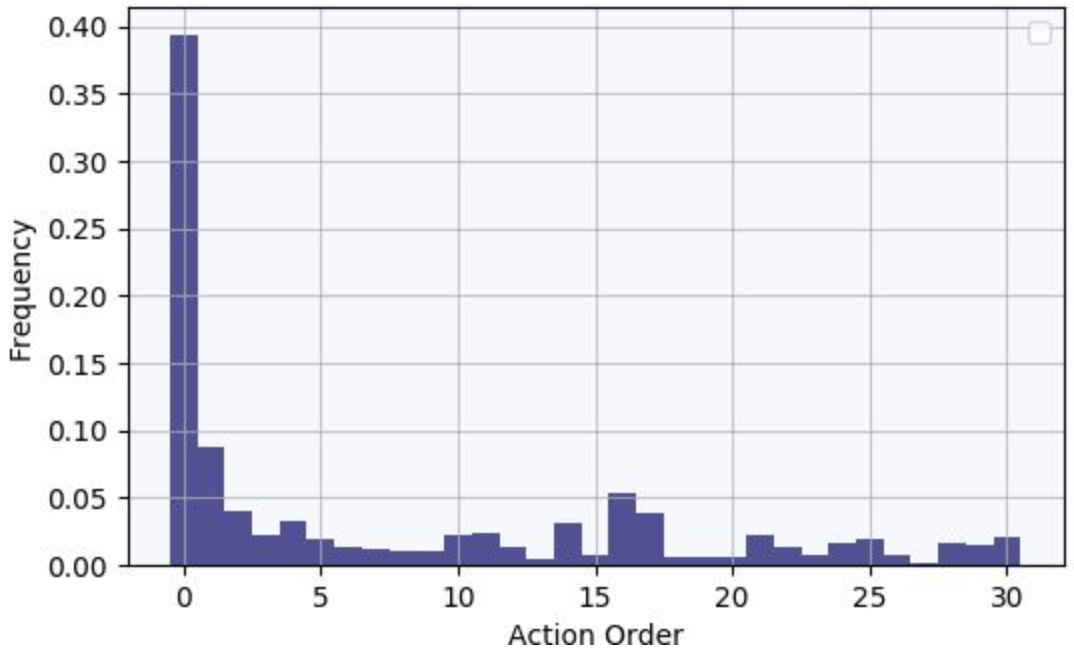
We would like to thank the reviewers for their insightful comments.

Addressing the common point of limiting our experimentation to a single-decision setting, our intent was to focus our analysis only on the effects of candidate generation. By removing the influences of other factors on the performance of search, for instance, rollout policies and state value function approximations, we can focus the evaluation. We are aware that the sequential-decision setting requires extra reasoning. We would argue, though, that the other components of learning algorithms for search try to ameliorate the amount of reasoning needed --- indeed, learning a perfect value function approximation would essentially reduce a sequential-decision problem to a single-decision problem. However, we do plan on examining our ideas in a full MCTS setting, which we think is a problem deserving its own investigation.

With regards to the ordering of actions produced by marginal utility, it is indeed true that the gradients use the ordering that comes from the action index. However, the actions are not pre-sorted before the marginal utility gradient is calculated. Rather, by optimizing the marginal utility objective the model learns an inherent ordering of the actions. Even though our search ignores this ordering at present, we see potential in its use for progressive widening or alpha-beta pruning in sequential search. Plotting out the frequency with which the actions are selected (see below) over test states in the curling domain shows that in expectation, the first action generated is most often selected (although still under 40% of the time), then the second, and so forth. We will add this to the appendix of the final version of the paper.



We had run preliminary experiments in the curling domain that explicitly tried to optimize for diversity in the action set (as a penalty on the set similarity added to the sum objective) before we even explored the marginal utility objective. Simply put, the results were unimpressive, performing no better than without the diversity term in the objective. Furthermore, we found the action set to be very sensitive to the coefficient on the penalty. Too low and actions would just cluster around the single best action; too high and actions would spread to the corners of the action space. We believe that while a lack of diversity makes for an ineffective candidate generator for search, diversity alone is also not the answer (note that the max objective in our experiments does produce diverse but still ineffective candidate sets). Marginal utility seems to produce a set of complementary actions and not just different ones (i.e., covering situations that the previous actions do not account for). We can include the diversity-modified objective as another baseline in the final paper to make this more clear.