Response to Reviewers for MS 2021-314, "Patterns of Panethnic Intermarriage in the United States, 1980-2018."

I thank the two reviewers for their close reading and helpful comments on the prior version of this manuscript. I have taken all of these comments seriously and attempted to address them in the revised manuscript. I believe the result is a much improved paper.

Before addressing the specific points of each reviewer below, I would like to note the major changes made to this revision that I will reference below:

- I have reworked the introduction to provide more justification for the current research and to outline the goals of the research.
- I have completely re-worked the section on "The Role of Birthplace and Language Endogamy." I have removed all of the theoretical discussion of boundary formation that was unhelpful and I have moved up the analysis of birthplace and language dissimilarity that originally was part of the results section to help motivate why these factors are so important to address.
- I have expanded the discussion of the model in the Data and Methods section to address some concerns and requests for clarification by both reviewers.
- I have created a subsection in Data and Methods named "Measuring Ethnoracial Exogamy" that goes into greater detail about how ethnoracial exogamy is measured and justifies the decisions that I have made in terms of coding. In addition, I have added a new table (Table 2 on p. 28) that gives a schematic representation of how ethnoracial exogamy is coded in the more parsimonious model. To remain within the limit on the number of figures and tables allowed, I have removed the MSA map.
- I have tidied up the language in terms of how I refer to "ethnoracial", "racial", and "ethnic" groups which I believe may have led to some confusion in the first draft. I have also explained my use of these terms in footnote 1 on p. 2.
- To address R2's concern about comparing across time when marriage market boundaries are shifting, I conducted a sensitivity analysis using states as the marriage market. The results of this sensitivity analysis are provided in the supplementary tables and figures and I reference them in the text. I discuss this issue in more detail below in response to R2's specific point.
- I have also made a variety of smaller changes to the text to address specific points below.
- Basic copy editing of grammar throughout the text to reduce bloat and remain within the word count limits.

Below I respond to each reviewer's points in detail. Reviewers' comments are shown in italics.

Reviewer 1

This is a very well written and well researched paper. I hope that the paper can be published in Demography after a major revision addressing the following issues.

Thank you for the kind words. I hope that the revisions made here are satisfactory.

The core analysis is a context-specific conditional logit model discussed by Zeng and Xie for interracial friendship (AJS 2008). It's an appropriate method, as it accounts for context-specific racial/ethnic composition.

Thank you for this note. I overlooked the Zeng and Xie article because of its focus on friendship nominations, but the model used in that article is indeed of the same form as the model I use here. I have included a citation to it on p. 10 of the revised manuscript. I have also included several citations to other work that has used this model to look at marriage patterns.

However, it is unclear how 20 counterfactual marriage partners are randomly drawn for each observed marriage. Is it conditional on the agent (one of the marriage partners) beyond the residence restriction? If it is, it is then a matching method akin to that of Xie and Dong (AJS 2021). If it is not, then comparison is far-fetched for most marriages.

The Xie and Dong article is an excellent article, but I believe the goals of that analysis are more ambitious than my goals here. Using techniques from the causal modeling literature, that paper uses a matching technique akin to propensity score matching to address a very specific question about status exchange. The goals of my paper are far more about describing trends and patterns in panethnic intermarriage. Thus, I do not restrict sampling in any way other than residence in the same marriage market. I do apply "model-based" restrictions in the form of parameters for age and education matching in all models to address the most "far-fetched" results. The resulting models are akin to more traditional log-linear models that estimate the odds of exogamy and endogamy across categorical combinations, but using a more flexible framework. Indeed, one of the model's features is the ability to easily account for multiple confounders (as I do for birthplace and language endogamy) which Xie and Dong note is "very challenging" with log-linear models (p. 1192). I also note that the sampling technique I use here is generally consistent with prior work using this method that I now cite.

Scale of the measurement concerns me. In the paper, the measurement of panethnicity marriage or interracial marriage (in odds ratios) is in reference to ethnic endogamy (less than 1). Thus, the results are sensitive to the variation in endogamy across ethnic groups. Endogamy is not the only choice. One may choose cross-racial marriage (saying marrying whites) as the reference (greater than 1), called "exogamy" by the author. That is, one may choose ethnic-specific white-exogamy or just all-exogamy as the reference. It may even be possible to combine the two references of endogamy and exogamy.

For example, the key finding reported in the paper "panethnic intermarriage is far more likely among Asians than among Latinos" (abstract) may be partly attributable to Asians' low endogamy (i.e., high interracial marriage with whites).

I believe the reviewer's concerns here are a result of insufficient description of the model in the initial manuscript. I have now substantially re-written and expanded my description of the model, the coding of ethnoracial exogamy terms, and the interpretation of coefficients. I hope that these improvements will address the reviewer's concerns.

I do use ethnoracial endogamy (e.g. black-black, Chinese-Chinese, Mexican-Mexican) as the reference category. The reviewer is correct that I could have used other reference points, but endogamy offers, in my view, the most satisfying interpretation. Because all non-endogamous combinations are fit with some dummy variable, all of the coefficients for these tables measure the odds of a particular form of exogamy relative to endogamy. As I note on p. 13 of the revised draft:

> Regardless of the specific model, the coefficient for each ethnoracial exogamy term has a similar interpretation. The exponentiated coefficient gives the ratio of the odds of a union between the two specified ethnoracial groups relative to the odds of ethnoracial endogamy. Values below one indicate that exogamy is less likely than endogamy. A relatively lower odds ratio indicates lower likelihood of this form of exogamy relative to other forms of exogamy. These odds ratios represent the likelihood of intermarriage net of group size differences in partner availability. These group size differences are accounted for by the sampling procedure which will draw alternate partners from different groups in proportion to their size in the designated marriage market.

These exogamy odds ratios should not be affected by "variation in endogamy" because there is no variation in endogamy apart from the inverse of the summation of different forms of exogamy which are already modeled. By using endogamy as the reference and measuring all off-diagonal cells as different forms of exogamy, I am explicitly making different forms of exogamy comparable to each other in strength. I hope that Table 2 of the revised draft which more explicitly shows the coding scheme will clarify this issue.

I am not sure I understand the reviewer's claim that "panethnic intermarriage is far more likely among Asians than among Latinos" (abstract) may be partly attributable to Asians' low endogamy (i.e., high interracial marriage with whites)." By definition, interracial marriage with whites is a form of exogamy, not endogamy. In actual fact, this form of exogamy is modeled in the paper and the odds of White-Asian exogamy are slightly lower than the odds of White-Latino exogamy (see Figure 3 of the revised manuscript). Note that these odds ratios account for group size differences and thus are much different than what one would estimate with crude outmarriage percentages, which generally show higher Asian exogamy rates because Asians are a smaller group.

I believe that the reviewer's concern here may reflect a similar concern expressed by R2 regarding how the exogamy of one group may affect the measure of exogamy to other groups. One of the features of this modeling approach (which it shares with log-linear models) is that the odds ratio of exogamy between any two groups is not constrained by the rate of exogamy to any other group, whether that be high or low. For example, it is theoretically possible for all of the ethnoracial exogamy odds ratios to be exactly one, in which case respondents would be selecting partners at random with regard to race.

To address this point more directly in the revised draft, I have added the following paragraph to p. 13:

> Each odds ratio is also unaffected by the degree of ethnoracial exogamy to other groups. For a given focal group, a high odds ratio of exogamy to one group does not entail that the odds ratio of exogamy to other groups must necessarily be low. Theoretically, for example, the odds ratio of exogamy to all outgroups could equal one, indicating no preference for endogamy and that partnering was conducted randomly with regard to ethnorace.

This is further complicated that Asian-White marriages are much more often (in fact, twice as often, see Xie and Goyette 2004, A Demographic Portrait of Asian Americans) for female-Asian and male-white marriages than for male-Asian and female-white marriages. This gender asymmetry should receive more attention than it is in the current version.

Reviewer 1 is absolutely correct that there is significant gender asymmetry in interracial marriage. Asian female/White male pairings are much more common than Asian male/White female pairings, as they note. White/black marriages have the opposite pattern. It is also true that I ignore this gender asymmetry in the models here by estimating a pooled parameter that basically averages across the two possible gender combinations for every form of exogamy. This was not well explained or defended in the first draft. The reasons however are straightforward. First, my primary goal here is to understand the tendency to panethnicity, not gender imbalance and so gender asymmetry is not particularly of interest. An average effect across gender combinations gives me a singular measure of social distance that I can use to compare different ethnoracial combinations. Second, given the number of terms under estimation here, accounting fully for gender asymmetry would both be computationally prohibitive and difficult to describe efficiently in the paper. My most expansive model includes 85 ethnoracial exogamy terms. That would have to be doubled to 170 terms to examine gender asymmetry and the results would be difficult to summarize given the sheer size and the fact that for many small groups, there would be little statistical power to confirm gender asymmetry. Furthermore, for the more parsimonious models that rely on a single panethnic exogamy term, there is no way to incorporate gender asymmetry.

In the revised draft, I have added the following paragraph on pp. 11-12 of the Data and Methods section to further explain and justify my choice to use gender symmetric terms:

In all models, patterns of ethnoracial exogamy, including panethnic intermarriage, are measured by a set of gender-symmetric dummy variables where the reference category is an ethnoracially endogamous union (e.g. a White-White or Chinese-Chinese marriage). Although substantial gender asymmetry exists in intermarriage for several important combinations (Xie and Goyette 2000; Gullickson 2006), the use of gender-symmetric terms more closely matches my goal of estimating social boundaries between groups. Such social boundaries are best measured by averaging across gender combinations, which is accomplished by gender symmetric terms. Furthermore, given the large number of parameters involved in some models, gender asymmetric terms would be impossible to fit in many cases.

Reviewer 2

The author does not make a particularly compelling case for this study, finding similar patterns of endogamy as those reported in earlier studies by Qian et al. (Demography 2012).

Unsurprisingly, I disagree with the reviewer about the value of this study. However, I do recognize that I did not do enough in the first draft to make a compelling case for it's value. In this draft, I have significantly reworked the introduction of the paper to much more clearly lay out that case on pp. 3-4. In short, the improvements of this paper are both in terms of the modeling strategy which can address shortcomings of prior work and the data which are better because they are (a) newer; (b) can identify a more broad array of Asian and Latino ethnic groups, including multiple South Asian ethnic groups, and (c) provide information on marriage timing which is crucial both to capture incidence rather than prevalence and to address the "immigrants married abroad" problem.

In terms of the modeling strategy, one of the most important contributions of this study is to more effectively address the confounding issues of birthplace and language endogamy on estimates of panethnic intermarriage. To help bring this issue out more clearly, I have now included an entire paragraph in the introduction on p. 3 that reads:

> Furthermore, prior research has paid insufficient attention to the role of birthplace and language endogamy in shaping patterns of panethnic intermarriage. Generally, individuals prefer partners from the same birthplace and who speak the same language. The considerable diversity in birthplace and language both within and between Asian and Latino ethnic groups has complex effects on the tendency toward panethnic intermarriage. Prior work has largely tried to address this issue indirectly by including an immigrant-native comparison, but this comparison does not adequately capture the complexity of birthplace and language diversity.

I have also completely re-written the section on "The Role of Birthplace and Language Endogamy" to highlight the complex confounding role that birthplace and language endogamy play on estimates of panethnic intermarriage.

Relatedly, the reviewer also notes later that they did not care for the "refresh" term to describe the contribution of the paper. I agree that this was probably not the best choice of words and my revised introduction removes that term amid the other changes. However, I would disagree with the reviewer that providing results with more contemporary data is not of significant value in and of itself even if those results largely confirm prior work. The Qian et. al. paper that the reviewer references uses Census 2000 data on four ethnic groups. I think many readers of *Demography* would like to see more contemporary estimates of panethnic intermarriage on a broader array of ethnic groups.

What's potentially new and different are results showing that East/Southeast Asians and South Asians have low rates of intermarriage with each other. The author was unable to "explain" this result, only saying that these patterns are not due to differences in birthplace and language endogamy. In the end, the author speculates, based on previous studies, that South Asians are phenotypically different from other Asian populations, which reduces panethnicity as measured by intermarriage (i.e., skin color matters). Religious differences may also play a role. These are plausible explanations but cannot be established with census-based data products. The author also claims that "only Qian et al. (2001) have analyzed panethnic intermarriage using data that includes South Asians." This is technically accurate (because of the groupings used in the paper), but also a bit misleading. For example, in one recent example using ACS data, Lichter et al. (The ANNALS 2015) showed that Asian Indian immigrants had the highest incidence of in-marriage. Nearly two-thirds of Asian Indian men and a quarter of Asian Indian women married their foreign-born co-nationals. Marriages with other Asians were very low—only about 5-10 percent. It is well known in the demographic literature that South Asians in general and Asian Indians in particular display much different marriage patterns than other Asian groups. Again, the authors will need to clearly differentiate their findings and conclusions from previous studies.

Thank you for the citation to the Litchter et. al. paper which I have now included in the revised manuscript on p. 6.

On the larger issues being addressed here, I believe the inclusion of South Asian ethnic groups allows me to make two important contributions. One of those is to establish the degree of intermarriage between East/Southeast Asian and South asian panethnic groups, as the reviewer notes. The reviewer also is correct that some prior work has been able to establish this to some degree in the case of Asian Indians. My results are consistent with that work in showing a low rate of intermarriage between these groups. They are also an improvement over prior work in that my estimates are not limited to Asian Indians but rather include several South Asian nationality groups.

The second contribution is that I am able to estimate the strength of panethnicity among South Asian ethnic groups themselves (at least in the more recent ACS data). This has not been possible in prior work because that work depends on a single South Asian ethnic group (i.e. Asian Indians). What I show here is that the degree of panethnicity among South Asian ethnic groups is high and similar in magnitude to that between East/Southeast Asian ethnic groups. When these two findings are combined, the conclusion is that we see two distinct "melting pots"

of panethnic affinity between these groups. I have highlighted this finding in both the abstract and conclusion of the revised draft.

I believe the prior draft was somewhat unclear on these contributions in the front end. I have made edits in several places to help address this issue. Most notably, I have changed the last paragraph of the "Panethnicity in Marriage" section on p. 6, which addresses the South Asian issue, to read as follows (changed part is highlighted in yellow):

> Related to this issue, most prior work has not addressed the issue of South Asian panethnic intermarriage. The placement of South Asians within the panethnic category of Asian is contested at best. Individuals from the Indian subcontinent are treated as part of the panethnic Asian group by governmental agencies such as the Census Bureau, but research has shown that in everyday practice they are generally treated as "ambiguous non-whites" (Kibria 1996; Morning 2001; Schachter 2014). Consistent with this view, prior work on intermarriage among Asian Indians has shown that this group has relatively low rates of intermarriage with other (non-South) Asian populations (Qian et al. 2001; Lichter et al. 2015). However, due to sample size limitations, Asian Indians are the only South Asian population included in prior analyses, making it impossible to explore the extent of panethnicity among South Asian ethnic groups.

The reviewer also seems bothered by my lack of explanation for these differences. My goals here are descriptive and I lean on prior theoretical and empirical work in terms of potential explanations for difference. Although I am not formally testing explanations, I do note in the conclusion that my results are more consistent with a racialization perspective on panethnic formation than one which emphasizes cultural similarity.

The trend data also are not especially convincing, in part because the study includes only 2 data points separated by more than 35 years. In my opinion, trends are not observable with only two data points. In fact, it could be that patterns of among Asians increased during the 1980s but then declined over the subsequent 25 year period. We just don't know how trends are fluctuating over the last several decades.

I take the reviewer's point that describing the change across two time points as "trends" is potentially misleading. Unfortunately, it is not possible to estimate many intermediate points here, because most of the data for the intervening time points lacks information on marriage timing. Nonetheless, I still think the comparison of these two time points is important and useful. I have removed reference to this as "trends" in the revised draft and note the following in footnote 4 on p. 15:

> The analysis of change over time relies upon two time points where the data are sufficient to identify recent marriages. It is possible that patterns of ethnoracial exogamy changed in non-linear ways in the interim between these periods.

This interpretative problem is compounded by the fact that the racial and ethnic data are not strictly comparable between 1980 and 2014-2015 and the spatial units from which "choice sets" are drawn are not the same. It is not clear how this might affect the marital opportunities of Asians or Latinos (and their various nationalities). At a minimum, it would be good to know the percentage shares of all U.S. Asians and Latinos that are actually included in the metropolitan areas identified in this study. Whether it is 50 percent or 90 percent matters when drawing inferences about changing patterns of ethnic intermarriage.

I should clarify that respondents are included in both time periods regardless of whether they were in an identified metropolitan area. If they are not in an identifiable metropolitan area, then they are given a "residual" marriage market of the state of residence. As the reviewer notes, this approach may cause some problems for comparisons across time because the metropolitan areas that are identifiable are not consistent across time. To address this issue, I conducted a sensitivity analysis in which I used state as the marriage market for all respondents. The results of these models are provided in the supplementary tables and figures. I also show the key Figure S.1 below which illustrates the change over time in all ethnoracial exogamy parameters, using the two different definitions of the marriage market. This figure is analogous to Figure 3 from the manuscript, but with two panels to compare the two different marriage market definitions. Although there are a few slight differences, the results are remarkably close across the two panels. The substantive conclusions would not be altered by shifting to a state-based marriage market definition.
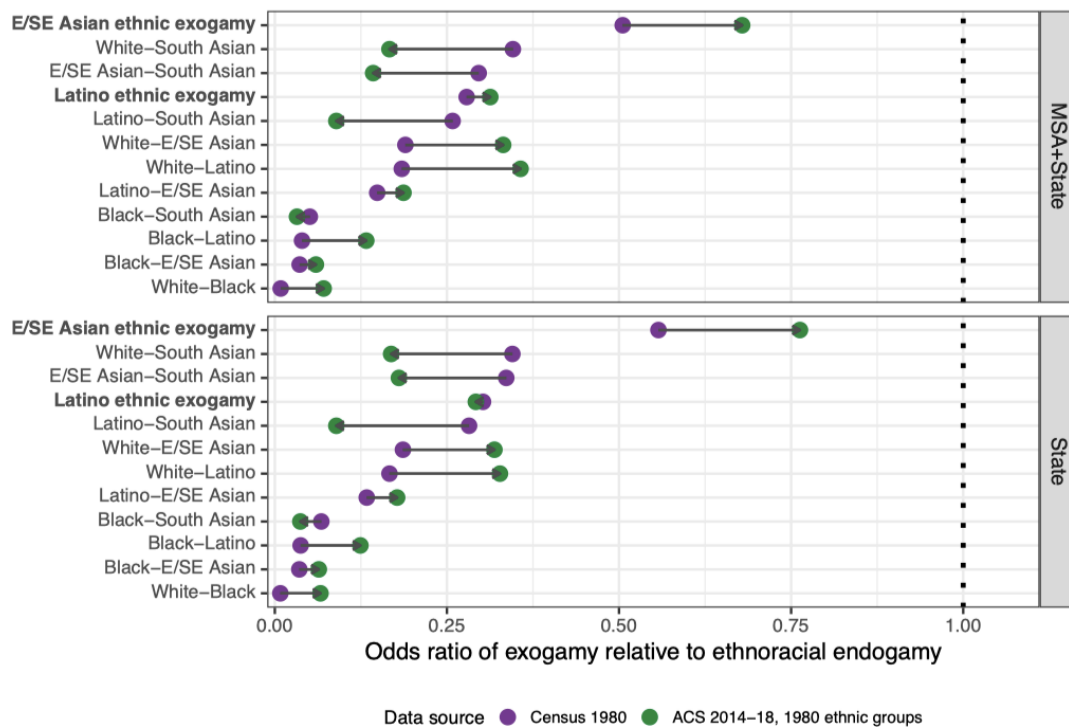
Figure S.1. Odds of ethnoracial exogamy relative to endogamy across two time periods. The panels compare the cases where marriage markets are defined either by a combination of MSA and state or by state alone. Results are based on models that control for age differences, educational differences, and birthplace and language endogamy. Values are sorted by ethnoracial exogamy in 1980. Arrows show the change across the two time periods based on comparable sets of ethnic groups. Results for American Indian/Alaska Native intermarriage are excluded due to sampling variability.

Other issues

1. The author should indicate the basic and specific objectives of this study and identify its main contributions to the literature much earlier in the paper. Simply reporting that "I refresh our understanding of panethnic intermarriage using recent data from the 2014-2018 American Community Survey" is not enough, nor is it enough to say that you use a counterfactual approach or control for a few additional variables in the models. How do these technical issues fix or adjudicate (competing) claims in the literature?

I agree that the contributions of this paper could have been more clearly stated in the first draft. As noted above, I have substantially revised the introduction in the revised draft to more clearly lay out the specific objectives and contributions of the study and I have removed the "refresh" language.

2. The author claims that the results have important implications for the future salience of panethnic categories. If so, the paper should include some additional discussion about why this is so. Much of the introduction seems to be focused generically on panethnicity or on particular

technical issues (e.g., using log-linear model) rather than on describing exactly why the author expects differences in panethnic marriages among Asian and Latino populations, or why panethnic marriages may have increased or decreased. The introduction is without any hypotheses of this sort. What does the author expect to find, and why?

As I now note more clearly in the introduction, I am trying to document the strength of panethnicity among Asians and Latinos (and change over time) while accounting for important issues such as birthplace and language endogamy as well as other issues that have plagued prior work. The goals of the paper are thus descriptive and formal hypotheses would be inappropriate in my view. Expectations are driven by the discussion of prior work in the front end of the paper.

3. It is never completely clear in the paper how outmarriage to Whites or Blacks are treated in the paper. For example, if Asians experience changing shares of outmarriage with other Asian co-ethnics, then questions arise about whether this is due to changes in outmarriage to these other groups. Quite mechanically, panethnic marriages may decline if there are growing shares of Asians or Hispanics who outmarry with Whites or Blacks. As a starter, how are Whites and Blacks treated in evaluating Asian or Hispanic exogamy in figure 3?

As I note at the top of this memo, I have now expanded the discussion of how ethnoracial exogamy is modeled and I have also included a table (Table 2 on p. 28) that shows schematically how this coding works. One feature of this modeling approach is that the strength of one form of exogamy places no necessary mechanical limit on the strength of other forms of exogamy. That is because the model is not estimating "shares" but rather odds ratios of actual unions from a pool of potential unions. This feature of the model is similar to that for log-linear models which may be more familiar to the reviewer. Please see my response to R1 above for more details regarding this issue.

It is still possible for panethnic marriages to decline in relation to outmarriage with whites or blacks. We see evidence of this on Figure 3 for the case of Latino/White intermarriage. However, this is an important empirical finding, not a mechanical artifact.

4. Relatedly, the language of ethnic exogamy is sometime unclear. Does this mean that Indians, for example, are marrying out to other Asians or that they are outmarrying with non-Asians? It also seems that endogamy sometimes refers to national origin groups and other times to panethnic groups. This make it difficult to fully understand or appreciate the more nuanced results.

I agree that my use of language in describing exogamy terms was somewhat inconsistent in the original draft. I have cleaned up this language in the revised draft. I have also sought to clarify terminology in two other ways in this draft. First, I have included footnote 1 on p. 2 that explains my use of terms. It reads:

The distinction between a racial and an ethnic group is often fuzzy and debated among scholars (Brubaker 2009; Valdez and Golash-Boza 2017; Lewis and Forman 2017). For analytical clarity, I use the term "ethnoracial" to refer to any group that may be identified either along racial or ethnic lines in popular practice, "racial group" to refer to the five major groups of White, Black, Indigenous, Asian, and Latino that constitute the highest level in the nested hierarchy of ethnoracial differences, and "ethnic group" to refer to different sub-populations among Asians and Latinos that are primarily defined in terms of national origin, such as Chinese, Korean, Mexican, and Colombian. Ethnic differentiation within the same national origin group exists as well, but is largely unmeasurable in the data that I use here.

Second, Table 2 on p. 28 also helps to address this issue by showing schematically how ethnoracial exogamy is coded.

I hope these changes will clear up the reviewer's confusion on this point. In direct response to the question above, because Indians belong to the South Asian panethnic category, they would be "ethnically exogamous" when marrying someone from another South Asian category (e.g. Pakistani). If they married someone from an East/Southeast Asian background, then this would be E/SE Asian-S Asian exogamy according to the modeling strategy.

More broadly, endogamy is in relation to the specific ethnoracial category. For some groups like Whites and Blacks, this is given by a broad racial group, but for Asians and Latinos, this is given by a specific ethnic category.

5. As I understand the pooling of different national origin groups, it seems that the author has already lumped different national origin groups into panethnic categories (i.e., South Asians, East/SE Asians and Hispanics). What is the empirical justification for making these groupings a priori? In fact, later the author looks at national origin group-specific affinities (p. 17) using the ACS and finds that social distance remains between East Asian groups (e.g., Chinese and Japanese) and the Vietnamese and Filipino ethnic groups (e.g., other East Asian populations). This entire section (p. 17-20) has a wandering character that is difficult to reconcile with the earlier results reported for the 3 panethnic groups.

I have reworked the Data and Methods section to better clarify how and when this grouping is done. I feel that the grouping of Asians and Latinos requires no justification here as the preceding discussion in the front end has made clear that these are the two categories of interest to this research, and are perfectly consistent with prior work. In terms of the separation of South Asians and East/Southeast Asians, I have added the following sentence on p. 11:

> I separate Asians into two separate panethnic blocs because prior work suggests social distance between these groups (Kibria 1996; Morning 2001; Schachter 2014). I estimate the likelihood of intermarriage between these two blocs in all models to test whether this separation is justified empirically.

Subsequently, on p. 17 of the revised draft, I note from the results:

> In the more recent ACS data, the odds of an East/Southeast Asian-South Asian intermarriage were only 14% as high as the odds of ethnic endogamy, making it one of the less likely forms of intermarriage and justifying the decision to separate out these panethnic blocs empirically.

I have also clarified in this section the two different approaches that I am using to model panethnic intermarriage and I further outline these two approaches at the beginning of the results section on p. 15. The first approach is the most parsimonious approach which allows me to make comparisons over time and to succinctly summarize the tendency toward panethnicity across multiple ethnic groups. The second approach then looks more closely at heterogeneity in panethnic interrmarriage across specific combinations which captures some important nuances at the cost of parsimony, and is also limited to the largest groups in the most recent time period. I believe that by clearly describing these two different approaches before proceeding to the results, the results section has less of a "wandering character." I thank the reviewer for pushing me to be more clear here.

5. The author makes this claim: "Rather than treat panethnicity as a form of assimilation/acculturation, I treat it as a form of ethnoracial boundary expansion." But how in practice does this conceptual claim actual influence the modeling approach used in the paper? I can't tell that it does.

I was attempting to contrast two different theoretical approaches to understanding panethnicity. However, in retrospect, the attempt here was too abbreviated to be useful and did not really advance the arguments that I am making in the paper. I have removed this entire discussion, which now focuses much more specifically on the empirical challenge of incorporating birthplace and language endogamy into our understanding of panethnic intermarriage. I think this results in a much cleaner presentation of this issue and I thank the reviewer for bringing this to my attention.