We thank the reviewers for their valuable and insightful feedback, and respond to their comments and questions below.

Reviewer 1. We are pleased that the different conceptual aspects of BoxE are clear, and that our experiments are illuminating in this regard. BoxE is indeed simple to train, only needing standard optimization (Adam) of its embeddings following loss computation, and is both fully expressive, and, quite uniquely, can jointly capture and inject a rich language of inference patterns, which is very desirable in practice for model interpretability and safety.

Model expressivity and compression: The bound in Theorem 5.1 is a worst-case bound that is only tight when all KB facts are fully independent, which is highly unlikely in practice. Indeed, real-world KB entities share several properties, and this dramatically reduces the required dimensionality. This is confirmed in our experiments, where all results use much smaller dimensionality than the bound (see Table 6). In fact, higher-arity experiments (see Section 6.2) are state-of-the-art with only 200 dimensions. Furthermore, we have evaluated model robustness in Appendix H.1, and observed that BoxE maintains strong performance even with just 50 dimensions on YAGO3-10. Thus, BoxE naturally compresses information within its embeddings, allowing it to perform well at lower dimensionality.

Reviewer 2. BoxE is indeed very different, as it handles arbitrary-arity KBs natively, and can capture and inject a rich class of logical rules. Please note that we present BoxE training details in Appendix G: We use cross-entropy loss, the Adam optimizer, and hyper-parameters (including negative samples) are in Table 6. We will mention this in the paper.

Novelty of the model: BoxE is substantially different from any existing box model. Box embedding models for entity classification cannot naturally scale beyond unary classification, and Query2Box yields a model like TransE on triples (and is primarily for querying rather than KBC). BoxE is novel in many ways: (i) it introduces translational bumps, without which existing box models are severely limited (see Section 4), (ii) it proposes a novel and unified way to represent multi-arity data, (iii) it has a powerful inductive capacity confirmed by a thorough theoretical analysis, and (iv) it allows for deductive inferences via rule injection, enabling a form of reasoning within gradient-based optimization. We will make these differences more explicit in the paper. Our response to reviewer suggestions is as follows:

(1) Rule injection on other datasets: We considered evaluating BoxE on other datasets, but unfortunately no rules/ontologies exist for standard KBC datasets. Hence, we evaluated rule injection on SportsNELL, a subset of NELL with a real-world ontology. We hope BoxE leads to the enriching of existing benchmarks with rule sets, so that future KBC works can be additionally evaluated on their ability to capture and/or inject rules.

(2) Interpretability: BoxE is a highly interpretable model, as a rich language of logical rules can be captured and read solely through box embeddings. Most importantly, inductive capacity (characterized logically) correlates highly with interpretability: the more rules a model can capture explicitly, the more interpretable it becomes. For example, we can simply read off hierarchies such as $r_1(x,y) \rightarrow r_2(x,y)$, as box subsumption in the space between $\mathbf{r_2}$ and $\mathbf{r_1}$. Box configurations can also inform us in various other ways, as we have empirically evaluated and observed in Appendix H.3: On YAGO3-10, BoxE captures symmetric relations through identical boxes, and its box sizes reflect many interesting relational properties (e.g., whether it is many-to-one, one-to-many, etc.). Overall, BoxE is significantly more interpretable than existing KBC models, but there is need for more interpretability also for box embeddings. We will explicitly mention the connection between inductive capacity and interpretability in the paper.

(3) Per-relation break-down: In Appendix H.3, BoxE learns identical boxes for symmetric relations HASNEIGHBOR and ISMARRIED. We have similar observations for symmetric relations in WN18RR. We will highlight these findings in the paper. Generally, our findings support the need for more systematic evaluations against inference patterns.

Reviewer 4. BoxE is indeed a strong unifying model for KBC that generalizes to arbitrary KBs and formalizes the study of inductive capacity. We will rectify all typing (equation numbering and element-wise operators) and notation (box centers and corners) malfunctions, and will transfer figures for loss and base model explanations from the appendix.

Number of possible configurations: Yes, "4^2 possible configurations" is a typing error, which should say "4^2 possible facts". Each fact can be true or false, so there are 2^{4^2} possible worlds/configurations (i.e., all subsets). We fixed this.

Power of bumps: Entity bumps are indeed powerful, but they are unique per entity and relation-independent, which enforces structure sharing and restricts their power implicitly, yielding a strong generalization. We also experimented with relation-specific bumps, which led to overfitting, suggesting that they were "too powerful" for the standard datasets.

Negative examples and regularization: Negative examples are indeed what is relied on to maintain reasonable box sizes. We have investigated both uniform and self-adversarial sampling, with the latter empirically delivering better performance in most cases. Therefore, we believe that other sampling techniques could indeed deliver better performance than uniform sampling. In terms of regularization, we have trained BoxE with fixed-size boxes (see Appendix G.2), and have also deployed regularizers on box size and position, but these approaches have not delivered any empirical gain relative to the presented setup. Nonetheless, we are confident that BoxE can further be improved with more sophisticated training and tuning techniques, which is a very interesting area for future work.