Recent Research in the Design and Analysis of Conjoint Experiments

Kirk Bansak

University of California, San Diego

April 29, 2021

Conjoint Experiments in Political Science

- Conjoint analysis $\rightarrow$ experimental technique introduced in 1970s to analyze preferences/choices about multi-dimensional objects
- Shift to computer-assisted administration $\rightarrow$ revolutionary
- Extensive use recently in political science; development of corresponding statistical tools
- Well matched to substantive problems in which people rate/rank multi-dimensional objects (vote choice, parties, immigrants, policy packages)

	CANDIDATE A	CANDIDATE B
Supports Creating Pathway to Citizenship for	All unauthorized immigrants with no criminal record	No unauthorized immigrants
Position on Climate Change	Ban the use of fossil fuels after 2040, reducing economic growth by 5%	Promote the use of renewable energy but allow continued use of fossil fuels
Previous Occupation	Doctor	Activist
Prior Political Experience	U.S. Representative	U.S. Representative
Sexual Orientation	Straight	Straight
Military Service Experience	Did not serve	Served in the Army
Gender	Male	Female
Supports Government Healthcare for	Only Americans who are older, poor, or disabled	All Americans
Race/Ethnicity	White	White
Age	37	45

Types of Spending Cuts/Tax Increases		Option 1		Option 2
<u>SPENDING CUTS</u>				
Pension cuts by		**29%**		**0%**
Education spending cuts by		**13%**		**24%**
Defense spending cuts by		**5%**		**5%**
Welfare spending cuts by		**14%**		**8%**
Public sector layoffs by		**23%**		**2%**
<u>TAX INCREASES</u>				
Corporate tax increase by		**7%**		**17%**
Income tax increase by		**16%**		**9%**
Sales tax increase by		**8%**		**25%**

Conjoint Experiments in Political Science

- Analysis of conjoint experiments in political science predominantly situated within causal inference (potential outcomes) framework

  Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto. 2014. "Causal Inference in Conjoint Analysis: Understanding Multidimensional Choices via Stated Preference Experiments." *Political Analysis*, Vol. 22, No. 1.

- Most common causal estimand in recent work is Average Marginal Component Effect (AMCE)
  - though the AMCE is often mis-interpreted and not the only estimand of potential interest (will revisit this later)

Defining the AMCE (Simplified Version)

- Consider common design context: forced choice between two profiles
- Assume three attributes, A, B, and C, and hence a random profile can be denoted by $[A, B, C]$
- Also assume each attribute is binary, e.g. $A \in \{0, 1\}$
- Let $Y_i([abc], [a'b'c']) \in \{0, 1\}$ denote i's potential outcome given paired forced-choice contest between $[abc]$ & $[a'b'c']$
- AMCE provides causally interpretable, summary measure of an attribute's effect; for instance, for attribute A:

$$AMCE_A \equiv \mathbb{E}\big[Y_i([1BC], [A'B'C']) - Y_i([0BC], [A'B'C'])\big] \qquad (1)$$

- expectation taken w.r.t.
  1. target population of individuals
  2. target distribution of other attributes

Recent Methods Research (with examples)

1. Design
   - Number of attributes (Bansak, Hainmueller, Hopkins, and Yamamoto 2021a)
   - Number of tasks (Bansak, Hainmueller, Hopkins, and Yamamoto 2018)
   - Number of profiles (Hainmueller, Hangartner and Yamamoto 2015)
2. Attention and Decision-Making
   - Using eye-tracking (Jenke, Bansak, Hainmueller, and Hangartner 2021)
   - Mitigation of social desirability bias (Horiuchi, Markovich, and Yamamoto 2019)
3. Unpacking the AMCE
   - Effect on vote shares in electoral setting (Bansak, Hainmueller, Hopkins, and Yamamoto 2021c)
   - Relationship with formal definitions of preferences (Abramson, Koçak, and Magazinnik 2019; Ganter 2020)
4. Other estimands, model-based approaches, and machine learning
   - Alternative estimands in electoral studies (Bansak, Hainmueller, Hopkins, and Yamamoto 2021c)
   - Leveraging machine learning techniques (Bansak, Bechtel, and Margalit 2021; Egami and Imai 2019; Horiuchi, Smith, and Yamamoto 2018)
5. Analysis of heterogeneity
   - Across respondent characteristics (Leeper, Hobolt and Tilley 2020)
   - Interactions between attributes and importance of profile distribution (Egami and Imai 2019; de la Cuesta, Egami, and Imai 2021)

Open Questions for Conjoint Analysis

- Lots of choices when implementing a conjoint experiment:
  - Which/how many attributes per profile?
  - How many profiles per task?
  - How many tasks per respondent?
  - Forced choice or rating?
- Conjoint experiments frequently used in political science, yet still open questions on how best to design them
- Recent research agenda: How do these choices affect the inferences we draw from conjoint designs? Which choices are optimal?
- Of primary concern: Satisficing
  - How many **attributes** should a conjoint profile include?
  - How many **tasks** should a conjoint study use for each respondent?

The Satisficing-Masking Trade-off

- Including **too many attributes** might cause excessive survey satisficing
  - Limits of working memory (Miller 1994)
  - Task difficulty $\uparrow \longrightarrow$ satisficing $\uparrow$ (Krosnick 1999)
- However, including **too few attributes** causes masking
  - Respondents use one attribute because of perceived correlation with unobserved attribute (also see Dafoe et al. 2015)
  - E.g. If conjoint table does not include issue positions, the effect of partisanship may go up
  - Masking does not invalidate the identification of the AMCE, but muddies interpretation

The Masking-Satisficing Tradeoff

Please carefully review the two candidates for President detailed below.

Which of these two candidates would you prefer to see as President of the United States?

	Candidate A	Candidate B
Highest education	graduated from high school	graduated from college
Largest campaign contributor	auto workers' unions	wall street firms
State of residence	Alabama	Ohio
Party affiliation	Republican	Republican
Your Choice:	○	○

NEXT

The Masking-Satisficing Tradeoff

Please carefully review the two candidates for President detailed below.

Which of these two candidates would you prefer to see as President of the United States?

	Candidate A	Candidate B
Highest education	graduated from high school	graduated from college
Largest campaign contributor	auto workers' unions	wall street firms
State of residence	Alabama	Ohio
Annual income	$75k	$32k
Race/Ethnicity	Asian American	Black
Profession	lawyer	farmer
Car	Ford pick-up truck	Toyota Sedan
Favorite professional sport	football	basketball
Military service	served in U.S. military	served in U.S. military
Age	72	63
Marital status	single	single

Marital status	single	single
Position on abortion	pro-life	neutral
Position on health care	government should do more	government should do more
Religion	Evangelical Protestant	Evangelical Protestant
Prior elected office	state attorney general	state attorney general
Favorite music	hip hop	country
Religious activity	occasionally attends church	attends church weekly
Gender	female	female
Position on gay marriage	opposes gay marriage	favors gay marriage
Party affiliation	Republican	Republican
Your Choice:	○	○

NEXT

The Satisficing-Masking Trade-off

- Ideally, one should include as many attributes as appropriate given the theoretical question...
- **But is there a limit because of satisficing?**

Number of Tasks and Fatigue

- Researchers typically ask respondents to complete multiple tasks in a conjoint study
- However, satisficing likely to increase as survey becomes lengthier
- "longstanding view that long questionnaires or interviews should be avoided" (De Vaus 2014)
- Conjoint tasks are more cognitively demanding than typical survey questions
  $\rightarrow$ Fatigue might be more problematic

Studies on Conjoint Robustness

Kirk Bansak, Jens Hainmueller, Dan Hopkins, and Teppei Yamamoto. 2021a. "Beyond the Breaking Point? Survey Satisficing in Conjoint Experiments." *Political Science Research and Methods*, Vol. 9, No. 1.

Kirk Bansak, Jens Hainmueller, Dan Hopkins, and Teppei Yamamoto. 2018. "The Number of Choice Tasks and Survey Satisficing in Conjoint Experiments." *Political Analysis*, Vol. 26, No. 1.

Contribution

A series of survey experiments to investigate the threat of satisficing with respect to:

- Number of attributes.
- Number of tasks.

Results in a nutshell

Conjoint designs are remarkably robust to satisficing on both of these dimensions.

Beyond the Breaking Point: Number of Attributes

- Goal: Investigate how many attributes one can include in a conjoint profile without making tasks overly prone to satisficing
- Challenge: Isolating the effect of satisficing from that of masking
- A two-stage design:
  - Begin with a set of "core" attributes of interest known to be important for respondent choice
  1. Stage 1: Identify other "filler" attributes not masked by the core attributes (i.e. attributes with no perceived association with the core attributes)
  2. Stage 2: Measure changes in effects of the core attributes as varying numbers of non-masking filler attributes (from stage 1) are added
- Domains:
  1. Hotel room choice
  2. Candidate choice

The Second Stage Experiment

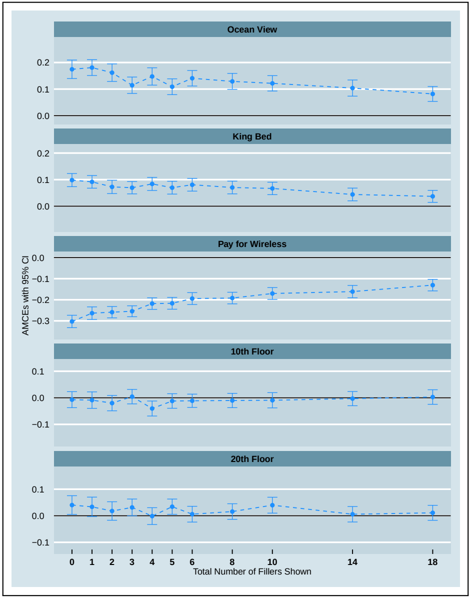
- After having identified pools of non-masking fillers in the first stage...
- Paired-profile conjoint experiments on MTurk; replication on SSI
- Test the effects of the same four core attributes from the 1st stage, with randomly varying # of non-masking fillers
- Metrics:
  1. Change in AMCEs of the four core attributes
  2. Change in (partial) R-squared of the core attributes
  - Both should be attenuated towards zero with satisficing

An Example Task with Four Non-Masking Fillers

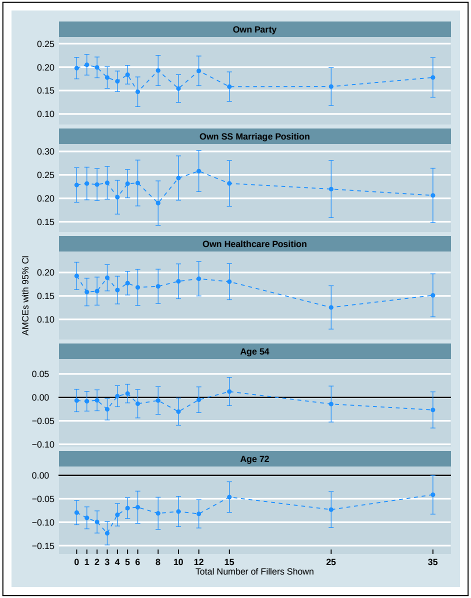
Please carefully read the descriptions of the two packages below. (3/15)

	PACKAGE A	PACKAGE B
Floor	Top floor (20th)	Top floor (20th)
View	Mountain view	Ocean view
Bathroom sinks	2 sinks with limited countertop space	2 sinks with limited countertop space
Bedroom furniture	1 king bed and 1 small couch	1 queen bed and 1 large couch
Wake-up calls performed by	Automated voice system	Live operator
Additional service provided	Free hot breakfast	Free dry cleaning
Complimentary chocolate on pillow	Mint chocolate	Dark chocolate
Internet availability	Pay for high bandwidth wireless	Free standard bandwidth wireless

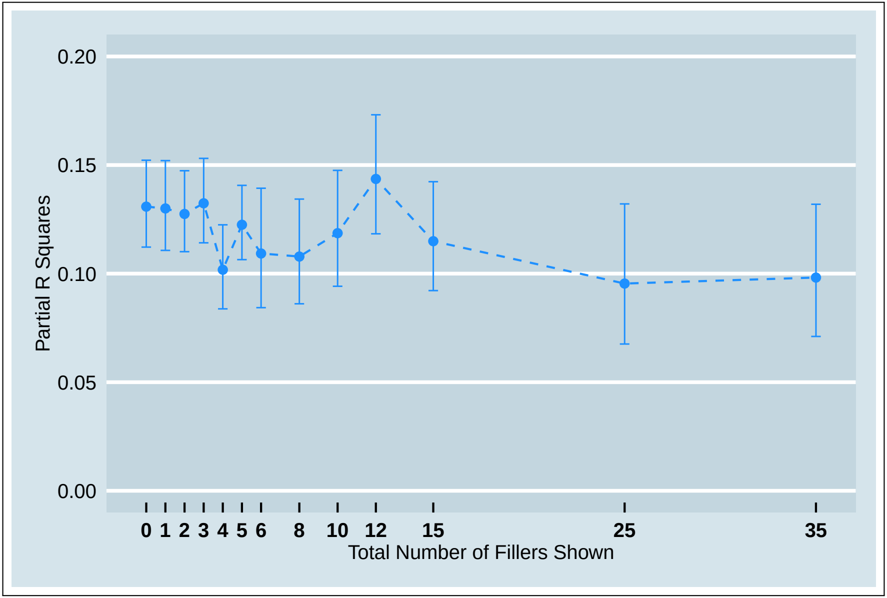
Results: Hotel AMCEs



Results: Candidate AMCEs



Results: Candidate Partial R^2 for Core Attributes

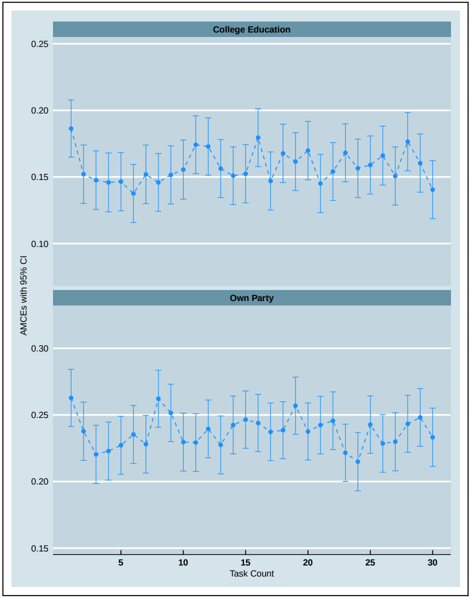


Study on the Number of Tasks

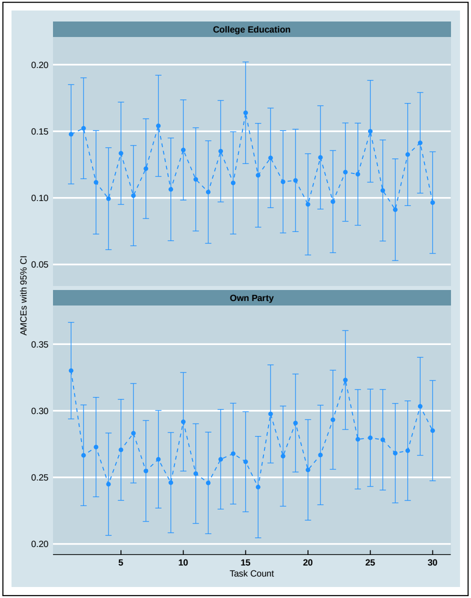
- **Do response patterns change as number of tasks increases?**
- Metrics: Core attribute AMCEs and R-squared

- Paired-profile conjoint experiments implemented on MTurk (4,921 respondents) and SSI (1,613 respondents)
- Domain: Presidential vote choice
- Two core attributes (always included): partisanship, education
- Randomly assigned up to 18 other attributes (e.g. income, issue positions, favorite professional sport, etc.)
- All respondents complete 30 choice tasks

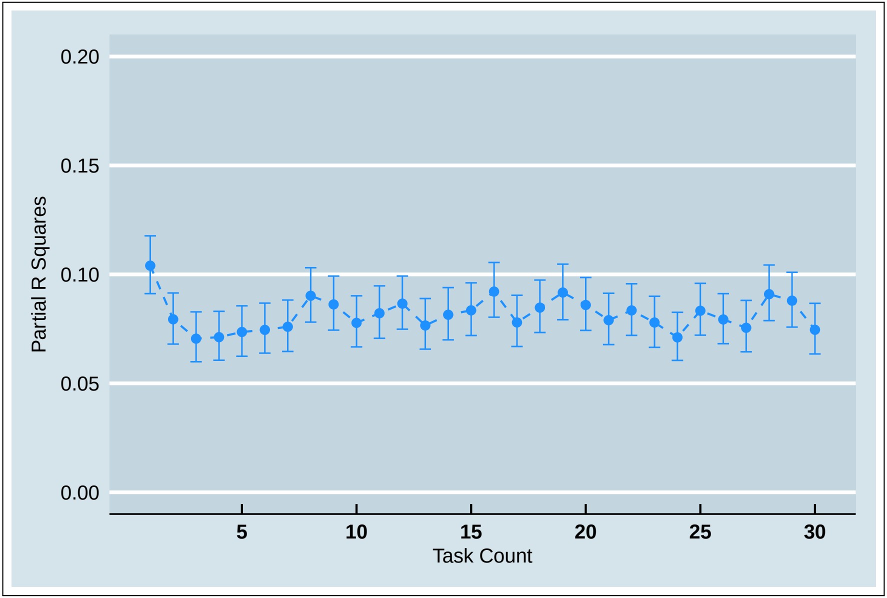
Results: AMCEs (MTurk)



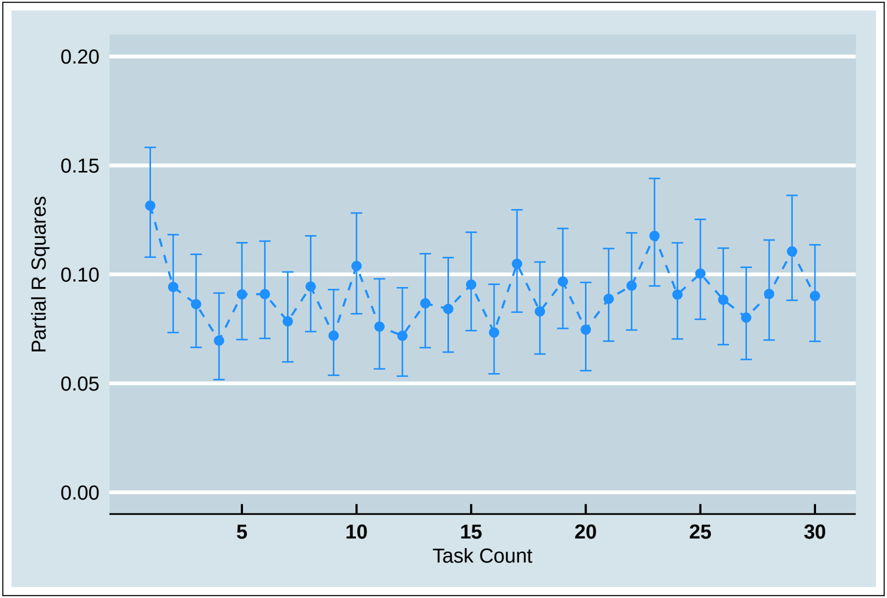
Results: AMCEs (SSI)



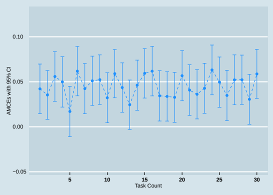
Results: R^2 (MTurk)



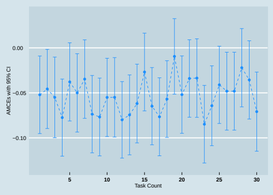
Results: R^2 (SSI)



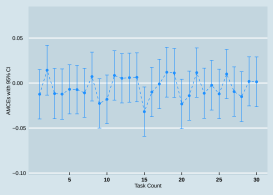
Results: Filler AMCEs (MT)



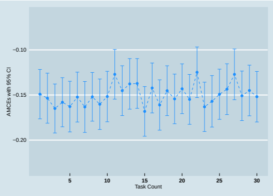
(a) Military Service



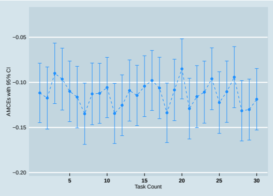
(b) Age



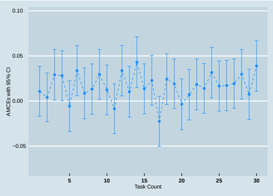
(c) Gender



(d) SS Marriage

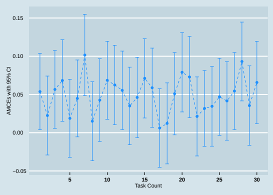


(e) Abortion

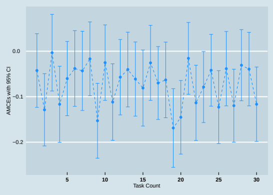


(f) Healthcare

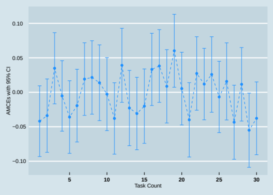
Results: Filler AMCEs (SSI)



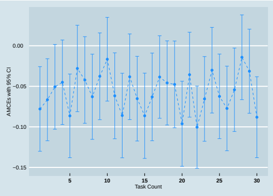
(g) Military Service



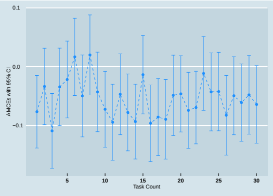
(h) Age



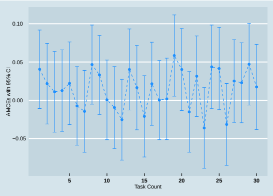
(i) Gender



(j) SS Marriage



(k) Abortion



(l) Healthcare

Summary and Conclusions

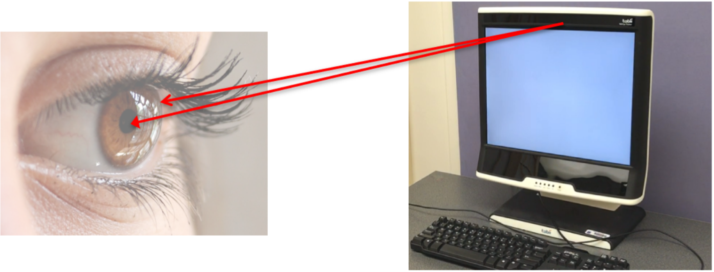
- Results indicate surprising robustness to satisficing, with respect to both number of attributes and number of tasks
- The "breaking points" appear to far exceed typical specifications
  - e.g. It is OK to use up to a dozen (and perhaps more) attributes
  - e.g. It is OK (and probably a good idea) to have respondents perform ten or even more tasks
- Let other considerations guide design decisions, not concerns over excessive satisficing
  - Theoretical goals and masking considerations
  - Budget and other practical constraints
  - Value of more data
  - Realism

Motivation

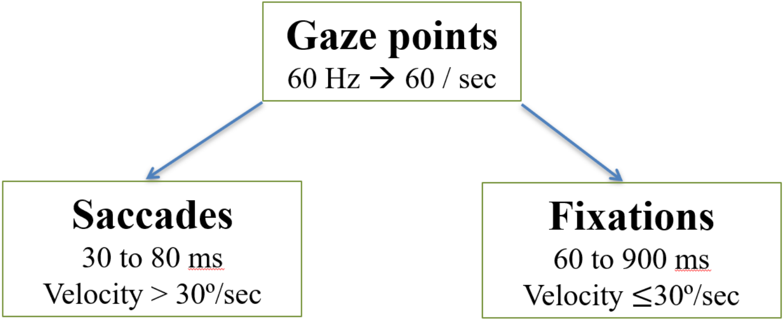
- Small but growing literature on how conjoint design affects choice behavior
- Our question: Why do results remain stable as design complexity increases?
  - Does attention decrease?
- Limited evidence/understanding of...
  - How respondents approach conjoints (cognitively, visually, etc.)
  - What information-processing and decision-making strategies they use
- Why we care: Implications for design, interpretation, external validity
- Idea: Leverage richer data from eye-tracking

Libby Jenke, Kirk Bansak, Jens Hainmeuller, and Dominik Hangartner. 2021. "Using Eye-Tracking to Understand Decision-Making in Conjoint Experiments." *Political Analysis*, Vol. 29, No. 1.

How an Eye-Tracker Works



Standard Fixation Classification Algorithms



Study Design

- Goal: Investigate design effects on choice behavior and visual attention
  - Assess eye fixation and movement patterns across conditions of varying complexity

- Conjoint decision task: Choose preferred candidate
  - Standard conjoint table presenting multiple candidate profiles
  - Profiles comprised on attributes with randomly varying values

- Six experimental conditions (blocks) comprising varying levels of complexity
  - Number of attributes per profile: 5, 8, 11
  - Number of profiles: 2, 3

- All subjects shown all six blocks, in random order
  - Sequence of 20 decision tasks per block
  - For each subject-block, attributes drawn from full list of 11 attributes

Gender	Male	Female
Occupation	Lawyer	Activist
Gun Control	Strongly oppose	Weakly support
Political party	Independent	Republican
Age	53	77

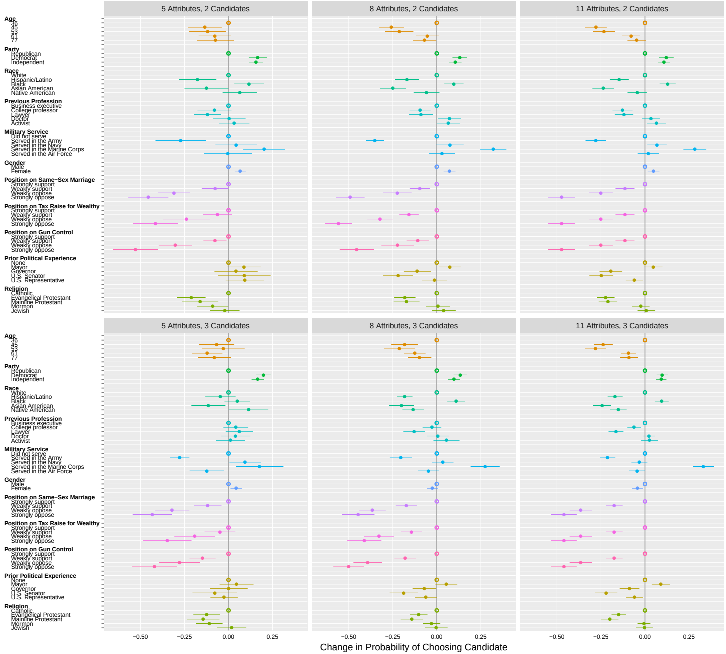
Political party	Republican	Democrat	Democrat
Gender	Female	Female	Male
Gun Control	Weakly support	Weakly oppose	Strongly oppose
Political experience	No prior political experience	U.S. Senator	U.S. Senator
Occupation	Activist	Lawyer	Business executive
Religion	Jewish	Mainline Protestant	Evangelical Protestant
Military service	Served in the Air Force	Served in the Navy	Did not serve
Same-sex marriage	Weakly support	Weakly oppose	Strongly oppose
Increase wealthy's tax	Weakly support	Weakly oppose	Strongly oppose
Age	77	53	45
Race	Black	Hispanic/Latino	Asian American

Attribute Pool

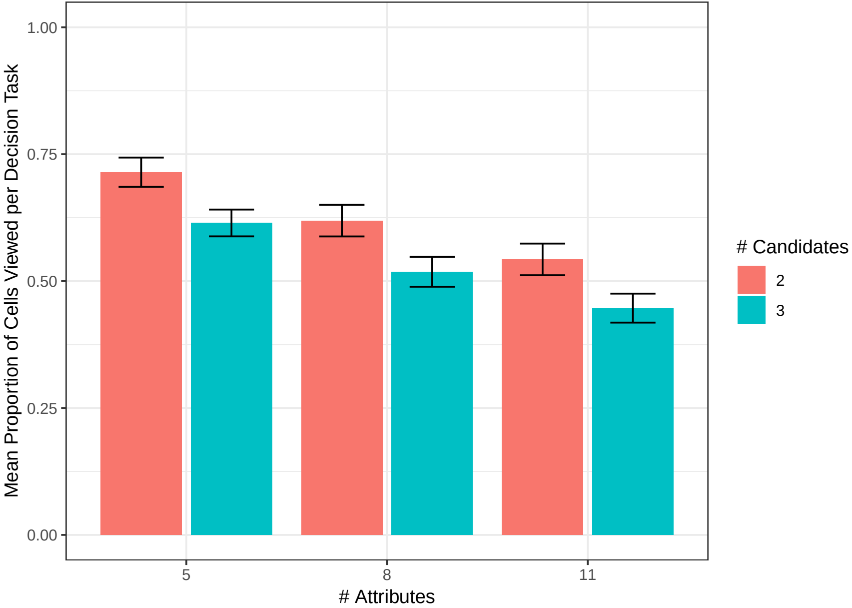
Attribute	**Values**
Age	37, 45, 53, 61, 77
Gender	Female, Male
Race/Ethnicity	White, Hispanic/Latino, Black, Asian American, Native American
Previous Occupation	Business executive, College professor, Lawyer, Doctor, Activist
Military Service Experience	Did not serve, Served in the Army, Served in the Navy, Served in the Marine Corps, Served in the Air Force
Prior Political Experience	Mayor, Governor, U.S. Senator, U.S. Representative, No prior political experience
Party	Democrat, Republican, Independent
Religion	Catholic, Evangelical Protestant, Mainline Protestant Mormon, Jewish
Position on Same-Sex Marriage	Strongly support, Support, Oppose, Strongly oppose
Position on Tax Raise for Wealthy	Strongly support, Support, Oppose, Strongly oppose
Position on Gun Control	Strongly support, Support, Oppose, Strongly oppose

Sample

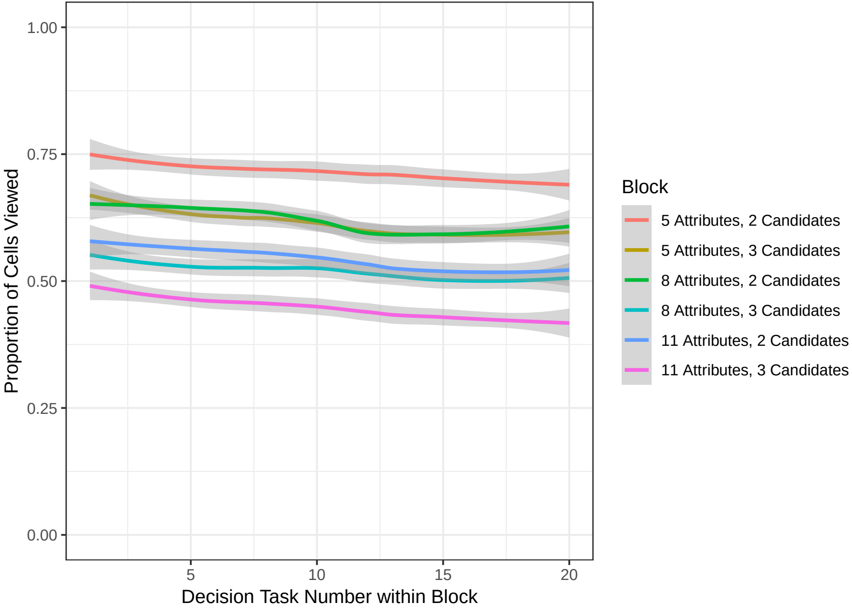
- 122 subjects run at Duke University in July 2019
- Subjects drawn from Duke Behavioral Research subject pool
  - 39% undergraduate/graduate students
  - 61% local residents
- Median completion time: 35 minutes


5 Attributes, 2 Candidates
8 Attributes, 2 Candidates
11 Attributes, 2 Candidates
5 Attributes, 3 Candidates
8 Attributes, 3 Candidates
11 Attributes, 3 Candidates
Age
36
45
53
61
77
Party
Republican
Democrat
Independent
Race
White
Hispanic/Latino
Black
Asian American
Native American
Previous Profession
Business executive
College professor
Lawyer
Doctor
Activist
Military Service
Did not serve
Served in the Army
Served in the Navy
Served in the Marine Corps
Served in the Air Force
Gender
Male
Female
Position on Same−Sex Marriage
Strongly support
Weakly support
Weakly oppose
Strongly oppose
Position on Tax Raise for Wealthy
Strongly support
Weakly support
Weakly oppose
Strongly oppose
Position on Gun Control
Strongly support
Weakly support
Weakly oppose
Strongly oppose
Prior Political Experience
None
Mayor
Governor
U.S. Senator
U.S. Representative
Religion
Catholic
Evangelical Protestant
Mainline Protestant
Mormon
Jewish
−0.50
−0.25
0.00
0.25
Change in Probability of Choosing Candidate

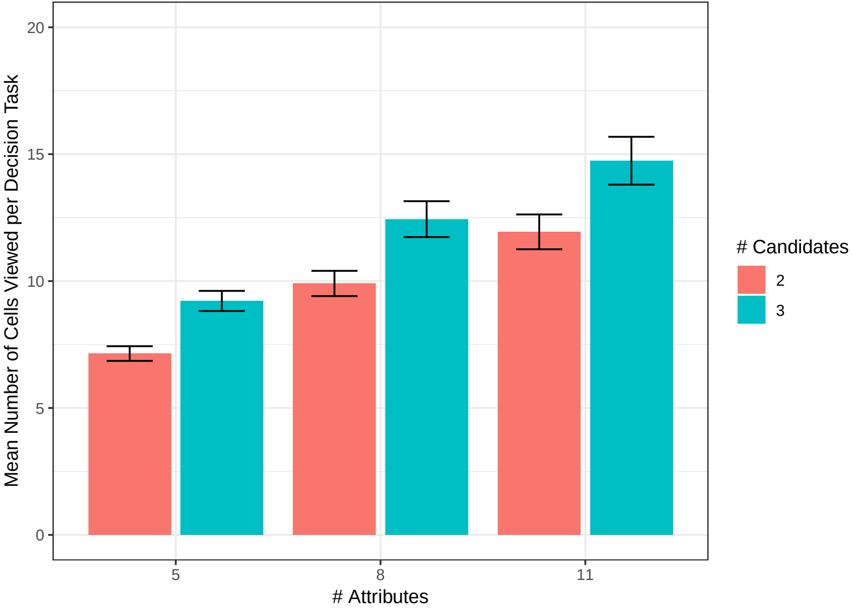
Changes in Attention Across Designs: Proportion of Cells Viewed



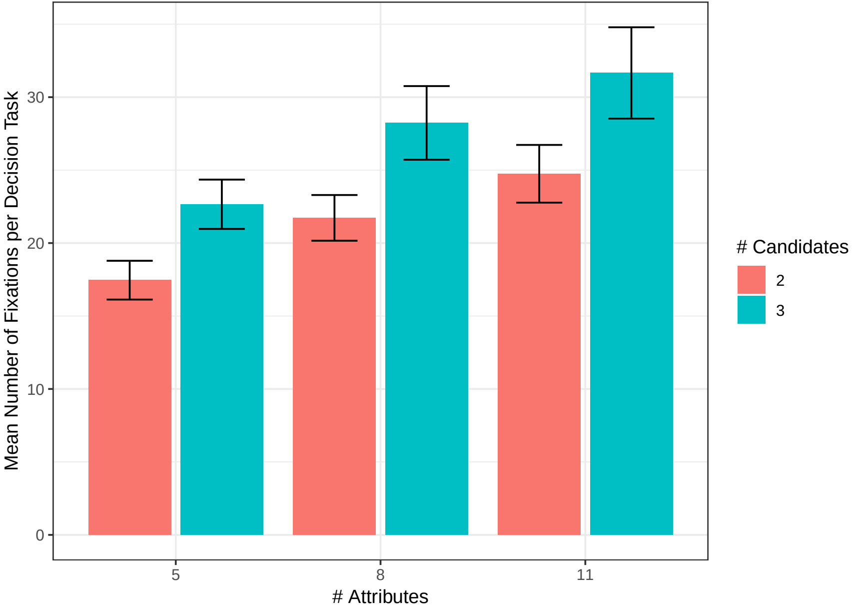
Proportion of Cells Viewed by Condition and Task (Pooled)



Changes in Attention Across Designs: Number of Cells Viewed



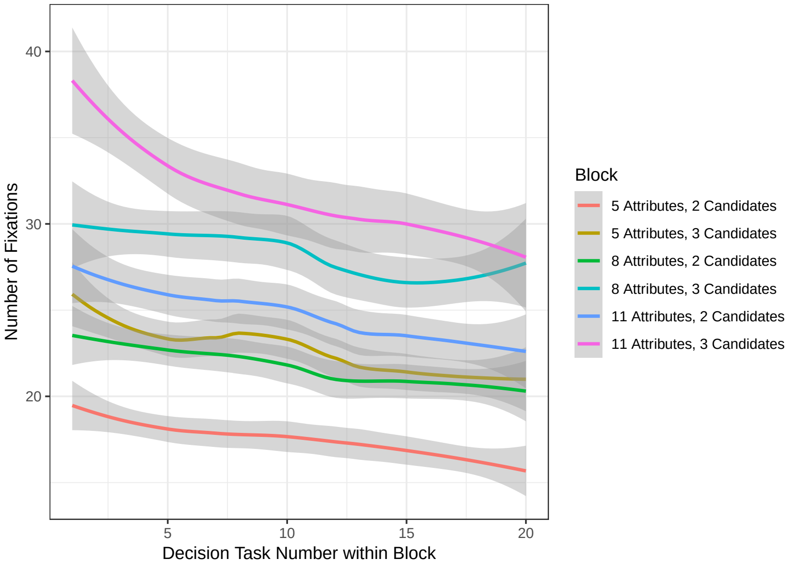
Changes in Attention Across Designs: Number of Fixations

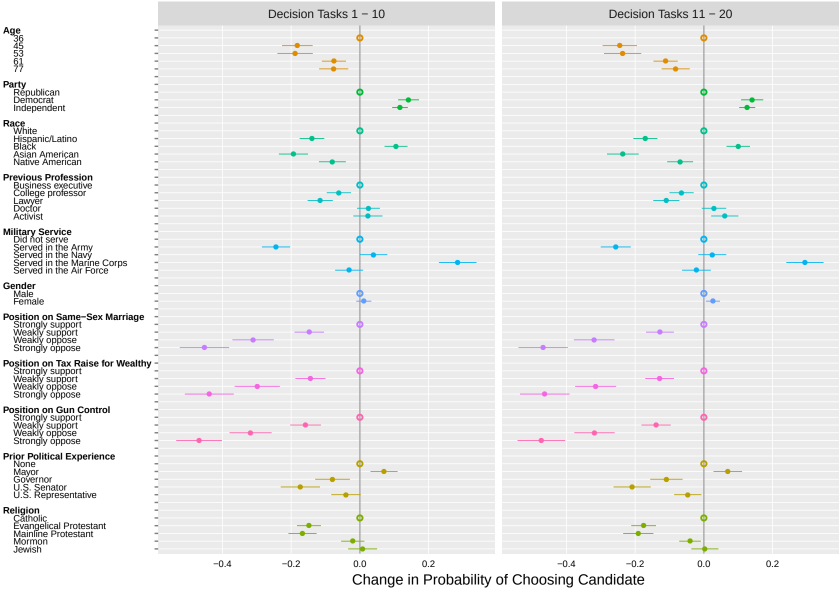


Interpretation

- Strategy of bounded rationality
- Respondents adapt to complexity and reduce their own cognitive processing costs by
  - Selectively incorporating relevant new information to focus on the important attributes
  - Ignoring information they deem less relevant
- Critically, respondents do not adjust to added complexity by simply paying less attention in general
- Stable learning process also occurs over tasks

Subjects Become More Efficient





Decision Tasks 1 − 10
Decision Tasks 11 − 20
Age
36
45
53
61
77
Party
Republican
Democrat
Independent
Race
White
Hispanic/Latino
Black
Asian American
Native American
Previous Profession
Business executive
College professor
Lawyer
Doctor
Activist
Military Service
Did not serve
Served in the Army
Served in the Navy
Served in the Marine Corps
Served in the Air Force
Gender
Male
Female
Position on Same−Sex Marriage
Strongly support
Weakly support
Weakly oppose
Strongly oppose
Position on Tax Raise for Wealthy
Strongly support
Weakly support
Weakly oppose
Strongly oppose
Position on Gun Control
Strongly support
Weakly support
Weakly oppose
Strongly oppose
Prior Political Experience
None
Mayor
Governor
U.S. Senator
U.S. Representative
Religion
Catholic
Evangelical Protestant
Mainline Protestant
Mormon
Jewish
−0.4 −0.2 0.0 0.2
−0.4 −0.2 0.0 0.2
Change in Probability of Choosing Candidate

Implications

- Sheds new light on how respondents actually engage with conjoint tables
  - They adapt to complexity, rather than getting overwhelmed by it
- Helps explain robustness with increasing number of tasks and attributes (and, to a more limited extent, also profiles)
- Highlights future research opportunities

Conjoint Designs, the AMCE, and Analyzing Elections

- Most common causal estimand in recent work: Average Marginal Component Effect (Hainmueller et al. 2014)
- Statistical, empirical development → outstripped theoretical attention to what quantities conjoints can, cannot recover (e.g. Abramson, Koçak & Magazinnik 2019; Leeper, Hobolt, & Tilley 2020)
- 27% of conjoint applications 2014-19 → voting (Bansak et al. 2021b)
- Recent work on illuminating the formal and conceptual underpinnings of the AMCE

Kirk Bansak, Jens Hainmueller, Dan Hopkins, and Teppei Yamamoto. 2021c. "Using Conjoint Experiments to Analyze Elections: The Essential Role of the Average Marginal Component Effect (AMCE)." Working Paper.

Defining the AMCE

- Assume profiles comprised of three binary attributes, A, B, and C.
- Let $Y_i([abc],[a'b'c']) \in \{0,1\}$ denote i's potential outcome given paired forced-choice contest between $[abc]$ & $[a'b'c']$.
- Focusing on the *AMCE* for attribute A:

$$AMCE_A \equiv \mathbb{E}\big[Y_i([1BC],[A'B'C']) - Y_i([0BC],[A'B'C'])\big]$$

Remarks

$$AMCE_A \equiv \mathbb{E}\big[Y_i([1BC],[A'B'C']) - Y_i([0BC],[A'B'C'])\big]$$

- One-number summary
- Double averaging
  - Averaging across attributes/profile distribution (e.g. candidates) and across target population (e.g. voters)
  - Empirically tractable; can be estimated straightforwardly via differences-in-means or OLS
- Importance of underlying attribute distribution (see also de la Cuesta, Egami & Imai 2019)
  - Though note the seriousness of this consideration depends on how strong interactions are between attributes
- AMCE compares change in attribute for the <u>same</u> profile, not difference in attributes across profiles
  - Not the difference between candidate with A=1 against another candidate with A=0

Preference Intensity

- AMCE incorporates preference intensity (Abramson, Koçak, and Magazinnik 2019)
- This means that the AMCE is not (and should not be interpreted as) a measure of majority preferences
- Similar to ATE, where a positive ATE does not necessarily mean a positive individual-level causal effect for the majority
- This is a feature, not a bug! The AMCE inherits several desirable properties as a result:
  - captures multi-dimensionality/trade-offs
  - maps onto meaningful empirical phenomenon of interest
  - empirically tractable with credible assumptions

If one were to ignore preference intensity...

- One might then be interested, for instance, in something like the fraction of voters who prefer $A = a$ over $A = a'$
- As a thought experiment, imagine A is handedness
- 90% of the world is right-handed
- Imagine everyone prefers a candidate who shares their handedness (but choice = multi-dimensional)
- Result: handedness would appear to dominate many other attributes despite its electoral irrelevance
- The multi-dimensionality of the profiles/choice task is a vital part of how people actually make their decisions, and preference intensity factors into that
- The AMCE takes this into account

AMCEs measure effects on vote shares

Proposition (Identification of the Expected Difference in Vote Shares with the AMCE)

Consider a J-profile conjoint experiment in which respondents are a simple random sample of size N drawn from $\mathcal{V}$. Then, the AMCE for attribute $A = a$ (versus the baseline level $A = a_0$) given the randomization distribution $\mathcal{A}$ identifies the difference in the expected vote share of a candidate with $A = a$ and a candidate with $A = a_0$ in the target election $\langle \mathcal{A}, \mathcal{V} \rangle$ with J candidates.

Do researchers care about vote shares?

- AMCE recovers vote share difference associated with presence or absence of an attribute in a profile
- Lit. review of four journals focused on voting behavior 2015-2019
- 82 non-conjoint articles (87%) include either aggregate vote shares or their individual-level analogs
- AMCEs
  - recover a meaningful quantity for election researchers (vote shares)
  - recover central quantity of interest to significant majority of applied electoral research

Alternative estimands in election studies: (different versions of) the probability of winning

Kirk Bansak, Jens Hainmueller, Dan Hopkins, and Teppei Yamamoto. 2021c. "Using Conjoint Experiments to Analyze Elections: The Essential Role of the Average Marginal Component Effect (AMCE)." Working Paper.

How likely is candidate with $A = a$ to win majority?

$$\mathbb{E}_{\mathcal{A}}\left[1\{\mathbb{E}_{\mathcal{V}}\big(Y_i([aBC],[A'B'C'])\big) > 0.5\}\right] \tag{2}$$

How likely is candidate $A = a$ to beat $A = a'$?

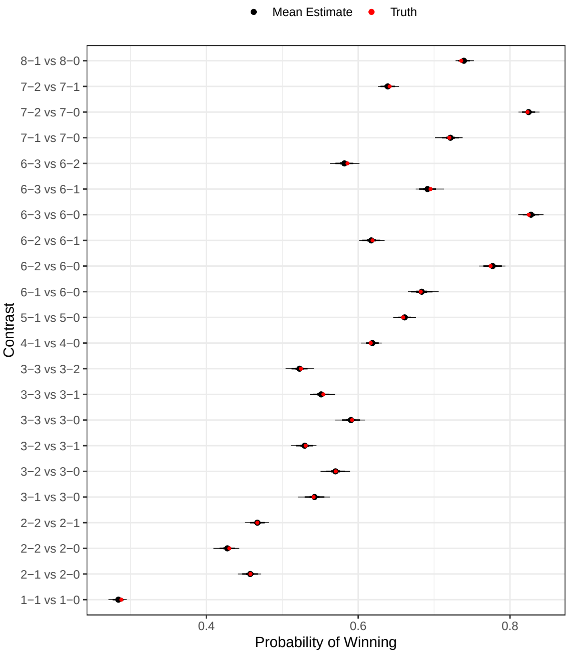
$$\mathbb{E}_{\mathcal{A}}\left[1\{\mathbb{E}_{\mathcal{V}}\big(Y_i([aBC],[a'B'C'])\big) > 0.5\}\right] \tag{3}$$

Estimating the Probability of Winning

- Because of non-linearity of indicator function and high-dimensionality, non-parametric (plug-in) estimator is intractable
- Instead, model-based approach that explicitly models inner expectation as function of attributes, followed by averaging over attribute distribution
- Is this feasible and effective? Preliminary simulation results suggest so
- Focusing on the following:

$$\mathbb{E}_{\mathcal{A}}\left[1\{\mathbb{E}_{\mathcal{V}}\big(Y_i([aBC],[a'B'C'])\big) > 0.5\}\right]$$

- Simulations with known estimand values, simulated conjoint data, and employment of conditional logistic ridge regression
- Can compare true values to estimates for all contrasts, probability of candidate with $X = x$ beating candidate with $X = x'$

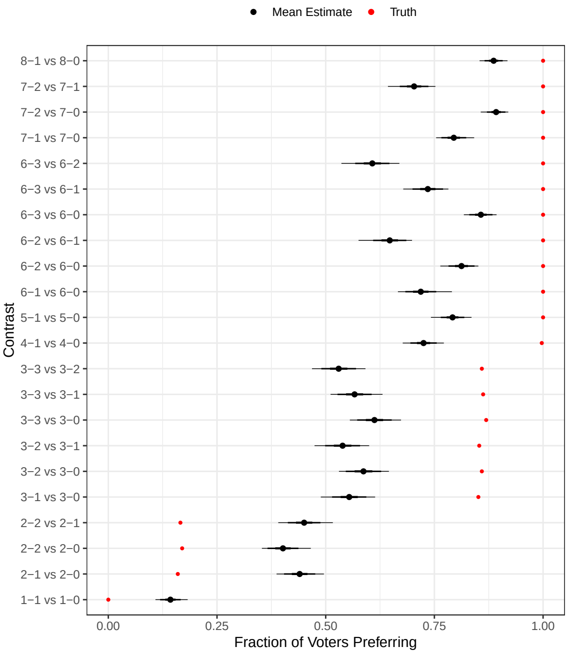

Mean Estimate
Truth
8−1 vs 8−0
7−2 vs 7−1
7−2 vs 7−0
7−1 vs 7−0
6−3 vs 6−2
6−3 vs 6−1
6−3 vs 6−0
6−2 vs 6−1
6−2 vs 6−0
6−1 vs 6−0
5−1 vs 5−0
4−1 vs 4−0
3−3 vs 3−2
3−3 vs 3−1
3−3 vs 3−0
3−2 vs 3−1
3−2 vs 3−0
3−1 vs 3−0
2−2 vs 2−1
2−2 vs 2−0
2−1 vs 2−0
1−1 vs 1−0
Contrast
0.4
0.6
0.8
Probability of Winning

Alternative estimands: fraction preferring

Fraction preferring:

$$\mathbb{E}_{\mathcal{V}}\left[\mathbf{1}\{\mathbb{E}_{\mathcal{A}}[Y_i([aBC],[a'B'C'])] > 0.5\}\right] \tag{4}$$

- Reversing the order of the expectations results in an entirely different estimand
- e.g. Classify all voters into those preferring female candidate, those preferring male candidate
- Calculate proportion of female preferers
- Inference about fraction preferred $\rightarrow$ intractable without extremely stringent and untestable assumptions


Mean Estimate
Truth
Contrast
8−1 vs 8−0
7−2 vs 7−1
7−2 vs 7−0
7−1 vs 7−0
6−3 vs 6−2
6−3 vs 6−1
6−3 vs 6−0
6−2 vs 6−1
6−2 vs 6−0
6−1 vs 6−0
5−1 vs 5−0
4−1 vs 4−0
3−3 vs 3−2
3−3 vs 3−1
3−3 vs 3−0
3−2 vs 3−1
3−2 vs 3−0
3−1 vs 3−0
2−2 vs 2−1
2−2 vs 2−0
2−1 vs 2−0
1−1 vs 1−0
0.00
0.25
0.50
0.75
1.00
Fraction of Voters Preferring

Other applications of machine learning to conjoint data

- Estimation of interactions between attributes, via regularization (Egami and Imai 2019)
- Estimating popularity of specific profiles, by flexibly modeling response
  - Ridge regression (Horiuchi, Smith, and Yamamoto 2018)
  - Stochastic gradient boosted trees (Bansak, Bechtel, and Margalit 2021)

Support for austerity and the role of strategic policy design

Kirk Bansak, Michael M. Bechtel and Yotam Margalit. 2021. "Why Austerity? The Mass Politics of a Contested Policy." *American Political Science Review*, Vol. 115, No. 2.

- Policy design: composition of austerity in terms of specific spending cuts and tax hikes
- Policymakers may select features that will generate majority support
- Implies/requires that voters have varying sensitivities to different types of austerity measures

High-dimensional austerity conjoint experiment

- Austerity package conjoint experiment fielded in Italy and Spain, 2019
- Austerity conjoint profiles specify...

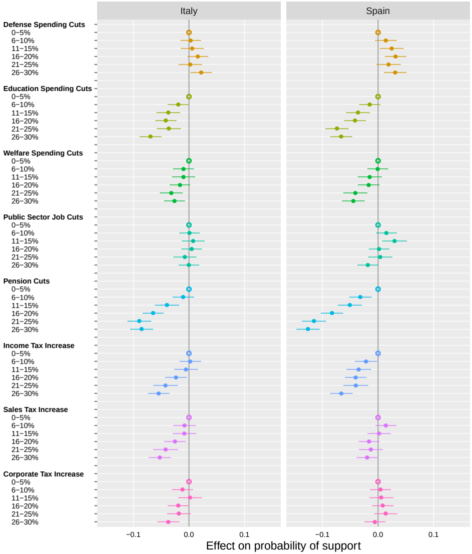
  % cuts to specific budget items (defense, education, pensions, welfare, public sector jobs)

  % increases to specific taxes (income tax, sales tax, corporate tax)

- Historically informed range of values: $\{0, 1, 2, ..., 30\}$
- Outcome: Support (or not) for a package (non-forced choice)
- Each respondent was shown ten pairs of profiles (i.e. 20 hypothetical packages total)

Types of Spending Cuts/Tax Increases		Option 1		Option 2
SPENDING CUTS				
Pension cuts by		**29%**		**0%**
Education spending cuts by		**13%**		**24%**
Defense spending cuts by		**5%**		**5%**
Welfare spending cuts by		**14%**		**8%**
Public sector layoffs by		**23%**		**2%**
TAX INCREASES				
Corporate tax increase by		**7%**		**17%**
Income tax increase by		**16%**		**9%**
Sales tax increase by		**8%**		**25%**

AMCEs



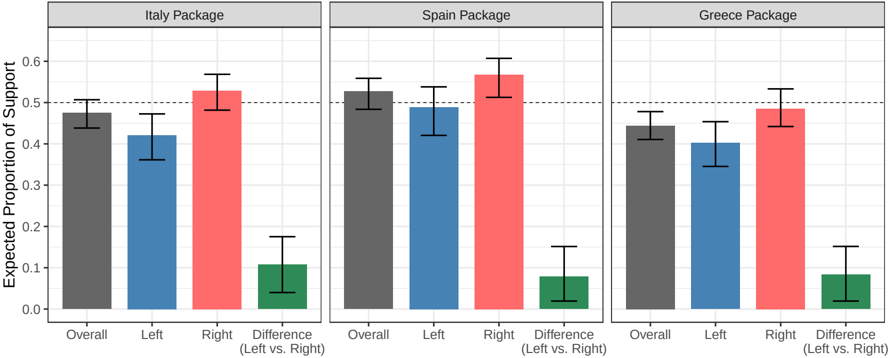
Italy
Spain
Defense Spending Cuts
0−5%
6−10%
11−15%
16−20%
21−25%
26−30%
Education Spending Cuts
0−5%
6−10%
11−15%
16−20%
21−25%
26−30%
Welfare Spending Cuts
0−5%
6−10%
11−15%
16−20%
21−25%
26−30%
Public Sector Job Cuts
0−5%
6−10%
11−15%
16−20%
21−25%
26−30%
Pension Cuts
0−5%
6−10%
11−15%
16−20%
21−25%
26−30%
Income Tax Increase
0−5%
6−10%
11−15%
16−20%
21−25%
26−30%
Sales Tax Increase
0−5%
6−10%
11−15%
16−20%
21−25%
26−30%
Corporate Tax Increase
0−5%
6−10%
11−15%
16−20%
21−25%
26−30%
−0.1
0.0
0.1
−0.1
0.0
0.1
Effect on probability of support

But also key was estimating support for real-world austerity packages...

- Estimate expected level of support for three actual austerity policy packages implemented in 2009-2014 in:
  - Italy
  - Spain
  - Greece (externally imposed example included for comparison)
- Machine learning approach: stochastic gradient boosted trees
- Predictors:
  1. Austerity policy features
  2. Individual-level characteristics

Among Italian respondents, support for package Italy implemented is statistically indistinguishable from 50%.

Results in this figure are for the Italian sample only.



Evidence of strategic design by policymakers to craft package that maximizes austerity while maintaining sufficient public support.

Analysis of Heterogeneity

- Area of great interest and active research; overlaps with other research strands
- Analysis of conditional (or individual-level) AMCEs
  - Subgroup analysis: Different baseline averages $\rightarrow$ Be careful with interpretation (Leeper, Hobolt, and Tilley 2019)
    - Comparing marginal means as alternative
  - Individual-level effects (Zhirkov 2021)
- Heterogeneity of AMCEs conditional on other attributes and profile distribution
  - Focus on interactions between attributes (e.g. Egami and Imai 2019)
  - Importance of profile distribution in definition of AMCE (de la Cuesta, Egami, and Imai 2021)
  - In absence of interactions, ACME is invariant to the profile distributions
  - How common are large interactions (substantively significant in magnitude relative to marginal effects)?
- Expect to see much more in the future, particularly leveraging ML methods

References

Abramson, Scott F., Korhan Koçak and Asya Magazinnik. 2019. "What Do We Learn About Voter Preferences From Conjoint Experiments?" Working Paper.

Bansak, Kirk, Michael M. Bechtel and Yotam Margalit. 2021. "Why Austerity? The Mass Politics of a Contested Policy." American Political Science Review 115(2).

Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins and Teppei Yamamoto. 2018. "The number of choice tasks and survey satisficing in conjoint experiments." Political Analysis 26(1).

Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins and Teppei Yamamoto. 2021a. "Beyond the Breaking Point? Survey Satisficing in Conjoint Experiments." Political Science Research and Methods 9(1).

Bansak, Kirk, Jens Hainmueller, Dan Hopkins, and Teppei Yamamoto. 2021b. "Conjoint Survey Experiments," in Advances in Experimental Political Science, J. Druckman and D. Green (eds.), Cambridge: Cambridge University Press.

Bansak, Kirk, Jens Hainmueller, Daniel J. Hopkins and Teppei Yamamoto. 2021c. "Using Conjoint Experiments to Analyze Elections: The Essential Role of the Average Marginal Component Effect (AMCE)." Working Paper.

de la Cuesta, Brandon, Naoki Egami and Kosuke Imai. 2021. "Improving the External Validity of Conjoint Analysis: The Essential Role of Profile Distribution." Political Analysis, forthcoming.

Egami, Naoki and Kosuke Imai. 2019. "Causal interaction in factorial experiments: Application to conjoint analysis." Journal of the American Statistical Association 114(526).

Ganter, Flavien. 2020. "Identification of Preferences in Forced-Choice Conjoint Experiments: Reassessing the Quantity of Interest." Working Paper.

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

Hainmueller, Jens, Dominik Hangartner and Teppei Yamamoto. 2015. "Validating Vignette and Conjoint Survey Experiments against Real-world Behavior." Proceedings of the National Academy of Sciences 112(8).

Horiuchi, Yusaku, Daniel M. Smith and Teppei Yamamoto. 2018. "Measuring Voters' Multidimensional Policy Preferences with Conjoint Analysis: Application to Japan's 2014 Election." Political Analysis 26(2).

Horiuchi, Yusaku, Zach Markovich, and Teppei Yamamoto. 2019. "Does Conjoint Analysis Mitigate Social Desirability Bias?" Working Paper.

Jenke, Libby, Kirk Bansak, Jens Hainmueller, and Dominik Hangartner. 2021. "Using Eye-Tracking to Understand Decision-Making in Conjoint Experiments," Political Analysis 29(1).

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

Zhirkov, Kirill. 2021. "Estimating and Using Individual Marginal Component Effects from Conjoint Experiments." Political Analysis, forthcoming.