1 We would like to thank all the reviewers for their time. All reviewers acknowledged the novelty of our paper. R1 is
2 mainly concerned with the lack of finite-sample theory. R2 seems to have misunderstood our motivation. R3 requested
3 additional experiments. We address these and the other questions below. All minor suggestions will be implemented.

4 **R1:** *The main weakness is the missing contribution.* The idea of utilizing common entropy for causality is new. The
5 algorithmic approach for approximating common entropy is also new. The proposal of using all these together to
6 improve the constraint-based methods is also novel. We agree it would be desirable to complement our results with
7 finite-sample theory. Note that for this, at the very least we need good analytical bounds on common entropy, which are
8 currently unknown. We believe our algorithmic contribution complemented with the experimental evidence will open
9 up future research and lead to a better understanding of this fundamental quantity. To clarify this, we will add a *Future*
10 *Directions* section that lays out these next steps. We kindly ask you to reevaluate your score in light of this.

11 *References and discussions.* We will add the citation to the reference you suggested and also add a detailed discussion
12 on [23], which was dropped due to space constraints. Please see: "In [23], authors identify an assumption on $p(Y|X)$
13 which implies that there does not exist any latent variable Z with small support which can make X, Y conditionally
14 independent. For discrete variables, this assumption implies that the conditionals $p(Y|x)$ lie on the boundary of the
15 probability simplex, which corresponds to the joint probability matrix to be sparse in a structured way." Even though
16 the high level idea here is similar, ours is a completely different approach.

17 **R2:** Thank you for your feedback. We believe that you may have misunderstood our motivation and contributions. We
18 kindly ask you to reevaluate your assessment in light of our responses below.

19 *the motivation ... is to avoid the issue of high dimensional covariate.* This is not true. The motivation is to understand if
20 the fundamental information-theoretic quantity of common entropy can help improve/complement the existing causal
21 discovery methods. We provide theory and experiments that answers this question in the positive.

22 *ADULT data has no ground truth.* Causal inference suffers from this problem. Please see synthetic experiments instead.

23 *Prior work not there.* We do cite all of the related work that we are aware of. We will add the one suggested by R1 and
24 the discussion around it. Please recommend the article that you think is missing.

25 **R3:** *Experimental evaluation and comparing w/ extensions of PC.* This is a good suggestion and we can definitely
26 implement. Note that our goal was to provide evidence to performance improvement due only to our method without
27 conflating with other improvements on the PC algorithm. We believe comparing with baseline PC was necessary to
28 showcase the performance improvement. Modifying any extension of PC is trivial and since the reviewer thinks this is
29 valuable, we will add these in camera-ready.

30 *Minor suggestions about writing.* These are very valuable and we will make sure to implement all for clarity.

31 *Related work discussions.* This has been an issue due to space constraints. We will make sure to include all the extended
32 discussions we had to drop for space in the camera-ready. Also, please see the answer to R1. Thank you for your time.

33 **R5:** *Direct graph discussions.* This is a great point and for completeness, we will provide more details. Note that proofs
34 for distinguishing direct graph from the latent graph are "easier" than the proofs for distinguishing triangle graph from
35 the latent graph. This is because triangle graph brings in more parameters which should be handled carefully (see Proof
36 of Thm. 2) This is the main reason we focused on triangle/latent distinction. For completeness, we will complement the
37 writing by including discussions around the direct graph and the associated proofs in the camera-ready.

38 *Renyi 1, Renyi 0, Shannon entropy* We will add the statement that Renyi 1 and Shannon entropies are identical. Our
39 goal for using Renyi entropy is to unify different notions of complexity of the latent variable. In some cases, it might
40 be reasonable to assume the support size of the confounder is small. In others, we might have to relax this and
41 Renyi 1/Shannon entropy is a way to do that. From a fundamental point of view, understanding which Renyi entropy
42 assumptions lead to identifiability seem important and although we focused on two special cases, in the future we hope
43 to gain a better understanding for the whole spectrum of Renyi entropies and flesh out their connections to identifiability.

44 *Clarifying limitations/assumptions*: Thank you. We will further emphasize the limitations of the method, and the fact
45 that it relies on the assumptions provided.

46 *Citing Kumar et al.* This publication has indeed been our starting point, as we cite in line 73 in pg. 2.

47 *Detailed comments.* We will use all the suggestions to improve writing/clarity. We are confident this should be doable
48 with the extra page provided in the camera-ready. Thank you.

49 *Testing Assumption 1 on real data.* As far as we know, there is simply no real data that is known to fit latent graph,
50 whereas Tubingen provides one for causally related variables.

51 *80k samples.* This is a sanity check and is used as a proxy for infinite samples. Performances coincide as expected.