CAMBRIDGE
UNIVERSITY PRESS

RESEARCH ARTICLE

Introducing the Visual Conjoint, with an Application to Candidate Evaluation on Social Media

Alessandro Vecchiato[1] and Kevin Munger[2]

[1]Stanford University, Palo Alto, CA, USA and [2]European University Institute, Florence, Italy
Corresponding author: Kevin Munger; Email: kevin.munger@eui.eu

Abstract
Conjoint experiments have enabled scholars to understand the preferences of citizens in a variety of political contexts. We propose a method to modify the standard text-only "box conjoint" to make the treatment higher in external validity *with respect* to a common target context. Citizens frequently encounter political information encoded as images and in particular in the form of politicians' social media posts and profiles. We deploy "visual conjoint" experiments where subjects select between two images that encode the same explicit information as is standard in the box conjoint. We conduct an experiment in which we randomize the modality of a conjoint experiment where subjects evaluate the Twitter profiles of hypothetical candidates. We demonstrate that the visual conjoint more effectively encodes image-based information and social endorsement information. The visual conjoint also allows the salience of different attributes to vary naturally the way they do on social media, in contrast to the artificially enforced uniformity of the box conjoint.

Introduction

The process by which voters become informed about political candidates is central to the practice of democratic elections. An idealized model of this process entails voters gathering relevant information about each candidate and consciously selecting the one who they believe will be best able to represent their goals while in office.

Political scientists have recently embraced the use of conjoint analysis as a tool for understanding the way that voters evaluate this crucial electoral information. Most notably, researchers have used conjoint experiments to study how candidates' attributes such as age or race influence voters' support (Hainmueller, Hopkins, and Yamamoto, 2014).

This article has earned badges for transparent research practices: Open data and Open materials. For details see the Data Availability Statement.

Given their popularity, there has been significant methodological effort invested in refining conjoints. De la Cuesta, Egami, and Imai (2022) argue that conjoints that use a uniform distribution of candidate attributes are lower in external validity than conjoints where the distribution matches the true distribution in the target context.

Using the Egami and Hartman (2023) framework for external validity, we argue that the treatments used in standard candidate choice conjoint experiments are only externally valid for a small number of unique treatments in the real world.

This limitation is highlighted in Bansak et al. (2021)'s recent review article: "conjoint designs have been administered primarily via tables with written attribute values, even though information about political candidates or other choices is often processed through visual, audio, or other modes... The table-style presentation may prompt respondents to evaluate the choice in different ways, and so hamper external validity" (p27).

We develop, therefore, a *visual conjoint* that maintains the advantages of an ordinary conjoint in terms of the full randomization of features but presents the profiles as images. Our visual conjoint takes the form of a Twitter profile. The design is flexible, however, and could be applied to contexts including political advertisements, synthetic images of politicians, and partisan stereotypes. The task of evaluating a politician (or a potential dating partner or employee) based on their social media profile is a nearly ubiquitous part of contemporary political, social, and economic life; we argue that the visual conjoint like the one we deploy here is thus higher in external validity of the treatments with respect to the target treatments delivered when citizens encounter the Twitter profiles of politicians.

Our design includes typical details for candidate studies allowing us to benchmark our experiment against similar experiments run with standard conjoints. But the Twitter case affords the capacity to compare the magnitude of these well-established drivers of candidate preference with crucial elements specific to social media that inflect real-world citizen evaluations. Here, we investigate the importance of *social feedback*, in the form of the number of followers of the hypothetical politician profiles and in the number of "likes" and "retweets" on a handful of tweets included as part of the design.

We implement both a standard conjoint and our new Twitter visual conjoint in a candidate preference experiment based on the now-classic conjoint in Hainmueller, Hopkins, and Yamamoto (2014). Given rising concerns about the attention paid by subjects recruited through online convenience samples (Ternovski and Orr, 2022), we recruited 500 respondents through YouGov America.[1] Crucially for our argument about external validity, the sample only includes respondents who indicated that they were Facebook or Twitter users.

This relevant and attentive sample allows us to test for the existence of treatment effect heterogeneity both in terms of the conjoint modality (i.e., standard vs visual conjoint) and along theoretically relevant subject characteristics like age and digital literacy.

Our preregistered hypotheses were *all* concerned with treatment effect heterogeneity. We expected that the characteristics that become *more* salient when

---

[1] Among nonprobability online panels, YouGov outperforms others in various benchmarking tasks, with high representativeness and an attentive panel (Kennedy et al., 2016). Note that attentiveness and representativeness may be in tension if an attentive panel is not a representative one.

rendered as part of a hypothetical Twitter profile – the race and gender of the candidate, displayed in their profile picture – would have a larger effect in the visual conjoint than in the standard conjoint. We find evidence that this is the case for candidate gender. However, in accordance with the results in Abrajano, Elmendorf, and Quinn (2018), we find that candidate race has a significantly smaller effect in the visual conjoint. We discuss the theoretical logic behind this finding below.

We also preregistered the hypothesis that the characteristics that become *less* salient when rendered as part of a hypothetical Twitter profile – the profession, military service record, education, religion and age of the candidate, displayed in the small "bio" field – would have a smaller effect than in the standard conjoint. Even with the conservative multiple hypothesis correction recommended by Liu and Shiraito (2023), we find evidence for the expected significant differences for profession and education. The results for military service and religion are in the expected direction but not significant with the correction.

Finally, we preregistered the hypothesis that we would find heterogeneity in the effect of social feedback in respondent age and respondent digital literacy, but only in the visual conjoint. We find support for this hypothesis in the case of age but not in digital literacy as operationalized in our pre-analysis plan.

To restate what we see as our central contribution: our empirical exercise, of randomizing the treatment modality of a candidate choice conjoint experiment, demonstrates that these different modalities produce significantly different treatment effects. Which of these modalities is *better*? According to recent work on external validity, this question is undefined when posed in the abstract.

What we *can* say is that the visual conjoint is higher in external validity than the box conjoint *with respect* to the real-world task of evaluating politicians insofar as it is more representative of how citizens experience that task.

Dimensions of external validity in conjoint experiments

The foundational paper in the development of this method in Political Science is Hainmueller, Hopkins, and Yamamoto (2014). In the ensuing years, conjoints have been applied broadly, including evaluations of immigrants, neighborhoods, and climate-related policies; see Bansak et al. (2021) for an overview. In this work, we primarily engage with an emerging literature about their methodological aspects to improve the validity of the inferences made from them.

The standard conjoint experiment presents each piece of information as text, filled into a small box. The subject scans each of the two columns of boxes that comprise the two candidate profiles and then makes a choice (Jenke et al., 2019).

The artificiality of this research design is not a drawback for all applications. For contexts in which a subject encounters a novel entity and needs to evaluate it, the standard conjoint is plausibly externally valid. This is the case, for example, when a consumer needs to decide which product to purchase. Conjoint experiments were originally developed by psychometricians and initially flourished in the field of market research (Green and Srinivasan 1978; Luce and Tukey 1964). Conjoints that vary, say, the price, size, and quality of boxes of cereal are thus externally valid – *to the target context of a consumer strolling down the breakfast aisle*. This context is

importantly distinct from the context in which US citizens learn about political candidates, however, making it unlikely that the same modality of treatment delivery would be high in external validity with respect to both of these contexts.

Our concern is related to – but distinct from – the challenge to the external validity of conjoints raised by De la Cuesta, Egami, and Imai (2022), who argue that the use of a balanced distribution of attribute levels can produce misleading inferences about the population. They advocate for population-level distribution of the attributes other than the dimensions of primary interest in order to maximize the information gained from the closest possible counterfactual. The uniform distribution of attribute levels, they argue, is an unnecessary artifice motivated primarily by convenience. We argue that the uniform distribution of emphasis in the visual design of the conjoint is a similarly unnecessary artifice, one upon which we can improve by looking to the distribution of relevant real-world contexts.

Again, we agree with Egami and Hartman (2023) that the validity of a conjoint is only defined with respect to some specific goal. De la Cuesta, Egami, and Imai (2022) study the target *distribution of attributes* used in the treatments. Our focus is on the target *treatment modality* used to present the attributes. Our argument can also be usefully thought of in terms of another framework for experimental validity more common in the lab sciences: standard conjoints do not use representative "methods, materials, and settings," making them less *ecologically valid* (Morton and Williams, 2010).

Scholars of media effects have been particularly attuned to issues of external (ecological) validity. Arceneaux and Johnson (2013), for example, set up an ersatz living room in a shopping mall in order to deliver treatments in the form of television news and ads, complete with the remote control essential to the experience of watching TV at home. Kim (2023) expends monumental effort to deliver treatments that consist of televised game shows in a controlled and realistic setting.

Our contribution

Visual conjoints have been in use for decades in fields like marketing and engineering.

Increasingly, political science research has begun to include digitally manipulated pictures as stimuli in experiments about preferences for politicians or other politically relevant actors. Valentino et al. (2019), for example, manipulate the skin tone of hypothetical immigrants across a large experimental sample that spans eleven countries. Schachter, Flores, and Maghbouleh (2021) perform a similar skin tone manipulation to disentangle this contribution to racial categorization from that of ancestry and sociocultural cues. McClean and Ono (2021) use advanced photo editing technology to manipulate the apparent age of politicians, appending these images to standard conjoints.

Mechanics of the visual conjoint

Our approach allows us to generate hundreds or even thousands of unique images automatically and embed them in a survey. The design of the visual elements of the composite image takes more care than simply plugging in different textual values into a standard conjoint, but this step only has to be taken once for each visual

Figure 1. Visual conjoint task with two hypothetical twitter profiles. The figure shows two random social media profiles as they would appear to respondents. On the left, Candidate A is characterized by the following features: White, Woman, Millennial, with High Feedback, an Ivy League Graduate, Jewish, and a Doctor (no military service). On the right, Candidate B is characterized by the following features: Black, Man, Boomer, with Low Feedback, and a College graduate working as a lawyer that served the military (no religion).

conjoint experiment. For more details on the construction of the visual conjoint, see Appendix D.

Application of the visual conjoint to social media

The "social media visual conjoints" we develop below communicate all of the information used in a standard conjoint in a more organic manner. The standard demographic details are encoded in the "bio" field of the account, and standard partisan issue positions are expressed in the few tweets we include with the preview. Figure 1 provides a visual overview. We opted to use Twitter as the social media platform for this experiment as it is the locus of contemporary political discussion, but the framework is flexible, and visual conjoints based on other platforms are possible as well.

In addition to the improved ecological validity of this modality, the social media conjoint allows the researcher to manipulate aspects of the hypothetical candidate that cannot be communicated through the objective, scientific esthetic of the standard conjoint. In the current example, we manipulate the social feedback (in terms of Likes and Retweets) that the politician received on their tweets, as well as the number of followers their account had. This feedback is a crucial aspect of the social media environment, and Messing and Westwood (2014) demonstrates that it is roughly as important for media choice as are source cues. Although this information is central to the experience of frequent Twitter users or people who are high in digital literacy, it is possible that people who are unfamiliar with Twitter or social media more generally might not pick up on it. This heteoegeneity is an

advantage of our design insofar as it is representative of the heterogeneity in our target population of American citizens, as Guess and Munger (2023) argues is the case.

On the other hand, there are certain elements of the standard candidate preference conjoint that cannot be easily communicated in our design. It would be unnatural for a politician's Twitter bio to contain an explicit description of how much money they earn, a feature manipulated in Hainmueller, Hopkins, and Yamamoto (2014). Again, we interpret this as an *advantage* of our approach: voters encountering politicians through their self-presentation on social media is increasingly central to electoral politics, so the kind of information that is naturally conveyed in these contexts *should* be given more weight by external-validity-minded researchers.

Experimental design

Respondents were asked to select which candidate they would prefer between two randomly selected candidates among 4,800 possible combinations.

Each respondent had five candidate-pair (tasks) to evaluate. The candidate profile features were randomly selected among a predetermined set along in following dimensions: generation, gender, race, social feedback, education, party, profession, military service, and religion. Figure 1 displays one such choice. Note that each profile also includes two hypothetical tweets. These tweets were randomly drawn from a list of either Republican or Democratic tweets that we generated to be distinctively partisan but otherwise benign. We do not analyze the effect of each individual tweet for reasons of power and because we lack theoretical expectations. These tweets were necessary, however, to encode the crucial social feedback information.

Note also that we represent candidate age in two ways: the bio, which lists their birthday, and the photo. The photos (drawn from real but low-profile Members of Congress) were not fully randomized but rather were selected to match the age of the profile; for more details, see Appendix D.

Table 1 lists all possible feature levels for all of the nine dimensions. To avoid confusing participants with ethnically diverse names, we included identical last names and no first names (Edwards and Caballero, 2008).

To determine the sample size, we followed Schuessler and Freitag (2020) and assessed the required sample for minimal detectable effects of 0.05, the median effect size for the conjoint studies done in Political Science so far. Our power calculation led us to a sample of 500 individuals that multiplied by each task and profile pair ultimately would lead to 10,000 observations per conjoint experiment and a minimal detectable effect of 0.045.

Recruitment and selection

YouGov interviewed 653 respondents who were then matched down to a sample of 500 to produce the final dataset. In order to focus on our relevant population, respondents needed to indicate that they were Facebook or Twitter users. Recent Pew surveys report that 23% of Americans use Twitter, and that people who are interested in politics are highly overrepresented on the platform (Bestvater et al., 2022). The respondents were matched to a sampling frame on

Table 1. Dimensions of candidate characteristics

Dimensions	Possible values
Generation	Boomer
	Millennial
Gender	Male
	Female
Race	White
	Black
Feedback	High
	Low
Party	Democrat
	Republican
Military	Served
	Not Served
Education	Ivy League
	College No College
Religion	Protestant
	Catholic
	Mormon
	No Religion Jewish
Profession	Used Car Dealer
	Farmer
	Lawyer
	Doctor
	Teacher

gender, age, race, and education, as displayed in Table 2. Further details about the survey design can be found in Appendix E.

Demographic characteristics

Before the experimental treatment, respondents were asked a battery of descriptive questions aimed at gauging their level of digital literacy and their social media use. Our theoretical expectations pointed toward the importance of two (preregistered) dimensions of treatment effect heterogeneity: digital literacy and age. Following the recommendations of Guess and Munger (2023), we (Figure 1) deployed two measures of digital literacy.

The full questionnaire is presented in Appendix A.

Table 2. Sample descriptive statistics

	Mean
Age	51
Gender	0.53
Race: White	66.8%
Race: Black	12.4%
Race: Other	20.8%
Education: High School or less	35.6%
Education: College	32%
Education: Graduate or more	24%
Ideology: Strong Republican	14.6%
Ideology: Lean Republican	16.4%
Ideology: Independent	16.2%
Ideology: Lean Democrat	22.4%
Ideology: Strong Democrat	28.4%

Empirical strategy

Following our design, the empirical strategy is straightforward. We structure the data so that, for each respondent, it includes as dummy variables the descriptive characteristics of the profiles selected and seen. We calculate the Average Marginal Component Effect (AMCE) through a simple regression.

For examining the relevant subgroup preferences, we follow Leeper, Hobolt, and Tilley (2020) and use their R package "cregg" to estimate the difference in effects with marginal means. Differently from AMCEs, marginal means are calculated for each feature level by computing the probability that a profile with such feature level is selected over all possible combinations of other feature levels. For this reason, marginal means are comparable across sub-populations, while AMCEs, being estimates based on reference categories, may not be if baseline preferences in the reference categories differ across subgroups.

Results

Figure 2 presents corrected AMCEs with 95% confidence intervals for feature levels indicating the nine categories. Effects are relative to the excluded level for each category, represented in the graph as a green dot on the central vertical line. AMCEs show the increase in population probability that a profile would be chosen if its level would change to the one under consideration, averaged over all the possible values of the other components. Following the recommendation in Liu and Shiraito (2023), we have corrected our estimates using an adaptive shrinkage ("Ash") estimation to avoid false positives, though we present both initial and adjusted results. Given that
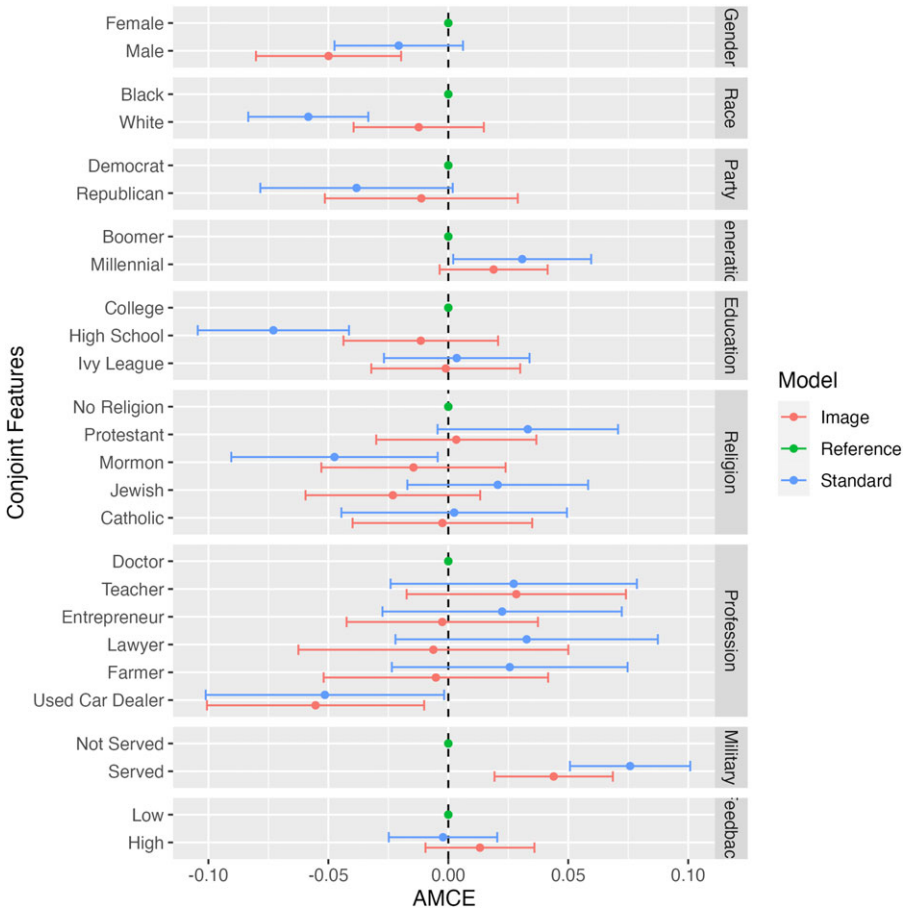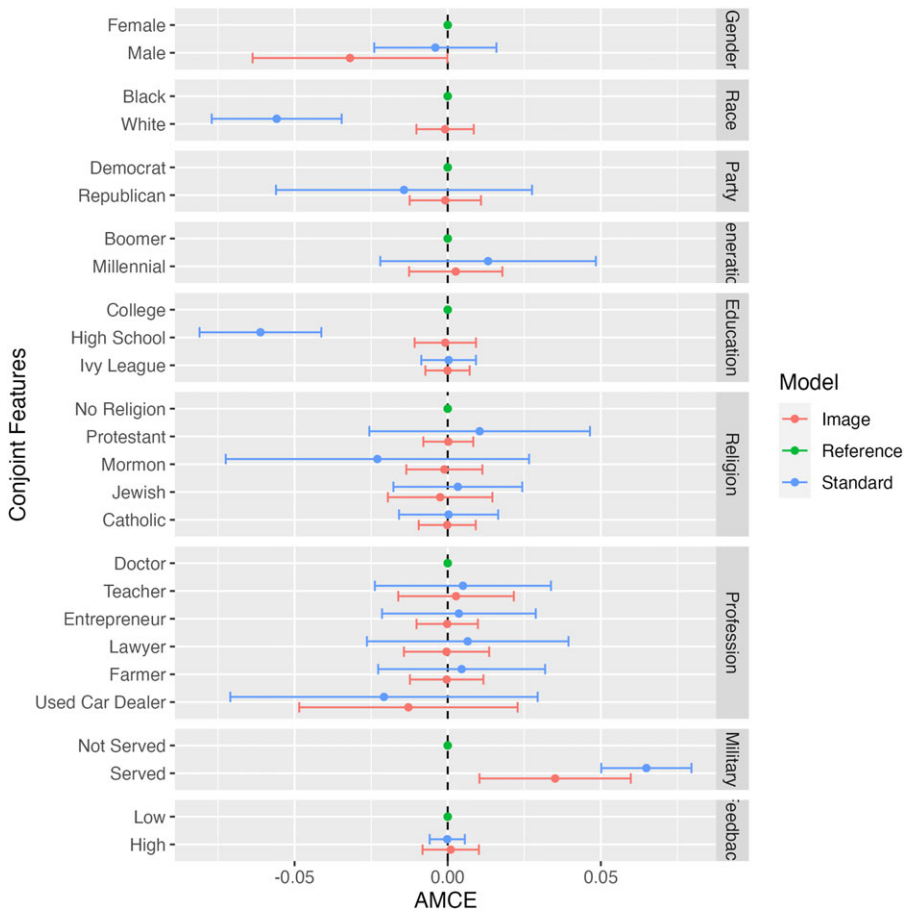
Figure 2. Preferences for politicians across two conjoints. Estimates of the change in probability of selecting the candidate, based on the reported characteristics.

all of our tests of interest were preregistered, we believe this to be a conservative adjustment.

Figure 2 displays the AMCE estimates from two different conjoint experiments with the same attributes displayed in either the standard "box" format (on top, in blue) or as part of our social media visual conjoint (below, in red). The former replicates several of the main findings in Hainmueller, Hopkins, and Yamamoto (2014). They find that Americans prefer political candidates who are veterans (vs non-veterans), Millennials (vs Baby Boomers), non-Mormon (vs non-religious), non-car dealers, with at least a BA degree; we find similar results for our sample of social media users.

Note that none of the conjoints published before 2023 used the Ash method for multiple comparisons. In our direct replication of Hainmueller, Hopkins, and Yamamoto (2014), the significant effects of generation, religion, and profession are no longer significant when we apply this multiple-comparison correction, in

Figure 3. Ash-adjusted preferences for politicians across two conjoints. Estimates of the change in probability of selecting the candidate, based on the reported characteristics. Includes the adaptive shrinkage adjustment per Liu and Shiraito (2023).

Figure 3. This raises interesting questions about the epistemology of replication and multiple-testing corrections; this difference could also, of course, stem from our different sampling frame, or a changing social environment.

Figure 4 tests our preregistered hypothesis: attributes that are made more salient by the visual conjoint (race, gender, and partisanship) will have a *larger* effect in that format, while attributes made less salient by the visual conjoint (profession, military service, education, religion, and age) will have a *smaller* effect link to preregistration here.

In other words, Figure 4 allows us to see whether the differences in the estimates of the effects of each attribute in the standard versus visual conjoint are themselves significant. The interpretation here is somewhat counter-intuitive given that this estimand is a "difference in differences"; the results in Figure 3 test the direction of the differences in our hypothesis, while those in Figure 4 show whether the differences are significant (again using the adaptive shrinkage adjustment).
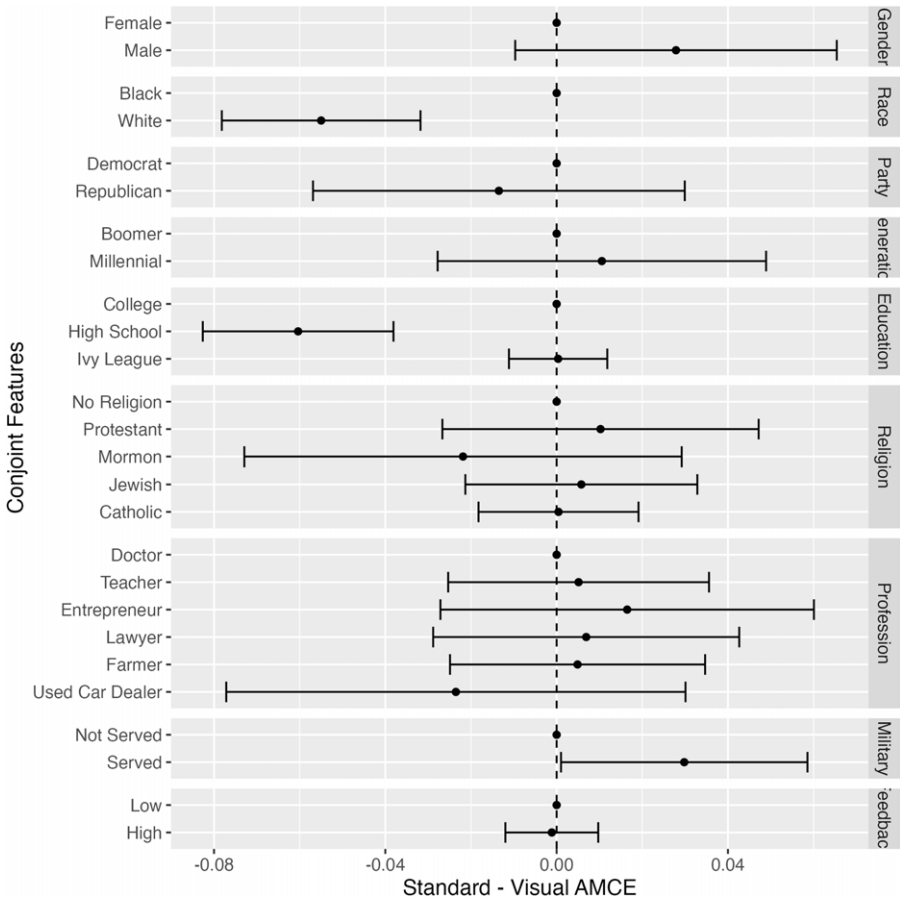
Figure 4. Difference in preferences for politicians across two conjoints. Estimates of the change in probability of selecting the candidate, based on the reported characteristics.

The effect of gender is significant ($p < .05$) in the visual conjoint but not the standard conjoint; these results are themselves significantly different from each other ($p < .10$). The opposite, though, is true of race, in contrast to our expectations; this result is significant ($p < .01$). We discuss this in detail below.

As expected, the effects of education, religion, age, and veteran status are smaller in the visual conjoint. Military veterans are still significantly preferred (with an effect size approximately two-thirds as large as in the standard conjoint) and used car dealers significantly punished, in contrast to our expectations – but in keeping with the long tradition of candidate choice conjoints that finds that used car dealers are robustly disliked. The significance of the last result does not survive the Ash adjustment.

However, these differences are only significant in the case of education ($p < .01$). There is a large difference in the expected direction for veteran status, but there is sizeable uncertainty in the image results. Generally, the effect of the Ash correction is more pronounced for the attributes that take five or six levels, like profession and religion.
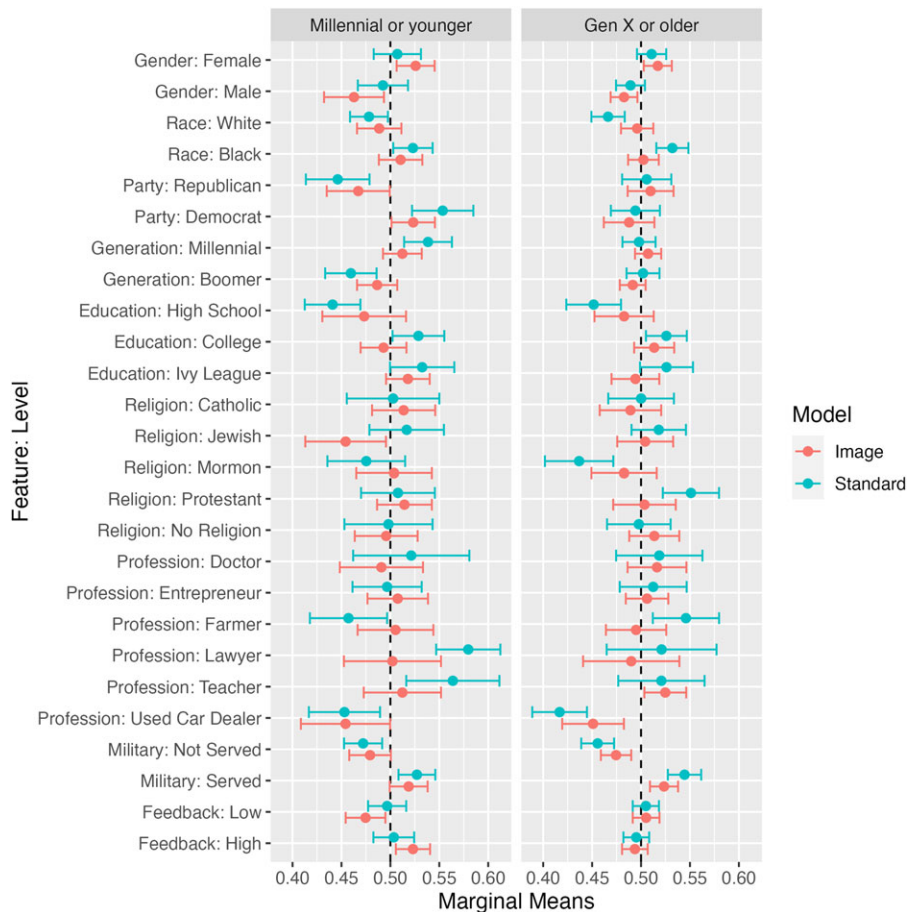
Figure 5. Preferences for Twitter profiles, by respondent generation. Estimates of the change in probability of selecting the candidate, based on the reported characteristics.

Finally, we find no main effect of the social feedback communicated in the follower counts and the likes and retweets on the accounts in either conjoint. Our preregistered hypotheses did not expect to find direct effects for this attribute; we only preregistered that this effect would be heterogeneous in respondent age and digital literacy.

Figure 5 displays the same results as Figure 2 with respondents divided along age.

In Figure 5, we find evidence that the null effect of social feedback in the overall sample is in fact masking offsetting effects among different generations: Generation Xers and Baby Boomers (forty-one or older) are less likely to support the politician with more social feedback (although the uncertainty estimates cannot rule out that the true effect on this subgroup is zero), while younger respondent are significantly *more* likely to support that politician (and *less* likely to support politicians with fewer Twitter followers, Retweets, and Likes). Again, as expected, this generational heterogeneity is observed only for the visual conjoint. Here, we do not perform the Ash adjustment because there is only a single preregistered effect.

In contrast, Figures B1 and B2 (in Appendix B) demonstrate very little effect heterogeneity for the social feedback condition in subject digital literacy, whether operationalized with the internet skills battery or the power user scale. This is evidence against our preregistered hypothesis.

Discussion

This paper extends recent work to increase the validity of conjoint experiments in political science. We argue that the visual conjoint is higher in external validity than is the standard conjoint *with respect* to a common target context in which Americans encounter and evaluate politicians.

Our preregistered hypotheses drew a key distinction between the attributes made more salient by the layout of the current visual conjoint (race and gender) and those made less salient (age, education, religion, veteran status, and occupation). Comparing the estimated treatment effects of these two groups of attributes provides support for this central hypothesis and motivation for the visual conjoint.

One key exception bears some discussion. Although we found the expected and substantively large differences in how respondents evaluated politicians' gender across the two conjoints, the effect of race was large and in the *opposite* direction we naively hypothesized based on salience alone. Our post hoc explanation for this surprising result is based on the theoretical arguments in Abrajano, Elmendorf, and Quinn (2018). They find that there are differential effects of signaling candidate ethnic identity with either an explicit conjoint attribute box or with an ethnically identifiable candidate photo.

Among White respondents who are actively trying to avoid using racial or ethnic cues (who are "internally motivated to control stereotyping"), treatment modality has no effect. However, white respondents who are not so motivated evince a large penalty for Latino candidates when ethnicity is cued by textual information but only a slight penalty when cued by candidate photographs. The theory is that communicating a candidate's race with a word like "Latino" or "Black" calls to mind negative stereotypes; however, when race is communicated with an image, it is bundled with other kinds of information. Since the images we used were professional headshots of actual members of Congress, the information explicitly counteracts those stereotypes.

Furthermore, we demonstrate the additional value of the visual conjoint with respect to the experience of evaluating social media profiles. We find that high levels of social feedback cause respondents to be more favorable – but only among respondents who are Millennials or younger.

Our results *do not* demonstrate that a visual conjoint created in the style of a Twitter profile is "better" than a box conjoint. We echo the emerging methodological consensus that concepts like "external validity" or "ecological validity" are only meaningful *with respect to a specific target.* No research design can generate a universally valid estimate of voter preferences; research on this topic, like all empirical research, can only be valid within specified scope conditions.

Future researchers who wish to implement a conjoint are thus encouraged to think carefully about the target context to which they hope to generalize their

results. All experimental research designs require trade-offs and simplifications for the purpose of control, but we believe that external validity should be prioritized in research designs which aim to inform decisions beyond their immediate context.

We have argued that the current visual conjoint is representative of an important and common experience in American political life, but digital media is constantly changing (Munger, 2023). "Twitter," for example, no longer exists; the platform is now called "X" and the way that citizens interact with the platform may have changed in a variety of as-yet untheorized and unmeasured ways. Our visual stimulus, which includes mention of "tweets," is no longer as representative of the experience which now would say "posts." As the use of visual conjoints expands, we hope to be able to develop a more precise understand of which of these visual aspects create meaningful differences in how citizens process information.

References

Abrajano, Marisa A., Christopher S. Elmendorf, and Kevin M. Quinn. 2018. "Labels vs. Pictures: Treatment-Mode Effects in Experiments about Discrimination." *Political Analysis* 26(1): 20–33.

Arceneaux, Kevin, and Martin Johnson. 2013. *Changing Minds or Changing Channels?: Partisan News in an Age of Choice.* University of Chicago Press.

Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins, Teppei Yamamoto, James N. Druckman, and Donald P. Green. 2021. "Conjoint Survey Experiments." *Advances in Experimental Political Science 19.*

Bestvater, Sam, Sono Shah, Gonzalo River, and Aaron Smith. 2022. *Politics on Twitter: One-third of Tweets from Us Adults are Political.* Report. Pew Research Centre.

De la Cuesta, Brandon, Naoki Egami, and Kosuke Imai. 2022. "Improving the External Validity of Conjoint Analysis: The Essential Role of Profile Distribution." *Political Analysis* 30(1): 19–45.

Edwards, Rosalind, and Chamion Caballero. 2008. "What's in a Name? An Exploration of the Significance of Personal Naming of 'Mixed' Children for Parents from Different Racial, Ethnic and Faith Backgrounds." *The Sociological Review* 56: 39–60. http://journals.sagepub.com/doi/10.1111/j.1467-954X.2008.00776.x

Egami, Naoki, and Erin Hartman. 2023. "Elements of External Validity: Framework, Design, and Analysis." *American Political Science Review* 117(3): 1070–1088.

Green, Paul E, and Venkatachary Srinivasan. 1978. "Conjoint Analysis in Consumer Research: Issues and Outlook." *Journal of Consumer Research* 5(2): 103–123.

Guess, Andrew M, and Kevin Munger. 2023. "Digital Literacy and Online Political Behavior." *Political Science Research and Methods* 11(1): 110–128

Hainmueller, Jens, Daniel J. Hopkins, and Teppei Yamamoto. 2014. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis* 22(1): 1–30.

Jenke, Libby, Kirk Bansak, Jens Hainmueller, and Dominik Hangartner. 2019. "Using Eye-Tracking to Understand Decision-Making in Conjoint Experiments." *Political Analysis* pp. 1–27.

Kennedy, Courtney, Keeter Scott, Andrew Mercer, Nick Hatley, Kyley McGeeney, and Alejandra Gimenez. 2016. *Evaluating Online Nonprobability Surveys*. Technical report.

Kim, Eunji. 2023. "Entertaining Beliefs in Economic Mobility." *American Journal of Political Science 67*(1): 39–54.

Leeper, Thomas J, Sara B Hobolt, and James Tilley. 2020. "Measuring Subgroup Preferences in Conjoint Experiments." *Political Analysis 28*(2): 207–221.

Liu, Guoer, and Yuki Shiraito. 2023. "Multiple Hypothesis Testing in Conjoint Analysis." *Political Analisys 31*.

Luce, R. Duncan, and John W. Tukey. 1964. "Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement." *Journal of Mathematical Psychology 1*(1): 1–27.

McClean, Charles T, and Yoshikuni Ono. 2021. "Too Young to Run? Voter Evaluations of the Age of Candidates." Report.

Messing, Solomon, and Sean J. Westwood. 2014. "Selective Exposure in the Age of Social Media: Endorsements Trump Partisan Source Affiliation when Selecting News Online." *Communication Research 41*(8): 1042–1063.

Morton, Rebecca B., and Kenneth C. Williams. 2010. *Experimental Political Science and the Study of Causality: From Nature to the Lab*. Cambridge University Press.

Munger, Kevin. 2023. "Temporal Validity as Meta-science." *Research & Politics 10*(3): 20531680231187271.

Munger, Kevin. 2024. "Replication Data for: Introducing the Visual Conjoint, with an Application to Candidate Evaluation on Social Media." https://doi.org/10.7910/DVN/TIBNLH

Schachter, Ariela, Ren´e D. Flores, and Neda Maghbouleh. 2021. "Ancestry, Color, or Culture? How Whites Racially Classify Others in the US." *American Journal of Sociology 126*(5): 1220–1263.

Schuessler, Julian, and Markus Freitag. 2020. "Power Analysis for Conjoint Experiments." *Mimeo*.

Ternovski, John, and Lilla Orr. 2022. "A Note on Increases in Inattentive Online Survey-Takers Since 2020." *Journal of Quantitative Description: Digital Media 2*.

Valentino, Nicholas A., Stuart N. Soroka, Shanto Iyengar, Toril Aalberg, Raymond Duch, Marta Fraile, Kyu S. Hahn, Kasper M. Hansen, Allison Harell, Marc Helbling, et al. 2019. "Economic and Cultural Drivers of Immigrant Support Worldwide." *British Journal of Political Science 49*(4): 1201–1226.