containment, composition, expression, grounding, and reference. Third, neither the concrete situation nor the abstracta it gave rise to is the sort of thing that can compose mechanisms for participant behaviour, which are constituted by causal and compositional relations between concrete objects and occurrences.

I use the term 'task' to refer to this partly-concrete, partly-abstract situation, which participants respond to in the course of psychological experiments. However, this conception of task may be a tough pill for some philosophers of science to swallow. After all, philosophers of science tend to reject that abstracta exist in any robust way. Deflationists might say that they are just sets of concreta subjected to abstract descriptions—as when nominalists say that numbers are just cardinalities. Or reductionists might say that they are, in fact, kinds of concreta for which we lack concrete descriptions—as when teleosemanticists say that reference is just a naturally selected effect to encode information (e.g., Millikan, 1984; Neander, 2017).

This scepticism about abstracta might lead many philosophers of science to give a deflationary or reductionist account of the task. Such an account would try to cash out a task in terms of a more "naturalistic" structure. An obvious candidate for this would be the mechanism that performs the task. Thus, scepticism about abstracta might lead us to look for a mechanism-dependent conception of task. For example, we might try to cash out the correctness conditions of possible responses to the task situation in terms of what a mechanism would actually do (a "causal role" conception of task), or what it is naturally selected to do (a "teleological" conception of task). In fact, I suspect that many cognitive scientists and philosophers of cognitive science would endorse a mechanism-dependent account of task something like either of these on reflection.

I personally find this move unpersuasive. I don't deny that we can partition the set of possible responses to a task situation into various categories—e.g., whether or not a mechanism would actually generate that response, whether or not a mechanism is naturally selected to generate that response, etc. I just deny that these categorisations are equivalent to a normative partition of the

[6] None of these relations are explicitly represented in Figure 1, which is a major reason why it is schematic.

set of possible responses to a task situation—i.e., whether or not the response is *correct*. The reason, of course, is my ontological commitments: I think a response is made correct by its non-causal, non-compositional dependencies on the concrete task situation, independent of the mechanism. Of course, though, I'm not asking the reader to endorse my view of tasks.

Rather, I'm just arguing that the position we take on the existence of abstracta will significantly constrain the position we take on the MSA. If we endorse a mechanism-*dependent* conception of task, then the task isn't a novel element that could be used to distinguish functional analysis from mechanistic explanation. That doesn't automatically mean there isn't another novel element on which functional analysis could depend, which could split it from mechanistic explanation. But it does rule out a serious candidate for what that novel element could be. Thus, a mechanism-dependent conception of task will strongly push us to accept that the MSA is sound.

By comparison, if we endorse a mechanism-*independent* conception of task, which is committed to a slew of abstracta, then the task is a deeply novel element that could easily split functional analysis from mechanistic explanation. All we'd need to show, then, is that functional analysis is, in fact, dependent on the task, conceived in this mechanism-independent way. That will be my goal in §4. Thus, a mechanism-independent conception of task may enable us to reject the MSA. In this way, I hope to show that disagreement about the MSA might come to turn on a deeper disagreement about the existence of abstracta—and, in particular, whether tasks are partly abstract things constituted independently of mechanisms or whether they are wholly concrete things constituted by mechanisms.

It's important to register that there is room for *reasonable* disagreement here. In particular, there is room for reasonable disagreement about the MSA (partly or fully) *because* there is room for reasonable disagreement about whether tasks should be conceptualised as partly-abstract, mechanism-independent structures or as fully-concrete, mechanism-dependent structures. However, the confidence with which many philosophers of science dismiss the ontological status of abstracta is out-of-step with the actual dialectical situation in the metaphysics of abstracta. Moreover, the confidence with which many philosophers of science assume that compelling deflationary or reductionist accounts of abstracta can be given is out-of-step with the extremely limited achievements that they've made on this front.[7]

§4. Task Involvement

I've argued that task can be conceived in a mechanism-independent way as a rich structure of abstracta that non-causally depend on the concrete task situation. Of course, it's unclear whether this is the correct way to conceive of tasks. If it is, though, would this make any difference to functional analysis? In this section, I'll argue that it would: different aspects of the task would feature on either side of a functional analysis, both explanandum and explanans. In other words, mechanism-independent tasks would be essentially involved in functional analysis. I'll consider the implications of this view for the MSA in §5.

[7] For instance, teleosemantic accounts have struggled for decades to reduce easy cases, like the reference of concrete particulars, to informational functions. Hence, even the most ardent reductionist should recognise it's reasonable to doubt that reductive accounts will succeed for harder cases, like reference to propositions or fictional entities.

Experimental psychologists generally purport to explain *correlations*. For Sternberg, the explanandum was the correlation between (a) the responses that participants selected and (b) the responses that Sternberg's task instructions implied were correct. A correlation is an extensional relation. If the correlations were weak, Sternberg might have concluded it was a purely extensional relation: that it was mere coincidence that his participants selected responses that happened to be correct. But the correlation was so strong that Sternberg inferred that it must have been *backed* by a non-accidental, intensional relation between participants and his task: "The low error-rates justify the assumption that on a typical trial the series of symbols in memory was the same as the series of symbols presented" (1966, p. 652).

However, functional analysis is a compositional form of explanation: it explains its explanandum by decomposing it into its parts (Cummins, 1983). Correlations are extensional relations, so they don't have parts and so they aren't amenable to compositional explanation. However, the intensional relations that back strong extensional correlations are the sorts of thing that could have parts. I propose that we describe each intensional relation of this sort in neutral terms as a relationship of *satisfaction*: participant responses are correct at rates significantly above chance because they *satisfy* the correctness conditions of the task. Given this, functional analysis takes up a further question: how do participants *satisfy* the correctness conditions of the task? In other words: what *composes* the satisfaction of correctness conditions by participant responses?

One answer would be to say that participant responses satisfied the correctness conditions *because* the participant responses *directly* depended on the correctness conditions. However, this is impossible, given the causal closure of the physical: concreta cannot directly depend on abstracta. Therefore, participant responses must *indirectly* depend on the correctness conditions. That is, I propose that the correctness conditions non-causally depend on the concrete task situation, participant responses causally depend on the concrete task situation, and these dependencies are "aligned" in some way such that participant responses end up satisfying the correctness conditions. This explanation is consistent with the causal closure of the physical.

For example, recall from §3 that the task involves the following chain of non-causal dependence: (a) ink markings *realise* digits, which (b) *denote* numbers, which (c) *instantiate* a membership (or non-membership) relation of the test number to the list of number, which (d) *grounds* the truth of the proposition *expressed* by one lever, which (e) *grounds* the correctness of any response that pulls that lever. Sternberg (1969) attributes a very similar structure to his participants: they (a) encode stimuli, (b) represent stimuli, (c) identify whether the test number featured on the list of numbers, (d) reach a verdict about which proposition to endorse, and (e) generate a behavioural response to indicate their verdict.

One respect in which there is "alignment" between the task and the mechanism is that there is a *homomorphism* from the task to the mechanism. A homomorphism is an operation-preserving mapping: f is a homomorphism from A to B that preserves operations '$\bullet$' and '$*$' iff, for any x and y in A, $f(x \bullet y) = f(x) * f(y)$. In other words, we get the same result if we (a) apply the operation '$\bullet$' to x and y in A and then map it to an element in B or (b) map x and y to elements in B and then apply the operation '$*$' to those elements in B. This does hold between task and mechanism, but just because the task and mechanism both involve chains of 5 units (at least under certain coarse-grainings of the mechanism). After all, a homomorphism will map the ith unit of the task chain

to the *i*th unit of the mechanism chain for all 5 units. Homomorphism is a weak condition in this respect (Rumana, 2025).

However, this task-to-mechanism mapping isn't just a homomorphism. After all, there is a deeper relationship between the units in the task and the units in the mechanism. Digit realisation in the task maps to the encoding of stimuli in the mechanism, which is just the *registration* of digit realisation. Number denotation in the task maps to the representing of stimuli in the mechanism, which is just the *registration* of number denotation. Membership of the test number to the list of numbers maps to the identification of whether the test number features in the list of numbers, which is just the *registration* of membership, and so on. Thus, the overall task-to-mechanism mapping is not only operation-preserving but also registration-preserving: it maps task units to the mechanism units that register them. I'll refer to this strong subtype of homomorphism as a *hypermorphism*.

What are these registration relations between task units and mechanism units? In actual instances of functional analysis in cognitive psychology, they are generally left as black-boxes: they go by descriptions like "encoding", "representing", "judging", etc. but they are never cashed out (more on this in §5). Thus, we could say that a functional analysis explains the satisfaction of correctness conditions by participant responses just by representing the task-to-mechanism *hypermorphism* from the task (i.e., the dependencies of the correctness conditions on the concrete task situation) to the mechanism (i.e., the dependencies of the participant responses on the concrete task situation). This hypermorphism preserves both operations (within-task and within-mechanism) and registrations (from mechanism to task), without analysing them.

I think this language is acceptable, just as long as we forestall a potential confusion. So far, we've considered a preliminary functional analysis, which Sternberg proposes before his experiment and which, I've suggested, is derivable from the task analysis. Ultimately, though, Sternberg proposes an empirical fine-graining of his task-derived functional analysis. In particular, recall from §2 that he argues that the reaction times of participant responses indicate that they use *exhaustive serial search*: they check for identity between the test number and every entry on the list before deciding whether the test number appeared on the list of numbers—even if they register identity before they reach the end of the list. This result isn't derivable from the task analysis, since participants could have reached a similar outcome using self-terminating serial search instead.

The lesson here is that operations in the task structure correspond to operations in a *coarse-graining* of the mechanism, but not necessarily to individual operations in a fine-graining of the mechanism. This is consistent with the task-to-mechanism mapping being a homomorphism (and a hypermorphism), but it's inconsistent with there being a homomorphic mapping from mechanism back to task. In other words, the task-to-mechanism mapping isn't an isomorphism (i.e., a two-way homomorphism). When we say that a functional analysis represents a task-to-mechanism hypermorphism, then, it's important to remember that it will generally involve a much finer-grained description of the mechanism than is possible for the task. In this way, empirical task analysis is task-*based* without being task-*derived*.

Recall how this fits in our overall argument. I didn't set out to defend the unconditional claim that tasks are constituted in a mechanism-independent way. Rather, I set out to defend a

conditional claim: if tasks are constituted in a mechanism-independent way, then functional analysis would be dependent on tasks. In this section, I've defended a version of this claim: functional analysis explains the satisfaction of correctness conditions by participants by representing "alignment" between the co-dependence of correctness conditions and participant responses on the concrete task situation. This "alignment" involves a hypermorphism from the (task) structure that non-causally determines the correctness conditions to the (mechanistic) structure that causally determines organism responses.

§5. Revisiting the MSA

We are finally in a good position to evaluate our reconstruction of the MSA. Recall from §2 that the following is the main premise of the MSA: singular (non-contrastive) functional analysis and singular (non-contrastive) mechanistic explanation (more or less explicitly) represent the dependencies of individual organism behaviour on occurrences involving objects that are parts of the individual organism. I've argued that this is false if we endorse a mechanism-independent conception of task. Given such an account of task, singular (non-contrastive) functional analysis represents a (non-isomorphic) task-to-mechanism hypermorphism.

This isn't an unconditional objection to the MSA, of course. For that, we'd need an unconditional defence of the mechanism-independent conception of task, which I have not provided here. However, it shifts the burden of proof onto the MSA. After all, two responses are available to friends of the MSA here. First, they could grant that functional analysis is task-dependent, but they could deny that this gives functional analysis independence from mechanistic explanation, because tasks themselves are mechanism-dependent. It's unclear how this response would go, so it is incumbent on friends of the MSA who endorse a mechanism-based conception of task to flesh out an account of this.

Second, friends of the MSA could reject my argument in §4 that functional analysis is task-dependent, even if tasks are constituted independently of mechanisms. One way to do this is to deny that a task-to-mechanism hypermorphism is an ontological dependency. If we insist that functional analysis is an explanation (or an explanation of a particular sort), it has to represent ontological dependencies. At pain of contradiction, then, it couldn't represent a task-to-mechanism hypermorphism. The thought, then, is that functional analysis had better not be task-dependent in the way that I've suggested or else it isn't an explanation at all—rather than just a type of explanation distinct from mechanistic explanation.

In response, I'd say that a task-to-mechanism hypermorphism is an ontological dependency, just not a *productive* one. An isomorphic task-to-mechanism hypermorphism (as in task-derived functional analysis) is a *necessary* condition for the satisfaction of correctness conditions by participant responses. A non-isomorphic task-to-mechanism hypermorphism (as in task-based empirical functional analysis) is, arguably, an insufficient but necessary part of an unnecessary but sufficient condition (an INUS condition) for the satisfaction of correctness conditions by participant responses (c.f., Rumana, 2025). Modal conditions (or constraints) like these aren't *productive* like grounding or causal conditions, but they are ontological dependencies nonetheless (Ross, 2021). Thus, task-dependence is no threat to the explanatory status of functional analysis (even on an ontic account of explanation).

Another response to my argument in §4 is to take issue with the notion of a task-to-mechanism hypermorphism itself. It's relatively clear what an operation-preserving mapping is, but they could complain that it's unclear what a registration relation is and hence, it's unclear what a registration-preserving mapping like a hypermorphism is. My "justification" for papering over an account of registration in §4 was that functional analysis in cognitive psychology generally treats registration relations as black boxes. Of course, though, this is unsatisfactory given a target-based approach to explanation: to individuate kinds of explanations, we need to know what in the world they represent. Thus, a target-based approach requires us to cash out registration relations.

Unfortunately, I don't have the space to develop a full account of registration relations. However, Rumana (2025) offers an account that might work for our purposes. She argues that mechanism-task fit (what we call "registration relations") consists in *constraint satisfaction*. Given a view like this, we could say that, e.g., the digit realisation relation in the task imposes constraints on any whole mechanism that would perform the task. These constraints are then satisfied by the mechanistic components that we describe as doing "digit registration" (or "stimulus representation"). In other words, for a mechanistic component to register a task relation just is for it to "take up" the constraints imposed on the whole mechanism by that task relation and satisfy them (on behalf of the whole mechanism).

I should clarify, though, that while registering components "take up" constraints on the whole mechanism and satisfy them, nothing I've said so far implies that they are *responsible* (in a normative sense) for doing so. If a mechanistic component had been registering a task relation in all previous trials but suddenly ceased to do so for the next trial, the mechanism might fail to satisfy the correctness conditions of the task, but it's unclear whether the mechanistic component is criticisable for this outcome. After all, a system-wide failure vis-à-vis the task can't automatically be charged against any individual mechanistic component, even if it had a history of preventing that failure. Thus, I propose, registration relations aren't apt for error: they are just making-possible relations (Rumana, 2025). There are no "mis-registrations".

To charge individual mechanistic components for system-wide failures, we'd need a division of responsibility, such that each mechanistic component is *responsible* (in a normative sense) for registering a particular task relation. I take this to be a potential job description for an account of representation, since representations are apt for error (there are misrepresentations). Even if there are normative relations like representation between the mechanism and the task, though, they aren't relevant to a functional analysis of correct responses.[8] After all, the functional analysis of correct responses is only interested in the *actual* way that an individual participant behaviour satisfied the correctness conditions, regardless of whether each contribution to this outcome was consistent or inconsistent with some division of responsibility.

Again, there is much left to say about registration relations. However, I hope this brief discussion is sufficient to make clear that registration relations are neither reducible to mechanism nor hopelessly mysterious (and probably much less mysterious than normative relations like representations). Therefore, whichever way we choose to cash out registration relations won't

[8] I suspect the functional analysis for how individual participant behaviours *fail* to satisfy the correctness conditions of the task will require a division of responsibility for system-wide failures to mechanistic components and so may require an appeal to representation relations.

make a difference to the fact that a mechanism-independent conception of task implies that the MSA reaches a false conclusion—that functional analysis and mechanistic explanation belong to different target-based kinds of explanation.

§6. Conclusion

Disagreement around the MSA has historically boiled down to disagreement about object involvement in explaining organism behaviour. Philosophers of psychology who reject the MSA maintain that there are advantages for representing the functional dependencies of organism behaviour separate from their structural dependencies. I've argued that there are no such advantages for singular (non-contrastive) explanations of individual organism behaviour. Thus, I've argued, singular explanations of individual organism behaviour should represent both its functional and structural dependencies. If functional analysis represented just the functional dependencies of organism behaviour, then P&C would be right to say that it would be a "sketch" of mechanistic explanation, which represents both structural and functional dependencies.

In this paper, though, I've suggested that task involvement in functional analysis is a more effective way of prying it away from mechanistic explanation. In particular, I've argued that if tasks are constituted independently of mechanisms, then the explanandum of functional analysis—i.e., how participants satisfy the correctness conditions for the task—can only be addressed by representing abstract relations between the co-dependencies of both participant behaviours and correctness conditions on their common determinant—i.e., the concrete task situation. These abstract relations are (non-productive) ontological dependencies but obviously different in kind from the (productive) structural and functional dependencies of participant behaviour on the mechanism for participant behaviour.

Many may find the mechanism-independent conception of task that I've developed counterintuitive at best and anti-naturalistic at worst. Again, my aim here isn't to say whether philosophy of science should be beholden to naturalistic intuitions (although I think it shouldn't). My aim has been to defend a pair of conditional claims. If your intuitions lead you to reject a mechanism-independent conception of task, then they lead you to endorse our reconstructed MSA. But if you wish to uphold the distinctiveness of functional analysis from mechanistic explanation, *contra* the MSA, then endorsing a mechanism-independent conception of task is the best available way to do this. Surprisingly, then, disagreement about the MSA may boil down to disagreement about whether tasks are constituted independently of mechanisms.

Acknowledgements: I thank David Barack, Jeremy Pober, Andrew Rubner, Alfredo Vernazzani, and other members of the Phil Neuro/Mind Writing Group for feedback on a draft of this paper.

References

Anderson, M. L. (2015). *After phrenology: Neural reuse and the interactive brain*. MIT Press.

Barrett, D. (2014). Functional analysis and mechanistic explanation. *Synthese*, *191*(12), 2695–2714. https://doi.org/10.1007/s11229-014-0410-9

Batterman, R. W., & Rice, C. C. (2014). Minimal model explanations. *Philosophy of Science*, *81*(3), 349–376. https://doi.org/10.1086/676677

Bechtel, W. (2008). *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Taylor & Francis Group/Lawrence Erlbaum Associates.

Bokulich, A. (2016). Fiction as a vehicle for truth: Moving beyond the ontic conception. *The Monist 99*(3): 260–79.

Bokulich, A. (2018). Representing and explaining: The eikonic conception of scientific explanation. *Philosophy of Science, 85*(5): 793–805.

Burnston, D. C. (2016). A contextualist approach to functional localization in the brain. *Biology & Philosophy, 31*(4): 527–50.

Chemero, T. & Silberstein, M. (2008). After the philosophy of mind: Replacing scholasticism with science. *Philosophy of Science, 75*(1): 1–27.

Craver, C. F. (2007). *Explaining the brain: Mechanisms and the mosaic unity of neuroscience*. Oxford University Press.

Craver, C. F. (2014). The ontic account of scientific explanation. In M. I. Kaiser, O. R. Scholz, D. Plenge, & A. Hüttemann (Eds.), *Explanation in the Special Sciences: The Case of Biology and History* (pp. 27–52). Springer Verlag.

Craver, C. F. (2019). Idealization and the ontic conception: A reply to Bokulich. *The Monist*, *102*(4), 525–530.

Craver, C. F., & Kaplan, D. M. (2020). Are more details better? On the norms of completeness for mechanistic explanations. *The British Journal for the Philosophy of Science*, *71*(1), 287–319. https://doi.org/10.1093/bjps/axy015

Cummins, R. C. (1983). *The nature of psychological explanation*. MIT Press.

D'Alessandro, W. (2020). Viewing-as explanations and ontic dependence. *Philosophical Studies, 177*(3), 769–92.

Dretske, F. (1988). *Explaining behavior: Reasons in a world of causes*. MIT Press.

Fodor, J. (1974). Special sciences (or: The disunity of science as a working hypothesis). *Synthese, 28*(2): 97–115.

Frigg, R. (2007). Models and fiction. *Synthese, 172*(2): 251–68.

Frigg, R. & Nguyen, J. (2020). *Modelling nature: An opinionated introduction to scientific representation*. OUP.

Godfrey-Smith, P. (2006). Theories and models in metaphysics. *The Harvard Review of Philosophy, 14*(1): 4–19.

Illari, P. (2013). Mechanistic explanation: Integrating the ontic and epistemic. *Erkenntnis*, *78*(2), 237–255. https://doi.org/10.1007/s10670-013-9511-y

Illari, P. & Williamson, J. (2011). Mechanisms are real and local. In P. Illari, F. Russo, & J. Williamson (Eds.), *Causality in the Sciences* (pp. 818–844). OUP. https://doi.org/10.1093/acprof:oso/9780199574131.003.0038

Kaiser, M. I., & Krickel, B. (2017). The metaphysics of constitutive mechanistic phenomena. *The British Journal for the Philosophy of Science*, *68*(3), 745–779. https://doi.org/10.1093/bjps/axv058

Kaplan, D. M., & Craver, C. F. (2011). The explanatory force of dynamical and mathematical models in neuroscience: A mechanistic perspective. *Philosophy of Science, 78*(4), 601–627. https://doi.org/10.1086/661755

Kennedy, A. G. (2012). A non representationalist view of model explanation. *Studies in History and Philosophy of Science, 43*(2), 326–32.

Kim, J. (1992). Multiple realization and the metaphysics of reduction. *Philosophy & Phenomenological Research* 52: 1–26.

Lycan, W. (2005). Explanation and epistemology. In P. Moser (Ed.), *Oxford Handbook of Epistemology* (408–433). OUP. https://doi.org/10.1093/oxfordhb/9780195301700.003.0015

Machamer, P., Darden, L., & Craver, C. F. (2000). Thinking about mechanisms. *Philosophy of Science*, *67*(1), 1–25. https://doi.org/10.1086/392759

Millikan, R. G. (1984). *Language, thought, and other biological categories: New foundations for realism*. MIT Press.

Neander, K. (2017). *Mark of the mental: A defence of informational teleosemantics.* MIT Press.

Piccinini, G., & Craver, C. (2011). Integrating psychology and neuroscience: Functional analyses as mechanism sketches. *Synthese*, *183*(3), 283–311. https://doi.org/10.1007/s11229-011-9898-4

Piccinini, G. (2021). *Neurocognitive mechanisms: Explaining biological cognition*. OUP.

Pylyshyn, Z. W. (1984). Computation and cognition: Toward a foundation for cognitive science. MIT Press.

Rosenberg, A. (2018). Making mechanism interesting. *Synthese*, *195*(1), 11–33. https://doi.org/10.1007/s11229-015-0713-5

Ross, L. N. (2021). Causal concepts in biology: How pathways differ from mechanisms and why it matters. *British Journal for the Philosophy of Science 72*(1): 131–58.

Rumana, A. (2022). Arbitrating norms for reasoning tasks. *Synthese*, *200*(6), 502. https://doi.org/10.1007/s11229-022-03981-8

Rumana, A. (2025). Explaining mechanism–task fit in neuroscience. *The British Journal for the Philosophy of Science*. https://doi.org/10.1086/736123

Salmon, W. C. (1989). 4 decades of scientific explanation. *Minnesota Studies in the Philosophy of Science, 13*: 3–219.

Shapiro, L. A. (2017). Mechanism or bust? Explanation in psychology. *The British Journal for the Philosophy of Science*, *68*(4), 1037–1059. https://doi.org/10.1093/bjps/axv062

Sternberg, S. (1966). High-speed scanning in human memory. *Science, 153*(3736): 652–4. https://doi.org/10.1126/science.153.3736.652

Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist, 57*, pp. 421–57.

Weiskopf, D. A. (2011). Models and mechanisms in psychological explanation. *Synthese*, *183*(3), 313–338. https://doi.org/10.1007/s11229-011-9958-9

Wright, C. & van Eck, D. (2018). Ontic explanation is either ontic or explanatory, but not both. *Ergo*, *5*. https://doi.org/10.3998/ergo.12405314.0005.038