Dear Editor,

Please find attached a new revised version of our manuscript "Why cooperation is not running away" following the additional revisions asked by the recommender Erol Akçay.

We thank the recommender for the very insightful comments that, we believe, have allowed us to clarify our manuscript.

We hope this new version of our manuscript will be worthy of the publication standards of Peer Community In Evolutionary Biology. In advance, we thank you very much for your kind attention in this matter.

On behalf of all authors, Félix Geoffroy.

Recommender: Erol Akçay

In all what follows, the recommender's comments are written in black font and our responses in blue.

In this revision, the authors have made a number of the changes I and the reviewers asked for and I think the paper has improved as a consequence. But there is still one major issue, which the authors have not adequately addressed. This relates to how the model builds on previous work, in particular McNamara et al. 2008. Reviewer 2 had pointed out in their review that it was unclear why the current model yields different results than McNamara et al, specifically, why McNamara et al do not get runaway in their model (and sometimes get efficient outcomes). The authors in their reply make it sound like this is purely a modeling artifact, b/c McNamara et al constrain cooperation (with linear costs and benefits) to be limited to a finite value. But this is clearly incorrect from Figure 3b of McNamara et al: it is true that for some parameter values the ES value of cooperation effort seems cut off at the boundary, but clearly for others there is an internal ESS that is not at the boundary.

We agree with the Recommender that the present work should address in greater details how it builds on McNamara et al. (2008)'s paper. We have modified the manuscript to show that our work relates to McNamara et al. in two ways.

First, in McNamara et al. (2008)'s article, the authors showed that a linkage disequilibrium between cooperativeness and choosiness arises when the mutation rate is high. This positive correlation leads to a positive assortative matching between cooperative types. Thus, we have compared their results with our second model where choosiness is plastic, since a positive assortative matching cannot arise in our first model where choosiness is hard-wired (lines 516-518, and Supplementary Information).

Second, we agree with the Recommender that McNamara et al. (2008)'s work can be used to strengthen our first model where choosiness is hard-wired. More precisely, two modelling features that are present in McNamara et al. can "counteract" the runaway process analyzed in our first model in the same way that market frictions and low phenotypic variability do. We have addressed the two features described by the Recommender (see below), and we thank him for this helpful suggestion.

This seemingly little detail about the discussion of McNamara et al made me look at their paper more closely, and realize that their model has two very crucial elements completely missing from the current work: (a) in McNamara et al, agents experience constant mortality across their lives (and also it's an overlapping generations model), (b) in McNamara et al, agents' current accept/reject decisions affects the supply of future partners. Both of these factors will work against run-away choosiness. I think these crucial differences need to be discussed in detail. Currently, they aren't mentioned at all and lead me — as well as the reviewers— think that there was a much closer semblance between the current work and McNamara et al. Substantively, they also make me less convinced about the bold statements about how "fixed" partner choice will always run away (which I already thought were somewhat overstated).

We have modified the manuscript to acknowledge that both features that are present in McNamara et al. are absent from our main text results but addressed in the SI (201-204, 294-297). The two features are (i) the assumption of a constant mortality across life, and (ii) the fact that the dynamics of acceptance and refusal affects the distribution of cooperative types in the solitary population. As suggested by the Recommender, we have investigated how both features may lead to different outcomes in our first model (hard-wired choosiness). More precisely, we have modified the Supplementary Information to present new simulation results. We have also presented these results in the "Results" section (lines 360-382).

On to the gritty details: the neglect of mortality in the current model basically amounts to no discounting of future payoffs. If there is a positive probability that I die before I get to choose again (and therefore leave no offspring), I will be more motivated to take a lower current match and get some payoff rather than have no payoff. This should reduce my choosiness threshold. Even if you don't think about literal mortality, humans (and other animals) clearly value present rewards more than future ones. The authors remove this discounting completely, and therefore create an unrealistic scenario that promotes the run-away dynamics.

We agree that discounting and/or mortality should "counteract" the runaway process towards ever larger values of cooperation and choosiness. In a sequential search setting, any mechanism that reduces the value of an individual's "outside options" (i.e. increases the cost of switching partner) should favour the evolution of lower levels of choosiness. This general statement includes market fluidity, variability of cooperative types, and, as suggested by the Recommender, mortality. We have run additional, more realistic, agent-based simulations, and we have shown that, as in McNamara et al. (2008), higher mortality rates result in lower levels of cooperation and choosiness at the evolutionary equilibrium (SI; lines 294-297 and 360-370). We thank the Recommender for this thoughtful suggestion.

The second piece important difference between the current model and McNamara et al is that if you are choosy and reject a match, chances are your next match is going to be worse. This is because the high quality individuals will not have been rejected, so the supply of new partners will be skewed towards lower quality individuals. This is true even in an infinite population, since even if there is always an infinite number of good quality matches, there will be proportionally more bad matches if the good ones are taken, so you're more likely to encounter the bad quality ones after the initial choice. McNamara et al do take this into account because their life-cycle only pairs individuals that are single (and therefore dismissed or "widowed" or newborn, page 3 of their supplement). The effect of this is also expected to work against run-away choosiness, since rejecting the current partner does not guarantee a better partner in the future (or from a different perspective, the steady state distribution of potential partners is going to be worse quality overall than the underlying distribution of quality). It is not clear whether this effect is taken into account in the agent-based model, since there isn't really an adequate description of it, or the code, but it seems clear at least the first effect is not there (since there is no mortality parameter anywhere).

The second feature proposed by the Recommender that could "counteract" the runaway process is the fact that, when individuals can choose their partner, the most generous individuals are more often chosen, and, therefore, are less likely to be found in the solitary population. This demographic mechanism increases the cost of switching partner, since the "outside options" are now less valuable on average (compared to a situation where the outside options are defined by the underlying distribution of cooperative types).

For the sake of simplicity, in the main text of our article, we have made the assumption that an individual's level of cooperation is randomized at every encounter she makes. Thus, our model does not capture this feature. We had previously studied the alternative assumption that an individual's level of cooperation is constant across her life in the Supplementary Information. However, we agree with the Recommender that it was not clear how this relates to the runaway of cooperation and choosiness. We have re-written the SI section "Complementary agent-based simulations", and we also have presented these results in the main text (lines 201-204 and 372-382).

In short, both of these effects would likely counteract the run-away dynamics of choosiness, and probably explain why McNamara et al don't get such a result. It is true that McNamara et al also don't always get efficient outcomes, but neither do the authors in their model. Rather, their main conceptual argument is only true quantitatively in the "frictionless" market regime (which as per above is actually quite a bit more frictionless than they make it out to be): there, runaway dynamics happen with fixed choice, but efficiency (or something close to it) obtains with plastic choice. Away from that regime, both conclusions are altered depending on parameters, and as I argue above, trying to make a "fair" comparison with McNamara et al would likely strengthen this conclusion.

We also agree with the Recommender that our framing about the runaway of cooperation in our first model was misleading. We have, therefore, modified the Abstract, Introduction, Results and Discussion (lines 19-24, 115-124, 313-355, 449-454 and 463-468) in order to emphasize that the runway can, indeed, stops before cooperation and choosiness reach non-efficient levels. The level at which the runaway stops, and, thus, the social efficiency of cooperation, depend on the market fluidity, as well as the mortality rate and the distribution of cooperative types in the solitary population. We have framed the Results and the Discussion to make it clear that the runaway leads to inefficient cooperation only in the "ideal" assumption of a "frictionless" market. Accordingly, we have also modified the claims concerning our second model (plastic choosiness, lines 28-29, 126-132, 419-426). We believe that our manuscript has gained in clarity by emphasizing that our two models mainly differ in their predictions when the market tends to be "frictionless". We also have changed Figure 2 in order to make the results more readable and transparent. We thank the Recommender for this clarification.

In conclusion, I still like the contribution of the paper, and believe that the plastic choice aspect is a welcome addition to the discussion of partner choice in the literature. But I think at a minimum the issues above need to be explicitly acknowledged and discussed. As it is, I think the ms creates a misleading view of how it differs from McNamara et al. but more importantly, it also glosses over some biologically/socially important aspects of partner choice that they take into account and the current paper doesn't. Ideally, one would take into account these two effects as well, but that might involve some substantial new modeling work.

We agree with the Recommender that it was necessary to take into account biologically relevant features of partner choice in our work. We hope to have implemented them in a satisfying manner. Furthermore, it helps shading light on the differences between our work and that of McNamara et al. (2008).

Minor comments: Incidentally, the authors talk (in their SI) of two potential but equivalent implementations of quality variation: one where it doesn't change through life, and one it "resets" after every interaction. But in fact the two would yield differences, since the latter implementation would not necessarily have this selection effect for the single-pool.

As mentioned above, we have followed the Recommender suggestion: we have investigated whether an alternative assumption over the phenotypic variability of cooperation changes the result of our first model (hard-wired choosiness). When individuals have the same cooperation level across their life, the cost of changing partner increases compared to the case where variability is applied at every encounter (see above). However, if the market fluidity is arbitrarily high enough ("ideal" case of a frictionless market), the runaway can go on up to socially inefficient cooperation.

The comment on Akcay and Van Cleve 2012 is also slightly off: what we showed there was in fact that you can build proximate mechanisms (e.g., families of preference functions) that can yield efficiency (or close to efficiency, see our Figure 2) with very low relatedness. So, it is a bit awkward to cite it as showing that one function requires relatedness as contrast to the current paper.

We have re-written the discussion of the Akcay and Van Cleve (2012)'s article to make it more accurate (lines 591-594). Relatedness is only one possible way explored by the article to evolve group-optimal behaviours.