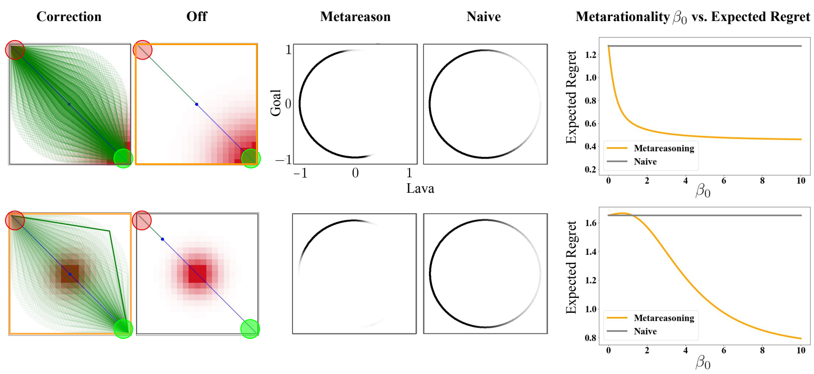
We thank the reviewers for their time and thoughtful feedback. They were kind to refer to the formalism as novel, elegant, and inspiring, and to point out that the instantiations demonstrated it is fairly universal for many existing and potentially new types of feedback. Even **R2**, our harshest critic, pointed that "I can imagine myself using this as a resource to cite both the diversity of feedback mechanisms available and to use this formalism to develop new ones." This is what we were hoping for! In what follows, we hope to alleviate **R2**'s main concern, and we take the opportunity to respond to other points the reviewers brought up.

Usefulness. **R2**'s main critique is that there isn't a new method falling out of the formalism. **R4** also asks "What are we able to do or think about that we were not able to do or think about prior to the framework?" As **R2** is actually aware, our discussion does point to several things, including the ability to combine and actively select the input types, but here we would like to emphasize the meta-choice. *The moment we said that there are multiple types of available feedback to a person, and that we should think the person is making an implicit choice within each type, it become clear that the type of feedback is itself an implicit choice* – our realization was that it too leaks information about the reward. We actually find this to be a really compelling example of exactly what **R2** and **R4** seem to be looking for! Now, **R2** does make a fair point that it'd be good to develop this further. Unfortunately there is no way to squeeze this in the paper (and still explain the formalism properly). We also really don't want it to distract from the formalism as the main contribution, which **R2** acknowledged can already be useful in developing new types of feedback. But we have run some experiments with meta-choice, and will put these in the appendix. The experiments simulate a user choosing between corrections and "off" when a robot was dealing with "lava". By understanding this as a reward rational implicit choice, the robot is able to understand more about the reward: on the bottom of the figure, if the human did a correction, and it knows the person had the "off" option and didn't use it, that tells it about the importance of reaching the goal. The plots in the center compare the belief ofer the weight on goal and lava for both naive and this "meta" inference, showing larger entropy reduction with the latter.



Actual implementation. **R1** and **R4** want to see the input types actually converted to the framework and implemented. We want to clarify that this is what is happening in Fig.1. Those are the actual reward inferences coming from each type, produced by our implementation (granted, in a simple domain, for illustration purposes to see how the types compare). **R4** might be suggesting taking this "evaluation" a step further, i.e. an analysis where each feedback type is used repeatedly.We've done experiments where the agent actively chooses which feedback is most informative which we could add to the appendix.But we do want to (respectfully) ask the reviewer to consider that this is one of the *many* things that would be useful to do with this framework, which is what makes the framework such a meaningful contribution.

Language. **R4** asks why language has to result in a uniform distribution over trajectories. We apologize, it does not have to: the formalism, as seen in eq. 1, maps choices to distributions over trajectories. Whether the distribution is uniform or not depends on the language model. This was our mistake, we will clarify! Thank you for bringing this up.

The rationality assumption. **R1** rightfully asks whether people are actually Boltzmann-rational. While this assumption has nice properties derived in our appendix and seems to have been useful in the works instantiating this formalism, it is also wrong, at least when applied naively. Recent work has explored how maybe people who seem to be irrational are actually rational, but under different assumptions. For instance, they might assume a different dynamics model, a different observation model, or use a different planning horizon. But any such improvements in human modeling can then translate to the formalism, now that we have all types of work under one unifying umbrella. One useful thing to note is that an agent can potentially detect when this assumption is wrong by detecting that no reward function explains the human's choice sufficiently well.

Cost. **R3** rightfully points out that different feedback types have different costs. When doing active learning, the agent could trade off between information gain and user cost, or have a cost budget. We'll be sure to discuss this in the paper! We also note that different types might be associated with different rationality parameters, which naturally affect their informativeness.