We believe that several concerns raised in the initial reviews stem from a misunderstanding about our paper's scope and motivation. Based on the Area Chair hCaX's metareview suggestions, we revised the paper to better highlight our main finding: **even in grounded mathematical tasks with objectively correct answers, LLMs show systematic response biases influenced by linguistic framing and demographic cues**. As suggested, we emphasized this key contribution more consistently throughout the paper by rewriting parts of several sections.

Moreover, we believe that most of the concerns raised by the reviewers are impossible to do with limited academic budgets, and we try to address them throughout the paper with new framing as well as adding new sections (a discussion and a new appendix)

Existing sections

We strengthened the articulation of the main mentioned contribution by the AC. Furthermore, as the Reviewer ZWSy was concerned, we added details on how the initial set of samples can end up in a very large set of evaluations, addressing the size concerns.
In the introduction, we also reframed the section in order to better introduce what we aim to do and focus more on the real reason and contribution of our work. We also detailed more about other contributions, such as the dataset and its existing size and its potential size with all the framing variations.
The main change that was made to the related work section is that we shortened the section, especially the subsections LLMs for Mathematical Reasoning, Demographic Bias in LLMs in order to be able to add new sections. We also modified the Framing Effects in Prompted Language Models subsection.
We did not change the content of the dataset section a lot, expect for the initial paragraph of the section, as it provides details on the dataset itself.
The evaluation section also remains mostly the same as it was previously, except for a few changes in word choices to put more attention toward the main contribution.
Throughout the results sections, we made changes both in wording and also rewrote some sentences to point out the main contribution.

Discussion

We also added a discussion section as well in order to better clarify some concerns. We explore possible strategies to mitigate the effects of linguistic framing. We mention different possible solutions, besides the CoT that is already tested, to cope with linguistic framing on Comparative math questions. We provide information on various strategies, from those that directly deal with the models to possible solutions without the need to change the model or prompt it. This part addressed the concerns of the Reviewer ZWSy.

In addition, we provided more detail about the number and scope of our experiments, addressing concerns raised by Reviewer ZWSy and Reviewer Pm7p. about the dataset and its dimensions.

We also address the effects of the ordering of the options on the results, as this was a point of concern for the reviewers rQLE, Pm7p, and ZWSy, by citing related literature and explaining that the results still show that the linguistic framing affects the results.

Appendix

Finally, as requested by the reviewers ZWSy, Pm7p, and rQLE, to have other metrics besides the introduced DirErr, we included additional metrics in the Appendix C.