We would like to thank both the editor and the reviewers for carefully evaluating our submission. We appreciate the extremely thoughtful and constructive suggestions, helping us to further improve the manuscript. We have addressed these comments. As a result,

(i) we have streamlined our text,
(ii) we have extended our discussion to cover the effect of cognition costs and of population structure, and
(iii) we have performed further analyses for the case that the game is a social dilemma different from the prisoner's dilemma.

In addition, we have incorporated several smaller changes. For all details, please find our point-by-point reply below.

Comments by Referee #1:

This manuscript proposes a modification to a widely employed evolutionary model, commonly utilized for the study of repeated interactions and direct reciprocity. The authors advocate for the adoption of a discounted expected payoff as the fitness metric for behavioral (or cultural, depending on the definition) transmission. Their argument posits that this approach offers a more realistic constraint, reflecting imperfect individual memory. Despite limited memory, the study demonstrates the feasibility of obtaining results akin to the original model without incorporating discounting.I commend the authors for the excellent work and clarity of their article. I believe this work is both relevant and novel and I recommend minor revisions. However, I believe there are a few things that could be improved.

Reply: We thank the reviewer for this encouraging feedback, and for the insightful and constructive feedback below.

Firstly, there is some redundancy in the arguments, such as the repetition in the abstract regarding the significance of individuals remembering recent payoffs. Similar redundancies exist in the introduction (see the penultimate and last paragraph of the introduction).

Reply and Changes: Thank you for making us aware of these redundancies. We agree, we have iterated some aspects more often than necessary. To address this comment, we have streamlined the final paragraph of the introduction.

Also, consider separating the last paragraph of the Discussion into a dedicated Conclusion section for more clarity.

Reply and Changes: We like this suggestion! In our revised manuscript, we now have a separate 'Discussion' and a 'Conclusions' section.

Secondly, I am curious at how much difference there would be in the results between the author's discounted payoff model and interactions in a regular graph where individuals are connected to t neighbors. While intuitively there should be differences, I the difference is bounded.

Reply: This is a great question. Indeed, throughout our article we consider well-mixed populations. This means that every individual is equally likely to interact with every other individual. By now,

there is a rich literature on the effects of population structure. In particular, this literature shows that games on regular graphs can often additionally promote cooperation (even for non-repeated games).

Given this background, it would be very interesting to explore the effect of population structure on the evolution of reciprocity with limited payoff memory. However, games in structured populations often require a number of additional modeling choices. Corresponding models need to specify which network topology to consider, and how strategies are updated in detail. While we are convinced that our qualitative results would continue to hold in such a setup, each of these possible modeling choices add further nuances that should be studied in detail. In light of this, we feel such a model extension should be treated as a scientific project in its own right, which we are planning to tackle in future. Thank you for raising this question!

Changes: To make the reader aware of this possible direction, we have extended our Discussion section to include a brief discussion of future research topics.

Concerning Figures 2 and 3, acknowledging the stochastic nature of the model and assuming that values are estimations, I suggest incorporating information on uncertainty. For clarity, provide the number of simulations conducted to generate each figure's results. Additionally, in Figure 3, indicate the duration of the simulations and the parameters used, as has been done in Figure 2.

Reply: That is a good point. In our original submission we did not specify how much variation is in our simulation results. The reason is the following: The evolutionary process we consider has the mathematical property of being 'ergodic'. This property implies that if the process is just simulated for long enough, results converge to a unique limit (independent of the initial population). In each case, we have simulated the process for long enough to make sure that different simulations give the same average cooperation rate (+/- an error of 1%). Therefore, the uncertainty in the depicted bar plots is very small, and not worth depicting with, say, error bars. Admittedly, however, we failed to communicate this aspect clearly.

Changes: We have revised the last paragraph of our Methods section to clarify how we obtained our simulation results. In the caption of Figure 3, we now also specify the duration of the simulations and the parameters used. Thank you for making us aware of these ambiguities!

Lastly, I concur with the authors on the importance of balancing the study of direct reciprocity and memory-n strategies (I propose [1] as a good reference for n-memory strategies in EGT).

Reply and Changes: Indeed, we have now included this reference on memory-n strategies in our manuscript.

Given that a memory of three interactions appears to be enough to recover results close to the classical model, I would like the authors to explain further why they believe that this theoretical result is crucial for making informed deductions about reciprocity in natural systems. Would it be possible that the effect of memory is also dependent on the game? That is, would the effect be stronger in an iterated snow-drift game than in a iterated prisoner's dilemma.

Reply: Again, this is a very good question. In our original submission, we entirely focused on the repeated prisoner's dilemma. While it may be fair to say that this game is the main paradigm to

explore reciprocity, conditional strategies can also evolve in other repeated games, such as the iterated snowdrift game, as mentioned by the reviewer.

Changes: To address the reviewer's question, we have added an extra section on the iterated snowdrift game to our Electronic Supplementary Material (ESM). Again, we first describe analytically under which conditions GTFT is stochastically stable. In a second step, we run evolutionary simulations similar to the ones displayed in **Figure 2** and **4** of the main text. Overall, this analysis shows that our earlier results on the prisoner's dilemma carry over. Interestingly, however, limited payoff memory has less of a negative effect on cooperation in the snowdrift game. For example, for $b = 2$, perfect payoff memory leads to a cooperation rate of 74.1%. This number only drops marginally, to 71.2%, under limited payoff memory (see the new **Supplementary Figures 8** and **9**). We discuss these new results in detail in the ESM. Moreover, we provide a short summary in the main text.

Comments by Referee: 2

It's a wonderfully simple and clever contribution to evolutionary game theory, with a robust supplement to complement a minimalistic analysis in the main text. The authors contend with the reasonable constraint that players may not remember all prior interaction partners, or even all prior rounds of interaction with a given partner (for iterated games), when choosing to learn or imitate strategies. The main result is that longer memories are generally more permissive for cooperative outcomes in evolving populations; but cooperation can still persist (with a more constrained set of strategies) even when memories are short.

Starting first from the simple case of GTFT vs ALLD, the authors analyze what level of generosity makes cooperation stochastically stable, and they find that having a very short memory (of only the last round of play) puts more stringent conditions on a strategy to make cooperation stable. The same basic result is observed in simulations, across the full space of reactive strategies under strong selection and weak mutation.

The authors also consider regimes that are intermediate to extremely short memories vs infinitely long memories -- such as memory of the last two rounds with a single partner, or the last round with two different partners, or last two rounds of two partners, or all rounds of one partner. The same basic trends hold in these intermediate regimes: cooperation is still possible with these intermediate memory types, but less so than with arbitrary memories.

Overall it's a nice paper and I recommend publication. I especially like the Discussion section on the mechanistic interpretation of "memory" in this context, and on cognitive constraints during learning.

Reply: Thank you for this kind feedback, and for the valuable and insightful suggestions further below.

For revision, I have some questions about extending the analysis outside of the strict regimes studies, and also some questions about interpretation:

1. (comment) All of the mathematical analysis is constrained to GTFT, whereas simulations are required for studying the full space reactive or all memory-1 strategies. Line 207 should be edited accordingly, because it seems to suggest analytical results for all reactive strategies.

Reply: We agree, for our analytical results, we explore the stochastic stability of a particular strategy (or rather, a particular family of strategies). We refer to this family as GTFT, for which the strategy values are given by $(y,p,q) = (1,1,q)$. We agree that this analysis is somewhat restricted: There are infinitely many other strategies whose stability we do not characterize.

Nevertheless, we believe that the family of GTFT is key to analyze the feasibility of cooperation, in the following sense. If a homogeneous population is to achieve full cooperation, it needs to adopt a strategy with $y = p = 1$. That is, it needs to adopt GTFT strategy. With our analysis we clarify which of these fully cooperative strategies are stochastically stable.

Changes: Having said that, we fully agree with the reviewer that we should be more precise about the scope of our analytical results. Unfortunately, at least in the version we have, line 207 does not seem to be relevant for this discussion. Instead, we carefully revised the description of our analytical results throughout the main text, to minimize the risk of misunderstandings.

2. (extension) What happens when selection is weaker? The entire analysis is done in the limit of strong selection and weak mutation, which makes things simple. But this leads to some pathologies, such as the non-dependence on payoff matrix in the case of a player who only remembers the last rounds of a single opponent. I believe that the simulations are also done in the limit of strong selection (is that correct? please clarify). Can the authors tell us anything about what happens when selection is not infinitely strong? What about the opposite limit of weak selection -- where the zeroth-order (in beta) analysis will predict no difference between a long and short memory?

Reply and Changes: This is an excellent question. In fact, when selection becomes sufficiently weak, there is no difference between the different kinds of memories from the perspective of evolutionary dynamics. This was established in Ref. (74), where it was shown that a model with random, realized payoffs has the same dynamics as a model with deterministic payoffs when the deterministic payoffs are obtained by taking expectations of the random payoffs. We have now made this point explicit in the paper as follows:

"With respect to the effect of different selection strengths, Figure 2*b* suggests that both perfect and limited payoff memory yield similar cooperation rates for weak selection ($\beta \ll 1$). This is not a coincidence: it is known that, under weak selection, stochastic payoffs can be replaced by their (deterministic) expectations without altering the evolutionary dynamics (74)—and perfect payoff memory corresponds to the expected value of the payoffs in the limited payoff memory model, due to the law of large numbers. Beyond weak selection, increasing selection has a positive effect under perfect payoff memory, but a negative effect under limited payoff memory."

We had not made this point clear in the original submission and we thank the referee for the suggestion.

3. (comment). The coincidence of the condition for stochastic stability in a scenario with memory of all rounds of the last co-player, with the result for all rounds of all co-players, is presumably not

coincidental, but a reflection of the fact that the calculation is made in the N → infinity limit with weak mutation. So a random last opponent is the same the average over all opponents. The authors should explain this logic (if they agree), or otherwise explain this result intuitively.

Reply: We fully agree. For our notion of stochastic stability, we consider the situation that everyone adopts the same resident strategy (GTFT), with exception of a single mutant (with strategy ALLD). This essentially mimics a situation with rare mutations. Because all possible co-players of ALLD adopt the same strategy, the expected payoff of ALLD against any one of them is the same as the expected payoff against all of them. Moreover, as we take the limit N → infinity, the last co-player of any GTFT player is almost certainly another GTFT player. As a result of these limits, the two notions for stochastic stability agree. Thanks for providing us with this intuition!

Changes: In the revised manuscript, we explain this aspect in full detail.

4. (extension). A very short summary of the main result is: longer memory, cooperation is more stable. But what if memory is costly (which it surely is)? Can the authors say anything analytic if there is a fixed cost to having a long memory, even in the simple case of GTFT vs ALLD (but when each strategic type can either pay a cost and have a long memory, or pay no cost and remember only the last round)? Can the authors say anything about the evolution of (costly) memory, especially as it provides for greater expected population mean fitness?

Reply: This is a great point. In our study, we take the players' memory capacities as given. We ask how these memory capacities affect the evolution of direct reciprocity. It is natural to ask how these memory capacities themselves co-evolve. From the outset, there could be an interesting tradeoff: On the one hand, better memory might allow players to adopt better strategies. On the other hand, maintaining a larger memory capacity may be more costly.

Exploring that tradeoff, however, is not straightforward. It seems to us the most natural model would involve a separation of timescales. In the short run, the players' memory capacities are fixed, and players optimize their strategies given their memory. In the long run, a player's memory capacity itself would be subject to (biological) evolution. Here, one could assume that the fitness of a certain memory capacity is the expected payoff that a player with this capacity gets in the long run, minus a small complexity cost. Computing the expected payoff of a player with a given memory capacity would require substantial computations, even when we assume that players can only choose among two possible strategies.

Changes: While this is a very exciting direction, eventually we felt that it goes beyond the scope of this article. Still, the issue of memory evolution is extremely interesting that we wish to tackle in future. We address this issue in our revised Discussion section, when we highlight possible directions for future work.

5. (interpretation). The comparison between temporally discounted future rewards vs actual rewards from the past is fascinating. It is mindblowing to realize that the actual payoff from the last round is an unbiased estimator of the expected (normalized) discounted payoff. And this gives a mechanistic reason for imitation based on last-round payoff. Great stuff.

Reply: We thank the reviewer for this positive feedback!

But there seems to be a subtle inconsistency in how imitation is implemented in the short-memory case -- meaning, when a player can only remember the very last round. Even in this case, when imitation is based only on the last round, a player can nonetheless imitate the entire strategy of their partner -- which requires knowledge of their entire strategy. But how could a player with one-round memory ever infer their co-player's strategy? This problem is discussed by authors a bit (lines 291-301), but I don't think they really address or resolve this issue directly. It seems like they assume that a short-memory player can just copy their co-players full strategy, which in principle would require observation and recall all game histories. Can the authors clarify this inconsistency, or at least acknowledge it more clearly?

Reply: Indeed, we implicitly assume that when players decide to imitate a co-player, they can easily infer that co-player's strategy. This assumption seems quite restrictive in the context of our model, in which we explicitly explore the effects of limited (payoff) memory.

Interestingly, even though this assumption seems particularly strong in our model, it similarly affects other models for the evolution of reciprocity by social learning. Also there, it is usually assumed that a role model's strategy can be inferred perfectly, even though not all strategy components may be equally observable. For example, in a population in which everyone fully cooperates, a player's conditional reaction to defection is impossible to infer through observations only. These observations suggest that strategy inference is another exciting topic that requires a careful analysis.

Changes: For simplicity, and to study the effect of limited payoff memory in isolation, we abstract from these issues in the current study. One could take our model as a reasonable approximation to the case when a learner can just ask possible role models which strategies they adopt. We discuss these aspects in our revised discussion section.