We thank the reviewers for the constructive comments. We will revise the paper accordingly. Below are the responses to the main concerns.

Reviewer #1.

- The assumption that the causal graph is given is common in the fairness research based on Pearl's structural causal models. In practice, there are quite a number of algorithms to build causal graphs from the data (and possibly some background knowledge), such as the PC algorithm, the GES algorithm, the FCI algorithm, and their variants.

- We admit that it is a limitation of the proposed method that requires all variables are discrete with finite domains. The barrier to continuous variables lies in how to parameterize a causal model for continuous variables with arbitrary distributions and how to solve the infinite-dimensional optimization problem when we estimate the bounds. For the first problem, there are some related work on causal graph learning and inference with continuous variables, under some model/distribution assumptions, e.g. the additive noise model. But relaxing those assumptions is challenging. For the second problem, there will be infinite response variables to parameterize the continuous causal model, thus the optimal solution to $P(\mathbf{r})$ is infinite dimensional. Hence, Eq. 4 (estimate the tight bound) is an infinite-dimensional optimization problem, which is also challenging. How to address these two challenges will be a future direction for our research.

- Constructing fair predictive models is another future research direction. One possible method would be to incorporate the bounding formulation into a post-processing method. The new formulation will be a min-max optimization problem, where the optimization variables will include response variables $P(\mathbf{r})$ as well as a post-processing mapping $P(\tilde{y}|\hat{y}, \mathsf{pa}_Y)$. The inner optimization is to maximize the path-specific counterfactual effect to find the upper bound, and the outer optimization is to minimize both the loss function and the upper bound. We plan to explore this method in future work.

- The proposed method can provide the tightest bounds because the response variables cover all possible domains of $\mathbf{U}$ so that we can explicitly traverse all possible causal models. We will add more explanations about how the proposed method works in the revised version.

- In Table 3, the results of the proposed method are either equivalent to or tighter than previous methods. The bold lines are to highlight the situation where the tighter bounds make differences in detecting discrimination, showing the practical meaning of the proposed method.

Reviewer #2.

- To the best of our knowledge, the notion of path-specific counterfactual effect has not been proposed in previou works. It is worthy to point out that a similar term has been used in paper "Path-Specific Counterfactual Fairness" (AAAI'19), but with a different meaning. The paper studied the causal effect along some specific pathways without conditioning on any observed values, which is equivalent to path-specific fairness, a special case of our proposed fairness notion where $\mathbf{O} = \emptyset$. In paper "A Potential Outcomes Calculus for Identifying Conditional Path-Specific Effects" (AISTATS'19), the conditional path-specific effect is different from our notion in that, for the former the condition is on the post-intervention distribution, and for the latter, the condition is on the pre-intervention distribution. We will add more references and discussions in the revised version.

- Our proposed notion is definitely practical. It can not only unify the previous notions but also resolve new types of fairness that the previous notions cannot do. A typical example is individual indirect discrimination, which means discrimination along the indirect paths for a particular individual. Individual indirect discrimination has not been studied yet in the literature, probably due to the difficulty in definition and identification. However, it can be directly defined and analyzed using our proposed notion by letting $\mathbf{O} = \{S, \mathbf{X}\}$ and $\pi = \pi_i$. Note that the condition here is on the pre-intervention distribution, i.e., we focus on a particular individual with certain observed values, and want to estimate the change of these values after the intervention is performed. Thus, individual indirect discrimination cannot be defined using the above conditional path-specific effect. We will add the above discussions and make our motivation clearer in the revised version.

Reviewer #3.

Thanks for the comments. We will incorporate all the comments into the revised version. We will add more deriving details for Section 4.2, reorganize this manuscript accordingly, and move some discussions into the supplementary file if necessary.