

When data seems to conflict with a theory there are two things we can do. The first is to revise the theory so that it is consistent with the data. There are complex issues around how much alteration a theory can take and still count as numerically the same theory¹. The second is to defend a theory from data by problematizing auxiliary assumptions since data doesn't support or disconfirm theory directly, but only by way of auxiliary assumption. In what follows I'll suggest a little of both. I'll suggest that both Cowan and Baddeley update the different kinds of functional / task distinctness that they posit in order to deal with behavioral dissociations and interference effects. Whether this will alter their identity as theories relies on the plausibility of my having interpreted the most significant difference between Cowan and Baddeley as that of whether the representational contents that are involved in the different kinds of activity are the same (Cowan) or different (Baddeley). I'll maintain that it is possible to defend Cowan's claim that working memory is mediated by long-term memory representations by maintaining that the neurological dissociations that have been found indicate task / activity independence rather than a difference in representational content. This will basically involve analyzing the localization assumption into two variants: Localization of representational contents (sensory cortex activity and motor cortex activity) and localization of function / task (more frontal activity). While it is possible to defend Cowan's claim about sameness of representational content in this way I do think that a more plausible view will consist in overlapping contents (same in some respects, different in some other respects) *all the way up*².

In cognitive psychology behavioral double dissociation is typically taken to be just the kind of finding that supports the claim that there are two distinct kinds of activity / task / function. Cowan's model posits different kinds of activity. There is a problem, however, in that there are different ways in which we can interpret the different kinds of activity that are posited by Cowan's model. One is that there are three kinds of *unstructured* activity as follows: declarative long-term remembering, actively remembering, attending remembering. If this is correct then finding double dissociations between each of these activities would seem to support Cowan's model rather than undermining it (though Cowan would need to posit further kinds of activity in order to handle other dissociation findings - as would Baddeley).

Behavioral double dissociations have been thought to be problematic for Cowan's model, however. The thought seems to be as follows: One way of understanding the claim that 'working memory is mediated by declarative long-term memory' is to understand it as a claim about the *structure* of tasks. So the thought is that rather than interpreting Cowan's model as making claims about *independent* tasks (where double dissociations are precisely what should motivate the claim of task independence), Cowan is making a claim about *dependency relations* between tasks as follows: declarative long-term remembering, activation of declarative long-term remembering, attending activation of declarative long-term remembering³. On this way of specifying the kinds of activity it would seem that finding a person who was unable to perform declarative long-term memory tasks who was able to perform working memory tasks would be inconsistent with the model⁴.

Another way of understanding the claim that 'working memory is mediated by declarative long-term memory' is to understand it as a claim about *representational contents* rather than *structure of tasks*, however. I think that the main point of difference between Cowan and Baddeley is that while both posit that there are different kinds of activity Cowan maintains that the representational contents that are involved in the different kinds of activity are the same whereas Baddeley maintains that the representational contents that are involved in the different kinds of activity are different. On this latter

¹The issue of whether we have the same theory or not isn't as interesting as developing a good one. People do seem to get hung up over who was right and who was wrong, however.

²My view is motivated mainly by philosophical considerations, admittedly (and my view would probably be denied by most philosophers). Still, executive processing has enough problems to deal with without being gifted the central controller - most unfair! Enriched structured contents (explained a little later) allow us to avoid the central controller and also, I maintain, have better prospects for the development of a science of consciousness, though I don't have the space to get into that here. Can HM learn to run a maze (by the way)? If so Lashley's failure to find localization doesn't have obvious consequences for declarative long-term representations

³Structured activities may not be able to account for combinatorial semantics the way that structured contents can.

⁴Though this would only be the case if we analyze 'x is mediated by y' into 'y is necessary for x' and thus the finding that 'not y and x' would disconfirm the hypothesis. 'Mediated by' seems to be precisely the kind of fudge word that renders a claim immune from this kind of disconfirmation, however. As such, even if we did read Cowan as committing to structured activities it is unclear whether finding dissociation between activities would disconfirm the model.

way of understanding the most important difference between Cowan and Baddeley’s models both of them seem comparably able to be modified to account for behavioral data (while retaining their identity as distinct theories). They simply need to posit different kinds of (unstructured) activity to handle the double dissociations that are found and task independence or overlap can explain dissociation or interference respectively.

So, if we now turn to neurological data then it seems that we need to decide whether sameness or difference of local activation is going to be interpreted as indicating sameness or difference of task / function / activity or sameness or difference of representational content. The localization assumption that seems to be employed in cognitive neuroscience is that activity in the sensory cortices indicates representational content whereas activity in other regions (particularly frontal regions) indicates function or activity. It seems to be what is employed when theorists claim that such functions as ‘attention’ or ‘consciousness’ are localized in frontal areas whereas the representational contents of those functions are localized in sensory cortical regions. The urge to localize representational content in sensory cortices comes from finding neural tuning curves in those regions. There are significant problems in individuating content, however. Does a heightened amplitude in attended sensory processing indicate sameness of content, overlap in content, or difference in content? How do we decide? As we move into more frontal regions the urge to equate different activation with different content and sameness or similarity of activation with sameness or similarity of content seems to disappear. Now what we seem to want to say is that sameness or difference indicates sameness or difference of task, activity, or process. Cowan’s model could thus be defended by interpreting the differences in neurological activation that have been found to reflect differences in functional activity rather than reflecting different kinds of content⁵.

The desire to take activation in frontal regions to indicate activity seems to be partly due to our (currently) not having much in the way of information about representational contents in those regions, however. Philosophers probably have a lot to answer for when they say (falsely IMHO) that ‘content’ refers to the representation of stimulus features where those features are *outside the head*. We basically need an *enriched* view of content. While it will be easier to find that neurons are tuned to features outside the head rather than being tuned to the contents of lower level neurons (so re-representing) our failure doesn’t imply that this isn’t the case. I think that taking activation in those areas to indicate differences in activity rather than content is problematic⁶. A controversial issue (in the philosophy of language and the philosophy of mind) is whether we are best to carve activity such that the same content can feature in different activities where the activities might even have a complex structure (as Cowan seems to have been interpreted as doing): representing perceptually(p) or representing, consciously, perceptually(p) and so on. Or alternatively, whether activities are something that is better to be understood as *certain kinds of representational content* as follows: representing(perceptually, p) or representing(consciously, perceptually, p) and so on⁷. The latter view seems less suggestive of a central controller, however. While it does seem that at this point we would have given up on both Cowan and Baddeley’s model it would seem fairly intuitive that contents can (sometimes) be the same (in some respects) and that contents can (sometimes) be different (in some respects) rather than taking extreme views as Cowan and Baddeley do (or interpreting the localization assumption differently for different neurological regions). It might seem counter-intuitive in reducing function to content, however (not sure how plausible or crazy that is going to seem⁸). It is also unclear to me how this model could be tested, though to be fair a full defense of Cowan’s model would seem to result in a similar problem.

⁵We could also defend Cowan’s model by maintaining that content can be distributed such that differences in neural activity don’t reflect differences in content though this might be thought to be more radical.

⁶The inevitable question that seems to arise from taking frontal processing to indicate task or activity is ‘what gets to allocate resources to task or activity, or what decides which takes priority?’ Onset of stimuli is useful for time locking trials. The representation of task instructions might be useful for the content we view as ‘central controller’.

⁷Of course the terms in the parentheses will need to be broken down properly into the contents. Individuating / carving contents up correctly is going to be a tricky matter. They will also need to be structured (to allow for inference / transformations etc).

⁸This comes up in the consciousness literature. Most grant that phenomenally experienced properties (e.g., subjective experience of brightness) entail certain kinds of representational contents but some maintain that representational contents won’t entail phenomenally experienced properties unless you build phenomenality into the representational content. Building it in is a way of getting a two-way entailment (which might be thought to be necessary for an identity). I’m wondering how much a similar strategy could generalize back to other forms of what are typically taken to be activity such as working memory or declarative long-term memory or attention.