Which Farm Animal Photos are Most Likely to Inspire People to Eat Vegan: A Reanalysis

Ryan Melvin and Jacob Peacock

Humane League Labs Report E005R06

May 4, 2018

Abstract

This document presents a reanalysis of the data and the conclusions reported by Humane League Labs on January 25, 2015, in the blog post titled Report: Which farm animal photos are most likely to inspire people to eat vegan? Using regression analysis of the provided data, our reanalysis finds that photos of sick or injured animals, large groups of animals, pigs, and mother and baby animals are most likely to increase self-reported desire to stop consuming animal products. However, in the course of our reanalysis, we found the provided data likely represents only a subset of all respondents to the survey. As this may impart significant bias, substantial uncertainty is warranted around these conclusions.

This research earned the Open Data badge for open science practices. All data is available on the Open Science Framework repository at `https://osf.io/k67ge/`.

DOI 10.17605/OSF.IO/K67GE

Contents

1 Introduction

This document presents a reanalysis of the data and conclusions described by Humane League Labs in the blog post titled "Report: Which farm animal photos are most likely to inspire people to eat vegan?" [1] and the accompanying report [2]. The data was generated as part of a study concluded in 2015 which sought to understand the effect of images used in animal advocacy literature on inspiring reduced consumption of animal products. This reanalysis is intended to serve two purposes: first, to address any statistical and numerical oversights and errors in the original analysis, and second, to identify any methodological limitations of the previously conducted study and clarify the range of possible inferences given the observations.

2 Methods

In keeping with our commitment to open code and data, the Python and R code used to produce this analysis can be found on the Humane League Labs Open Science Framework account at `https://osf.io/k67ge/`. All references to files, folders and paths herein refer to the files hosted at this address.

The specific survey questions analyzed here were administered as part of a larger survey on vegan activism and dietary habits. The full survey is described in [3] and the following is excerpted from that report. Briefly, survey participants were recruited from the crowdsourced marketplace Mechanical Turk, the social media site Reddit, and the classified advertisement site Craigslist. The survey itself was administered on another platform, although we do not know which platform, and ran from approximately July 22 to August 8, 2014. At least some respondents were given a code to redeem on Mechanical Turk, presumably to collect an incentive of some sort.

Respondents were shown ten photographs depicting farm animals in a range of conditions from happy, idyllic, grassy settings to injured or deceased animals confined to a factory farm. All photographs are available in the original report [2]. Photographs were selected to be representative of animal advocacy literature and included photographs from Mercy for Animals, Animal Equality, The Humane Society of the United States and Compassion Over Killing as well as generic stock photographs. The photographs were randomly selected from ten sets of three photographs, one photograph drawn from each set, for a total of 30 possible photographs. Respondents were asked three questions about each photograph and the animals depicted therein, with responses on a 10-point Likert scale:

1. How much does this photo make you want to STOP eating animal products like poultry, red meat, fish, and eggs?
    - 1 (Not at all), 2, 3, 4, 5 (Somewhat), 6, 7, 8, 9, 10 (Extremely)
2. How much do you think the animal(s) in this picture suffered?

    - 1 (Did not suffer at all), 2, 3, 4, 5 (Suffered somewhat), 6, 7, 8, 9, 10 (Suffered extremely)

3. How human-like do you think the animal(s) in this picture is (are)?

    - 1 (Not at all human-like), 2, 3, 4, 5 (Somewhat human-like), 6, 7, 8, 9, 10 (Extremely human-like)

The survey responses can be found in `data/raw/omnibus_data_anon.csv`. To facilitate more detailed analysis of the responses, each photograph was coded with four factors taking the following levels:

- Age: Baby, Mom/Baby, Adult
- Grouping: Individual, Small group, Large group
- Animal: Cows, Poultry, Pigs, Goats/sheep, Fish
- Category: Small cage/confinement, Happy, Farm overview, Dead/killing, Sick/injured

For example, a photograph of a mother and baby pig laying in a field would be coded with the values: Mom/Baby, Small group, Pigs and Happy, respectively. The values of these attributes for each photograph can be found in `data/raw/images_attributes.csv`.

3 Assessment of methodology

The exact methodology of this study is unclear, which makes it difficult to assess. If different incentives were provided for participants from different sources, this could represent a significant source of bias. The inclusion of this study as one part of a larger study also introduces opportunities for priming and other bias from the early questions. The study's design is not optimized to evaluate the effects of the four factors of interest because the number of images in each category is small and sometimes zero. For example, there are no photographs representing sick/injured fish. This design produces highly uncertain estimates of the effects of different levels of each factor. Furthermore, the study design does not include an awareness check or control that might detect insincere or unreasonable answers. Lastly, the primary outcome of this study is a self-report of desire for dietary change, which may not be an accurate proxy for actual diet change given the likely social desirability bias present in the study. Thus, from the results, we cannot draw conclusions on which photographs affect dietary change but only conclude on the reported desire to change.

4 Verification of numerical claims

We attempted to verify all numerical claims in the main body of the original report, namely tables 2–9. The eight tables consider the effects of each of the four factors describing the photographs (age, grouping, animal, category) on responses to how much a photograph motivated the viewer to reduce animal product consumption (Question 1). This analysis used both descriptive and inferential approaches. In the descriptive approach, the responses to photographs corresponding to each level of the factor (for example, photographs depicting pigs) were aggregated and the mean and standard deviation calculated. In the inferential approach, the means of each pair of levels were compared for each factor using a 2-sample t-test. To account for multiple comparisons, the Bonferroni correction was used to set the significance threshold for each factor, based off a significance threshold of 0.05. Responses to Questions 2 and 3 were not analyzed extensively.

In general, we could not reproduce the original numerical claims. For the descriptive approach, we found agreement usually within a few percent for the means and standard deviations calculated for the levels of each factor (Tables 2, 4, 6, 8). We consider the source of these small deviations below. For the pairwise comparisons of levels within each factor, we found similar results for the difference in means but significant disagreements for the pooled standard deviations and thus the Cohen's d effect sizes (Tables 3, 5, 7, 9). The source of these large disagreements is not clear and may be attributable to a computational error.

In the course of our reanalysis, we found the provided data set is likely only a subset of all survey respondents. Two pieces of evidence suggest this is the case. First, four columns in the survey results show great uniformity which suggests the provided data was filtered to exclude some respondents: (1) a column labelled "Include" is set to "Include in overall analysis" for all responses; (2) a column indicating whether respondents were from the US or Canada or elsewhere ("Q1_Country") showed only respondents from the USA or Canada; (3) a column indicating whether respondents completed the survey in under 7 minutes and 30 seconds ("TimeToComplete") showed only respondents completing the survey in more than the specified time; and (4) a column with responses on dietary pattern ("Q99_SelfCat") included no vegans or vegetarians. Taken together, this data strongly suggests some respondents were excluded based on their answers to these questions and time taken to complete the survey.

Second, images in the original blog post show a few examples of photographs used in the study accompanied by a number labeled "viewed" and a percentage in parenthesis [1]. In preparing this reanalysis, images like these were found for all photographs used in the study and the values tabulated in `data/raw/screenshots.csv`. Recalling the randomization process for the questions in this survey, the percentages are found to represent the fraction of respondents randomly assigned to each photograph in a given question. Thus, summing the percentages of sets of three consecutive images yields 100% and summing the "viewed" numbers gives the number of viewers per question. The number of viewers for the first question is 1,205 and declines with each subsequent question to 1,174 viewers for the tenth question. This pattern is consistent with attrition over the course of the survey. We also see this pattern in the food photo questions that directly precede this section of the survey. This again suggests there were more respondents to the survey than represented in the provided data.

Given the provided data seems to represent only a subset of respondents, it seems feasible the original report filtered the respondents to analyze a different subset than that provided in the data. Supporting this notion, while the descriptive statistics across levels of factors differ, the differences are generally small, suggesting the distributions of the original results are similar to those of the current analysis. We tested this using a series of two-sided one-sample t-tests to compare the means of the levels of each factor to those described in the original report. These tests found no significant differences at the 0.05 level.

Although the distributions are similar, depending on the exact differences in analysis between the two reports, this could have significant implications for bias in this study. The 7 minutes 30 second threshold apparently used to filter the data requires careful investigation to ensure results are robust to this filtering. Furthermore, with only 813 recorded respondents but possibly more than 1,205 respondents having participated in the survey, the magnitude of bias could be large. Lastly, the steep attrition rate suggested by the viewership figures may be indicative that the survey was too time demanding of its respondents.

5 Assessment of inferential claims

In the original report, the inferential analysis was performed through a series of pairwise t-tests, which does not allow for controlling for confounding variables. Using regression analysis, we can control for confounders and attempt to answer more directly just how much one factor contributes to the impact of a photograph. Four models are considered, described in Table 1, and reach broad agreement both internally and with the original report.

First, we apply a linear ordinary least squares (OLS) model with the four factors described above as independent categorical variables and the image responses as the dependent variable. Based on the fitted coefficients, this model at times disagrees with the conclusions from pairwise comparisons in the original analysis. While the original analysis found photos of individual animals more compelling than groups of animals, this model found large groups most compelling, followed by individual animals and small groups. Furthermore, while the original analysis found photos of baby animals more compelling than photos of adult animals, this model suggests mom and baby animals together were the most impactful followed by lone baby and adult animals.

Next, we consider interactions of the factors; that is, situations where the factors describing a photograph have a synergistic effect not adequately accounted for by either factor independently. For example, an interaction between “Category” and “Animal” might manifest as photographs of individual cows performing better than the expected sum of being an individual and a cow. To do this, we created a model with all the original covariates in addition to all possible pairwise combinations of factor levels. We then used stepwise Akaike information criteria-based model selection to determine which covariates sufficiently added to the explanatory ability of the model without over complicating it. Here we found that the interactions of "Category" and "Animal" along with the interactions of "Category" and "Grouping" warranted inclusion. When controlling for these interactions, our model indicates that individual animals are the least impactful grouping with large groups being the most impactful and the "mom/baby" combination has a larger effect than the "baby" alone. Otherwise, this model generally agrees with the OLS model without interaction terms.

Both of these OLS models treat the response variable as continuous, with equal spacing between steps. However, since the responses were on a Likert scale, the data is more properly treated as ordinal categories with no guarantee of equal spacing. To this end, we consider a proportional odds linear regression (POLR) model. This model uses the "parallel lines assumption", where the probability of a response at any given level is related to the independent variables by the same coefficients with only a constant difference distinguishing the response levels. Using the Brant test, we found our data conforms to this assumption [4]. Using the same covariates as the OLS model without interactions, we find a very similar result from the POLR model with the relative rankings of levels remaining the same. When including the interactions as above in the POLR model, we again find similar results to OLS model with interactions.

Calculating the relative likelihood of the four models, we find the POLR model with interactions is by far most likely to minimize the information lost even accounting for the extra parameters. However, since the qualitative implications of this model are fairly similar to the much simpler OLS model without interactions, we prefer the latter model in drawing conclusions.

6 Conclusion

In general, our reanalysis agreed with the findings of the original report. However, with more sophisticated modeling, we can be more confident that these claims describe the data well and are free from internal confounders. To summarize, we present the original claims as bullet points followed by our revised findings, based on the OLS model without interactions:

- *Original:* Photos of sick or injured animals are in general most compelling at causing people to want to stop eating animal products, followed by photos of animals who are dead or being killed, followed by photos of animals being kept in tight confinement.
    - *Revised:* When controlling for grouping, species and age, photos of sick and injured animals are the most compelling, agreeing with the original report. However, the grouping of remaining categories is not the same in our reanalysis. Small cage/confinement is the next most impactful followed by dead/killing. Happy animals are the least compelling.

<table>
<tr><th>Model</th><th>Assumptions</th><th>Conclusions</th></tr>
<tr><td>OLS
$y = \beta_0 + \beta\mathbf{x}$
RL = 0.00</td><td>• Linear relationship
• Multivariate normality
• Little multicollinearity
• Independent residuals
• Constant variance of residuals</td><td>• Most impactful image properties are
– Sick/ injured
– Pigs
– Large groups
– Mom and baby together</td></tr>
<tr><td>OLS + Interactions
RL = 0.00</td><td>• Same assumptions as OLS
• Interpretation changes since a category has both primary and interaction effects</td><td>• Significant interactions are
– Category ∗ Animal
– Category ∗ Grouping
• Individuals least impactful
• Baby animals most impactful
• Qualitative conclusions same as OLS</td></tr>
<tr><td>POLR
$\log \frac{\pi_1 + \ldots + \pi_j}{\pi_{j+1} + \ldots + \pi_j} = \beta_{0j} + \beta\mathbf{x}$
RL = 2.20×10^{-13}</td><td>• Response is ordinal categories
• Independent observations
• Little multicollinearity
• Linearity of independent variables and log odds
• Large sample size (> 10 cases of least frequent case)
• Parallel regression assumption</td><td>• Qualitative conclusions same as OLS</td></tr>
<tr><td>POLR + Interactions
RL = 1.00</td><td>• Same assumption as POLR
• Same change in interpretation as OLS + Interactions</td><td>• Qualitative conclusions same as OLS + Interactions</td></tr>
</table>

Table 1: The four models employed to analyze the data are presented along with their assumptions, relative likelihood (RL) and qualitative implications of each model.

- *Original:* Photos of pigs are more compelling than photos of other species. Photos of chickens/turkeys are next most compelling, possibly due to the abundant availability of photos of diseased, dead/dying, and intensely confined chickens and turkeys.
  - *Revised:* When controlling for condition, grouping and age, pigs are the most compelling species followed closely by poultry, cows and goats/sheep. Fish are the least compelling.
- *Original:* Photos of individual animals are more compelling than photos of groups of animals.
  - *Revised:* When controlling for condition, species and age, photos of animals in large groups were the most compelling images followed by individual animals. Small groups were the least impactful.
- *Original:* Photos of baby animals who are suffering are more compelling than photos of adult animals (or animals that appear to be adults).
  - *Revised:* When controlling for condition, species and grouping, photos of mom and baby animals together were the most impactful followed by baby animals alone. Adult animals were the least impactful.

Among these factors, the "Category" of photograph had the largest range of effect, where using a "Sick/injured" photo could confer an additional 2.5 points on the Likert scale of impact compared to a "Happy" photo. "Age" and "Grouping" had the lowest ranges of effect, respectively, with the latter conferring only 0.8 points of impact.

Unfortunately, while other methodological issues in the study design may have only a small impact on these results, the revelation that this data may represent a biased subset of the total responses makes application of these results difficult. It is not clear if these results represent a credible perspective on how different photographs of farmed animals may impact viewer's expressed desire to stop eating animal products and these results should be treated as weak evidence.

7 Acknowledgements

We thank the Sentience Institute (`https://www.sentienceinstitute.org/`) for reviewing the availability of the data for the Open Data Badge.

References

[1] Nick Cooney. *Report: Which Farm Animal Photos Are Most Likely to Inspire People to Eat Vegan?* Jan. 25, 2015. URL: `http://www.humaneleaguelabs.org/blog/2015-01-25-which-farm-animal-photos-inspire-vegan-eating/` (visited on 11/21/2017).

[2] Sabine Doebel, Susan Gabriel, and The Humane League. *Report: Which Farm Animal Photos Are Most Likely To Inspire People To Eat Vegan?* Humane League Labs, Jan. 2015. URL: http://www.humaneleaguelabs.org/static/reports/2015/01/compelling-farm-animal-pictures-study-results1.pdf (visited on 03/21/2018).

[3] Jacob Peacock. *Description of Omnibus Survey Used for Humane League Labs Study E005.* E005R08. Humane League Labs, May 2, 2018. DOI: 10.17605/OSF.IO/RYN7M. URL: http://www.humaneleaguelabs.org/static/reports/E005R08-description-of-omnibus-survey.pdf.

[4] Rollin Brant. "Assessing Proportionality in the Proportional Odds Model for Ordinal Logistic Regression". In: *Biometrics* (1990), pp. 1171–1178.