Political Analysis (2026), **34**, 541–568
doi:10.1017/pan.2026.10040

PA

ARTICLE

Adaptive Randomization in Conjoint Survey Experiments

Jennah Gosciak[1], Daniel Molitor[1] and Ian Lundberg[2]

[1]Information Science, Cornell University, USA; [2]UCLA, USA

Corresponding author: Jennah Gosciak; Email: jrg377@cornell.edu

(Received 28 April 2025; revised 9 January 2026; accepted 10 January 2026; published online 13 April 2026)

Abstract

Human choices are often both multi-dimensional and interactive. For example, a person deciding which of two immigrants is more worthy of admission to a country might weigh their education, and the weight placed on education may depend on other factors, such as their age, country of origin and employment history. We develop a response-adaptive experimental design that summarizes the range of effects of one attribute as a function of all other attributes. Our approach changes several aspects of the experimental design based on the ex ante choice to study the heterogeneous effects of one focal attribute (i.e., education). We update treatment assignment probabilities over the course of the experiment to search for the attribute vector at which the focal attribute has the most positive and most negative effects. By summarizing the full range of effects that exist, our approach complements existing approaches to conjoint experiments that typically aggregate over heterogeneity by marginalizing. We illustrate through two online experiments and provide customizable code infrastructure via a Docker container that other researchers can use to deploy adaptive randomization in online conjoint experiments.

Keywords: experimental design; multiple treatments; adaptive randomization; conjoint experiments; human preferences

Edited by: Daniel J. Hopkins

1. Introduction

Multi-dimensional choices abound in social life. When a voter chooses between two political candidates, those candidates may differ along several dimensions, such as age, gender and party affiliation. When a citizen considers which of two immigrants is worthy of admission to their country, they might weigh the immigrants' education, country of origin and employment history. Conjoint experiments, which present respondents with a choice between two fictitious profiles comprised of several randomized attributes, have rapidly improved scientific understanding of multi-dimensional human choices (Bansak *et al.* 2021; Hainmueller, Hopkins, and Yamamoto 2014).

Often, these choices are interactive. For example, a long literature in political science claims that voters prefer co-ethnic candidates. In a study of Ugandan voters, Carlson (2015) shows that the preference for a co-ethnic candidate is stronger when that candidate has a good performance record. That may not be the only interaction; a reanalysis by Egami and Imai (2019) further demonstrates that the effect of co-ethnicity varies by the candidate's platform (e.g., promoting jobs vs. improving education). What if co-ethnicity has effects that vary by an interactive function of even more attributes? When multi-dimensional choices involve interactions among several variables, these interactive preferences become difficult to summarize. A popular strategy is to summarize the interaction between two or three variables over the joint distribution of all other variables. We consider a different strategy to provide two summary statistics: the most positive and most negative effects of a focal attribute, conditional on values

© The Author(s), 2026. Published by Cambridge University Press on behalf of The Society for Political Methodology.
This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (https://creativecommons.org/licenses/by-nc-nd/4.0), which permits non-commercial re-use, distribution, and reproduction in any medium, provided that no alterations are made and the original article is properly cited. The written permission of Cambridge University Press or the rights holder(s) must be obtained prior to any commercial use and/or adaptation of the article.

https://doi.org/10.1017/pan.2026.10040

	Standard Conjoint	Adaptive Conjoint
Participants choose between two profiles	✓	✓
Each profile has many randomized attributes	✓	✓
Attributes are conceptualized	with all attributes analyzed in parallel	with one focal attribute and other context attributes
Estimand is the effect of one attribute	marginalized over all other attributes	at a particular value of all other attributes
Attribute assignment probabilities are	fixed throughout the experiment	changing over the experiment

Figure 1. Standard and adaptive conjoint designs: A comparison.

of the other attributes. We develop an experimental design that adaptively searches for these summary statistics. Our design summarizes the full range of effect sizes that an attribute may have, given particular values that the other attributes in the study can take.

To apply our design, a researcher first chooses a focal attribute of theoretical interest (e.g., candidate co-ethnicity). This choice differs from traditional conjoint experiments, which typically provide parallel evidence on the causal effects of many attributes. By prioritizing one attribute, our approach is more similar to a broad social science literature on human choices. Experiments often begin from a strong theoretical interest in one particular attribute. As one class of examples, audit studies have shown how race (a focal attribute) shapes human choices in hiring (Bertrand and Mullainathan 2004; Correll, Benard, and Paik 2007) or housing (Yinger 1986). For researchers most interested in one attribute, our approach enables a new form of inquiry: how to search for strong causal interactions between the focal attribute and other attributes.

Our design is most helpful when the focal attribute interacts strongly with other attributes. These settings may be common in both industry and the social sciences. One historical example from industry is an experiment by Honda which revealed that consumers preferred sliding doors (as opposed to swing doors) in their vehicle only if that vehicle was a minivan. The attribute (door type) interacts with the attribute (vehicle type) (Sawtooth Software 2024). Interactions are also widespread in social science studies of human choices, as illustrated by an example study on racial inequality. Steffensmeier, Ulmer, and Kramer (1998) analyzed judicial sentencing decisions and showed that judges sentence young Black men more severely than would be predicted by an additive model for the effects of race, gender and age. By focusing on attribute interactions, researchers can discover new patterns in human judgments.

Our approach is complementary to existing strategies for interactive preferences that discover low-order interactions (Egami and Imai 2019) or that test for the presence of any effect of one attribute conditional on any vector of values for all other attributes (Ham, Imai, and Janson 2024). Once a study finds interesting interactions or rejects the null that an attribute has no effect, it is natural to ask questions about the total range of effects that attribute may have. Models involving two-way interactions may poorly approximate this total range if higher-order interactions exist (Appendix E). Our method searches efficiently for the total range of causal interaction using adaptive randomization.

The article proceeds in several sections. We first formalize our experimental design and causal estimands, allowing us to draw connections to past work on conjoint experiments (Figure 1). We then

https://doi.org/10.1017/pan.2026.10040

present our approach to adaptive randomization, drawing connections to past work on adaptive experiments. We illustrate our approach with two empirical examples before concluding with a discussion.[1]

2. Causal estimands and experimental design

Our causal estimand and experimental design differ from those of a typical conjoint experiment in several ways, which we introduce in this section through a motivating example and mathematical notation.

2.1. A motivating example

We ground our method in a well-known example from the conjoint literature on attitudes among U.S. citizens toward immigrants. Hainmueller and Hopkins (2015) test the effects of nine different applicant attributes that each take on between 2 and 10 values. Hainmueller and Hopkins (2015) summarize each attribute's average marginal effect, for example, estimating that a bachelor's degree increases the likelihood of admission compared to no formal education by 20 percentage points.

We focus on interactions: what if the effect of having a bachelor's degree differs by the values of the other attributes? Instead of averaging away the heterogeneity, one may want to summarize the range of heterogeneous effects. We answer this question with a three-part experimental design that we implement in a simplified replication of Hainmueller and Hopkins (2015), randomizing a binary indicator of education and four binary indicators of other attributes.

2.2. Notation and design elements

We present each study participant i with a single pair of profiles for fictitious prospective immigrants and ask which is more worthy to be admitted to the country. Let $y_i \in \{L, R\}$ indicate whether the participant selects the left or right profile. Profiles involve several elements, which we introduce mathematically here and visually in Figures 2 and 3. Let $a_i \in \{L, R\}$ indicate whether the left or right profile features a prospective immigrant with a college degree (vs. no formal education). Because we are interested in the

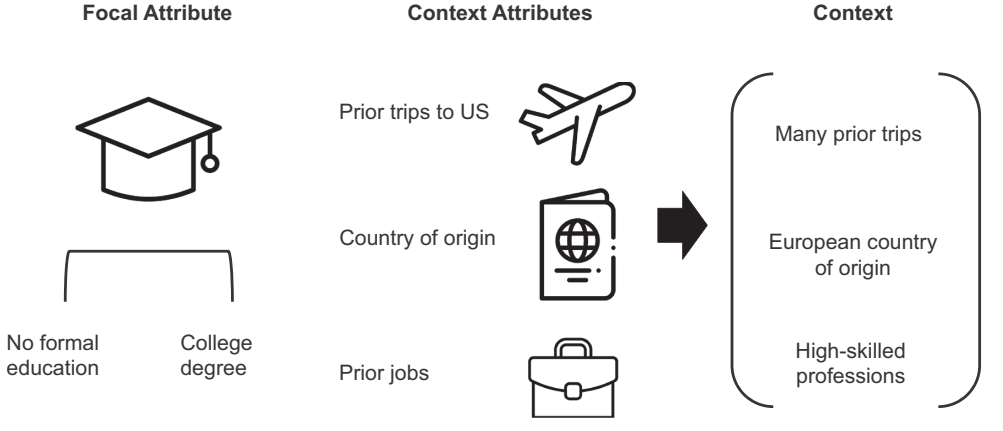


Figure 2. Elements of an adaptive conjoint design. Our design focuses on the causal effect of a randomized focal attribute, as it causally interacts with randomized context attributes. Each context attribute takes a value, and we refer to a vector of context attribute values as a context within which the focal attribute may have an effect.

[1] All replication materials for our empirical examples are archived and publicly available on Dataverse (Gosciak, Molitor, and Lundberg 2026).

https://doi.org/10.1017/pan.2026.10040

Context Attribute An element of a profile	**Attribute Values** Randomized across profile pairs	**Two Signals of Each Value** Randomly permuted within profile pairs
Prior trips to the U.S.	Many	• Has traveled to the U.S. every other summer since childhood • Has visited the U.S. many times as a tourist
	Few	• Has visited once for vacation • Two trips, both during the holidays
Profession	Skilled	• Information technology specialist • Construction worker
	Working Class	• Waiter • Janitor
Reason for application	Employment	• Seeking better employment opportunities • Greater access to high-paying jobs
	Persecution	• Escaping political/religious persecution • Escalating violence in home country
Country of origin	Europe	• Germany • Poland
	Northern Africa	• Sudan • Somalia

Figure 3. Context attributes, values and signals. Every profile has a set of context attributes: all attributes other than the focal attribute. Our design adaptively randomizes a vector of attribute values across profile pairs; within a pair, the two profiles share identical attribute values. To ensure that the two profiles in the pair do not appear identical to each other, we randomly permute two signals of each value across the profiles. Our experimental design therefore requires the researcher to specify two signals for each attribute value.

causal effect of a_i on the choice y_i, we refer to a_i as the *focal attribute*. We refer to all other attributes as *context attributes*. Our context attributes include prior trips to the United States, profession, reason for application and country of origin. Let $\vec{x}_i$ denote the vector of context attributes assigned to respondent i, for example, an immigrant from Europe applying for employment reasons who has many prior trips to the United States and works in a skilled profession. Each element of $\vec{x}_i$ takes one of two possible values in our applications. It is straightforward to generalize to attributes with more than two values, but doing so produces more possible profiles. With two possible values for each of four attributes, there are $2^4 = 16$ unique values of $\vec{x}_i$. With three possible values, there are $3^4 = 81$ unique values of $\vec{x}_i$. The dimensionality of the task grows with both the number of attributes and the number of values per attribute.

In the experiment, respondents view two profiles that differ on the focal attribute a_i but share the same context attribute vector $\vec{x}_i$. For example, both profiles may show immigrants from European countries of origin. To reduce suspicion, we create two *signals* of each context attribute value. One profile is from Poland and one is from Germany, for example (see Figure 3 for more examples). Each of the two profiles in a choice pair presents the respondent with a different vector of signals, even though the context attribute value is the same for both. Because signals are randomly assigned within contexts and our estimands (defined in the next section) marginalize over the random permutation of signals, these estimands are well-defined and identified even if signals have effects (e.g., respondents prefer immigrants from Poland over Germany). However, in practice, researchers should choose signals that indicate the same underlying context relevant to the choice. Optimally, the signals have minimal effects so that an estimand marginalized over random signals is marginalized over a nearly constant potential outcome.

https://doi.org/10.1017/pan.2026.10040

When randomizing the signals within a context, there are at least two viable strategies. One strategy is to pre-specify two vectors of signals and make a single randomization for whether a particular vector appears on the left or the right, $s_i \in \{L,R\}$. Advantages of this strategy include notational simplicity and potentially simpler deployment in software. We use this strategy for our main empirical illustration. A second strategy is to independently permute the signal possibilities for each attribute in the profile, so that the vector would be $\vec{s}_i \in \{L,R\}^p$ for p attributes. An advantage of independent randomization is that causal effects of individual signal elements are identified. We use independent randomization in our second empirical illustration.

The respondent's choice is a function of the focal attribute, context attributes and signals presented to that respondent. Let $y_i(\vec{x},s,a) \in \{L,R\}$ denote the potential choice outcome for respondent i if exposed to profiles with context attribute vector $\vec{x}$, signal permutation s and focal attribute permutation a. Each respondent has many potential outcomes, one for each combination of $\vec{x} \in \text{support}(\vec{X})$, $s \in \{L,R\}$ and $a \in \{L,R\}$. We assume consistency such that $y_i = y_i(\vec{x}_i,s_i,a_i)$. Let capital letters without subscripts refer to random variables that are random across respondents, so that $\{Y,\vec{X},S,A\}$ are the stochastic analogs of the particular values $\{y_i,\vec{x}_i,s_i,a_i\}$ taken for respondent i.

To preview our randomization (discussed in Section 3), a central phase of our experiment assigns new participants to a context $\vec{X}$ as a function of past data, and randomly permutes the focal attribute A and signal S across the two profiles in a pair. The purpose of the design is to study how the focal attribute A affects the respondent's choice within a context $\vec{X} = \vec{x}$.

2.3. Causal estimands

Define the choice probability within context $\vec{x}$ as follows: marginalized over the random permutation of the focal attribute and signals, with what probability is the chosen profile Y the same as the profile A that signals a selected value of the focal attribute?

$$\underbrace{\theta(\vec{x})}_{\substack{\text{Choice probability}\\ \text{in a particular}\\ \text{context } \vec{x}}} \equiv \underbrace{\mathbf{E}_{S,A}}_{\substack{\text{On average}\\ \text{over permuted}\\ \text{signal } S \text{ and}\\ \text{focal attribute } A}} \left[\mathrm{P}\Big(\underbrace{Y(\vec{x},S,A)}_{\substack{\text{Respondent}\\ \text{choice}\\ \text{(left or right)}}} = \underbrace{A}_{\substack{\text{Where selected}\\ \text{value of focal}\\ \text{attribute appears}\\ \text{(left or right)}}} \Big) \right]. \tag{1}$$

In our motivating example, $\theta(\vec{x})$ corresponds to the probability that a respondent presented with a pair of profiles with context attribute values $\vec{x}$ would choose the applicant with a college degree over the applicant with no formal education. Because of the forced-choice design, a value of $\theta(\vec{x}) = 0.5$ would imply that the focal attribute has zero effect on the choice (analogous to a marginal mean, Leeper, Hobolt, and Tilley 2020). Appendix F formalizes connections to common conjoint estimands. A key difference is that our estimand involves the effect of A holding all other attributes $\vec{X}$ at a particular value $\vec{x}$, rather than marginalized over a distribution of $\vec{X}$. A benefit is that one need not specify a population distribution of $\vec{X}$ (as in De la Cuesta, Egami, and Imai 2022).

Our non-marginalized estimand creates a different challenge: many unique $\vec{x}$ values (many contexts) produce correspondingly many unknown parameters $\theta(\vec{x})$. With four binary attributes, there are $2^4 = 16$ parameters $\theta(\vec{x})$. It may be costly to gather enough data to estimate them all precisely. An experiment with $n = 800$ participants would on average allocate only $\frac{800}{16} = 50$ participants to each context $\vec{x}$. High-variance estimates $\hat{\theta}(\vec{x})$ might leave the researcher with little confidence about which $\theta(\vec{x})$ is largest and which is smallest. Too few samples are spread across too many estimands. When the support of $\vec{X}$ is truly high-dimensional, for example, with hundreds of thousands of unique values, there is no hope of exploring the full space without assumptions that pool information across $\vec{x}$-values. But when the support of $\vec{X}$ is only moderately large, such as 16 unique $\vec{x}$-values, adaptive randomization can explore the space more efficiently than fixed randomization. We search for two summary estimands: the choice probabilities in contexts $\vec{x}_{\text{Max}}$ and $\vec{x}_{\text{Min}}$ that maximize and minimize $\theta(\vec{x})$:

https://doi.org/10.1017/pan.2026.10040

$$\theta(\vec{x}_{\text{Max}}) = \max_{\vec{x}} \ \theta(\vec{x}) \tag{2}$$

$$\theta(\vec{x}_{\text{Min}}) = \min_{\vec{x}} \ \theta(\vec{x}). \tag{3}$$

These estimands summarize the total range of effect sizes by which the focal attribute A may affect the choice Y across values of the context vector $\vec{X}$.

2.4. Design considerations for the structure of an adaptive conjoint

Several key design decisions differ under our approach compared with a standard conjoint.

The researcher must choose the focal attribute. Theory may guide the choice, as when audit study experiments focus on the effect of race as an attribute of particular importance (e.g., Bertrand and Mullainathan 2004). Past research may also guide the choice, for example, if low-order interactions are present (Egami and Imai 2019) or if statistical tests point toward large, unexplored heterogenous effects (Ham *et al.* 2024). Ideally, the focal attribute will have large causal interactions with other attributes. But if it has minimal causal interactions or even zero effect, the result may still be interesting, as we show in our second empirical illustration. In this case, $\theta(\vec{x}_{\text{Max}}) \approx \theta(\vec{x}_{\text{Min}})$. One can more confidently conclude that an attribute has nearly homogeneous effects if one searches explicitly for heterogeneity and finds nothing.

A second choice is how to select signals of context attributes. For each context attribute value (e.g., Eastern Europe), our design requires two signals (e.g., Poland and Germany) so that the two presented profiles are not identical. It is ideal to choose distinct signals that have minimal implications for the respondents' choices. For example, it is simplest if respondents are thinking about Europe (the context) rather than about the particularities of Poland and Germany (the signals). Researchers should choose signals that avoid social desirability bias; Germany and Poland would be a bad signal pair if one was a more socially desirable preference than the other, alone or in combination with other attributes. It is of course not possible to know in advance when designing the experiment. Two considerations offer comfort: (1) the estimands $\theta(\vec{x})$ remain well-defined and identified even if signals have effects and (2) one can estimate signal effects directly (Appendix B and C). Researchers should choose signals that they believe will not affect the choice and should assess the effects empirically.

A third design choice further reduces suspicion: we show each participant only one pair of profiles. A participant presented with a sequence of choices might come to recognize that they all differ on the focal attribute. Our choice comes at a cost, because traditional conjoints get much more data per respondent through repeated choices (Bansak *et al.* 2018). An advantage that partially compensates for this change is that the survey is shorter and thus costs less per respondent.

3. Adaptive randomization

Recall that our goal is to discover the contexts $\vec{x}_{\text{Max}}$ and $\vec{x}_{\text{Min}}$ that maximize and minimize an unknown parameter $\theta(\vec{x})$. The problem is analogous to the well-studied multi-armed bandit problem of choosing among many experimental arms to maximize an unknown payout. This section reviews existing work in adaptive experimentation and then formalizes the adaptive design we follow in our experiment.

Experimental research classically follows fixed (non-adaptive) designs. For example, a pre-specified number of units are randomized to treatment A or B by known, fixed probabilities. The fixed design brings many advantages. It is easy to understand. Researchers can pre-specify the design; nothing changes as a function of the data collected. Yet these advantages also correspond to limits of the fixed design: a researcher typically (though not always) selects a small number of treatment values (often two) and collects a pre-specified amount of data even if the answer gradually becomes clear before the end of the study.

In adaptive designs as we consider them in this article, the treatment assignment probabilities change over the course of data collection. In one of the earliest statements of adaptive designs, Thompson (1933)

https://doi.org/10.1017/pan.2026.10040

considered two interventions A and B. Over the course of an experiment that randomizes cases to A or B, imagine that it gradually becomes clear that A leads to better outcomes. Must we continue assigning new research participants to B? Thompson (1933) instead proposed a Bayesian updating procedure in which each participant's probability of assignment to treatment A equaled the posterior probability that A was the more effective treatment, given evidence from all participants who had come before. Under the Thompson sampling design, the probability of assignment to treatment A rises smoothly to 1 as evidence builds that A is the more effective treatment.

Adaptive designs are especially helpful in randomized trials with high-stakes consequences, such as those in which assigning a person to an inferior treatment may result in death. The gains to social welfare that come from assigning units to effective (rather than ineffective) treatments in real time over the course of a study are one reason adaptive designs have become increasingly common in clinical trials (Chow and Chang 2008; Pallmann *et al.* 2018; Rosenberger and Lachin 1993; Villar, Bowden, and Wason 2015; Yao *et al.* 2021).

Even in settings with lower stakes, adaptive designs are desirable because they can yield substantial efficiency gains, reducing the cost of data collection and improving the selection of the best treatment arm. Thus, adaptive experiments have begun to appear in economics (Kasy and Sautmann 2021) and political science (Offer-Westort, Coppock, and Green 2021).

3.1. Our design: Adaptive randomization of contexts in a three-phase experiment

In our experimental design, the arms of adaptive randomization are the contexts $\vec{x}$, and we seek to find the maximum and minimum of $\theta(\vec{x})$, the choice probability which is defined for each arm. We carry out randomization in phases—warm-up, adaptive and validation—to avoid some difficulties that otherwise threaten inference in adaptive experiments. Appendix G formalizes the procedure in pseudocode.[2]

3.1.1. Warm-up phase

At the beginning of the experiment, we have no knowledge of the unknown $\theta(\vec{x})$ probabilities. We begin by collecting a set of n_{Warmup} observations equally distributed across the contexts. The warm-up phase provides initial evidence about all $\theta(\vec{x})$ parameters, enabling one to proceed to adaptive randomization with some evidence rather than solely a subjective prior.

3.1.2. Adaptive phase

After the warm-up phase, we begin assigning treatments with unequal probabilities adaptively as a function of past data. There are two adaptive searches: one for $\theta(\vec{x}_{\text{Max}})$ and one for $\theta(\vec{x}_{\text{Min}})$. One can carry out these searches in parallel, using only the warm-up data as the initial input to each adaptive randomization search. We follow this strategy in our second illustration with job candidate profiles. One can also carry out these searches in sequence, for example, searching first for $\theta(\vec{x}_{\text{Max}})$ and then for $\theta(\vec{x}_{\text{Min}})$. We follow this strategy in our first illustration with immigrant profiles. We recommend the sequential (rather than parallel) approach because it makes the second search more efficient: the data from the first search support better treatment assignments in the second search. An open direction for future research would be to identify an optimal mixing of the two searches over time, using all cumulative data in each search.

The adaptive phase begins with a prior on the unknown choice probabilities $\theta(\vec{x})$. We use a Beta distribution as the conjugate prior to a Bernoulli choice, but generalizations to categorical choices (Dirichlet prior, Categorical likelihood) or continuous ratings (Normal prior, Normal likelihood) are analogous. For an example of adaptive updating with a categorical response, see Deliu (2024). In our setting, we assume a uniform Beta(1,1) prior on each unknown choice probability:

[2]The pseudocode and our discussion here emphasize our recommended design decisions, which we arrived at through two empirical applications. We note in several places where one of our empirical applications diverges slightly from our recommended design.

https://doi.org/10.1017/pan.2026.10040

$$\theta(\vec{x}) \sim \text{Beta}(1,1). \tag{4}$$

Using conjugacy, we update the posterior probabilities after seeing a set of observations indexed by i,

$$\theta(\vec{x}) \mid \mathbf{X}, \vec{Y} \sim \text{Beta}\Bigg(\underbrace{1+\sum_i \mathbb{I}(\vec{x}_i = \vec{x}, y_i = a_i)}_{\substack{\text{Number of past responses}\\ \text{in this context choosing}\\ \text{the profile with a}\\ \text{college degree}}}, \quad \underbrace{1+\sum_i \mathbb{I}(\vec{x}_i = \vec{x}, y_i \neq a_i)}_{\substack{\text{Number of past responses}\\ \text{in this context choosing}\\ \text{the other profile}}}\Bigg), \tag{5}$$

where $\mathbf{X}$ and $\vec{Y}$ correspond to the matrix of assigned contexts and to the vector of observed outcomes, respectively.

For each respondent, we simulate many values of each $\theta(\vec{x})$ from its posterior distribution given data from all previous respondents. For each simulation, we identify which context yields the highest and lowest values of the unknown probability. Averaging these results over all simulations provides our Monte Carlo estimates of the probabilities that each context maximizes or minimizes the value of $\theta(\vec{x})$:

$$\pi_{\text{Highest}}(\vec{x}) \equiv \mathbb{P}\left(\theta(\vec{x}) = \max_{\vec{x}'} \theta(\vec{x}') \mid \mathbf{X}, \vec{Y}\right) \tag{6}$$

$$\pi_{\text{Lowest}}(\vec{x}) \equiv \mathbb{P}\left(\theta(\vec{x}) = \min_{\vec{x}'} \theta(\vec{x}') \mid \mathbf{X}, \vec{Y}\right). \tag{7}$$

In our empirical illustration, we base these estimates on 100,000 simulated draws.[3] We use the estimated posterior probabilities $\hat{\pi}_{\text{Highest}}(\vec{x})$ (or $\hat{\pi}_{\text{Lowest}}(\vec{x})$ when searching for the minimum context) to assign new survey participants to contexts.

The adaptive phase can be carried out in continuous fashion (updating posteriors after each respondent) or discrete fashion (updating posteriors after each batch of respondents). Our first empirical illustration and accompanying software support the continuous approach. Our second illustration uses discrete batches. We stop when a budgeted sample size has been reached. Future research could explore alternative stopping rules, such as halting the process when a target posterior probability threshold is reached.

At the end of the adaptive phase, we select two key contexts, $\vec{x}_{\text{Max}}$ and $\vec{x}_{\text{Min}}$. These are the contexts with the highest probability of having the most positive and most negative focal attribute effects, respectively, across posterior draws given all data. Researchers should report these posterior probabilities as evidence for their degree of certainty. We also report posterior mean estimates $\hat{\theta}(\vec{x}_{\text{Max}})$ and $\hat{\theta}(\vec{x}_{\text{Min}})$ from all data collected to this point (warm-up + both adaptive phases). However, for reasons explained below, our preferred estimates of these parameters come from the validation phase, which we describe next.

3.1.3. *Validation phase*

The final phase of our experiment is a validation phase: we collect new data in which respondents are assigned to the two selected contexts with fixed, equal probabilities.[4] The validation phase addresses a problem known as the winner's curse.

After the adaptive phase, there remains statistical uncertainty about the true (and unknown) values of $\theta(\vec{x})$. The variance $\text{V}(\hat{\theta}(\vec{x}))$ may be non-negligible. Suppose there were dozens of contexts and a

[3] When the goal is simply to assign a new case and not to draw inference about π_{Highest} and π_{Lowest}, a more computationally efficient and equivalent strategy is to make one draw of the unknown $\theta(\vec{x})$ parameters and assign the next respondent to whatever context $\vec{x}$ has the highest (lowest) simulated value of that single draw.

[4] It would be ideal to assign contexts at random. The actual implementation for our first empirical illustration used two simultaneous online experiments hosted on Prolific, one in which all respondents saw $\vec{x}_{\text{Min}}$ and one in which all respondents saw $\vec{x}_{\text{Max}}$. Because these two experiments happened at the same time with identical calls, it is reasonable to think of $\vec{X}$ as randomized.

https://doi.org/10.1017/pan.2026.10040

comparably small sample size in the adaptive phase. The context with the highest estimated value is not necessarily the context with the highest true value: it is possible that $\arg\max_{\vec{x}}\left(\hat{\theta}(\vec{x})\right) \neq \arg\max_{\vec{x}}\left(\theta(\vec{x})\right)$. In fact, which context is chosen depends partly on the true signals in each context and partly on the luck of which context happens to have a high estimated value in the particular adaptive sample analyzed. On average, over many trials, the winner has positive noise. The answer is to re-estimate the winner's parameter $\theta(\vec{x})$ in validation data where the noise is independent of the choice.[5]

The winner's curse may seem unfamiliar, but it is the same as the well-known motivation for training and validation sets to evaluate the performance of predictive algorithms. In such settings, a researcher selects among many algorithms using a training sample. But predictive performance metrics in a training sample are biased estimators of out-of-sample performance due to the winner's curse: the fact that an algorithm wins in the training sample is correlated with the outcomes in the training sample. Just as an independent test sample solves the winner's curse in evaluations of predictive algorithms, an independent validation sample solves the winner's curse in our setting.

As an added benefit, the validation phase simplifies statistical inference. Inference on the adaptive phase is made challenging by the fact that treatment assignment probabilities are unequal across respondents as a function of past data. Unbiased estimators are possible through inverse probability of treatment weighting, yet challenges to inference can remain due to issues that arise with asymptotic normality (Hadad *et al.* 2021). Several approaches have been proposed to address this inferential challenge (Hadad *et al.* 2021; Zhang, Janson, and Murphy 2020, 2021). These complex solutions are not necessary in our design: the validation phase is a classic experiment with fixed assignment probabilities. Unweighted mean estimators are unbiased and asymptotically normal.

3.2. Summary: Design considerations for adaptive randomization

To summarize, the goal of our adaptive experimental design is to search for $\theta(\vec{x}_{\text{Min}})$ and $\theta(\vec{x}_{\text{Max}})$. A central reason these estimands are of interest is that $\theta(\vec{x}_{\text{Min}}) \leq \theta(\vec{x}) \leq \theta(\vec{x}_{\text{Max}})$ for all $\vec{x}$, so that these two estimands bound the total range. There are two important limitations that should inform design choices.

The first limitation arises because of statistical uncertainty: it is possible that the wrong contexts are selected, for example, $\arg\max_{\vec{x}}(\hat{\theta}^{\text{Warmup + Adaptive}}(\vec{x})) \neq \arg\max_{\vec{x}}(\theta(\vec{x}))$. If you pick the wrong context, then the validation estimates will be unbiased for the selected context while being downwardly biased for the parameter of the context that is the true maximum: $\mathbf{E}(\hat{\theta}(\vec{x}_{\text{Selected Max}})) = \theta(\vec{x}_{\text{Selected Max}}) < \theta(\vec{x}_{\text{True Max}})$. By allocating more sample to contexts with high probabilities of being the true maximum, adaptive randomization reduces the risk of selecting the wrong context (simulation in Appendix D). Researchers can also be transparent about their uncertainty by reporting the posterior probabilities that the chosen contexts are the minimum and maximum.

A second limitation occurs even when the warm-up and adaptive phases lead to correct choices for the extreme contexts. After selection, the parameters of those contexts must be estimated in the validation phase. While validation estimates are unbiased, they will still be statistically uncertain. One can reduce this risk by allocating more samples to the validation phase and can make this risk transparent by reporting estimates of uncertainty (e.g., 95% credible intervals) for estimates from the validation phase.

These limitations can inform choices about how to allocate sample size across phases. On the one hand, allocating more samples to the warm-up and adaptive phases reduces the chance of selecting the

[5]The winner's curse arises from the researcher's selection of two contexts out of many candidates, not from the adaptive randomization. The problem would also exist in a fixed design where a researcher chooses among many estimands as a function of their estimated values.

https://doi.org/10.1017/pan.2026.10040

wrong context. On the other hand, allocating more samples to the validation phase supports precise estimation for the selected contexts. The correct balance will depend on the application.

When allocating sample across the warm-up and adaptive phases, a few anchoring points are useful. To allocate all cases to warm-up is equivalent to running a fixed-randomization trial. To allocate all cases to the adaptive phase is a purely Bayesian approach, allowing the posterior to update treatment assignments from the start. Allocating some cases to warm-up and some to the adaptive phase is a middle ground. The warm-up phase then serves to provide initial observations so that the prior at the start of the adaptive phase is partially data-driven. We recommend this middle-ground approach.

A further design choice is when to stop the trial early. In our second empirical illustration (below), there was little evidence of variation in $\theta(\vec{x})$ across context values $\vec{x}$ at the end of the warm-up and adaptive phase. As discussed below, in that example, we decided not to carry out a validation phase because the range of point estimates from the first two phases suggested insufficient variation to be of scientific interest. In general, we recommend that researchers after the warm-up and adaptive phase consider whether the likely results of the validation phase are of sufficient interest to justify the cost of collecting data in that phase.

4. Empirical illustrations

We demonstrate our approach in two applied illustrations: one with respondents choosing which of two fictitious immigrants is more worthy for admission to the country, and one with respondents choosing which of two fictitious job applicants to hire for a hypothetical job. We discover substantial heterogeneity in the first illustration and almost no heterogeneity in the second.

4.1. Illustration 1: Prospective immigrants

We conducted an online experiment based on Hainmueller and Hopkins (2015). We recruited 10,000 participants using Prolific, a web-based survey platform with a non-probability sample. The Prolific data collection took place between June 24, 2024 and June 30, 2024. We administered the survey with a custom Python Shiny App deployed via AWS.[6] Our implementation is standalone software that researchers can use to administer an experiment. It replicates the survey functionality of tools like Qualtrics, but enables continuous adaptive updating, which Qualtrics does not easily support.

The average age in the sample was 35 years old, 68% were white and 56% were female. We allocated 2,000 respondents to the Warm-up Phase (approximately 125 per each of 16 arms). We then recruited 6,000 respondents for the adaptive phase: 3,000 for finding $\vec{x}_{\text{Max}}$ and 3,000 for finding $\vec{x}_{\text{Min}}$. We allocated 2,000 respondents to the validation phase (1,000 for each $\vec{x}_{\text{Max}}$ and $\vec{x}_{\text{Min}}$).

Similar to Hainmueller and Hopkins (2015), we asked each participant: "Please read the descriptions of the potential immigrants carefully. Then, please indicate which of the two immigrants you would personally prefer to see admitted to the United States." Respondents then viewed a table similar to Figure 4. Figure 3 enumerates the full set of 16 contexts.

Figure 5 presents the results at the end of the warm-up and adaptive phases. For each context $\vec{x}$, we report the posterior mean probability $\hat{\theta}(\vec{x})$ of preferring the college-educated immigrant over the immigrant with no formal education, along with a 95% Bayesian credible interval. Recall that a value of $\theta(\vec{x}) = 0.5$ corresponds to no causal effect of the focal attribute, and that the signal of a college degree increases the probability that a profile is chosen to the degree that $\theta(\vec{x})$ exceeds 0.5. A preference for the college-educated profile is apparent in all contexts. There is also clear heterogeneity across contexts, with estimates ranging from below 65% to above 75%.

Two estimates in Figure 5 are more important than the others: the contexts highlighted in red with the maximum and minimum estimates. The Bayesian credible intervals for these contexts are the most narrow; over the course of the experiment, the adaptive phase came to allocate sample to these contexts

[6]The code for the survey deployment is available at https://github.com/dmolitor/adaptive-infra.

https://doi.org/10.1017/pan.2026.10040

Please read the descriptions of the fictional immigrants carefully. Then, please indicate which of the two you would personally prefer to see admitted to the United States.

	Immigrant 1	**Immigrant 2**
Education	College degree	No formal education
Profession	Waiter	Janitor
Country of origin	Germany	Poland
Reason for application	Escaping political/religious persecution	Escalating violence in home country
Prior trips to the U.S.	Has traveled to the U.S. every other summer since childhood	Has visited the U.S. many times as a tourist

Which of these two immigrant profiles would you select for admission to the United States?

○ Immigrant 1

○ Immigrant 2

Figure 4. Example of two fictional immigrant profiles shown to respondents. The two values of education, the focal attribute, are college degree versus no formal education. The values of the other context attributes are identical for both immigrants 1 and 2 (e.g., both from Eastern Europe), though each immigrant has a different signal of each attribute value (e.g., Germany and Poland).

with high probabilities. The chosen maximum arm has the highest posterior draw of $\hat{\theta}(\vec{x})$ in 79% of posterior draws. The chosen minimum arm has the lowest in 49% of posterior draws. These quantities summarize our degree of confidence that these are the arms of interest. With a larger budget, we could have continued the experiment to become more certain of our selections.

The validation phase collected new independent data on these selected contexts. Figure 6 compares estimates from the (warm-up + adaptive) and (validation) phases. Estimates are very similar, though validation estimates are slightly closer to 0.5 as one might expect given the winner's curse from the adaptive phase. Overall, this illustration demonstrates that our design can successfully discover contexts across which the effect of the focal attribute is substantially heterogeneous.

4.2. *Illustration 2: Job applicants*

Our second illustration shows a setting in which we estimate precise effect sizes near zero in all contexts.

An extensive literature has documented a negative association between motherhood and women's labor market outcomes (Budig and England 2001; Kleven, Landais, and Søgaard 2019; Lundberg and Rose 2000). One source of this disparity may be employer discrimination against mothers compared with childless women. Some studies randomize signals of motherhood in lab studies and real-world audits to assess labor market effects (Correll *et al.* 2007; Ishizuka 2021). We designed an online experiment to investigate how the effect of motherhood differs across contexts defined by two background attributes: the applicant's race and the ranking of the applicant's educational institution. We collected data on Prolific from February 8, 2024 to March 13, 2024. An important caveat to our results is that our evidence among online survey respondents may not generalize to the population of employers making actual hiring decisions.

Each respondent is randomized to a pair of job applicant resumes who have degrees in marketing and are applying for a human resources position. One resume signals motherhood while the other does not. Figure 7 presents an example pair of resumes. There are two context attributes (race and educational institution rank) and one focal attribute (motherhood). Figure 8 shows all context attribute values and signals. Appendix H discusses how we chose these signals drawing on prior research. For

https://doi.org/10.1017/pan.2026.10040

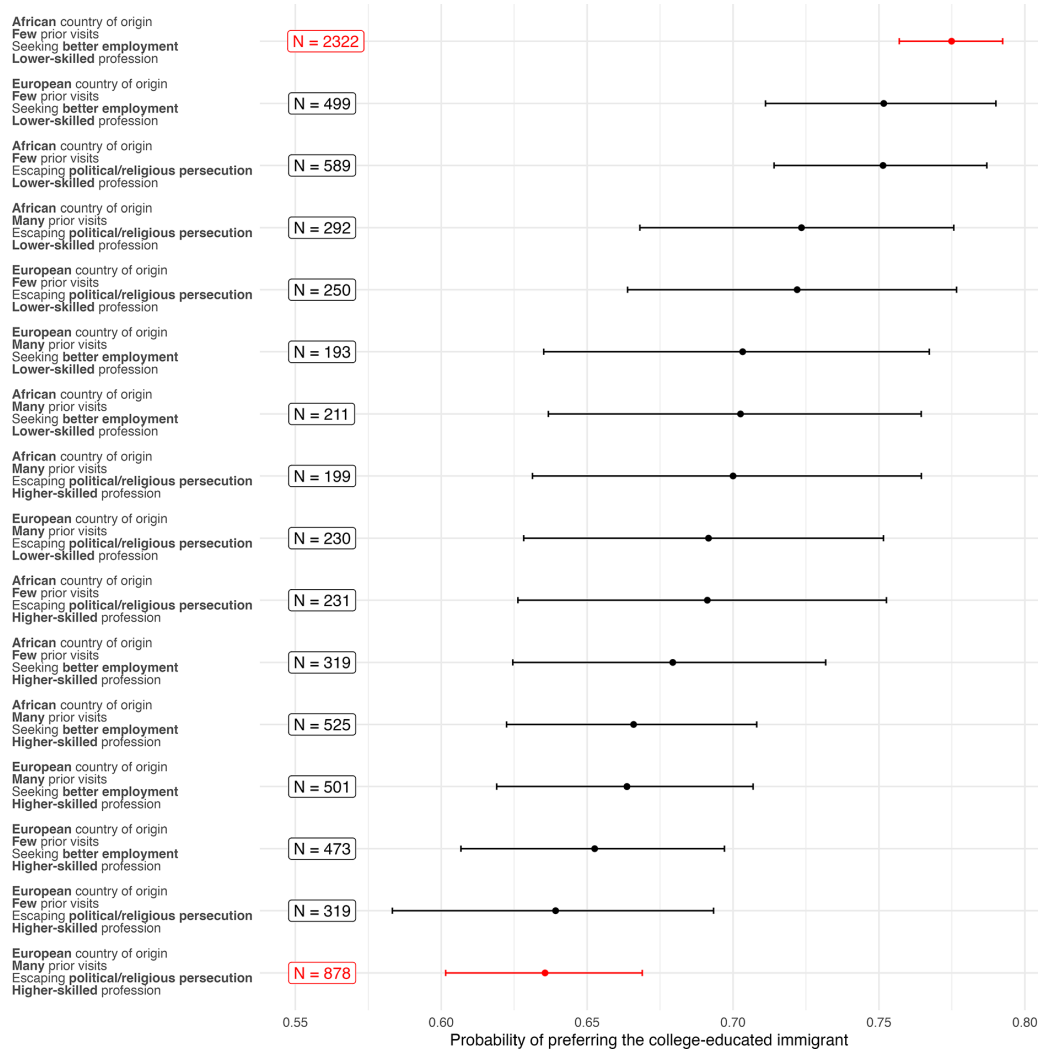


Figure 5. Results after the warm-up and adaptive phases: Estimated preference for the college-educated immigrant profile. The *x*-axis depicts the estimated probability of choosing the college-educated immigrant within each of those contexts, along with 95% credible intervals. The *y*-axis shows the full set of context attributes for all 16 contexts. The contexts highlighted in red are the contexts discovered as having the highest and lowest posterior probabilities of a respondent choosing the more-educated profile.

this illustration, the adaptive phase proceeded in batches of 200 participants with posterior probabilities updated between batches.

Figure 9 presents results from the warm-up and adaptive phases. We report the estimated values of $\theta(\vec{x})$, or the posterior mean probability of preferring the non-mother job applicant over the equally qualified mother. Estimates for all four contexts are close to 0.5, indicating minimal effects of motherhood regardless of context. If we were to select $\vec{x}_{\text{Max}}$ and $\vec{x}_{\text{Min}}$, we would be highly uncertain: the context with white job candidates from mid-ranked educational institutions context has a 39% posterior probability of being $\vec{x}_{\text{Max}}$, and the context with the black job candidates from highly ranked educational institutions has a 47% posterior probability of being $\vec{x}_{\text{Min}}$. Because we detected effectively no evidence that the focal attribute had any effect in any context, we stopped the study after the adaptive phase and did not carry out a validation phase.

https://doi.org/10.1017/pan.2026.10040

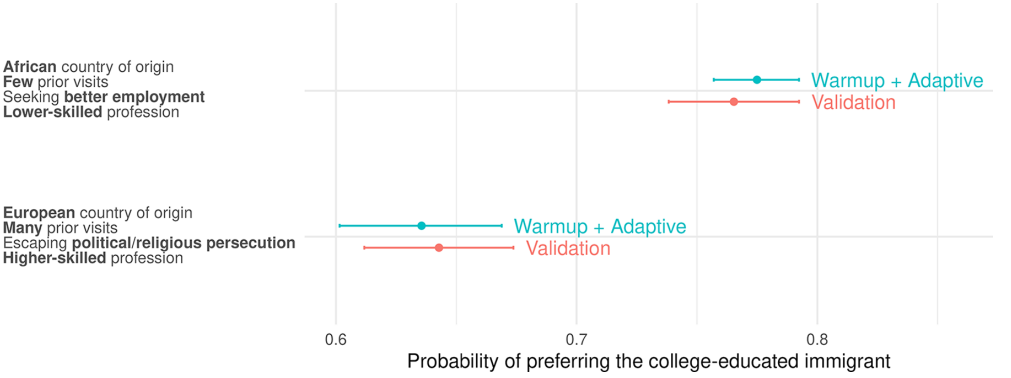


Figure 6. Validation results: Preference for the college-educated immigrant. The y-axis shows the contexts $\vec{x}_{\text{Max}}$ and $\vec{x}_{\text{Min}}$ that were identified in the adaptive experimental phase as having the highest and lowest posterior probabilities that the respondent would choose the more-educated prospective immigrant within the pair. The x-axis depicts the estimated probability of choosing the more educated immigrant within each of those contexts. Warm-up and adaptive estimates are posterior mean estimates with 95% credible intervals, and validation estimates are frequentist mean estimates with 95% confidence intervals.

*Laurie Schmitt is a 36 year-old marketing coordinator who graduated with a B.S. in Business Administration and Marketing. She's been working in her current field for seven years and has experience directing all phases of both the creative and technical elements of marketing initiatives. Her hobbies include swimming and tennis. Laurie volunteers with the Parent Teacher Association at her child's elementary school.

Candidate 1
Laurie Schmitt
Marketing Coordinator
(303) 555-0146 | laurie.schmitt@gmail.com | **LinkedIn.com/Laurie.Schmitt**

Education
B.S. in Business Administration; Marketing | University of Pennsylvania | 2010
Professional Experience
Marketing Coordinator | Goods Delivered | 2021 to Present
Report directly to Marketing Manager, assisting in the development of the overall marketing strategy.
Marketing Coordinator | Live Healthy | 2017 to 2021
Supported a team of four marketing account executives responsible for 12 client accounts in the healthcare industry.
Volunteer Experience
Parent Teacher Association Volunteer | Creekside Elementary | 2022 – present

Allison O'Connell is a 36 year-old marketing coordinator who has a degree in Marketing and minored in Digital Marketing. She's been working in this field for seven years. She is an experienced and accomplished marketing professional with a track record of proven results. In her free time, she likes to run and play soccer. Allison also works as volunteer treasurer with her local neighborhood association.

Candidate 2
Allison O'Connell
Marketing Coordinator
720-555-0182 | allison.oconnell@gmail.com | **LinkedIn.com/Allison.OConnell**

Education
B.A. in Marketing, Digital Marketing Minor | Massachusetts Institute of Technology | 2010
Professional Experience
Marketing Coordinator | Local Guiding | 2021 – present
Crafted SEO-friendly website, blog, and social media content, telling the organization's story across platforms.
Digital Marketing Coordinator | Cloud Sales Support | 2017 – 2021
Supported a team of four marketing account executives responsible for 12 client accounts in the healthcare industry.
Volunteer Experience
Volunteer Treasurer | Riverview Neighborhood Association | 2022 - present

Figure 7. Example of two fictional resumes shown to respondents. Both applicants have similar backgrounds and amount of experience. The two resumes differ by the signal of motherhood. In the resume on the left, the applicant is a member of the parent–teacher association while the applicant on the right volunteers with a neighborhood association.

Context Attribute An element of a profile	**Attribute Values** Randomized across profile pairs	**Two Signals of Each Value** Randomly permuted within profile pairs
Race	Black	• Keisha Mosely • Tanisha Rivers
	White	• Allison O'Connell • Laurie Schmitt
Educational institution rank	High Rank	• University of Pennsylvania • Massachusetts Institute of Technology
	Mid Rank	• California State University, Long Beach • San Diego State University

Figure 8. Context attributes, values and signals: Motherhood illustration. Analogous to Figure 3, each profile has a set of context attributes: all attributes other than the focal attribute. Our design adaptively randomizes a vector of attribute values across profile pairs; within the two profiles share identical attribute values. To ensure that the two profiles in the pair do not appear identical to each other, we randomly permute two signals of each value across the profiles. Our experimental design therefore requires the researcher to specify two signals for each attribute value. Appendix H explains how we chose these signals.

https://doi.org/10.1017/pan.2026.10040

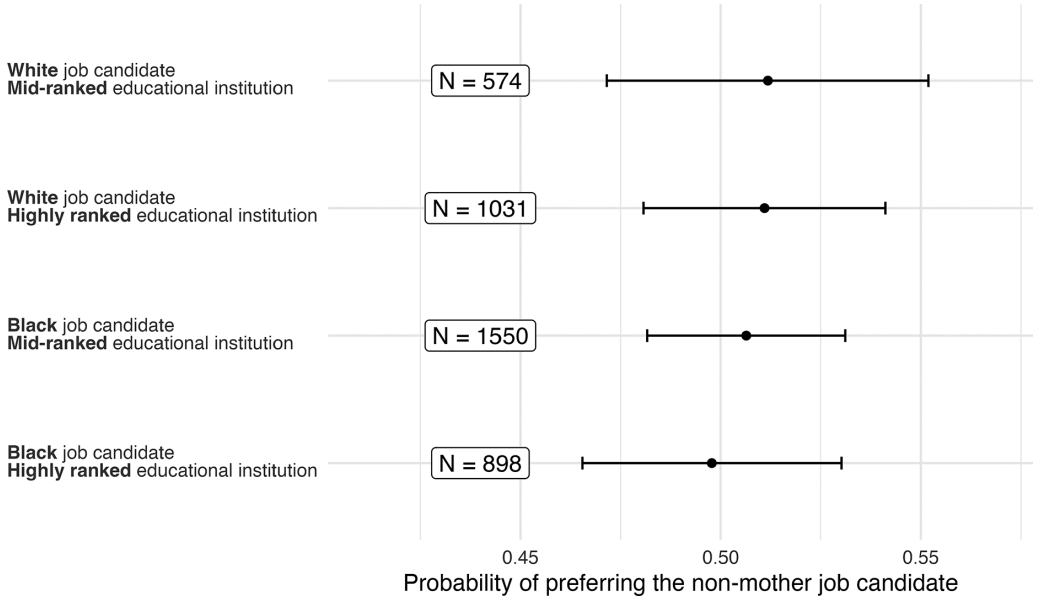


Figure 9. Estimated preference for the non-mother job candidate. The *y*-axis shows the labels for each context. The *x*-axis shows the posterior probability within each of those contexts, along with 95% credible intervals. The estimates in this figure are all from the Warm-up + Adaptive phases of our experiment.

The motherhood illustration shows both limitations and strengths of the adaptive randomization design. Adaptive randomization discovers heterogeneous effects. In scenarios where little or no heterogeneity exists, our method will not discover heterogeneity. But this is also a strength. In a single-context experiment, critics may argue that a null finding would have been significant in a different context. But when an adaptive design indicates homogeneity, there is compelling evidence that none of the contexts have especially large effects.

5. Discussion

This article introduces an adaptive experimental design for conjoint experiments. By adaptively adjusting randomization probabilities via Thompson sampling, our method efficiently identifies the contexts where the focal attribute has the most positive and most negative effects.

The adaptive randomization design complements the standard conjoint design. The standard design produces an additive summary of marginal attribute effects, or perhaps marginal interactions. The adaptive design searches for the particular contexts (e.g., work experience attribute values) at which the effect of the focal attribute (e.g., education) is most positive and most negative. Because of their distinct goals, the standard and adaptive designs have very different structures. While the standard conjoint design considers all attributes in parallel, the adaptive conjoint design defines one focal attribute (e.g., education) and explores its heterogeneous effects across multiple contexts (e.g., different values of work experience and country of origin). The adaptive design randomizes the context values by Thompson sampling and then randomly permutes signals of the focal attribute and the context value signals across two profiles.

The adaptive conjoint solves the problem of a moderately high-dimensional estimand in a different way from a standard conjoint. A standard conjoint solves the curse of dimensionality by marginalizing: scatter sample uniformly over a large space of attribute values and then summarize by estimands that marginalize over most attributes. An adaptive conjoint solves the curse of dimensionality by changing the randomization process: begin scattering uniformly but gradually concentrate the sample on two

https://doi.org/10.1017/pan.2026.10040

chosen contexts, ultimately gathering enough information to produce two context-specific estimates. The adaptive design works best in settings, where the dimensionality is large but still small enough that all contexts can be explored with data. In these settings, the adaptive design searches for heterogeneity at a cost: we do not obtain well-powered estimates for all attributes. Instead of the additive average effects of all attributes, we estimate the effect of one focal attribute in two specific contexts discovered data-adaptively.

To create the adaptive conjoint design, this article makes three main contributions: (1) We conceptualize conjoint experiments in a new way, with one focal attribute and other context attributes. (2) We define two signals (e.g., Germany and Poland) for each context attribute value (e.g., Europe), so that profiles differ within one context. (3) We provide a three-phase framework for randomization in which the warm-up reduces sensitivity to the prior, the adaptive phase discovers the estimands of interest, and the validation phase yields valid statistical inferences. We illustrate the utility of the framework through two illustrations and simulated evidence (Appendix D).

A noteworthy tradeoff involves participant suspicion. A traditional conjoint infers potentially sensitive preferences from observed choices. It avoids social desirability bias because preferences emerge only in the aggregate; any individual choice could be attributed to any of several differing attributes (Hainmueller, Hangartner, and Yamamoto 2015; Hainmueller *et al.* 2014). Our design risks greater suspicion; profiles within a pair are similar except the focal attribute. We reduce suspicion in two ways. First, we construct two signals of each context so the pair are not identical. Second, while a traditional conjoint participant may complete several choice tasks (Hainmueller *et al.* 2014), we present only one task to each respondent. Researchers must weight these conceptual and statistical difficulties in our design against the benefit of discovering causal interactions through adaptive randomization.

Our approach speaks indirectly to the broader literature on factorial and multi-arm experiments (Kasy and Sautmann 2021; Offer-Westort *et al.* 2021; Villar *et al.* 2015). This literature often emphasizes the search for an arm that maximizes an unknown parameter or payoff. Researchers might also search for the arm that minimizes the value of that parameter. By searching for both, researchers can summarize the total range of parameter values. Our approach also illustrates how to address inferential challenges of adaptive experiments through a simple yet powerful solution: carry out an adaptive phase to discover estimands of interest, followed by a validation phase to estimate the values of those estimands.

The adaptive conjoint design points to many areas for future research within the domain of conjoint experiments. First, there are open questions about the optimal allocation of sample size to the warm-up, adaptive and validation phases. Second, adaptive randomization could apply in real-world audit studies to discover heterogeneous discrimination by applicant attributes. A difficulty is the time lag: the researcher submits a fictitious resume to a job and then waits a month to see if there is a callback. Future methodological research that improves the speed of data collection in audit studies is needed to make these designs adaptive. Third, future research could explore model-based strategies to pool information across contexts, thereby extending our approach from moderately-large settings with ≈ 16 contexts to truly high-dimensional settings with hundreds or thousands of contexts.

Acknowledgements. We thank Elizabeth Moon for research assistance and helpful conversations. For feedback on this project, we thank Allison Koenecke, Soonhong Cho, Nanum Jeon, Nina Naffziger, Tanvi Shinkre and attendees at the 2024 American Causal Inference Conference, the 2024 Conference on Digital Experimentation and the 2024 Annual Meeting of the American Sociological Association.

Funding statement. Research reported in this article was supported by the Department of Information Science at Cornell University. This material is also based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 2139899. Any opinion, findings and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors also benefited from facilities and resources provided by the California Center for Population Research at UCLA (CCPR), which receives core support (P2C-HD041022) from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). The content is solely the responsibility of the authors and does not necessarily represent the official views of the Eunice Kennedy Shriver National Institute of Child Health and Human Development or the National Institutes of Health.

https://doi.org/10.1017/pan.2026.10040

Data availability statement. A replication package is archived on Dataverse at doi.org/10.7910/DVN/7EBDDY (Gosciak *et al.* 2026) and is also available at github.com/jennahgosciak/adaptive_conjoint. Generalizable code infrastructure for hosting an adaptively randomized conjoint experiment is available at github.com/dmolitor/adaptive-infra.

Competing interests. The authors declare none.

Ethical standards. The collection and analysis of data for this project were reviewed by the Cornell University IRB under protocols IRB0148142 and IRB0147962. Our study is pre-registrated at osf.io/ctd54. The second empirical illustration is pre-registered at osf.io/ekjbm.

References

Bansak, K., J. Hainmueller, D. J. Hopkins, and T. Yamamoto. 2018. "The Number of Choice Tasks and Survey Satisficing in Conjoint Experiments." *Political Analysis* 26 (1): 112–119.

Bansak, K., J. Hainmueller, D. J. Hopkins, T. Yamamoto, J. N. Druckman, and D. P. Green. 2021. "Conjoint Survey Experiments." *Advances in Experimental Political Science* 19: 19–41.

Bertrand, M., and S. Mullainathan. 2004. "Are Emily and Greg more Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination." *American Economic Review* 94 (4): 991–1013.

Budig, M. J., and P. England. 2001. "The Wage Penalty for Motherhood." *American Sociological Review* 66 (2): 204–225.

Carlson, E. 2015. "Ethnic Voting and Accountability in Africa: A Choice Experiment in Uganda." *World Politics* 67 (2): 353–385.

Chow, S.-C., and M. Chang. 2008. "Adaptive Design Methods in Clinical Trials—A Review." *Orphanet Journal of Rare Diseases* 3 (1): 1–13.

Correll, S. J., S. Benard, and I. Paik. 2007. "Getting a Job: Is there a Motherhood Penalty?" *American Journal of Sociology* 112 (5): 1297–1338.

Dasgupta, T., N. S. Pillai, and D. B. Rubin. 2015. "Causal Inference from 2^k Factorial Designs by Using Potential Outcomes." *Journal of the Royal Statistical Society Series B: Statistical Methodology* 77 (4): 727–753.

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

Deliu, N. 2024. "Multinomial Thompson Sampling for Rating Scales and Prior Considerations for Calibrating Uncertainty." *Statistical Methods & Applications* 33 (2): 439–469.

Egami, N., and K. Imai. 2019. "Causal Interaction in Factorial Experiments: Application to Conjoint Analysis." *Journal of the American Statistical Association* 114 (526): 529–540. https://doi.org/10.1080/01621459.2018.1476246.

Gosciak, J., D. Molitor, and I. Lundberg. 2026. "Replication Data for: Adaptive Randomization in Conjoint Survey Experiments." Harvard Dataverse V1. DOI: https://doi.org/10.7910/DVN/7EBDDY.

Hadad, V., D. A. Hirshberg, R. Zhan, S. Wager, and S. Athey. 2021. "Confidence Intervals for Policy Evaluation in Adaptive Experiments." *Proceedings of the National Academy of Sciences* 118 (15): e2014602118.

Hainmueller, J., D. Hangartner, and T. Yamamoto. 2015. "Validating Vignette and Conjoint Survey Experiments against Real-World Behavior." *Proceedings of the National Academy of Sciences* 112 (8): 2395–2400.

Hainmueller, J., and D. J. Hopkins. 2015. "The Hidden American Immigration Consensus: A Conjoint Analysis of Attitudes toward Immigrants." *American Journal of Political Science* 59 (3): 529–548.

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

Ham, D. W., K. Imai, and L. Janson. 2024. "Using Machine Learning to Test Causal Hypotheses in Conjoint Analysis." *Political Analysis* 32 (3): 329–344.

Ishizuka, P. 2021. "The Motherhood Penalty in Context: Assessing Discrimination in a Polarized Labor Market." *Demography* 58 (4): 1275–1300.

Kasy, M., and A. Sautmann. 2021. "Adaptive Treatment Assignment in Experiments for Policy Choice." *Econometrica* 89 (1): 113–132.

Kleven, H., C. Landais, and J. E. Søgaard. 2019. "Children and Gender Inequality: Evidence from Denmark." *American Economic Journal: Applied Economics* 11 (4): 181–209.

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

Lundberg, S., and E. Rose. 2000. "Parenthood and the Earnings of Married Men and Women." *Labour Economics* 7 (6): 689–710.

Milkman, K. L., et al. 2021. "A Megastudy of Text-Based Nudges Encouraging Patients to Get Vaccinated at an Upcoming Doctor's Appointment." *Proceedings of the National Academy of Sciences* 118 (20): e2101165118.

Milkman, K. L., et al. 2022. "A 680,000-Person Megastudy of Nudges to Encourage Vaccination in Pharmacies." *Proceedings of the National Academy of Sciences* 119 (6): e2115126119.

Offer-Westort, M., A. Coppock, and D. P. Green. 2021. "Adaptive Experimental Design: Prospects and Applications in Political Science." *American Journal of Political Science* 65 (4): 826–844.

Pallmann, P., et al. 2018. "Adaptive Designs in Clinical Trials: Why Use them, and How to Run and Report them." *BMC Medicine* 16: 1–15.

https://doi.org/10.1017/pan.2026.10040

Rosenberger, W. F., and J. M. Lachin. 1993. "The Use of Response-Adaptive Designs in Clinical Trials." *Controlled Clinical Trials* 14 (6): 471–484.
Ruggles, S., et al. 2025. "IPUMS USA: Version 16.0 [Dataset]." Minneapolis, MN: IPUMS. https://doi.org/10.18128/D010.V16.0
Sawtooth Software. (2024). "5 Examples of Conjoint Analysis Studies in the Real World." https://sawtoothsoftware.com/resources/blog/posts/5-conjoint-analysis-examples
Steffensmeier, D., J. Ulmer, and J. Kramer. 1998. "The Interaction of Race, Gender, and Age in Criminal Sentencing: The Punishment Cost of Being Young, Black, and Male." *Criminology* 36 (4): 763–798.
Thompson, W. R. 1933. "On the Likelihood that One Unknown Probability Exceeds another in View of the Evidence of Two Samples." *Biometrika* 25 (3–4): 285–294.
U.S. Census Bureau. 2016. "Frequently Occurring Surnames from the 2010 Census." https://www.census.gov/topics/population/genealogy/data/2010_surnames.html
Villar, S. S., J. Bowden, and J. Wason. 2015. "Multi-Armed Bandit Models for the Optimal Design of Clinical Trials: Benefits and Challenges." *Statistical Science* 30 (2): 199.
Voelkel, J. G., et al. 2024. "Megastudy Testing 25 Treatments to Reduce Antidemocratic Attitudes and Partisan Animosity." *Science* 386 (6719): eadh4764.
Yao, J., E. Brunskill, W. Pan, S. Murphy, and F. Doshi-Velez. 2021. "Power Constrained Bandits." In *Proceedings of the 6th Machine Learning for Healthcare Conference*, vol. 149, 209–259. Available from https://proceedings.mlr.press/v149/yao21a.html.
Yinger, J. 1986. "Measuring Racial Discrimination with Fair Housing Audits: Caught in the Act." *The American Economic Review* 76 (5): 881–893. http://www.jstor.org/stable/1816458.
Zhang, K., L. Janson, and S. Murphy. 2020. "Inference for Batched Bandits." In *Advances in Neural Information Processing Systems*, vol. 33, 9818–9829.
Zhang, K., L. Janson, and S. Murphy. 2021. "Statistical Inference with M-Estimators on Adaptively Collected Data." In *Advances in Neural Information Processing Systems*, vol. 34, 7460–7471.

Appendices

A. Poststratified estimates

The estimates reported in the main text are unweighted estimates that speak to the population of Prolific respondents. To make better inferences toward a population of U.S. adults, we alternatively produce poststratified estimates using weights derived from the 2022 American Community Survey (ACS), a nationally representative probability sample that we accessed via the Integrated Public Use Microdata Series (IPUMS; Ruggles *et al.* 2025).

In the ACS, we define population strata by age, sex, race (Black or African American, White, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, Other, Two or More Races, Prefer not to disclose) and Hispanicity (Hispanic and not Hispanic). We sum the person weight in each cell defined by the interaction of these variables to produce population counts.

In the Prolific survey responses, we first estimate an additive logistic regression model for the outcome given age, sex, race and Hispanicity. Then we make predictions for every stratum in our poststratification weights. Each time we make a

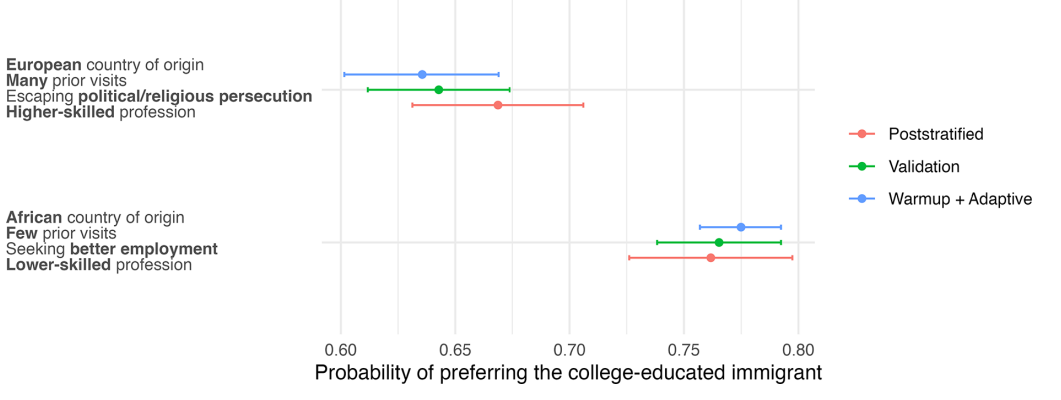


Figure A1. Validation results with post-stratification: Estimated preference for the college-educated immigrant profile. The *x*-axis shows the contexts $\vec{x}_{\text{Max}}$ and $\vec{x}_{\text{Min}}$ that were identified in the adaptive experimental phase as having the highest and lowest posterior probabilities that the respondent would choose the more-educated prospective immigrant within the pair. The *y*-axis depicts the estimated probability of choosing the more educated immigrant within each of those contexts, along with 95% credible intervals. We also report post-stratified estimates and 95% confidence intervals using U.S. population weights for demographic characteristics from the 2022 American Community Survey, accessed via IPUMS (Ruggles *et al*. 2025).

https://doi.org/10.1017/pan.2026.10040

poststratified estimate, we drop any ACS cells that involve a predictor that is not observed among units with the treatment of interest in the Prolific data (e.g., race = prefer not to disclose). Thus, our poststratified estimates correspond to the subpopulation for whom inference is feasible. We report the weighted mean estimate over the poststratification cells, weighted by population size.

This approach is valid to the degree that whether a U.S. adult participates in our Prolific survey is independent of the value of Y that we would observe if they were to participate, conditional on the measured demographic variables. As with all population inferences from non-probability samples, this identifying assumption is likely imperfect. We nonetheless note that our post-stratified estimates (Figure A1) are very similar to the unweighted sample mean estimates reported in the main text.

B. Causal identification proof

This section proves that the context-specific causal estimands $\theta(\vec{x}_{\text{Max}})$ and $\theta(\vec{x}_{\text{Min}})$ are identified by the validation phase of the experiment:

$$\theta(\vec{x}) = \mathbf{E}_{S,A}\Big[\mathrm{P}\big(Y(\vec{x},S,A) = A\big)\Big]. \tag{B.1}$$

By definition of expectation,

$$= \sum_{s,a} \mathrm{P}(S = s, A = a)\mathrm{P}\big(Y(\vec{x},s,a) = a\big). \tag{B.2}$$

By randomization, $Y(\vec{x},s,a) \perp\!\!\!\perp \{X,S,A\}$ in the validation phase:

$$= \sum_{s,a} \mathrm{P}(S = s, A = a)\mathrm{P}\Big(Y(\vec{x},s,a) = a \mid \vec{X} = \vec{x}, S = s, A = a\Big). \tag{B.3}$$

By consistency, $Y = Y(\vec{X},S,A)$

$$= \sum_{s,a} \mathrm{P}(S = s, A = a \mid \vec{X} = \vec{x})\mathrm{P}\Big(Y = a \mid \vec{X} = \vec{x}, S = s, A = a\Big). \tag{B.4}$$

C. Signal effects

Our research design involves two signals for each context. For example, immigrant profiles from Europe are randomized to one of two signals: Germany or Poland. Arbitrarily selecting one of the vector of signals within each context to be the primary vector, recall that $S \in \{L,R\}$ indicates which of the profiles shows this primary value (with the other value appearing on the other profile).

Just as the causal effect of the focal attribute is identified, our design also identifies the causal effect of the signal permutation. We focus here on the probability that a respondents choice $Y \in \{L,R\}$ equals the profile $S \in \{L,R\}$ that shows the primary signal value. To reduce complications from adaptivity, we carry out the proof conditional on $\vec{X}$.

$$\mathbf{E}_{S,A}\Big[\mathrm{P}\big(Y(\vec{x},S,A) = S \mid \vec{X} = \vec{x}\big)\Big]. \tag{C.1}$$

By definition of expectation,

$$= \sum_{s,a} \mathrm{P}(S = s, A = a \mid \vec{X} = \vec{x})\mathrm{P}\big(Y(\vec{x},s,a) = s \mid \vec{X} = \vec{x}\big). \tag{C.2}$$

By ignorability, $Y(\vec{x},s,a) \perp\!\!\!\perp \{S,A\} \mid \vec{X}$

$$= \sum_{s,a} \mathrm{P}(S = s, A = a \mid \vec{X} = \vec{x})\mathrm{P}\Big(Y(\vec{x},s,a) = s \mid \vec{X} = \vec{x}, S = s, A = a\Big). \tag{C.3}$$

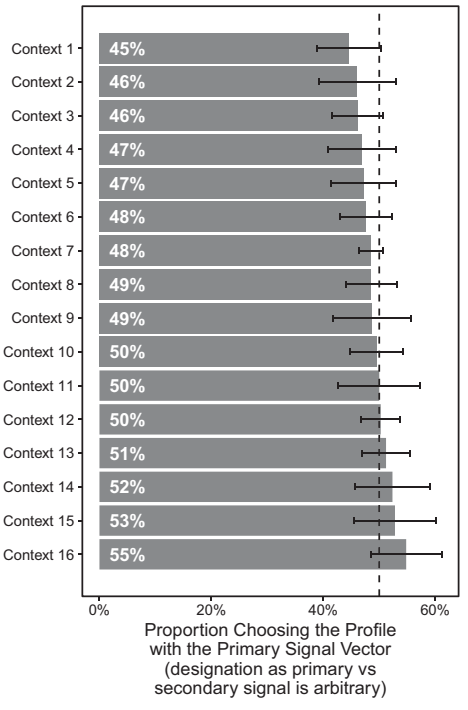
By consistency, $Y = Y(\vec{X},S,A)$

$$= \sum_{s,a} \mathrm{P}(S = s, A = a \mid \vec{X} = \vec{x})\mathrm{P}\Big(Y = s \mid \vec{X} = \vec{x}, S = s, A = a\Big). \tag{C.4}$$

Figure C1 shows that in all 16 contexts of the immigrant profile illustration, between 45% and 55% of respondents chose the primary signal value. This suggests that whether a profile shows the primary or secondary value of the context signal has little bearing on respondent choices, on average. By contrast, the focal attribute (education) had a large effect on respondent choices: in every context, more than 63% chose the profile signaling a college degree.

In the job applicants illustration, there was a slight difference in how randomization occurred. In the immigrant profile illustration, the entire vector of signals was randomized together to the left or right profile. In the job applicants illustration, each individual element of the vector of context signals was randomized to the left or right profile. This difference makes it possible in the job applicant illustration to look at the choice probability as a function of each individual signal element.

https://doi.org/10.1017/pan.2026.10040



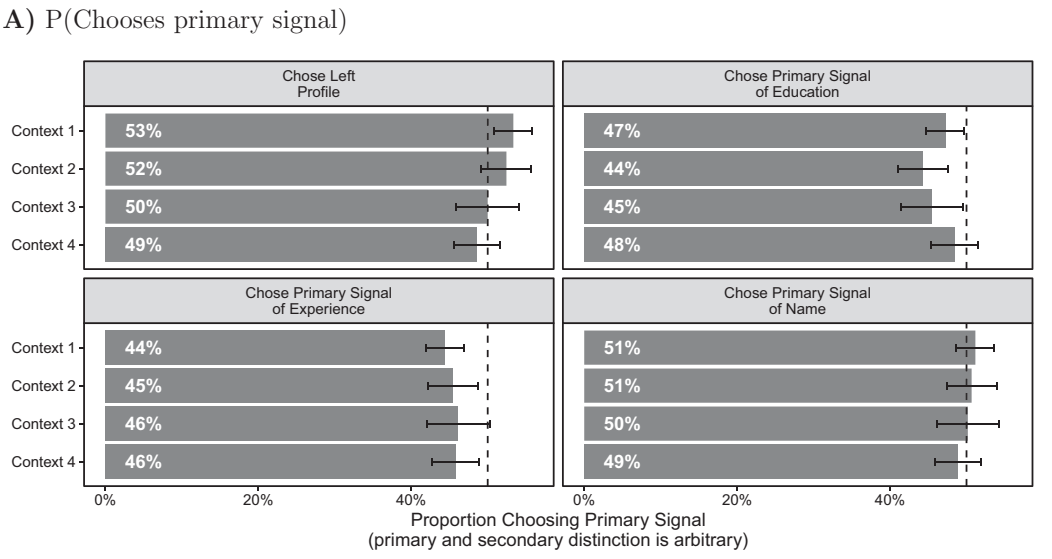
B) Signal definitions

Context	Primary Signal	Secondary Signal
Context 1: 45% Chose Primary Signal	Germany Construction Worker Escalating violence in home country Two trips, both during the holidays	Poland Information Technology Specialist Escaping political/religious persecution Has visited once for vacation
Context 2: 46% Chose Primary Signal	Somalia Construction Worker Escalating violence in home country Two trips, both during the holidays	Sudan Information Technology Specialist Escaping political/religious persecution Has visited once for vacation
Context 3: 46% Chose Primary Signal	Somalia Construction Worker Greater access to high−paying jobs Has visited the U.S. many times as a tourist	Sudan Information Technology Specialist Seeking better employment opportunities Has traveled to the U.S. every other summer since childhood
Context 4: 47% Chose Primary Signal	Somalia Janitor Escalating violence in home country Has visited the U.S. many times as a tourist	Sudan Waiter Escaping political/religious persecution Has traveled to the U.S. every other summer since childhood
Context 5: 47% Chose Primary Signal	Somalia Construction Worker Greater access to high−paying jobs Two trips, both during the holidays	Sudan Information Technology Specialist Seeking better employment opportunities Has visited once for vacation
Context 6: 48% Chose Primary Signal	Germany Construction Worker Greater access to high−paying jobs Has visited the U.S. many times as a tourist	Poland Information Technology Specialist Seeking better employment opportunities Has traveled to the U.S. every other summer since childhood
Context 7: 48% Chose Primary Signal	Somalia Janitor Greater access to high−paying jobs Two trips, both during the holidays	Sudan Waiter Seeking better employment opportunities Has visited once for vacation
Context 8: 49% Chose Primary Signal	Germany Janitor Greater access to high−paying jobs Two trips, both during the holidays	Poland Waiter Seeking better employment opportunities Has visited once for vacation
Context 9: 49% Chose Primary Signal	Somalia Janitor Greater access to high−paying jobs Has visited the U.S. many times as a tourist	Sudan Waiter Seeking better employment opportunities Has traveled to the U.S. every other summer since childhood
Context 10: 50% Chose Primary Signal	Germany Construction Worker Greater access to high−paying jobs Two trips, both during the holidays	Poland Information Technology Specialist Seeking better employment opportunities Has visited once for vacation
Context 11: 50% Chose Primary Signal	Germany Janitor Greater access to high−paying jobs Has visited the U.S. many times as a tourist	Poland Waiter Seeking better employment opportunities Has traveled to the U.S. every other summer since childhood
Context 12: 50% Chose Primary Signal	Germany Construction Worker Escalating violence in home country Has visited the U.S. many times as a tourist	Poland Information Technology Specialist Escaping political/religious persecution Has traveled to the U.S. every other summer since childhood
Context 13: 51% Chose Primary Signal	Somalia Janitor Escalating violence in home country Two trips, both during the holidays	Sudan Waiter Escaping political/religious persecution Has visited once for vacation
Context 14: 52% Chose Primary Signal	Germany Janitor Escalating violence in home country Has visited the U.S. many times as a tourist	Poland Waiter Escaping political/religious persecution Has traveled to the U.S. every other summer since childhood
Context 15: 53% Chose Primary Signal	Somalia Construction Worker Escalating violence in home country Has visited the U.S. many times as a tourist	Sudan Information Technology Specialist Escaping political/religious persecution Has traveled to the U.S. every other summer since childhood
Context 16: 55% Chose Primary Signal	Germany Janitor Escalating violence in home country Two trips, both during the holidays	Poland Waiter Escaping political/religious persecution Has visited once for vacation

Figure C1. Signal effects are minimal: Immigrant illustration. Within every context, approximately 50% of respondents chose the profile that showed the primary signal value (the choice of which is primary and which is secondary is arbitrary). Meanwhile, more than 63% chose the profile signaling a college degree (Figure 5). The random permutation of signals of the same context is far less consequential than the random permutation of the focal attribute in our illustration.

Figure C2 presents these results. First, across contexts, participants chose the left profile over the right profile between 49% and 53% of the time, suggesting minimal effects of profile ordering. For each context attribute, the probability of choosing any particular signal was always within ±6 percentage points of 50%. This also suggests that the signals had effectively zero effect on the choices.

https://doi.org/10.1017/pan.2026.10040



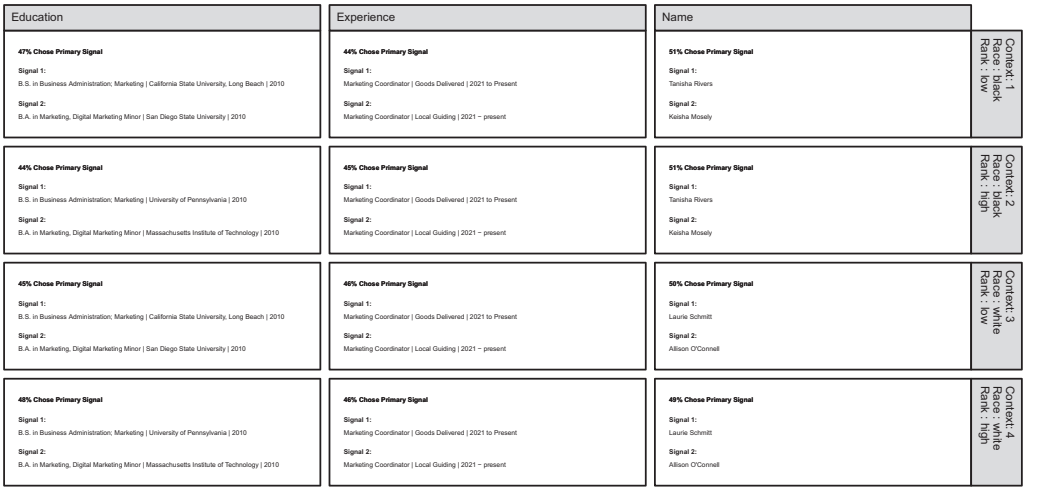


Figure C2. Signal effects are minimal: Job applicant illustration. Within every context, the rate of choosing a profile was close to 50% regardless of the signal used for each element of the context.

To summarize, the absence of signal effects is not a requirement of our research design, as average effects marginalized over signals are well-defined and identified even in the presence of substantial signal effects. But interpretations are simplest when this marginalization is inconsequential, that is, when the randomization of signal vectors has no effect on choices, as in our example.

D. Efficiency gains from adaptive randomization

This section presents simulation results to demonstrate both the efficiency and effectiveness of the adaptive approach relative to fixed randomization. To understand how the adaptive algorithm scales to higher-dimensional treatment settings, we allow the number of contexts to take values of $9, 12, 15, \ldots, 30$. Every context has a unique value for the probability that a respondent assigned to that context will choose the profile signaling a particular value of the focal attribute, ranging from 30% to 70%. Since the adaptive algorithm is sensitive to heterogeneity, we fix the difference in probabilities between the two contexts with the largest probabilities even as we vary the total number of contexts in the simulation.

https://doi.org/10.1017/pan.2026.10040

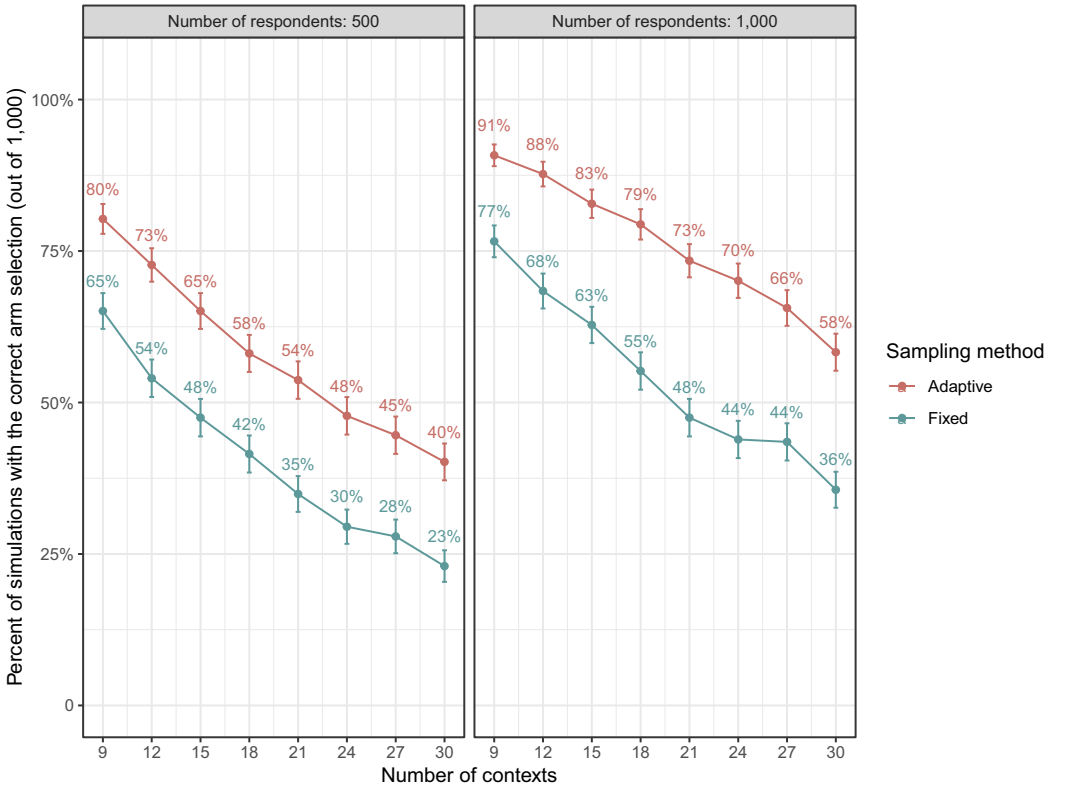


Figure D1. Percentage of simulations with the correct arm selection. Figure shows the mean number of times over 1,000 simulations in which the model selects the correct arm using either the adaptive or fixed randomization approach. For the same number of sample respondents, adaptive randomization selects the correct arm with greater accuracy. Note that all percentages can be made higher by increasing the fixed sample size; a trial with 30 contexts might call for a sample size much larger than 1,000 respondents.

$$
\begin{aligned}
&\text{Setting with 9 contexts:} && \theta(\vec{x}_1),\theta(\vec{x}_2),\ldots,\theta(\vec{x}_8),\theta(\vec{x}_9) = [0.3, 0.34, \ldots, 0.65, 0.7] \\
&\text{Setting with 12 contexts:} && \theta(\vec{x}_1),\theta(\vec{x}_2),\ldots,\theta(\vec{x}_{11}),\theta(\vec{x}_{12}) = [0.3, 0.33, \ldots, 0.65, 0.7] \\
&\vdots && \vdots \\
&\text{Setting with 30 contexts:} && \theta(\vec{x}_1),\theta(\vec{x}_2),\ldots,\theta(\vec{x}_{29}),\theta(\vec{x}_{30}) = [0.3, 0.31, \ldots, 0.65, 0.7]
\end{aligned}
$$

We simulate data gathered by two randomization procedures. Let K represent the total number of contexts in a simulation. Under fixed randomization, respondents are assigned to every arm with equal probability:

$$\mathrm{P}(\vec{X} = \vec{x} \mid \mathbf{X}, \vec{Y}) = \frac{1}{K} \qquad \forall \quad \vec{x}. \tag{D.1}$$

Under adaptive randomization, we first conduct a warm-up phase for 100 observations following equal randomization. Thereafter, we assign contexts proportional to the posterior probability that a context has the largest value of $\theta(\vec{x})$,[7]

$$\mathrm{P}(\vec{X} = \vec{x} \mid \mathbf{X}, \vec{Y}) = \mathrm{P}\Big(\theta(\vec{x}) = \max_{\vec{x}} \ \theta(\vec{x}) \mid \mathbf{X}, \vec{Y}\Big). \tag{D.2}$$

We simulate a draw of $\vec{x}$ by drawing from the posterior beta distribution across all $\theta(\vec{x})$ and assigning profiles to the context with the largest draw.

[7]This is the same procedure described in Section 3.1.2. Otherwise, we make two modifications. To reduce the runtime, we only draw from the Beta distribution 1,000 times when fully estimating the posterior probability. Additionally, the first draw from the adaptive phase is initialized with equal randomization; all draws thereafter use the posterior probability that a context has the largest value of $\theta(\vec{x})$ based on data from both the warm-up and adaptive phases.

https://doi.org/10.1017/pan.2026.10040

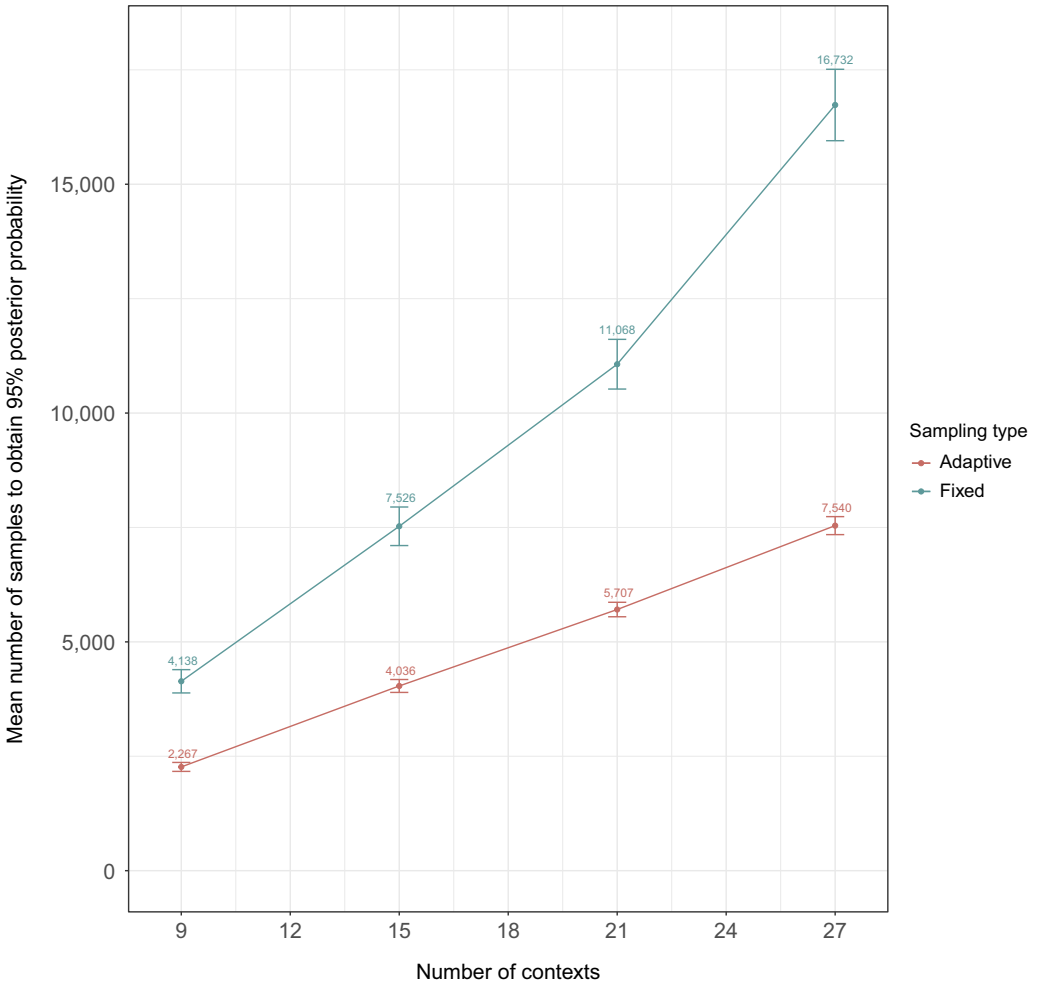


Figure D2. Average sample size in adaptive versus fixed randomization approaches. The adaptive approach requires fewer sample respondents on average to select the arm with the largest value of $\theta(\vec{x})$ with 95% posterior probability.

We evaluate both the efficiency and effectiveness of the adaptive approach relative to fixed randomization. First, we consider how often both the adaptive and fixed randomization approaches correctly identify the context with the largest value of $\theta(\vec{x})$ for a fixed number of respondents. Figure D1 visualizes the percentage of simulations with the correct context arm selection across 1,000 simulations for 500 and 1,000 respondents. In this illustration, the adaptive approach is substantially more accurate than the fixed randomization approach and, with only 1,000 respondents, maintains accuracy of 70% even as the number of contexts increases to 24. Readers should note that both the adaptive and fixed randomization designs can be made more certain by collecting larger samples; for a trial with 30 contexts, one may want many more than n = 1,000 participants, and perhaps would want to continue the trial until crossing a pre-determined threshold, such as 95% posterior probability of having chosen the true maximum arm.

Second, and following the alternative strategy above, we estimate the average sample size needed to obtain at least 95% posterior probability estimates of selecting the maximum arm. We simulate both the fixed and adaptive randomization approaches and calculate the posterior probability that each context has the largest value of $\theta(\vec{x})$ at increments of 100 respondents. We record the sample size when the posterior probability for any context is greater than 95%. For the adaptive randomization approach, we allocate 100 respondents to a warm-up phase in which contexts are assigned with equal probability.

Figure D2 demonstrates the average sample sizes needed to obtain 95% posterior probability estimates that a context has the largest value of $\theta(\vec{x})$. At all values for the number of contexts, the adaptive approach achieves this posterior probability at roughly half the number of respondents required by the fixed approach.

https://doi.org/10.1017/pan.2026.10040

In high-dimensional treatment settings where the number of treatments may be 21 or more—settings similar to large megastudies, such as Milkman *et al.* (2021, 2022) and Voelkel *et al.* (2024)—adaptive randomization can lead to substantial efficiency gains. These gains may be particularly useful to researchers who face resource constraints and may not be able to run a large megastudy, but who have reason to believe there is substantial effect heterogeneity.

E. Low-order interactions may understate the range of effects

Our approach searches for the maximum and minimum contexts in a world that may be characterized by high-order interactions between the focal attribute A and the attribute vector $\vec{X}$. One might ask: would estimates be nearly as good under a linear regression model that assumes low-order interactions?

As an illustration of this point, consider a hypothetical setting with three binary context attributes x_1, x_2 and x_3. Suppose all $2^3 = 8$ vectors $\vec{x}$ are assigned to respondents with equal probabilities. Suppose the choice probability follows a functional form involving three-way interactions among all these variables:

$$\theta(\vec{x}) = \begin{cases} 0.1 & \text{if } x_1 + x_2 + x_3 = 0, \quad \text{i.e., all } x_j = 0 \\ 0.8 & \text{if } x_1 + x_2 + x_3 = 1, \quad \text{i.e., two } x_j = 0, \text{ one } x_j = 1 \\ 0.2 & \text{if } x_1 + x_2 + x_3 = 2, \quad \text{i.e., one } x_j = 0, \text{ two } x_j = 1 \\ 0.9 & \text{if } x_1 + x_2 + x_3 = 3, \quad \text{i.e., all } x_j = 1. \end{cases} \tag{E.1}$$

Now suppose a researcher considers two regression specifications: an additive linear model and a model with two-way interactions:

$$f_{\text{Linear}}(\vec{x}) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \tag{E.2}$$

$$f_{\text{Two-Way Interactions}}(\vec{x}) = \gamma_0 + \gamma_1 x_1 + \gamma_2 x_2 + \gamma_3 x_3 + \eta_{12} x_1 x_2 + \eta_{13} x_1 x_3 + \eta_{23} x_2 x_3, \tag{E.3}$$

where both $f_{\text{Linear}}(\vec{x})$ and $f_{\text{Two-Way Interactions}}(\vec{x})$ are meant as parametric approximations to the unknown $\theta(\vec{x})$ function.

For the data generating process from Equation (E.1), the best linear approximation is the same under either parametric model (E.2) or (E.3). The intercept is $\beta_0 = \gamma_0 = 0.425$ for both models, and the main effect coefficients are $\beta_j = \gamma_j = 0.05$ for $j = 1,2,3$. The interaction coefficients η are all zero. While the true value of $\theta(\vec{x})$ ranges from 0.1 to 0.9, the model-approximation values $f_{\text{Linear}}(\vec{x})$ and $f_{\text{Two-Way Interactions}}(\vec{x})$ range only from 0.425 to 0.575. The total range of the model-based predictions ($0.575 - 0.425 = 0.15$) is only 19% of the range of the true values ($0.9 - 0.1 = 0.8$).

In this example, what would a researcher who relied on low-order interactions conclude? The researcher would conclude first that there are no two-way interactions, since the interaction terms in Equation (E.3) are asymptotically zero. The researcher would also conclude (mistakenly) that the total range of $\theta(\vec{x})$ seems to be very tight around 0.5 based on their models, while in fact the range extends from 0.1 to 0.9.

This example illustrates a simple yet general point: models that assume low-order interactions can miss patterns that require high-order interactions. Analogously, a researcher who wants to study the full range of effects that a focal attribute may have across context attribute values might miss some of this range if they rely on approaches that summarize by low-order interactions.

There are many scientific settings when one might want summaries focused on low-order interactions, such as when a theoretical argument involves the interaction of a few attributes, and in those settings, researchers would be well-served by existing methods for interactions in conjoints. In other settings, such as when one wants to summarize the full range of effects that a focal attribute may have, one would be better served by the adaptive randomization approach developed in this article.

F. Mathematical comparison to marginal estimands

Researchers who study conjoint experiments often summarize results by a set of marginal estimands. This section defines the analogs of those estimands in our design, focusing on derivations that apply to our validation phase. These derivations show that several of these estimands have values that are fixed by an element of our experimental design: forced choice between pairs with fixed $\vec{X}$ and differing A. An overall summary of this section is that adaptive randomization serves different research goals than traditional conjoint randomization. Researchers who want to study marginal effects and marginal interactions should follow the traditional design, whereas researchers who want to study the full range of heterogeneous choice probabilities should follow the adaptive design. Neither design answers the other's research question well.

F.1. Modified notation

Modified notation is necessary to make these connections. Our main text notation for the adaptive conjoint assumes that the pair of profiles share the context attribute vector $\vec{X}$, with one outcome per pair $Y \in \{L, R\}$ indicating choice of the left or right profile, and one focal attribute $A \in \{L, R\}$ defined for the pair. Because traditional conjoints independently randomize all

https://doi.org/10.1017/pan.2026.10040

attributes, many conjoint estimands are defined using notation that indexes on the profile within the pair. Therefore, for this section only, we define variables for each profile. Let $\vec{X}_j$ indicate the context attributes of profile $j \in \{1,2\}$, where in our design, $\vec{X}_1 = \vec{X}_2$ always but in a typical conjoint design, these two are not necessarily equal.

For profile $j \in \{1,2\}$, let $A_j \in \{0,1\}$ indicate the focal attribute in profile j, let $S_j \in \{0,1\}$ indicate whether profile j displays the primary value of the signal (where primary and secondary designations are arbitrary) and let $Y_j \in \{0,1\}$ indicate whether profile j was chosen. Note that $\sum_{j=1}^{2} Y_j = 1$ due to the forced choice design. In this modified notation, the potential outcome for profile j is $Y_j(\vec{x}_j, s_j, a_j)$, which is a function of the context value $\vec{x}_j$, signal value s_j and focal attribute value a_j for profile j. In this modified notation, our choice parameter is

$$\theta(\vec{x}) = \mathbf{E}\Big(Y_j(\vec{X}_j = \vec{x}, S_j, A_j = 1)\Big), \tag{F.1}$$

where the expectation is taken over S_j and over Y_j. This is the probability profile j would be chosen if it were the treated profile ($A_j = 1$) showing context vector $\vec{X}_j = \vec{x}$.

Note that because of the forced-choice design with one of two profiles treated, the event "chooses the profile untreated on the focal attribute" is the complement of the event "chooses the profile treated on the focal attribute." Thus, the probability under focal attribute $A_j = 0$ is the complement of the outcome under focal attribute $A_j = 1$:

$$\mathbf{E}\Big(Y_j(\vec{X}_j = \vec{x}, S_j, A_j = 0)\Big) = 1 - \theta(\vec{x}). \tag{F.2}$$

Thus, the conditional causal effect of A_j within context $\vec{x}$ can be written in terms of $\theta(\vec{x})$:

$$\tau_{\text{ConditionalEffect}}(\vec{x}) = \mathbf{E}\Big(Y_j(\vec{X}_j = \vec{x}, S_j, A_j = 1)\Big) - \mathbf{E}\Big(Y_j(\vec{X}_j = \vec{x}, S_j, A_j = 0)\Big) \tag{F.3}$$

$$= \theta(\vec{x}) - \Big(1 - \theta(\vec{x})\Big). \tag{F.4}$$

As a concrete example, if 60% of respondents choose the college-educated profile within context $\vec{x}$, then the causal effect of a profile's college degree on the probability of being chosen within this context is $60\% - 40\% = 20\%$.

F.2. Connections to the average marginal component effect

The average marginal component effect (AMCE) in a traditional conjoint is the average effect of a particular attribute, marginalized over the joint distribution of all other attributes (Hainmueller *et al.* 2014). With the caveat that differences in the structure of randomization make the joint distribution very different in our design than in a traditional conjoint, an analogous causal estimand can be defined in our design:

$$\tau_{\text{Marginal},A} = \mathbf{E}\Big[Y(\vec{X}_j, S_j, A_j = 1) - Y(\vec{X}_j, S_j, A_j = 0)\Big] \tag{F.5}$$

$$= \mathbf{E}\Big[\theta(\vec{X}) - (1 - \theta(\vec{X}))\Big]. \tag{F.6}$$

If this estimand were defined within the validation phase of our design, the marginal estimand would be the unweighted average of $\tau_{\text{Conditional}}(\vec{x}_{\text{Min}})$ and $\tau_{\text{Conditional}}(\vec{x}_{\text{Max}})$, since these contexts are assigned with equal probability in the validation phase:

$$\tau_{\text{Marginal},A,\text{Validation}} = 0.5\Big[\theta(\vec{x}_{\text{Min}}) - (1 - \theta(\vec{x}_{\text{Min}}))\Big] + 0.5\Big[\theta(\vec{x}_{\text{Max}}) - (1 - \theta(\vec{x}_{\text{Max}}))\Big] \tag{F.7}$$

$$= \theta(\vec{x}_{\text{Min}}) + \theta(\vec{x}_{\text{Max}}) - 1 \tag{F.8}$$

$$= \theta(\vec{x}_{\text{Max}}) - (1 - \theta(\vec{x}_{\text{Min}})), \tag{F.9}$$

where the final line is arranged for comparison with a future subsequent proof below about interactive effects.

In general, the marginal estimand will be different from the conditional estimands to the degree that the conditional estimands are heterogeneous. In fact, debates about the distribution of other attributes over which to marginalize when estimating the effect of a chosen attribute (e.g., De la Cuesta *et al.* 2022) are premised on the assumption that the $\theta(\vec{x})$ estimands are heterogeneous over $\vec{x}$, so that the way one weights different attributes shapes the result.

The average marginal effect of $\vec{X}$ is zero by design because every pair of profiles shares $\vec{X}$ and involves a forced choice. At every $\vec{x}$-value, 50% of profiles are chosen. The proof below derives the zero effect for the comparison in the validation phase:

https://doi.org/10.1017/pan.2026.10040

$$\tau_{\text{Marginal},\vec{X}} = \mathbf{E}\Big[Y(\vec{X}_j = \vec{x}_{\text{Max}}, S_j, A_j) - Y(\vec{X}_j = \vec{x}_{\text{Min}}, S_j, A_j)\Big] \tag{F.10}$$

$$= \sum_{a=0}^{1} \mathrm{P}(A = a)\mathbf{E}\Big[Y(\vec{X}_j = \vec{x}_{\text{Max}}, S_j, A_j = a) - Y(\vec{X}_j = \vec{x}_{\text{Min}}, S_j, A_j = a)\Big] \tag{F.11}$$

$$= 0.5\Big(\mathbf{E}\Big[Y(\vec{X}_j = \vec{x}_{\text{Max}}, S_j, A_j = 0)\Big] - \mathbf{E}\Big[Y(\vec{X}_j = \vec{x}_{\text{Min}}, S_j, A_j = 0)\Big]\Big)$$
$$+ 0.5\Big(\mathbf{E}\Big[Y(\vec{X}_j = \vec{x}_{\text{Max}}, S_j, A_j = 1)\Big] - \mathbf{E}\Big[Y(\vec{X}_j = \vec{x}_{\text{Min}}, S_j, A_j = 1)\Big]\Big) \tag{F.12}$$

$$= 0.5\Big((1 - \theta(\vec{x}_{\text{Max}})) - (1 - \theta(\vec{x}_{\text{Min}}))\Big)$$
$$+ 0.5\Big(\theta(\vec{x}_{\text{Max}}) - \theta(\vec{x}_{\text{Min}})\Big) \tag{F.13}$$

$$= 0. \tag{F.14}$$

F.3. Connections to the average combination effect

One might also define the average combination effect (ACE) as the expected difference if both A and $\vec{X}$ are changed, similar to Egami and Imai (2019) and Dasgupta et al. (2015). The ACE in our setting takes a simple form:

$$\tau_{\text{Combination}} = \mathbf{E}\Big[Y(\vec{X}_j = \vec{x}_{\text{Max}}, S_j, A_j = 1) - Y(\vec{X}_j = \vec{x}_{\text{Min}}, S_j, A_j = 0)\Big]\Big] \tag{F.15}$$

$$= \theta(\vec{x}_{\text{Max}}) - (1 - \theta(\vec{x}_{\text{Min}})), \tag{F.16}$$

which is mathematically equivalent to the marginal effect of A (Equation (F.9)). To build intuition for this equivalence, note that the first component of $\tau_{\text{Combination}}$ is greater than 0.5 by the amount $\theta(\vec{x}_{\text{Max}}) - 0.5$. The second component is less than 0.5 by the amount $0.5 - (1 - \theta(\vec{x}_{\text{Min}}))$. The degree to which the first component exceeds 0.5 equals half the conditional effect of A in the max context, and the degree to which the second is below 0.5 equals half the conditional effect of A in the min context. Thus, the combination effect equals the unweighted mean of the two conditional effects of A, which thus equals the marginal effect of A.

F.4. Connections to the average marginal interaction effect

The average marginal interaction effect (AMIE; Egami and Imai 2019) is the average conditional effect with the marginal effects removed. But in our setting, the average marginal effect of $\vec{X}$ is zero by design and the average marginal effect of A equals the average combination effect, so that the AMIE is zero by design:

$$\tau_{\text{MarginalInteraction}} = \tau_{\text{Combination}} - \Big(\tau_{\text{Marginal},A} + \tau_{\text{Marginal},\vec{X}}\Big) \tag{F.17}$$

$$= \theta(\vec{x}_{\text{Max}}) - (1 - \theta(\vec{x}_{\text{Min}})) - \Big(\theta(\vec{x}_{\text{Max}}) - (1 - \theta(\vec{x}_{\text{Min}})) + 0\Big) \tag{F.18}$$

$$= 0. \tag{F.19}$$

F.5. Mathematical connections: Discussion

The goals of a traditional conjoint and an adaptive conjoint are not the same, and neither approach produces useful nor interesting estimates of the estimands targeted in the other. A traditional conjoint speaks only indirectly to the conditional choice probabilities $\theta(\vec{x})$ because it is rare for both profiles in a pair to match along all elements of $\vec{X}$ when those elements have been independently randomized. In fact, a traditional conjoint with this comparison would be suspicious to respondents. An adaptive conjoint involves a design constructed to estimate selected values of $\theta(\vec{x})$, but the construction of the design with forced choices between profiles matched on $\vec{X}$ leads to non-interesting fixed results for estimands involving the causal effect of $\vec{X}$.

While the mathematical estimands of each design do not translate well to the data collected by the other, the substantive motivations of the two are linked in ways that could lead to complementary studies by both designs. The AMCE estimands of a traditional conjoint (Hainmueller *et al.* 2014) can help researchers find attributes with potentially large effects. Researchers might be further motivated to analyze those attributes through an adaptive conjoint if a traditional conjoint yields evidence of strong interactions from estimands like the AMIE (Egami and Imai 2019) or through joint statistical tests that reject the null in the presence of interactive effects (Ham *et al.* 2024). Once a traditional conjoint discovers large marginal effects and evidence of interactions, there are opportunities to further explore heterogeneous choices through an adaptive design.

https://doi.org/10.1017/pan.2026.10040

G. Adaptive conjoint algorithm

Algorithm 1 presents the steps carried out during the data collection for an adaptive conjoint. We wrote this algorithm as an ideal case of the design, but our applications deviate slightly from the algorithm as noted below.

First, the code below is written as though each participant provides a response before the next participant enters the study. But in fact, sometimes a participant may enter the study while one or more previous participants have been assigned profiles but have not yet made a choice. Thus, in the adaptive phase of the first illustration (immigrant profiles), each participant's context is determined by the posterior distribution based on all previous completed observations. Because our experiment takes a very short time to complete, for practical purposes, the distinction is trivial. In the adaptive phase of the second illustration (job applicants), the adaptive phase proceeds in batches of 200 respondents so that each respondent's posterior draw is based on data through the end of the previous batch.

Second, in lines (11) and (23), the algorithm shows a single draw from a Beta distribution for each context, with line (13) assigning the context with the maximum simulated value and line (25) assigning the context with the minimum simulated value. In practice, our application code instead takes 100,000 draws from each Beta distribution and estimates π_{Highest} (or π_{Lowest}) by the proportion of draws in which each $\bar{x}$-value is the highest (lowest). We then assign by a categorical draw with these probabilities. Both approaches equivalently assign a context based on the posterior probabilities. The advantage of the approach shown below is computational efficiency. An advantage of the approach that we used in our empirical examples is transparency: the probabilities of treatment assignment are recorded consistently throughout the experiment. Another advantage of calculating the full posterior probabilities at each step is that a researcher could then use these posterior probabilities for stopping rules dynamically, at every step rather than in batches.

Third, in the validation phase, line (39) contexts are randomized to units. In fact, the validation phase of the immigrant illustration was carried out with two simultaneous experiments carried out online, with one experiment per selected context. Each experiment randomized the focal attribute and signals across every pair of profiles. Because the two experiments happened at the same time and had identical calls on Prolific, it is also reasonable to think of $\vec{X} \in \{\vec{x}_{\text{Min}}, \vec{x}_{\text{Max}}\}$ as randomized to respondents in the validation phase.

https://doi.org/10.1017/pan.2026.10040

Algorithm 1 Adaptive conjoint with warm-up, adaptive and validation phases.

Warmup Phase

Input: $n_{\text{Warmup}}, \vec{X}$

for $i = 1, 2, \ldots, n_{\text{Warmup}}$ **do**
  $\vec{x}_{w,i} \sim \text{DiscreteUniform}(\text{support}(\vec{X}))$ ▷ Assign context with equal probabilities
  $y_{w,i} \sim \text{Bernoulli}(\theta(\vec{x}_{w,i}))$ ▷ Observe response
end for
for $\vec{x} \in \text{support}(\vec{X})$ **do** ▷ Initialize posterior after warmup
  $\alpha(\vec{x}) \leftarrow 1 + \sum_{i=1}^{n_{\text{Warmup}}} \mathbb{I}(\vec{x}_{w,i} = \vec{x}, y_{w,i} = 1)$
  $\beta(\vec{x}) \leftarrow 1 + \sum_{i=1}^{n_{\text{Warmup}}} \mathbb{I}(\vec{x}_{w,i} = \vec{x}, y_{w,i} = 0)$
end for

Adaptive Phase Part 1: Search for Max Context

Input: $n_{\text{AdaptiveMax}}, \vec{X}, \vec{\alpha}, \vec{\beta}$
for $i = 1, 2, \ldots, n_{\text{AdaptiveMax}}$ **do**
  for $\vec{x} \in \text{support}(\vec{X})$ **do**
    $\theta^*(\vec{x}) \sim \text{Beta}(\alpha(\vec{x}), \beta(\vec{x}))$ ▷ Simulate from posterior
  end for
  $\vec{x}_{\text{Assigned}} \leftarrow \arg\max_x \ \theta^*(\vec{x})$ ▷ Assign context to new respondent
  $y_{\text{Observed}} \sim \text{Bernoulli}(\theta(\vec{x}_{\text{Assigned}}))$ ▷ Observe response
  if $y_{\text{Observed}} = 1$ **then** ▷ Update posterior
    $\alpha(\vec{x}) \leftarrow \alpha(\vec{x}) + 1$
  else if $y_{\text{Observed}} = 0$ **then**
    $\beta(\vec{x}) \leftarrow \beta(\vec{x}) + 1$
  end if
end for

Adaptive Phase Part 2: Search for Min Context

Input: $n_{\text{AdaptiveMin}}, \vec{X}, \vec{\alpha}, \vec{\beta}$
for $i = 1, 2, \ldots, n_{\text{AdaptiveMin}}$ **do**
  for $\vec{x} \in \text{support}(\vec{X})$ **do**
    $\theta^*(\vec{x}) \sim \text{Beta}(\alpha(\vec{x}), \beta(\vec{x}))$ ▷ Simulate from posterior
  end for
  $\vec{x}_{\text{Assigned}} \leftarrow \arg\min_x \ \theta^*(\vec{x})$ ▷ Assign context to new respondent
  $y_{\text{Observed}} \sim \text{Bernoulli}(\theta(\vec{x}_{\text{Assigned}}))$ ▷ Observe response
  if $y_{\text{Observed}} = 1$ **then** ▷ Update posterior
    $\alpha(\vec{x}) \leftarrow \alpha(\vec{x}) + 1$
  else if $y_{\text{Observed}} = 0$ **then**
    $\beta(\vec{x}) \leftarrow \beta(\vec{x}) + 1$
  end if
end for

Select Contexts for Validation

Input: $\vec{X}, \vec{\alpha}, \vec{\beta}, n_{\text{Sim}}$
for $\vec{x} \in \text{support}(\vec{X})$ **do** ▷ Simulations to select contexts
  $\theta^*_1(\vec{x}), \ldots, \theta^*_{n_{\text{Sim}}}(\vec{x}) \overset{\text{iid}}{\sim} \text{Beta}(\alpha(\vec{x}), \beta(\vec{x}))$
end for
$\vec{x}_{\text{Max}} \leftarrow \arg\max_{\vec{x}} \sum_{r=1}^{n_{\text{Sim}}} \mathbb{I}\left(\theta^*_r(\vec{x}) = \max_{\vec{x}'}(\theta^*_r(\vec{x}'))\right)$ ▷ Assign selected max context
$\vec{x}_{\text{Min}} \leftarrow \arg\max_{\vec{x}} \sum_{r=1}^{n_{\text{Sim}}} \mathbb{I}\left(\theta^*_r(\vec{x}) = \min_{\vec{x}'}(\theta^*_r(\vec{x}'))\right)$ ▷ Assign selected min context

Validation Phase

Input: $n_{\text{Validation}}, \vec{x}_{\text{Min}}, \vec{x}_{\text{Max}}$
for $i = 1, 2, \ldots, n_{\text{Validation}}$ **do**
  $\vec{x}_{v,i} \sim \text{DiscreteUniform}(\{\vec{x}_{\text{Min}}, \vec{x}_{\text{Max}}\})$ ▷ Assign context with equal probabilities
  $y_{v,i} \sim \text{Bernoulli}(\theta(\vec{x}_{v,i}))$ ▷ Observe response
end for
for $\vec{x} \in \{\vec{x}_{\text{Max}}, \vec{x}_{\text{Min}}\}$ **do**
  $\hat{\theta}(\vec{x}) \leftarrow \frac{\sum_{i=1}^{n_{\text{Validation}}} \mathbb{I}(\vec{x}_{v,i} = \vec{x}) y_{v,i}}{\sum_{i=1}^{n_{\text{Validation}}} \mathbb{I}(\vec{x}_{v,i} = \vec{x})}$ ▷ Estimate by validation sample proportion
end for

https://doi.org/10.1017/pan.2026.10040

H. Methods for signal selection: Job applicants

In the second illustration, described in Section 4.2, we selected signals of both race and educational institution rank that appeared realistically on two fictitious resumes.

We signaled race with names. We selected first names randomly from those used by Bertrand and Mullainathan (2004): Keisha and Tanisha to signal a Black applicant and Allison and Laurie to signal a White applicant. To select last names, we used a list of the top 10 last names by percent for each racial group from the 2010 U.S. Census[8] (U.S. Census Bureau 2016), from which we randomly selected two names for each: Rivers and Mosely to signal a Black applicant and O'Connell and Shmitt to signal a White applicant. As signals of educational institution rank, we relied on the U.S. News and World Report Rankings[9] similar to prior work (Ishizuka 2021). Filtering for only national universities with programs in marketing, accounting and business, we chose two of the highest-ranked programs (University of Pennsylvania and Massachusetts Institute of Technology) and two mid-ranked programs (California State University, Long Beach and San Diego State University). We chose two mid-ranked programs that are very similar because our design has the best power when the two signals of any given attribute value lead to similar assessments of candidate strengths.

[8] https://www.census.gov/topics/population/genealogy/data/2010_surnames.html

[9] https://www.usnews.com/best-colleges/rankings/national-universities?study=Business%2C+Management%2C+Marketing%2C+and+Related+Support+Services&major=Accounting+and+Related+Services%7CMarketing&_sort=rank&_sortDirection=asc

Cite this article: Gosciak, J., Molitor, D. and Lundberg, I. (2026). Adaptive Randomization in Conjoint Survey Experiments. *Political Analysis*, 541–568. https://doi.org/10.1017/pan.2026.10040

https://doi.org/10.1017/pan.2026.10040