The authors would like to extend our sincere thanks to the reviewer, Florian Pargent, for his careful and thorough review of our article. The reviewer found no errors within Section 1 (Methods, Measurement, and Design), or Section 3 (Claims, Presentation, and Interpretation) of his ERROR review. The reviewer did identify several inconsistencies and recommended analytic improvements within Section 2 of the review (Data, Code, and Statistical Analyses), which we will comment on below.

**Discrepancies in Reporting**

The Reviewer found four small discrepancies in values reported within the paper. Two had been reported correctly but not sufficiently clearly (Points 1 and 2) and two appear to be typos (Points 3 and 4).

1. **Gender breakdown**

The reviewer notes that our descriptive statistics report 187 undergraduate students in Sample B (93 women and 94 men). However, some of the data files have only 93 men (one gender value is missing). Upon investigation, we found that this is because one of the male participants is missing a key DV (own desire). This male participant is therefore missing from analyses that require this variable (e.g., models predicting own desire or dyadic desire). However, this participant still matched with other, female participants, and is therefore still included in other models (e.g., those predicting partner desire). That is why the participant is still included in the sample statistics. We should have documented this missing data and subsequent exclusion in the manuscript.

1. **Total variance explained**

The reviewer notes that we claimed in the abstract and in the discussion that Random Forests predicted a minimum of 4% of actor variance and 7% of partner variance. However, the numbers in Table 2 show their lowest values as being 5% and 18%, respectively. The reviewer speculates that we may have obtained these lower values (4% and 7%) by taking the squared correlation between predicted and actual scores presented in Table 3. The reviewer is correct: we considered the results of all the models we ran (both Tables 2 and 3) when describing the full range of obtained values in the abstract and in the discussion. We should have noted the source of these values more clearly in the manuscript.

1. **Number of predictors**

The reviewer identifies a sentence on page 1485 that says that Sample B included “38 predictors (20 male and 20 female predictors)”. The reviewer is correct that this is a typo and the sentence should have said 40 predictors.

1. **Variance in women’s desire explained**

The reviewer identifies a discrepancy in how much variance the stringent variable selection was able to predict in women’s desire for men. The text says 1.30%, but Table 2 says 1.34%. The reviewer ran the analyses and obtained a value of 1.33%. They note that updates to the relevant packages make it impossible to say which value was ‘correct’ at the time the paper was published, however, the discrepancy within the paper was likely a copy error. We agree with this assessment.

**Missing Preprocessing Information**

The reviewer notes that only preprocessed data were included in the project, leading to a general lack of transparency and reproducibility at the data processing step. There are three main reasons for this. First, a pragmatic reason: the current manuscript reports on speed-dating data that had been collected some years ago at the time (2007 and 2008), and which were being re-used for the current project. Thus, most of the data pre-processing had already occurred long before the current project took place (and indeed, before OSF existed).

The remaining, project-specific data processing steps are also not transparent, for a reason that I (Sam, the first author) am rather sheepish about. This was my very first research project using R rather than SPSS, and I was not yet comfortable using the program to manipulate data frames. That is, I was unsure of how to properly calculate new columns in a data file within R, so I made changes to the dataset within excel instead. This explains why there are so many different, near-identical versions of the same data file. Today, I would of course use a more streamlined approach (tidyverse is a thing of beauty).

Finally—and perhaps most compellingly—there are ethical concerns surrounding the sharing of these data, both because they include sensitive topics and because they contain data from multiple people who may know each other (see Joel, Eastwick & Finkel, 2018 for discussion). An already cleaned and de-identified dataset poses fewer ethical risks than a raw, unprocessed data file, especially when many of the initial cleaning steps involve the use of identifiable variables (e.g., matching email addresses or names). This concern seems likely to be broadly relevant to researchers who study sensitive topics.

**Recommended Analytic Improvements**

Finally, the reviewer notes three places where recent advances and current best practices would suggest using a different approach to the analyses.

1. **Handling of Missing Values**

The reviewer notes that missing data were handled with listwise deletion, by simply removing those rows from the analyses. Although the amount of missing data in these studies was quite minimal, we agree with the reviewer that 1) we should have made note of the missing data within the manuscript, and 2) that there are more optimal ways to handle the missing data. If we were testing these hypotheses today, we would indeed use a more sophisticated solution, such as variable imputation with a package like ‘mice’.

1. **Use of the Same Data for Variable Selection and Performance Testing**

In the key analyses reported, we used the full sample for both variable selection (using VSURF), and to evaluate the predictive performance of the model. This approach introduces a risk of overfitting because the same data were used for both steps. To address this concern, we also manually separated the data into testing and training sets. The reviewer notes that doing so effectively mitigated the overfitting concerns (and, indeed, it produced more conservative predictive estimates). However, if we were to conduct these analyses today, we would be expected to address the problem using current best practices such as the use of nested resampling (e.g., with k-fold cross-validation, which we used in the more recent paper Eastwick et al., 2023).

1. **Handling of Dependent Observations**

The data used in the current manuscript had a nested format, whereby each participant met with multiple potential dating partners. At the time, we knew of no technique for accounting for this non-independence within the context of a random forests model. The reviewer notes that today, blocked resampling (Roberts et al., 2017) would likely be used as a way to address this concern, although block resampling had not yet been popularized at the time the focal paper was published. We agree with this assessment.

**Conclusion**

In sum, we want to once again extend our thanks to Florian Pargent for reviewing our work. We feel grateful to have had a machine learning expert carefully review our analyses and resulting claims, particularly given the rapidly evolving landscape surrounding ML techniques. Knowing that the conclusions still hold up gives us peace of mind.

Sincerely,

Samantha Joel, Paul Eastwick, and Eli Finkel