sociological science

Supplement to:

Pessin, Léa, Kevin Munger, 2026. “Beyond Text: Using AI-Generated Visual Conjoints to Study Gender and Housework Attribution” Sociological Science 13: 661-684.

ONLINE APPENDIX

A Description of Image Generation Process

We experimented with several cutting-edge generative AI models to create the experimental stimuli. At the time we were designing the experiment, we found Gemini 3 to have the most desirable properties: high-quality and replicable images, rapid iteration, and zero cost (at our usage rates). This is of course a rapidly evolving space and we recommend authors do their own experimentation.

The primary issue that we encountered was spillovers between facets. For example, when we attempted to generate two identical living rooms in which one was high SES and the other low SES, the latter image generated was much darker than the former – not a relevant dimension of variation. The layout of the rooms was also much different in ways that seemed likely to distract the respondent. Our solution to this problem was to iterate with the generation of the images, adding more specific details each time we discovered some unintended variation introduced by the generative models. Figure 8 displays one of our example prompts, in this case for the high-SES kitchen with kids mess.

After many rounds of iteration on the images of the room, we asked three LLMs (ChatGPT 5, Gemini 3, and Claude Sonnet 4.5) to evaluate the images of the rooms in terms of messiness, SES, and the likelihood that a child lived in the house, on a scale of 1-10. For images in which the desired value different from the average AI-evaluated value for that type of image, we re-generated the image until the variance within image types as labelled by the three models was low.

We performed a similar procedure for the images of the occupants. The two dimensions that we tested were professionalism and attractiveness.

```
Create a photorealistic interior photograph of a suburban kitchen.

- Stainless fridge; stainless glass-top range; stainless microwave in upper cabinet; stainless dishwasher panel beneath
right side of sink.
- Medium-tone wood plank floor; subway-tile backsplash; light quartz-like countertops.
- Shaker/flat-panel cabinets with clean hardware.
- Fixed props (positions don't change): coffee maker to the right of the sink; vase with sunflowers to the left of the sink,
wood table in the room.

Single soft window light only. Curtains "sheer." No ceiling lights. White balance fixed at 4200K.
Exposure targets: countertop ≈ 45% luminance; white highlights ≈ 85–90%; deep cabinet shadows ≈ 10–15%.
No hard sun patches. Same tone mapping/contrast as every other image.

Camera:
Tripod. 38mm full-frame, f/8. Camera height aligned with drawer-pull line under the counter.
Frame with the window perfectly centered; the left fridge edge and right oven edge just inside the frame. No tilt or roll.

add the following objects. make them look  MESSY, IN PILES on top of each other, CHAOTIC. DO NOT spread them evenly across
surfaces:

On floor:  ring-stackers,  soft blocks, open picture books,  stuffed animals,  and other toys, action figures, dinosaurs,
legos, balls, kites, blankets, children's clothes

On the table: dozens of crayons and coloring books

On the countertop: sippy cups, pacifier

If I asked you to rate how messy this room would look from 1-10, it should be an 8 (very messy)
```

Figure 8: Example Prompt

B Marginal Means Analysis

We present here the marginal means analysis for the two outcomes which are not presented in the body of the manuscript.

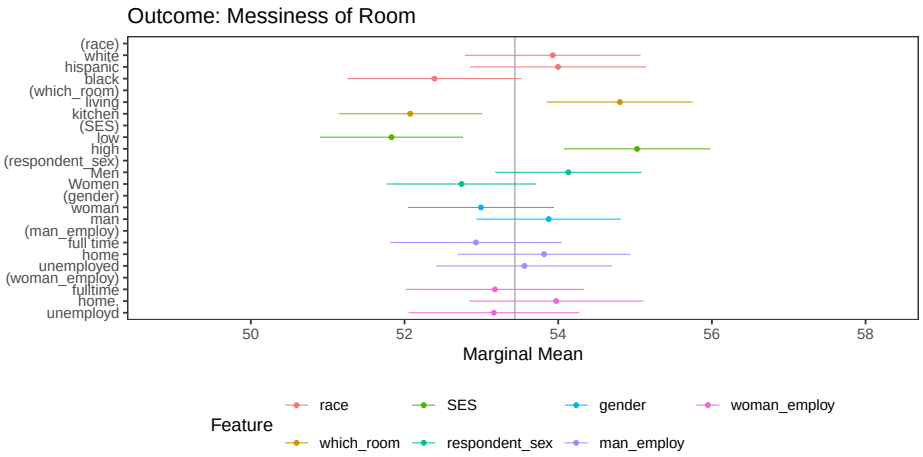

Figure 9: Marginal means for all factors predicting perceived messiness of the room. Marginal means represent the average predicted messiness level for each factor level, marginalizing over all other factors. The vertical line indicates the overall mean. N=14,810.

C Interaction Effects Between Facets

C.1 Perceived Messiness

We tested for interactions between occupant gender and several other factors: race/ethnicity (Figure 11), room type (Figure 12), and SES (Figure 13). None of these interactions reached conventional levels of statistical significance, providing little evidence that gender effects on perceptions vary by these contextual factors.

C.2 Social Consequences

We tested interactions between gender and race/ethnicity (Figure 14), room type (Figure 15), SES (Figure 16), and source of mess (Figure 17). Consistent with the perception results, none reached statistical significance.

D Results with Adaptive Shrinkage Estimator

Following our pre-registration, we also include the AMCE results calculated with the adaptive shrinkage estimator proposed by Liu and Shiraito (2023). Each of Figures 3, 4, 5 and 6 from

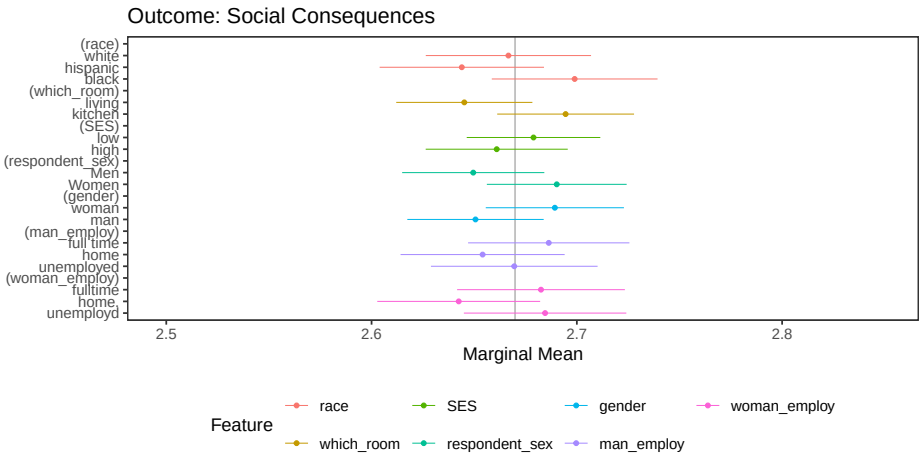

Figure 10: Marginal means for all factors predicting social consequences for the occupant. Marginal means represent the average predicted social consequences level for each factor level, marginalizing over all other factors. The vertical line indicates the overall mean. N=14,810.

the main text are replicated here with this procedure applied. We discussed any discrepancies in inference between the

E Full Set of Stimulus Images

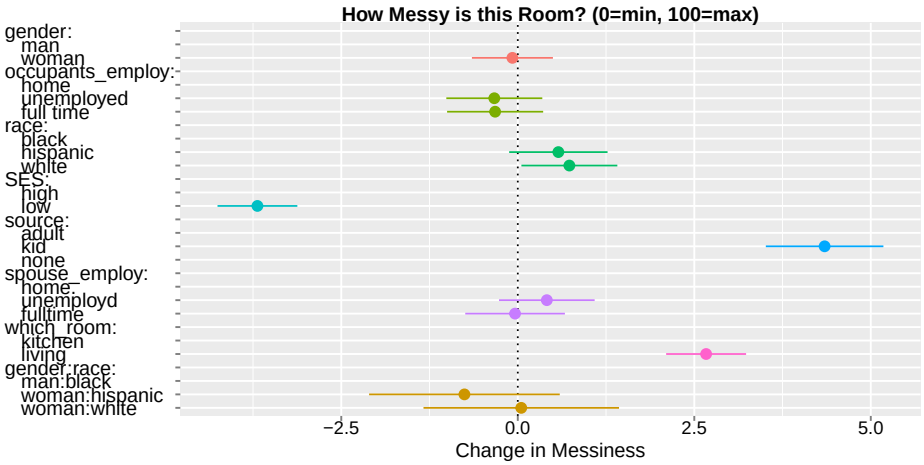

Figure 11: Interaction between occupant gender and race/ethnicity on perceived messiness. N=14,970.

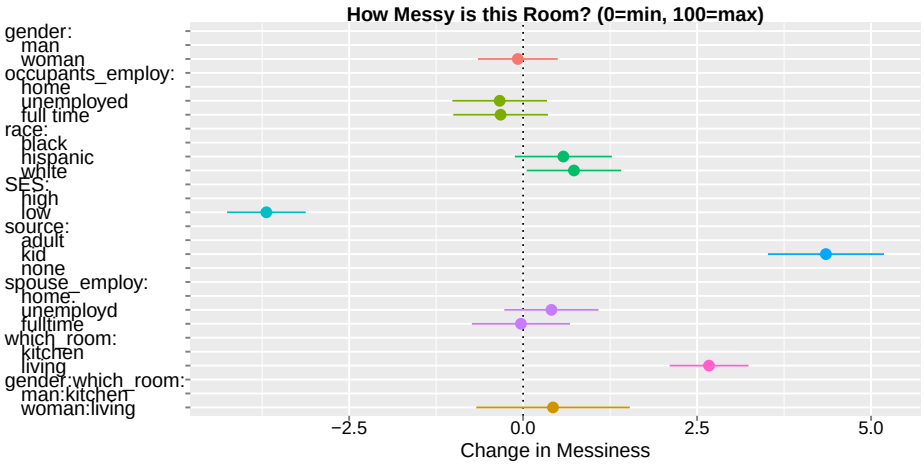

Figure 12: Interaction between occupant gender and room type on perceived messiness. N=14,970.

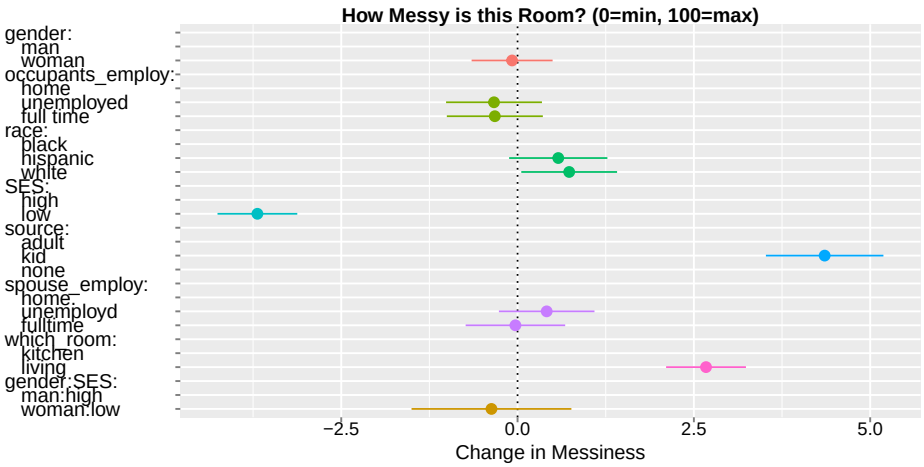

Figure 13: Interaction between occupant gender and SES on perceived messiness. N=14,970.

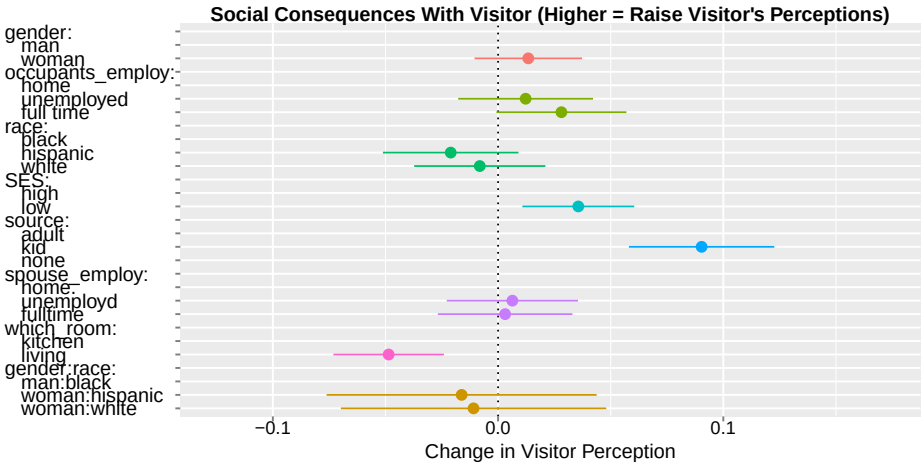

Figure 14: Interaction between occupant gender and race/ethnicity on social consequences. N=14,970.

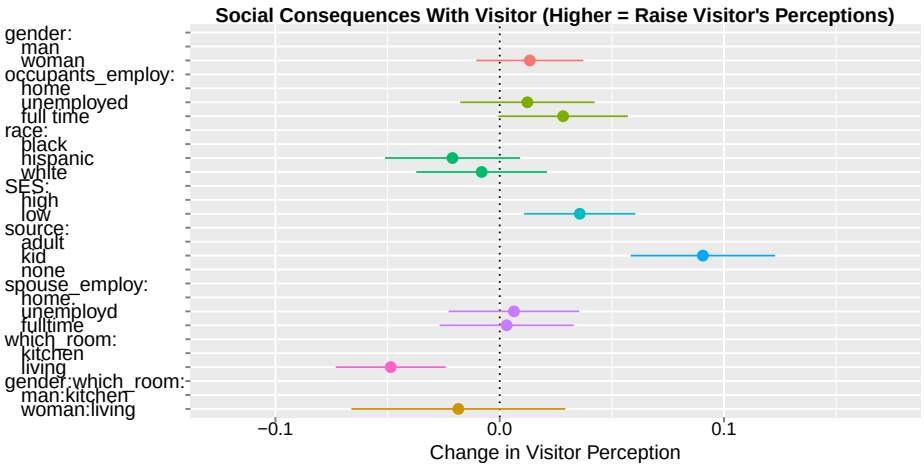

Figure 15: Interaction between occupant gender and room type on social consequences. N=14,970.

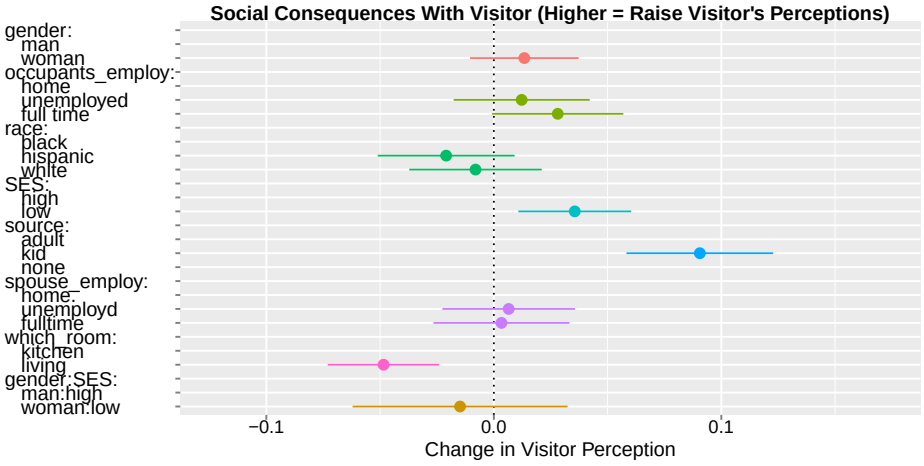

Figure 16: Interaction between occupant gender and SES on social consequences. N=14,970.

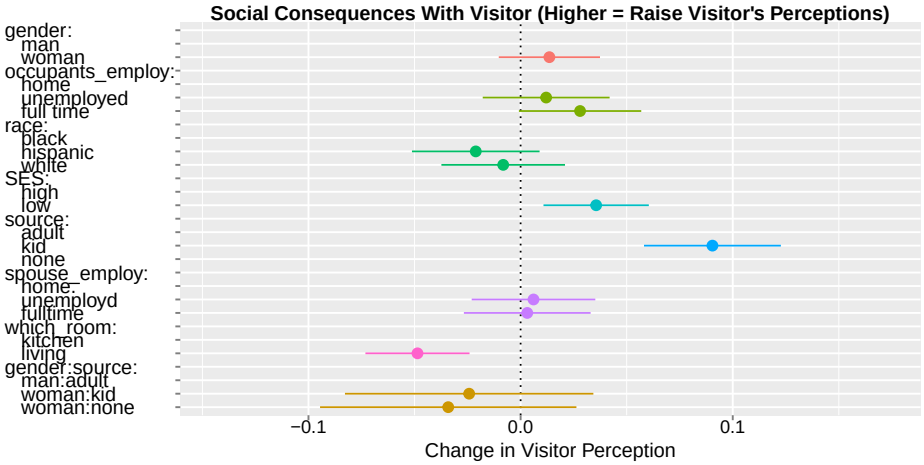

Figure 17: Interaction between occupant gender and source of mess on social consequences. N=14,970.

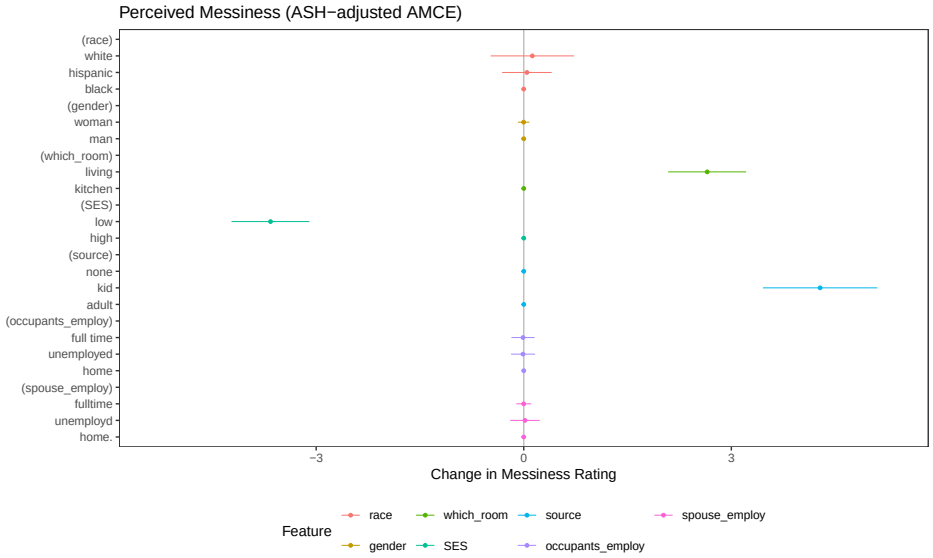

Figure 18: Average marginal component effects on perceived messiness (0-100 scale). Points represent point estimates with 95% confidence intervals. Reference categories are the first listed below the facet name. The room cleanliness factor is excluded from visualization but included in the regression model. Standard errors clustered at the respondent level. Adaptive shrinkage estimator applied. N=14,970.

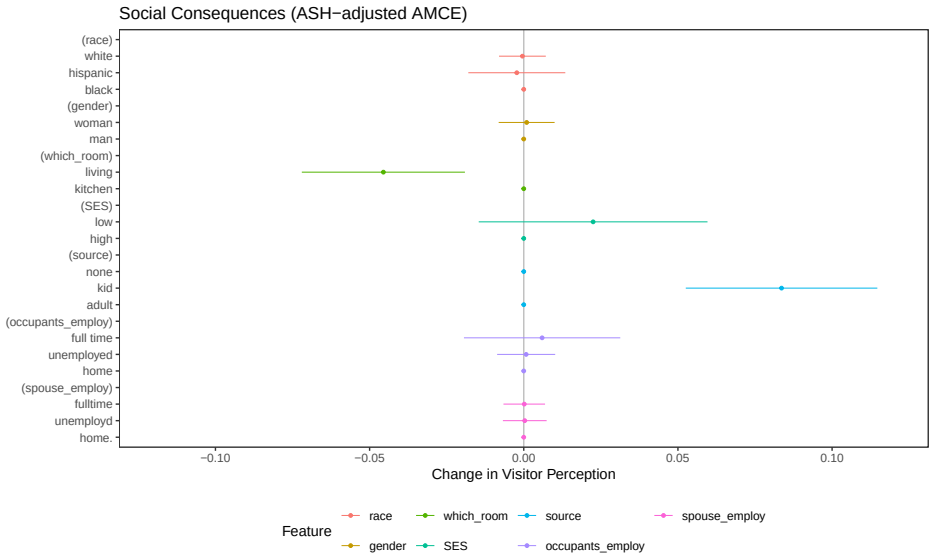

Figure 19: Average marginal component effects on anticipated social consequences (5-point scale from much more negative to much more positive visitor opinion). Points represent point estimates with 95% confidence intervals. Reference categories are the first listed below the facet name. The room cleanliness factor is excluded from visualization but included in the regression model. Standard errors clustered at the respondent level. Adaptive shrinkage estimator applied. N=14,970.

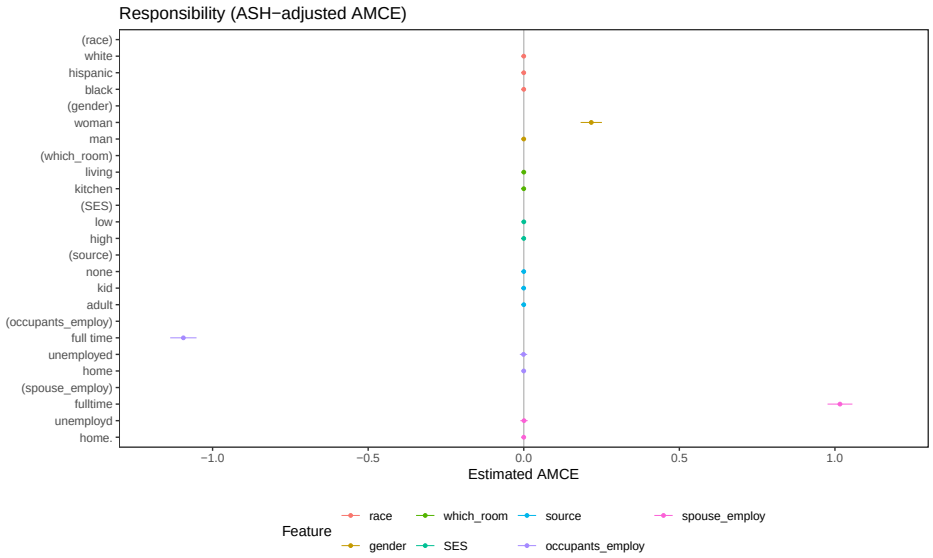

Figure 20: Average marginal component effects on responsibility attribution (5-point scale from pictured person's spouse to pictured person). Higher values indicate more responsibility assigned to the pictured person. Points represent point estimates with 95% confidence intervals. Reference categories are the first listed below the facet name. The room cleanliness factor is excluded from visualization but included in the regression model. Standard errors clustered at the respondent level. Adaptive shrinkage estimator applied. N=14,970.

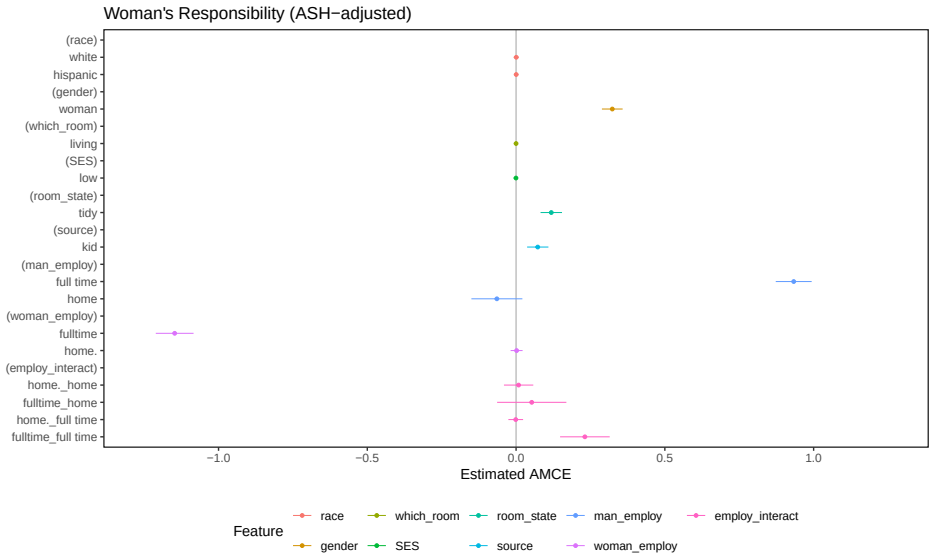

Figure 21: Average marginal component effects on responsibility attribution (5-point scale from men's responsibility (1) to woman's responsibility (5)). Points represent point estimates with 95% confidence intervals. Reference categories are the first listed below the facet name. The room cleanliness factor is excluded from visualization but included in the regression model. Standard errors clustered at the respondent level. Adaptive shrinkage estimator applied. N=14,970.

Figure 22: Kitchen, High SES, Adult Mess

Figure 23: Kitchen, High SES, Kid Mess

Figure 24: Kitchen, High SES, Tidy

Figure 25: Kitchen, Low SES, Adult Mess

Figure 26: Kitchen, Low SES, Kid Mess

Figure 27: Kitchen, Low SES, Tidy

Figure 28: Living Room, High SES, Adult Mess

Figure 29: Living Room, High SES, Kid Mess

Figure 30: Living Room, High SES, Tidy

Figure 31: Living Room, Low SES, Adult Mess

Figure 32: Living Room, Low SES, Kid Mess

Figure 33: Living Room, Low SES, Tidy