Dear Professor Keller,

Thank you very much for the chance to revise our manuscript. We would also like to thank both reviewers for their comments. Below we deal with the points they raised.

We've also made some other changes:

- We moved the results on causality of polygenic scores to the appendix. We thought they were a distraction from the key point of Section 4, which is to test our theory.
- Code is now publicly available at https://github.com/hughjonesd/why-natural-selection.

With best wishes,

David Hugh-Jones
Abdel Abdellaoui

Editor

> In particular, Reviewer 2 makes several good points that I believe can improve the clarity and flow of the manuscript and I suggest you consider these carefully. Reviewer 1 was highly favorable but also makes several excellent suggestions – including pointing out that a similar observation has been made (remarkably) almost 100 years ago by Fisher. This seems important to include.

See below.

Reviewer 1

> This is an outstanding paper, perhaps the most important that will be published in the field of behavioral genetics this year.

Thank you!

> Most curious is the failure to cite Fisher (1930)…. The overlap in theory and observations with the present work is striking, so much so that the omission is equally striking.

We thank the reviewer for giving us this reference. Indeed it is striking. Fisher's theory about the positive correlation he mentions is that it's driven by selection bias: Yale parents who can afford to send 6 children to Yale are "more able" than those who can only send 1, and as a result, among Yale students, children from bigger families are smarter. This is related to the income effect in our theory. We now mention this work in our conclusion ("Strikingly, some of these results were predicted by Fisher (1930), pp. 253-254.")

The authors do not have to respond to my next point. This paper already has a lot. But I am curious as to whether the authors have an opinion about Fisher's proposed solution to this problem (if it is a problem). He in essence suggested that the population be stratified into classes based on "human capital," as we might say now, and that within each stratum redistribution should be performed so as to reward parents of larger families. The basic idea is that if I have four children, while an academic colleague who is the same age as me has none (which has allowed him to publish twice as many papers), the tax-and-spend system will take away some of his income and give it to me, the total amount of redistribution being such as to remove the fundamental tradeoff. Can this work in theory? Can it be made politically feasible?

Indeed, we don't want to talk about policy in the paper - we think we do not yet know enough about the underlying forces. The proposed policy sounds close to some things that already exist e.g. child benefit in the UK, which is a non-means-tested per-child benefit. For what it's worth, my personal view is that government schemes to encourage fertility have a poor track record, see e.g. Gauthier (2007), though see also the recent Bergsvig et al. (2021). I wrote more informally on the policy issue here.

Another general point is that there is some uncertainty over whether the fertility-IQ correlation is negative in entire populations. Some suggest that its negative value in unrepresentative samples is the result of ascertainment bias (e.g., Higgins, Reed, & Reed, 1962; Kolk & Kieron, 2019). In fact, I have reviewed a manuscript where the authors declined using the UK Biobank for a particular analysis (while using it for all others) on the grounds that the fertility-IQ relationship cannot be accurately inferred in that dataset. My position is that the fertility-IQ correlation has been negative, at most times since relevant data have been collected, while being less so at certain atypical times (e.g., the Baby Boom) and in men. In the present work, the authors use sample reweighting to address ascertainment bias in the UK Biobank, but this is not guaranteed to be effective. A brief literature review addressing the sign of the fertility-IQ correlation, in most Western populations in recent decades, therefore seems warranted to provide some reassurance. My suggestion is to begin with Lynn and Van Court (2004), Meisenberg (2010), and Reeve, Heeney, and Woodley (2018).

We agree this is an important topic, but we are more focused on human capital and earnings than IQ specifically. (Of course the polygenic score for educational attainment also captures a portion of the polygenic effects on IQ.) The negative fertility-income relationship, in developed countries, is a very well-established stylized fact, and we mention some papers that also show income is causal. We now mention two of the cited papers: "A related literature shows negative correlations between IQ and fertility (e.g. Lynn and Van Court 2004; Reeve, Heeney, and Menie 2018)".

Regarding ascertainment bias, the reviewer is right that there are no guarantees we can eliminate it. When we do correct for ascertainment bias, every correction increases our estimated effect sizes. So we expect that any remaining ascertainment is also probably biasing our estimates towards zero. We now say this: "There may be remaining sources of ascertainment bias after our weighting; if so, we expect that, like the sources of ascertainment we have controlled for, they probably bias our

results towards zero."

abstract: "Consistently over time, polygenic scores associated with lower (higher) earnings, education and health are selected for (against)." This is hard to understand. How about: "Consistently over time, polygenic scores predicting higher earnings, education, and health also predict lower fertility."

Done.

p. 2: "natural selection may substantially increase the correlation between genetic scores and income, i.e. the"genetic lottery" (Harden, 2021)." Although the meaning eventually becomes clear, this sentence may be very cryptic.

We have made it more explicit: "natural selection may substantially increase the correlation between genetic scores and income, increasing genetic differences between different social groups, and thus making the"genetic lottery" (Harden 2021) more unfair."

p. 4, footnote 1: That the covariance between the number of children and the polygenic score for a given phenotype is equal to the evolutionary change in the phenotype between generations is, after some adjustment of terminology, Robertson's (1966, 1968) Secondary Theorem of Natural Selection (see also Walsh & Lynch, 2018, pp. 165-167). I think this should be pointed out.

We now cite Robertson (1966).

p. 4: It should be "stabilizing" selection, not "balancing" selection. The latter refers to any kind of selection that favors the coexistence of two or more alleles at a site. The former refers to selection that favors an intermediate value of a quantitative trait, and this is clearly what the authors mean. While I believe the term "diversifying" selection is sometimes uses for selection favoring both extremes of a quantitative trait, "disruptive" selection is probably more common. A similar point applies on p. 20.

We have changed vocabulary accordingly, using "stabilizing" and "disruptive" throughout.

p. 10: I am interested in the functional form chosen for u(y). This has sensible properties for sigma < 1 near 1, including being increasing in y for y somewhat greater than zero and being concave. Intuitively, this means that utility increases with income (net of costs of childrearing) but less so as income increases. Is there any particular reason for this functional form as opposed to others that also satisfy these properties? Also, is there any empirical evidence that "utility" obeys this relationship?

The functional form is a standard one in economics. It has the specific property that relative risk aversion is constant. That is, if you have CRRA utility and £100, and you are indifferent to a 50-50

bet between winning £30 and losing £20, then if you had £1000 you would be indifferent between a 50-50 to win £300 or lose £200.

The literature on estimating utility functions is vast (see e.g. Chuang and Schechter 2015). One issue is that of context. CRRA may make sense for risky financial decisions. Here, as Reviewer 2 points out, the "curvature" of the utility function is used not to explain attitudes to risk, but to trade off income against children at different income levels. This author wishes there was a single utility function that worked to explain all human behaviour across contexts… in reality, economists tend to use "what works". We use the CRRA form because it is tractable; we expect that using other forms, intermediate levels of risk aversion would still generate our empirical results, but the maths would be uglier. For a useful discussion of microeconomic theories of fertility, including functional forms, see Jones et al. (2008) and more recently Doepke et al. (2022) which uses a very similar form to ours to capture birth timing.

p. 30: I feel that more thought might be given to the choice of covariates. For example, if fluid IQ is on the causal path between the PGS and EA, then controlling it will bias downward the estimate of the the PGS effect on EA. But since this mediation analysis seems tentative and not crucial to the paper, perhaps some looseness here can be tolerated.

Yes. Conversely, fluid IQ might affect fertility choices via other pathways than human capital in the labour market, and not controlling for it might leave these to be picked up by EA. We prefer to be "conservative" in controlling for it, in the sense that we'd rather bias the estimate indirect effect of EA towards zero than away from zero. We now mention this issue in the appendix: "Note that controlling for fluid IQ is conservative, in the sense that IQ might also measure human capital; nevertheless, we control for it because fluid IQ might affect fertility decisions via other channels than human capital."

That said, the mediation analysis is really the central test of our theory. So we have tried to make it seem less tentative. See our response to Reviewer 2.

p. 34: For various reasons I think this section (and the bit in the main text that it supports) ought to be eliminated. The point is to show that natural selection increases the correlation between the EA PGS and income from one generation to the next, correcting for the noise in the PGS. I think a qualitative statement about this point is adequate; the more quantitative attempt by the authors has some problems.

(1) "PSEA" is not defined, although I'm guessing this means "polygenic score for educational attainment."

(2) It happens that there is a more recent meta-analysis of relevant twin studies (Silventoinen et al., 2020).

(3) There appears to be substantial passive gene-environment correlation ("genetic nurture," some now call it), as a result of highly educated parents providing both favorable genes and a social status promoting educational attainment (e.g., Kong et al., 2018). I am sure this complicates the analysis in some way. The authors might respond that they deal with this by using within-family estimates, but see below.

(4) The proof of Equation 5 uses the "standard errors-in-variables formula" plim \hat\beta = \lambda\beta. While this relationship seems plausible, it may not be standard to behavioral geneticists.

(5) When there is genetic nurture, what is the right PGS effect to use in order to address the question here? The effect in the population, or the effect within families? If it is the latter, a problem is that the estimate of a 142.6% change in R^2 between generations seems far too high and indicative of low statistical precision.

(6) The assumption that eta is independent of p_star seems very strong. It includes, for example, the assumption that the PGS over rare variants is uncorrelated with the PGS over common variants. I think this is unlikely to be true, as a result of assortative mating (Yengo et al., 2018).

We accept point (6) in particular. We had pushback from people saying "this is no big deal", so we want to show that at least under some assumptions, yes it might be. The within-family estimate should remove many of the worries about exogeneity of p*, but they are estimated with a smaller sample. We agree that the estimate seems high, and indeed, the bootstrapped 95% confidence interval is from around 0% to 300%, which is not really informative, though it does suggest that genetic nurture is more likely to bias our results downwards than upwards. We've removed the discussion.

p. 40 and following: I confess to not following the derivations completely. The authors seem adept at what they do, and I trust that they have made no serious errors.

(1) This is my first encounter with the Karush-Kun-Tucker conditions or the Inada condition.

KKT are the standard first order conditions for constrained optimization (see e.g. Sundaram 1996). We think people who want to follow the proofs will know about them. The Inada condition is a bit more specific to economics. We are now more explicit about what it means: "the Inada condition (that marginal utility of income grows without

bound as income approaches zero, $\lim_{x\to 0} u'(x) = \infty$)". It can be derived in this case by differentiating u to give $u'(y) = y^{-\sigma}$ which for $\sigma > 0$ approaches infinity as $y \to 0$. An intuitive interpretation is that if you have extremely low income, your need for income overrides other considerations.

> (2) Below Equation 25, I did not verify the second derivative of N_2 with respect to h. This seems to require simple (if tedious) applications of the product and chain rules. The important conclusion, that N_2 is convex in h for sigma close to 1, is something that I will accept by analogy to the case N_1, N_2 > 0.

This is correct.

> (3) I do not understand in what way Equations 13 and 14 determine when the case N_1 = 0, N_2 > 0 holds. Nevertheless, I accept the intuitive conclusion that it holds for intermediate values of h because too small a value of h (which determines s*) will make N_1* in Equation 17 greater than zero and because too high a value will make N_2* in 18 negative. In a similar way, I accept that no children at all is the optimum for very high values of h; this seems reasonable for this model.

That is the right intuition. Here are the details. Equation 13 requires

$$-bY_1^{-\sigma} + a \leq 0$$

since λ_1 must be non-negative. The LHS is increasing in Y_1. $Y_1 = 1 - s$ since $N_1 = 0$. So the LHS is decreasing in s and the inequality puts a minimum value on s. Lastly, optimal choice of education s^* increases in h for $\sigma < 1$ (and decreases for $\sigma > 1$). Hence for $\sigma < 1$, the inequality puts a minimum on h, and for $\sigma > 1$ it puts a maximum on h.

Equation 14 similarly bounds h, as follows. Equation 22 rearranges equation 14. Equation 25 differentiates equation 22 to show how N_2 changes in h; for $\sigma < 1$, it decreases, for $\sigma > 1$ it increases. Therefore, the requirement that $N_2 > 0$ (for this solution) means that when $\sigma < 1$, h cannot be too high, and when $\sigma > 1$ it cannot be too low.

We have tweaked the discussion slightly to reflect the above. The best way to gain intuition is probably to try a few numerical examples.

Reviewer 2

> 1. The authors have improved the presentation of their results, but it can still be further improved. There are still so many results, plus a model, and as a result the paper still reads like a disconnected set of findings and lacks an overall coherent narrative (although a bit less so than before).

We have tried to work on this. Unconnected results and minor points have been pushed to footnotes and/or the appendix.

We don't want to go too far, by pretending that the theory came to us fully-formed in our ivory tower, and then we tested it. In fact, the motivating results led us to develop the theory, first informally and then more precisely; only the results in section 4 should be considered a clean test.

> In my opinion, the paper's most important finding is that human capital mediates natural selection. I suspect the authors agree with me on this point, since the title of their manuscript says just that. Still, when reading the paper, one does not get to the topic of human capital and natural selection until p. 9 (and the paper has only 15 pages)! The lengthy discussion of the stratified analysis (Figures 3-5) is interesting, and the authors have edited the text (vs. in the earlier version) to explain that these results are related to predictions of their model, but that discussion is too long and distracts from the main point of the paper. Plus, the stratified analyses are
> limited and only so much can be learned from them. As the authors acknowledge, "these categories are not necessarily exogenous to the polygenic scores. For example, both in the data and in our theoretical model, age at first live birth is a choice variable, which is endogenous to human-capital-related polygenic." Thus, I recommend that that discussion be shortened and the figures consolidated (the authors should find a way to show the must important results in Figures 3-5 on ~1 page.

We think these analyses are important. First, they motivate our theory, in which education, income, AFLB etc. are indeed endogenous choice variables which are co-determined with fertility. If these mediators were *not* endogenous, we'd need a different theory.

We also think the results matter in their own right. Any theory of contemporary natural selection needs to explain why correlations with fertility are lower at higher levels of income, education, etc. Non-causal explanations are fine, but theories which don't generate these correlations must be missing something. Note that Fisher, cited by Reviewer 1, used observed correlations quite like ours to differentiate between theories of the fertility-income relationship. Generally, significant correlations have often led to important theoretical literatures, even when the direction of causality is initially uncertain (e.g. education/wages in labour economics, or democracy/GDP levels in political economy).

As described above, we have shortened Section 2 to focus on the key results of variation by income, education, number of partners, partner presence and AFLB. These empirics now take up just 39 lines of text, plus Figures 3-5. We no longer split respondents by sex in Figure 4. This saves two figures, and should reduce space in the final version.

2. The authors test their human-capital-mediation theory in two ways in Section 4, but I think their analysis can and should be improved, since this section contains the core empirical results supporting the authors' main claim about human capital mediation. First, the authors show that a PGS's correlation with two measures of human capital negatively correlates with selection effects for the PGS (Figure 6); second, they run a formal mediation analysis.
> Can the authors introduce a basic econometric framework and assess whether their results in Figure 6 are consistent with those from the mediation analysis (i.e., could the amount of mediation reported in Table 5 for each phenotype explain the results shown in Figure 6)?

The result in Figure 6 is that the size of a score's selection effect correlates with the score's correlation with education and earnings. Intuitively, the fact that part of the selection effect comes from the mediation indeed helps to explains table 6. That is, if

PGS effect on fertility = PGS correlation with education X education effect on fertility + direct effect of PGS on fertility

then indeed PGS effect on fertility, and PGS correlation with education, will be correlated. In that sense, the mediation analysis is just a more precise version of Figure 6 (more precise because it also estimates the education effect on fertility). We're not sure if there is an obvious way to work out "how much" of the Figure 6 correlation is explained by mediation. It may be a bit like asking "how much of a bivariate correlation is explained by the same correlation but with controls": these are really two alternate models of the same process, one less controlled than than the other. We are probably hitting the limits of our econometric expertise, so we are open to suggestions.

> In the mediation analysis, why do the authors only use education as the mediator? The mediation framework they use can easily accommodate many mediators (see Section 6 of the Supplementary Information of the "EA2" paper by Okbay et al., Nature Genetics 2016). I would like to see a more full-fledged mediation analysis that also includes other proxies for human capital as mediators, including earnings in a respondent's first job as well as household income (though the later is not an ideal variable for that, especially for females).

We would like to have more mediators, but appropriate variables are hard to find. Earnings in first job reduces our N by a factor of 10, and this makes results too noisy to be informative. As the reviewer suggests, current household income is very much endogenous not only to human capital, but also to previous fertility, which makes it inappropriate as a mediator.

> Also, in the mediation analysis, some of the ratios of indirect to total effects are negative or larger than 100%. This could indicate that the indirect and direct effects have opposite signs, but could also be due to the imprecision of the estimates. The authors should estimate 95% CIs for the ratio of indirect/total effects (these can be obtained by bootstrapping). They could then focus their discussion on the ratios that are significant and not on those that are just noisy estimates.

Bootstrapping ratios is not so simple, since if the confidence interval for the denominator contains

0, the ratio confidence set may contain infinity and may not be a single interval, and the standard approach of taking quantiles can be misleading (see Franz 2007). Alternatives like the Fieller method have the same problem. Another issue is that to estimate Bonferroni-corrected significance at p = 0.05/33 we'd need 0.00075- and 0.99925-quantiles of the distribution, which would require many thousands of bootstraps (unfeasibly slow, given our sample size).

Following the advice of Franz (2007), we now compromise: we bootstrap uncorrected 95% CIs of the ratio, but only when the denominator (i.e. the total effect) is significantly different from 0 at $p < 0.05/33$. This is substantively sensible anyway: if we can't reject that a PGS has no effect on fertility, it does not make much sense to ask whether the effect is mediated. We continue to report p values for the indirect effect only, following Zhao et al. (2010).

> Also, in the older version, the authors had a nice figure illustrating how education mediates selection (their old Appendix Figure 17, which showed results from re-estimating Equation 1 for each PGS while controlling for education vs. the baseline estimates). The mediation analysis is in a way more thorough version of that figure, but that figure (ideally also showing the effects of controlling for earnings and any other human capital variables) was nice and helpful and could go in the main text

We have introduced a new figure in the main text, showing the results of the mediation analysis with the bootstrapped confidence intervals.

> More generally, I think the authors can improve their empirical analysis of human capital as a mediator of natural selection (the above comments should help with this).

We welcome any further suggestions.

3. To adjust for ascertainment bias, the authors adjust their estimates using three weighting schemes. Why not combine all the variables from the three schemes and add other observables that could impact ascertainment (including also sex, proximity to a recruitment center, number of individuals in the household, and household income)? If the authors' goal is to adjust their estimates for ascertainment in the UKB, they should use all relevant observables to create UKB weights so that their estimates more closely resemble those they'd obtain in a more representative sample. This is particularly important here since we're interested in selection, and estimates can be severely biased in a selected sample like the UKB.

Also, instead of presenting a small subset of the results with various weighting schemes in Figure 2, it'd be preferable for the authors to pick what they think is the best weighting scheme and then present all results with that scheme. Then, they can briefly discuss the robustness of their results to various weighting schemes.

We agree with this. Combining all the variables is hard: age at first live birth is only available for females, and some of the data comes from different sources. In the new version, we use weights kindly

provided to us by Sjoerd van Alten and coauthors. Their working paper is specifically dedicated to weighting UK Biobank, so it seems good to use the "industry standard" solution rather than rolling our own. We now use these weights throughout, but detail results from alternative weighting schemes in the appendix.

> 4. On p. 4, line 13, the authors write that "Beta is the expected polygenic score among children of the sample", with a footnote explaining the reason. However, since the authors use N=number of children rather than relative number of children (=number of children/mean number of children), I don't think that's correct. See Beauchamp PNAS 2016. More generally, why not use relative lifetime reproductive success (rLRS) instead of number of children as the dependent variable in equation 1?

Oops, yes. We hope the corrected footnote is now right.

We now use RLRS as the dependent variable for both respondents and parents. Beyond the obvious change in absolute effect sizes, results are not much changed, except there is now slightly more evidence for larger selection effects over time: 8 polygenic scores show a significant increase in effect size on respondents' RLRS when we median-split them by birth year. Probably this was masked when we used absolute fertility, by the fall in fertility over time. In turn, this makes us less sure that we can rule out that the welfare state is a relevant explanation for natural selection, so we have slightly qualified our comments on this topic.

> 5. How were the 33 PGSs selected? For instance, why didn't the authors include a PGS of risk preferences? (I'm not requesting that the authors include that specific PGS, but only that they explain their selection process.)

We address this in the Materials and Methods section: "Polygenic scores were chosen so as to cover a reasonably broad range of traits, and based on the availability of a large and powerful GWAS which did not include UK Biobank."

> 6. Did the authors use an age cutoff to include males and females in their selection analyses. To ensure measured fertility is completed fertility, the authors should only use males who were at least 50 years old and females who were at least 45 years old at the time of measurement.

We used 45 years old. We have adjusted this to 50 male/45 female as suggested. Results are little changed, except that EA3 is no longer undergoing disruptive/diversifying selection in the respondents' generation at the p = 0.05/33 significance threshold.

> 7. The authors write that sigma captures risk aversion in income in their model. But that's confusing and not a good way of describing this for their model, because as far as I understand there's no income risk! They should clarify this and say that sigma regulates the curvature of the utility function, and explain what this means in their riskless model.

It is a fair point. We have rewritten accordingly: "$\sigma > 0$ measures the curvature of the utility

function, i.e. the decline in marginal utility of income as income increases".

OTHER COMMENTS AND SUGGESTIONS
8. Most figures are hard to read. For example, Figures 3-5 are hard to read. In Figure 3a, the dot color indicate education and the dot border color indicates significance; that's very difficult to read quickly (for instance, a blue dot with a yellow border looks green). Instead, the authors could use different symbols (crosses, squares) for education and colors for significance, or vice-versa.

We have changed our figures to use both shape and colour for categories. We hope this is an improvement. We've also worked on Figure 1.

9. The authors should provide more info to help readers understand the details of the analyses whose results are shown in Appendix Tables 3-4.

We now say: 'We report the number of scores showing significant changes over time (i.e. a significant interaction between polygenic score and the "late born" dummy): either a significant change in sign, a significant increase in size, or a significant decrease in size.'

10. The authors do not refer to Figure 10 in the main text.

Fixed.

11. As mentioned above, only so much can be learned from the stratified analyses, and I recommend that less space be devoted to them in the main text. Still, the authors should consider showing the results (probably in the Appendix, given space restriction) stratified by geographical variables, which are more plausibly exogenous (though not perfectly so), such as being born in the North vs. South, in coal mining area vs. not, in rural areas vs. in towns. This may not related directly to their model, though.

See above for our thoughts on exogeneity versus theory. We have added in results for respondents' RLRS, stratified by Townsend deprivation of birthplace. These look very similar to the results for parents' RLRS. Note, though, that since polygenic scores are inherited, and correlate with parents' mobility (Abdellaoui et al. 2019), geographic variables are indeed not exogenous.

References

Abdellaoui, A., Hugh-Jones, D., Yengo, L., Kemper, K.E., Nivard, M.G., Veul, L., Holtz, Y., Zietsch, B.P., Frayling, T.M., Wray, N.R. and Yang, J., 2019. Genetic correlates of social stratification in Great Britain. *Nature human behaviour*, *3*(12), pp.1332-1342.

Bergsvik, J., Fauske, A. and Hart, R.K., 2021. Can Policies Stall the Fertility Fall? A Systematic Review of the (Quasi-) Experimental Literature. *Population and Development Review*, *47*(4), pp.913-964.

Chuang, Y. and Schechter, L., 2015. Stability of experimental and survey measures of risk, time, and social preferences: A review and some new results. *Journal of development economics*, *117*, pp.151-170.

Doepke, M., Hannusch, A., Kindermann, F. and Tertilt, M., 2022. *The economics of fertility: A new era* (No. w29948). National Bureau of Economic Research.

Franz, V.H., 2007. Ratios: A short guide to confidence limits and proper use. *arXiv preprint arXiv:0710.2024*.

Gauthier, A.H., 2007. The impact of family policies on fertility in industrialized countries: a review of the literature. *Population research and policy review*, *26*(3), pp.323-346.

Jones, L.E., Schoonbroodt, A. and Tertilt, M., 2008. *Fertility theories: can they explain the negative fertility-income relationship?* (No. w14266). National Bureau of Economic Research.

Sundaram, R.K., 1996. *A first course in optimization theory*. Cambridge University Press.

Zhao, X., Lynch Jr, J.G. and Chen, Q., 2010. Reconsidering Baron and Kenny: Myths and truths about mediation analysis. *Journal of consumer research*, *37*(2), pp.197-206.