We thank all the reviewers for their time and effort.

Conference peer review setup is not a fully randomized controlled trial (i.e., the reviewers are not assigned at random) and hence past approaches fail due to idiosyncrasies of the process. With respect to the specific work mentioned by Reviewer 1 (Bertrand & Mullainathan, 2004), their method assigns identities of authors to (fabricated) documents at random. In our setup, random assignment of author identities to real (i.e., non-fabricated) documents at random. In our setup, random assignment of author identities to real (i.e., non-fabricated) submissions is problematic due to various logistical and ethical issues including reviewers guessing actual authors thereby causing biases, not all authors/researchers agreeing to have their paper/name modified, and others — this opens a separate can of worms which should be rigorously addressed before using it in the peer review setting. We are definitely happy to add a discussion of this and other relevant papers (including those mentioned by Reviewer 3) in the final version.

The standard way of performing the permutation test would fail to control for the Type-I error because of the additional confounding due to quality of submissions. Our test is a careful modification of the permutation test which provably controls for the Type-I error rate even in presence of such confoundings. In the final version, we are happy to detail the shortcomings of the standard permutation test in our setup.

Unfortunately, the data from the Tomkins et al. experiment is not available to us. Tomkins et al. mention in their work that releasing this data (even in an anonymized format) would make it possible to deanonymize reviewers. Through our developed toolkit, we are happy to assist any program chairs who are interested to conduct tests of biases in their respective research community.