We thank the reviewers for their generous comments on the manuscript which we will update to address their concerns and improve the paper. We respond to their concerns in detail below.

R1: We are glad to hear that you find this environment appealing for many RL researchers and that you enjoyed the discussion of failure cases. You suggest that we should have train/test splits to demonstrate robustness and generalization: we already use an explicit train/test split, since the environment is procedurally generated and novel seeds (therefore unseen observations, starting conditions, environment dynamics etc.) are used for testing. Furthermore, unlike MetaWorld, NetHack offers a radically different game experience depending on the seed (special dungeon features, bosses, or creature types don't occur in every seed). In Figure 3, we show the performance of agents trained on small sets of seeds, demonstrating the effect of overfitting (poor generalization) when training with small sets of seeds, and increasingly better generalization behavior for larger sets of seeds. We hope this addresses your key concern.

Additionally, you suggests more RL algorithms should be tested: we argue that training a large set of known RL algorithms is not in the scope, or indeed common, for environment papers. We chose to run experiments with IMPALA, one of the strongest distributed deep reinforcement learning approaches currently available. Therefore, we do not expect that alternative learning algorithms like A2C or DQN would reveal interesting insights into the dynamics of our environment. Instead, we carried out experiments with Random Network Distillation, a common curiosity-driven exploration approach, to investigate how intrinsic motivation might aide learning in our environment.

R2: We thank you for your supportive comments, and are somewhat surprised by the low score in light of them. We hope to address outstanding issues here to the point you will reconsider your assessment.

As per your suggestion, we will release our full research code reproducing the paper's results as an additional supplement on top of the example agent implementation and environment in the submitted supplements. We respectfully disagree with your claim that "novelty is not really present". Related RL environments such as *gym_nethack* expose a heavily simplified version of NetHack. As you acknowledge, our proposal is the only environment exposing the entire game in all its complexity, allowing for larger-scale experimentation to push the boundaries of RL research. We refer you to the list of changed environment dynamics on the *gym_nethack* project on GitHub, which in effect strips down NetHack to be much closer to its simpler predecessor, Rogue. In contrast, our version includes greater complexity, which we argue is critical for challenging modern RL techniques as well as enabling novel lines of research such as imitation learning from existing human replays, as well as NLP-augmented approaches using the game's wiki.

R4: We would like to thank you for your thorough and supportive review, and the suggestions contained therein. To respond to your individual points: **(1)** We will update the paper to include a screenshot of Medusa's island in place of Figure 1, and update the link to the referenced video to include the exact time we refer to. **(2)** *"I would like though a better justification* [. . .] *that Obstacle Tower does not provide the necessary depth for a longterm RL benchmark.":* Existing environments like Obstacle Tower or OpenAI ProcGen Benchmark, while excellent environments for testing systematic generalization of RL agents, do not require agents to deal with hundreds of items and monsters each behaving differently. As we emphasize throughout the paper, the complexity of the interactions between the large number of entities (items, monsters etc.) in NetHack will provide a greater long-term challenge to RL algorithms. The existence of large community-curated resources like the NetHackWiki, which explain NetHack's environment dynamics, are both a testament to exactly how complex the environment is (human players need these resources to achieve high scores) as well as enabling research on using external knowledge with reinforcement learning agents. **(3)** As per your suggestion, we ran experiments of our CNN model without the "cropped" parts of the input. The results confirm that the cropping contributes to the improved performance of our benchmarks. We will include the full data in the appendix of our updated manuscript. Our assumption is that the generalization behavior of the two CNNs in our original model are different due to the simpler "hero is in the middle" property of the cropped input, as well as the different sizes of the two CNNs, which in turn were chosen for performance reasons. **(4)** *curricula*: Thanks for flagging this. Our agent sees some easy tasks along the way which it may learn, helping it to do better on the harder ones. No temporal order of the presentation of the tasks was implied; hence we referred to it as implicit curriculum. To avoid confusion, we will update the manuscript to not include this statement, as per your observation that this is not necessarily how the term curriculum is understood.

R5: You state we did a reasonably good job in describing the tasks and environment to someone unfamiliar with NetHack and that we submitted a well written manuscript. Your sole criticism is that more information is needed for learning algorithm designers to select our proposed environment. Without specifics as to why, we can only say that we patently disagree: Throughout the paper, we make a detailed argument as to the benefits of NLE over popular benchmark environments. Some such key benefits are: superior runtime performance with low resource requirements, bountiful human play data for imitation learning, rich and plentiful natural language resources, a symbolic observation space with rich hierarchical structure, and a far larger quantity of entities (items, monsters, etc), actions, and more copmlex interactions between them than exist in other environments. We believe that all of these properties make the NetHack Learning Environment an attractive choice for researchers compared to other commonly used benchmarks.