## Collins 2018 most likely captures declarative knowledge more than reinforcement learning.

Collins (2018) presents a two-component model of learning that fits behavioral data well, but it is more probable that it is capturing only the dynamics of learning with the declarative memory system characterized as working memory with little if any influence of reinforcement learning. A three-pronged reason for this assertion, that considers the time scales of declarative memory formation, a hierarchical view of task components in the RLWM experiment and the limitation of working memory is presented below.

Firstly, Collins (2018) does not acknowledge that declarative memory system controls behavior in the early stages of skill learning, and there are perhaps not enough learning trials for engagement of reinforcement learning to occur. Several studies in humans and animals show that declarative and procedural memory systems acquire new information in parallel but the former controls behavior before the procedural or reinforcement learning system (Poldrack and Packard, 2003; McDonald and Hong, 2013). Normally, following experience with more trials, as the learner gains expertise, control of behavior largely shifts to procedural learning mechanisms, but it is questionable if there are enough stimulus-response pairing trials for this shift to occur in the RLWM task.

Building on the first point above, procedural learning, on the other hand, might apply to different parts of the RLWM task. Let us take a multi-level view of the task based on its structure to illustrate this point.

Consider just two levels of the RLWM task — a lower, with-in block level of learning image-letter associations, and a higher block-by-block learning level. I argue that the actual stimulus-response pairings with-in the blocks are refreshed so often between blocks, that is, every block presents novel images, that new associations always must be acquired declaratively. But many other bits of the task such as reading instructions, viewing stimuli, and making responses, and all other necessary cues that are consistent throughout the task, at the higher level, might provide the learner with enough experience to lead to proceduralization. Therefore, all the modeled data, which is only the stimulus-letter associations could only be acquired as declarative information with reinforcement learning affecting the models only marginally.

Secondly, the non-RL learning component of the model, that Collins (2018) characterizes as working memory, cannot be working memory for two related reasons: all the information that survives from decay is already in declarative memory, and active maintenance of associations in working memory is discouraged by the dynamics of the learning environment. In other words, the RLWM task is probably too disruptive to allow for working memory maintenance of associations during response making because the learner has to process stimuli and keep track of responses and feedback. So, offloading to a more stable declarative storage is very likely. This is likely because usually, declarative long-term memory is engaged along with working memory in similar tasks (Baddeley, 2012) and, rehearsal of information in the early stage of working memory engagement is enough to write information to declarative memory (Ranganath et al., 2005).

Lastly, reinforcement of declarative memory contents improves retention of the associations if the task allows for sufficient practice. The RLWM task might do so because the task presents each novel stimulus 13 times. This reinforcement might explain why some of the learned associations survive the 10-minute distracting break that learners demonstrate by responding correctly to a surprise test. The 10-minute distracting break is very likely unnecessary to clear working memory contents because stimulus-response associations have long been purged from working memory, probably during the transitions between blocks if not during learning in the blocks themselves. So, what remains is most likely declarative long-term memory.

Taken together, the pieces of evidence above suggest that the task is too dynamic for performance by working memory maintenance; the individual blocks are too short to allow for proceduralization of associations but, have enough trials to strengthen declarative representation of the stimulus-response pairings.