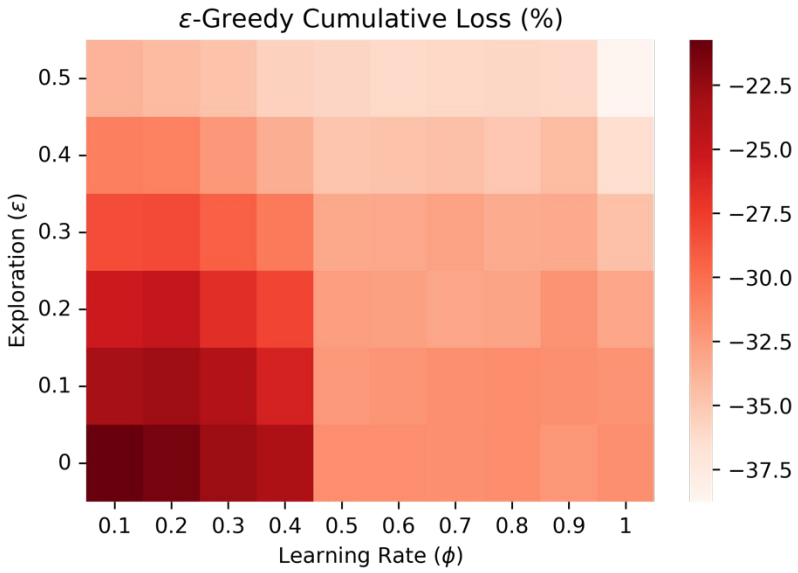
APPENDIX

Robustness to an ε-greedy choice rule:

While we employ the softmax choice rule (Luce, 1959) in our main analysis, it is important to assess the robustness of our baseline results to alternative implementations of exploration. Specifically, we consider an ϵ-greedy choice rule (Sutton and Barto, 1998) where the organization randomly selects an alternative with a probability of ε and is greedy with probability $(1 - \varepsilon)$. A key difference between softmax and ε-greedy is the role of beliefs in driving exploration. Specifically, beliefs are an input into the softmax choice operator such that beliefs inform the choice probability of each alternative and the dispersion on beliefs affects the level of exploration (Posen and Levinthal, 2012). Conversely, when exploration is ϵ-greedy, the dispersion in beliefs does not directly affect the exploration process. That is not to say, however, there is no role for beliefs in driving search as changes in beliefs will still result in changes to what is perceived as the most favored alternative. However, the level of exploration and the probability of selecting any given alternative conditional on exploration is independent of the organization's beliefs.

The results of this analysis are attached below where we replicate our baseline results with 100 alternatives and 100 periods for various values of the learning rate (ϕ) and exploration (ε). Specifically, we note that our primary finding regarding the superiority of slow learning in conjunction with a greedy choice rule is robust to this change in specification, with the best performing policy being a learning rate of ϕ=0.1 and a maximally greedy (ε=0) strategy of exploitation/exploration. However, we do note a change in results regarding the interaction between the learning rate and exploration at higher values of ϕ and/or ε. Specifically, for each value of ε, performance is eroded as ϕ increases. This is driven by the tendency of increases in both ϕ and ε to independently drive search. The result of this is that the lowest performing policy shifts from a slow learning rate and exploratory strategy under softmax (as a function of the dispersion of beliefs interacting with the level exploration) to fast learning coupled with an exploratory strategy.