We thank all three reviewers for their detailed and thoughtful reviews. We were glad to see a consensus that the paper was “well-written” (R1, R2) and constituted an interesting interdisciplinary research direction (R1, R2, R3). Below, we address each reviewer’s specific questions in turn.

Reply to R1

“Why this baseline? Use baseline from literature?” We created this baseline guided by the intuition that by setting a high positive reward at the end of the path, with a negative reward along the route, we might identify those users willing vs unwilling to endure the obstacle. We are currently exploring ways to adapt pre-existing approaches from the prior literature to our problem. However, we note that because much of the related work does not address MDPs, adapting these methods faithfully to our setting can be non-trivial.

“How about if the slopes differ?” Per your feedback, we ran new experiments where the slopes differ. In this setup, loss-neutral players have both slopes as 1, gain-seeking players slopes 2 (positive) and .5 (negative), and loss-averse slopes are .5 and 2, respectively. In the Path setting, these experiments yielded qualitatively similar results. We will add this analysis to the final version. With these players, the baseline achieves mutual information loss of .922, classification accuracy of .474 (.051 std dev). While the design optimized by our proposed framework has lower mutual information loss, .735, and much higher classification accuracy of .772 (.037 std dev).

“Do the players learn from previous experience?” We do not model the player’s learning but plan to in future work.

Re: Gumbel-max and softmax

Indeed our player policies do use softmax to produce a distribution over actions (Eq 9). Gumbel-max is a way to sample from this distribution with injected noise. Gumbel-softmax provides a continuous approximation to this sampling step, enabling end-to-end learning. We will clarify this in the revised text.

Reply to R2

“A larger evaluation of the baselines?” To understand the variance of mutual information loss and classification accuracy for different choices of baseline, we considered three more baseline rewards where \( r_{pos} \) and \( r_{neg} \) vary. As a brief reminder, the baseline is defined as \( R(s) = r_{neg} \) if \( s = 3 \) and \( r_{pos} \) if \( s = 6 \), otherwise zero. For \((r_{pos}, r_{neg}) = (-3, 3), (-5, 5), \) or \((-5, 3)\), the mutual information loss is .133, .111, .133, respectively. The classification accuracy (and std dev) is .472(.054), .460(.058), .492(.047), respectively. We also experimented with five more random baselines, where each state in one such baseline is a random number between \(-5\) and \(5\). The averaged mutual information loss is \(.150 \pm 0.02\) and classification accuracy is \(.427 \pm 0.11\). Different baselines achieve similar performance with small variation. This indicates that search in the game space is necessary.

“Why no baseline for the Grid? setting” Per your suggestion, we added a baseline for a Grid of size \(3 \times 6\) where the last state \((s = 18)\) has a positive reward 5 and the middle state \((s = 9)\) has a negative reward \(-3\), all other states have zero reward. Comparing with the reward found by optimization, this baseline has higher mutual information loss .115 and lower classification accuracy .482(.052).

“Why these 3 types of human?” We chose these 3 types based on the risk preference patterns discussed in the seminal paper on prospect theory by Kahneman and Tversky. Among the many applications of these patterns, loss aversion has been widely studied in behavioral economics, e.g., [Benartzi and Thaler 1995].

Reply to R3

Re: contextualizing our work with respect to related literature Thanks for this rich list of related papers and for highlighting important connections that escaped our attention. We will incorporate these references, with discussion, in the final version. Notably, Nielsen et al. also study how to generate games for the purpose of differentiating players, but use a different objective function based on Relative Algorithm Performance Profiles. In contrast, our framework optimizes mutual information to distinguish player types. Reading Tekofsky et al., Yee et al., Canossa et al. we learned that there can be complex relationships between players’ in-game and outside-game personality traits, which we assume are the same in this current work. In future work, we look forward to addressing this distribution shift.

References

Alessandro Canossa, Josep B Martinez, and Julian Togelius. Give me a reason to dig minecraft and psychology of motivation. In Conference on Computational Intelligence in Games (CIG), 2013.