October 29, 2023

**Response to Reviewers**

*Manuscript RSOS-230650*

Editorial Office

Royal Society Publishing

Dear Editor,

Attached please find our revised manuscript, “Prestige bias in cultural evolutionary dynamics We have made several changes to the manuscript in response to the reviewers’ comments. Details of the changes appear below, with our responses in blue. In addition, we attach a version of the revised manuscript with colored changed (red for deletions, blue for additions).

We thank reviewers for their comments, and we hope you find our response satisfactory.

Sincerely,

Yoav Ram, PhD

Senior Lecturer

School of Zoology, Faculty of Life Sciences

Sagol School of Neuroscience

Tel Aviv University

**Reviewer #1:**

This paper investigates the evolutionary dynamics of prestige bias by modeling it as an additive combination of “influence bias” and “success bias”. Both continuous and dichotomous cases are discussed. I like the gist of the paper, and I think rigorous, formal analysis of the evolutionary dynamics of prestige bias is urgently needed in the cultural evolution literature.

We are glad the reviewer liked our paper.

While I don’t feel competent enough to examine the mathematical details (the editor should therefore rely on other reviewers), I do have a few larger, conceptual comments that the authors may find helpful.

Major comments:

1) My first substantive comment concerns the definition of “prestige bias”. This is important because there’s quite a bit of confusion in the literature regarding how this term is used, which creates problems in not only how theories are constructed but also “prestige” itself gets operationalized and tested. Jimenez & Mesoudi (2019) provide some general comments on the current state of affairs (see Chellappoo (2021) for a philosophical critique of the concept of prestige itself), and I think these issues should be more explicitly acknowledged in the paper.

Thank you for pointing this out. We now include in the introduction a paragraph about the literature on the different definitions of “prestige bias” and reference the papers highlighted by the reviewer (line XXX): COPY THE PARAGRAPH.

The authors make frequent reference to Boyd and Richerson 1988 book, which is fine, but notice that “prestige” was used rather casually and not explicitly formulated as a theoretical concept, To my knowledge “prestige” was first elevated to a “bias” status in Henrich & Gil-White (2001), and much subsequent scholarly discussion around “prestige bias” has been based on that 2001 paper. Given this reality, I really think it would make more sense to take Henrich & Gil-White’s 2001 paper as central. My suggestion is therefore to call “influence” (as defined by the authors) “prestige”, and G\_i,j in equation (10) would be the joint contribution of success and prestige as some proxy of copying probability. This formulation would have the advantage of defining prestige in a narrow yet clear manner, and make it easier to analyze the interaction of different biases in human decision-making. This change would not affect the model itself, only how it’s described and the discussions.

Thank you for the suggestion, it makes sense. We now center the definition of the biases around Henrich & Gil-White 2001. Furthermore, we now changed “influence” to “prestige” for the effect of number of copiers, or “clientele size”; and change “prestige” to “influence” for the joint contribution to the probability that an individual will be chosen as a model. We cite the relevant places in Henrich & Gil-White 2001 for these definitions. The main changes are in the following paragraph (line XXX): COPY PARAGRAPH HERE.

the most skilled/knowledgeable models will, on-average, end up with the biggest and most lavish clienteles, so the size and lavishness of a given model's clientele 1the prestige) provides a convenient and reliable proxy for that person's information quality. Thus, selection favors clients who initially pick their models on the basis of the current deference distribution,

Prestige rests on merit in the eyes of others (rather than force deployed against them), and promotes the admiration of inferiors (not their fear), a desire for proximity not distance), and periods of sustained observation not furtive glances).

Prediction: Skilled individuals have higher status

Prediction: People preferentially copy skilled/successful individuals.

Prediction: Prestigious individuals are influential, even beyond their domain of expertise.

2) My other comment is on better situating this work within the larger literature. Based on my first comment, if we treat the models in the paper as modeling the evolutionary dynamics of multiple learning biases, then it would make sense to reference how this has been done in previous work. Hong (2022), for example, models how conformist and payoff bias may be additively combined in influencing social learning. In reality, many different biases may collectively contribute to social learning and this may be done in different ways: in addition to the additive model, learning decisions can be determined in a step-wise manner (Enquist et al., 2007). The authors alluded to this in the discussion, but I think the relationship between the present modeling choices and previous approaches should be made more explicit and perhaps appear earlier in the paper.

We now include a paragraph on previous cultural evolutionary models with multiple learning biases, including references to Hong 2022, Enquist et al., 2007, Denton et al. 2022. ANOTHER??? (line XXX): COPY PARAGRAPH HERE.

3) The adaptive success-bias weight is in my opinion the most interesting part of the paper, and in my opinion could benefit from a bit more clarification.

Sorry for the confusion. By “adaptive” we meant that each individual choses a success-bias weight that minimizes the difference between the optimal trait value A-hat and the expected chosen trait value. To avoid confusion, we now changed “adaptive success-bias weight” to “Optimal success-bias weight” in the relevant sub-section (line XXX) and in Figure 3 CHECK THOUGHOUT PAPER.

First, equation (24) needs more explication. In evolution, “adaptive” typically means fitness maximizing, and the authors should specify how this argmin term relates to fitness in more detail.

Formally, in our model there is no natural selection and hence no “fitness” per se. Indeed, if the trait value would have been correlated with fitness, the optimal success-bias weight would increase fitness. We added a statement to clarify this line XXX: COPY PARAGRAPH HERE.

Second, my reading of Figure 3 is that both panel A and B shows change within one generation (and because of this I wouldn’t call such change evolutionary). I was expecting some analyses of evolutionary dynamics of this weight in the form of long-term (over many generations) evolutionary trajectories. I don’t how much extra coding this would require, but it certainly seems a worthwhile effort.

We agree that the kind of change within one generation shown in Fig. 3 cannot be called “evolutionary change”, and we made sure that we did not call it “evolutionary change”. We revised the caption of Fig. 3 to highlight this to prevent confusion: “Both success-bias weight 𝛼 (A) and estimation error (B) decrease during the role-model choosing process (within a single generation), demonstrating that influence becomes more favored by copiers as more copiers have made their choice.”

Minor comments:

Line 402: “...includes both indirect success and influence biases, where the latter is a bias towards role-models with many copiers and hence is the same as conformity bias.” Influence bias as the authors define it is not the same as conformity bias. They are similar with important qualifications.

We removed “and hence is the same as conformity bias” (line XXX).

The paragraph that starts from line 459. This seems to me a casual use of the term “prestige” not to be confused with the more technical use here. Authors should at least point out that such use of “prestige” differs from how prestige is modelled in the paper.

You are correct. We decided to remove the paragraph from line 459.

Typo:  
Line 56: Heinrich and Broesch... Line 99: The copier than copies...

Fixed.

**Reviewer #2:**

Strategies concerning from whom to learn are termed “model biases” and include skill bias, success bias and prestige bias. The manuscript “Prestige bias in cultural evolutionary dynamics” uses a model that is a combination of within- and between- generation cultural dynamics to investigate the effect of prestige bias on cultural evolutionary dynamics. The authors derive mathematical frameworks and approximations for the evolutionary processes discussed here and use these approximations, together with simulation results to study probabilities and times to fixation for new traits entering the population.

Specifically, the authors define prestige bias as indirect success bias plus conformity bias, and through mathematical approximation they show that this model is approximately equivalent to a Wright Fisher model in which indirect success is captured through effective selection and influence bias is captured through effective population size. This means indirect success is analogous to constant natural selection and effective selection is analogous to genetic drift.

Main comments:

At a big picture level, the novelty and significance of the paper is hard to understand. Mostly because the authors use a variety of different models (constant or variable, constant or variable ‘environment’) and the differences and significance of each of these modeling choices is not discussed or really thoroughly compared to previous work.

For example, the authors introduce the concept of ‘influence’ bias, but then state, in line 405, that influence bias is equivalent to conformity bias - not sure if this is true. If this is indeed true, why introduce a whole other term/ concept which can only confuse the reader? On line 440, the authors claim ‘their prestige is similar to our influence bias’, but really to me it seems that the paper defines prestige bias as a combination of indirect success and conformity bias. A thorough rewrite and clarity on the definitions (that the authors themselves use) of the biases discussed, as well as how they are all different from each other (for the purpose of this model) would greatly improve the paper. These contradictory claims make it difficult for the reader to ultimately understand the paper.

It would be worthwhile to give a summary of the 3 different models used (in the introduction section maybe) and thoroughly discuss how the modeling choices differ from previous approaches: constant/changing environment (which, side note, is not mentioned at all until the middle of Results section, not even in the Model description sections), constant and variable alpha. What is the purpose of each, how do you compare and contrast them and what is the significance of the results presented?

Separately, currently, the discussion section, for the most part, argues for the existence of prestige bias in experimental settings and discusses previous papers, but doesn’t really discuss the results observed here. The choice to choose role models one by one versus all at the same time is also not thoroughly discussed. Ultimately, the authors should use the Introduction section to argue for the importance to study these biases and the Discussion section to actually discuss their observed results and place them in context of previous work, as well as highlight the novel dynamics observed.

Moreover, the figures and the results need a point of comparison, such that the novelty and significance of the paper can be understood by the reader. What dynamics do you observe that are different from previous models? What is the specific change or mechanism that is driving these differences in the dynamics?

I found the results on lines 436 (fig 3) potentially interesting and worth more exploration, but are only studied superficially. What drives these lower values? Also, the fact that the value seems to stabilize rather quickly after a sharp decrease is not mentioned or discussed, but seems important for understanding of what might cause the observed dynamics.

Specific comments, typos, etc:

Equation 11 is wrong: Numerator needs K\_i-1.

Fixed.

Wright-Fisher mean change equation in line 307 we computed should read sx(1-x), unclear why the extra term, please double check.

Fixed, it is now sx(1-x)+o(s).

Equation 19: plugging s and Ne into the exact equation cited (eq 8 in ref 19) gives an extra factor of 2, please double check

In ref 19, Kimura assumes Ne = 2N due to “randomly mating population of size N”, and therefore there is an extra factor of 2 in Eq 8. Thus, our eq 19 remains unchanged. Note that the reference just above result 4 said “Kimura and Ohta”, this was a mistake and corrected now to “Kimura”.

Figure 1: Why does the probability of fixation increase with alpha in Fig1 but the opposite trend can be observed in Fig2? This comment falls in the general trend that the results of the model and the mechanisms driving them, their significance, implications are not discussed.

Figure 1: caption states A/A\_hat varies between 0.01 and 0.99, but x axes in C and D don't show this, very unclear how the numbers would match.

Are both Fig 2C and Fig 2D needed? 2D does not contain new information.

TRUE. WE SHOULD DECIDE WHICH ONE TO KEEP.

In addition, the paper contains many typos, omissions as well as confusing statements or sentences that are hard to parse. A thorough rewrite is necessary. We write a subset of the typos encountered here:

Line 98, than to then

Fixed.

Line 100 transmission models are

Fixed.

Line 124: say that you’ll describe this new distribution in what follows.

We added “We formulate this assumption in the following.”

Remove comma line 151.

Fixed.

Line 152: a simplified version of what? Of your previous model?

We meant a simplified version of the previous model—instead of a continuous trait, we focus on a dichotomous trait. We revised the text to “We introduce a simplified version of the above model where the trait has only two phenotypes”.

The paper keeps switching between individual-based alphas and constant alpha and it’s hard to keep track of how all the assumptions change as the various results are presented. Moreover, even though it’s introduced in line 156, alpha is technically never formally defined. See also line 174, confusing, since these don’t appear in eq 11.

Line 160: Without loss of generality: not sure this is true, especially when just two phenotypes, error rates could very well change the dynamics significantly here.

We removed the remark “without loss of generality”

Line 162: which rest of the details?

We removed the remark about the rest of the details.

Line 186: confusing to say order of copiers doesn’t matter when it’s baked into the model.

We removed the remark about the order of copiers.

Line 274: The sentence that starts with that has no verb? Unclear what the meaning is.

Right, we now revised the sentence to say “We find that the number of simulations needed to sufficiently approximate our model with the DM approximation is roughly 1,000”

Line 302: why is this surprising?

We changed “surprising” to “interesting”.

Result 6 needs citation. Why?

The line before Result 6 does indeed give a citation to Ram et al 2018 (ref 21). It says: “Using the drift and diffusion terms and following [21], we can approximate the fixation probability in a changing environment.”

Caption Fig1: “bounded by blue”. There is no blue line in this figure.

Changed from “blue” to “dashed line”.

Overall the captions sometimes lack all the parameters necessary to understand the results:

For example Fig 2A, B: what is k, l? That figure also needs more runs of the simulation.

Figure 3: what does copier mean on x axis? Number of copiers?

Line 404: typo, the these

Fixed.

Line 456: what is ‘it’ here?

Changed to: “They show that the social networks representing copier--role-model relationships are centralized, suggesting that it which is consistent with the prediction that people substantially share notions about who is a good cultural model”.

Also I would reword ‘they’ studied: The researchers studied. The authors studied…

Done.