We thank the reviewers for their constructive and thoughtful reviews of our work. We are excited about these results and we are pleased that they share our sentiments.

Reviewer #1 made an excellent point regarding the fixation requirement; we considered this problem at length and have two general responses. First, we ran additional versions of our experiments in which fixation was required, and indeed all agents failed to learn this variant (which was expected given that they learn through random exploration of the action space, which is unlikely to yield fixation strategies from which to discover rewards). We also ran variants wherein the agent *trained* normally but was tested with fixation requirements. Still, the agents failed to maintain fixation at test time. From this we conclude that a carefully designed training curriculum is necessary to enable RL-based discovery of this strategy, not unlike the extended behavioral shaping procedures in the animal literature. We think this is an important next step to consider but did not implement it at this stage. More generally, we hypothesize that both agents and animals will seize upon the action spaces available to them; such as limb movements in the rodent study we cited, gaze trajectories in our study, and possibly non-experimentally monitored movements (e.g. finger tapping) in studies requiring fixation. We have included the aforementioned variants in the code we will open-source before the conference.

Reviewer #1 also made an important point regarding scalar variability. In response, we fitted a generalized power law to the data, and the coverage of the posterior over c is indeed around 1, though the best-fit value is 0.7, as the reviewer suspected. At this point we view the finding as an approximation of scalar variability with the possibility of a more complex relation to be explored. Regarding the possibility of this scaling emerging from the reward structure: this is an important point, and one we considered when setting task parameters prior to analysis; in fact, both this and the subsequent point concerning the dimensionality of beta are explained by a typo on our part: alpha and beta were swapped in the task description. In fact, alpha was set to 8 frames, and beta was set to 0, thus eliminating reward scaling (though that option exists in the task architecture in case future users choose to set it to a non-zero value). In summary, the approximate scalar variability we observe is indeed real and not related to task reward structure.

Finally, we are also releasing all the code we are able to in order to support reproducibility. The agent can be reproduced using the open-source IMPALA implementation, and we are happy to provide advice by email to anyone seeking to do so. We also thank the reviewer for pointing us to these important references that we will now discuss in the paper: the Karmarkar and Buonomano (2007) result presents important evidence that timing can be achieved without the often-proposed centralized clock mechanism; our result similarly points to strategies for timing requiring no explicit clock. The Orhan and Ma paper is also highly relevant, since our task can be framed as a particular form of short-term memory, and they address the timely topic of persistent neural activity vs. sequences in short-term memory.

Reviewer #2 raised the important consideration of learning dynamics and how the agent converges on its strategy. We always trained the agent end-to-end (from pixels to actions) and it therefore acquired its strategy through RL-based exploration of the action space and environment. Most of our work focused on placing strict *environmental* (i.e. task) constraints on the agent. We agree that an important extension of the work is to place specific constraints on the agent architecture in order to more thoroughly identify the mechanisms by which the agent develops strategies. We have attempted preliminary experiments in which we perturb the forget gates of the LSTM model causing it to be forgetful, hypothesizing that it may then converge on a different behavioral strategy. We also point to our result in which a frozen LSTM can learn the task; this demonstrated the non-necessity of trained LSTM gates to achieve the task. There is much more exploration that can be done here, and we take that to be one of the main results of our work: agents can find many solutions using the abilities given to them, so careful task design is crucial to understanding agent behavior.

In response to Reviewer #3: we have made some attempts to determine how general these results are. Figure 7 shows a sweep across various agent architectures, and so far the purely feedforward agent was the only one to strongly develop the stigmergic strategy. From this we conclude that recurrence in any form is likely to be used for timing purposes when available to the agent; whereas in its absence, agents learn more behaviorally linked strategies. In the context of biology, we agree that it is difficult to compare animals and agents. Nevertheless, we think that the similarities to some biological data may suggest that animals rely on neural strategies with relatively low memory capacity (feedforward systems being one such example). More importantly, we consider it an important cautionary tale for the study of animals: modern systems neuroscience often approaches these questions with the prior that neural activity will most directly explain cognitive phenomena, whereas results like these demonstrate that a behavioral phenotype may be a more direct mechanistic explanation that should be analyzed in concert with neural activity.

Regarding the changing of target locations: we indeed attempted these experiments, and agents failed to learn task variants with random target placements. We suspect that development of more detailed curricula will overcome this barrier. Interestingly, in a binary temporal discrimination task which we are also open-sourcing ("report whether the stimulus was short or long"), agents learned to match intervals to randomly placed colored targets, but failed when the rule was a pro-anti rule instead. We think these task complexity barriers to learning are fascinating, and one lesson from this study is the importance of detailed analyses concerning which requirements cause agent learning to break down.