

# Evaluating the Effectiveness of AI-Powered Personalized Persuasion on Support for Transgender Rights: Pre-Analysis Plan

Mitchell Bosley<sup>1</sup> Charles Crabtree<sup>2</sup> John Holbein<sup>3</sup> Semra Sevi<sup>1</sup>

<sup>1</sup>University of Toronto

<sup>2</sup>Dartmouth College

<sup>3</sup>University of Virginia

## Overview

- **Kalla et al (2023)** show evidence that aligning messages with individuals' moral values can enhance persuasion on contentious topics like abortion rights.
  - Field experiment where canvassers were trained to learn voters' implicit moral values, then tailor persuasive messages regarding abortion rights.
  - They show that personalized conversations increase interest in taking action, and affect policy attitudes.
- However, field experiments are costly. In the vein of recent work by **Costello et al. (2024)** that shows that dialogues with **AI-powered chatbots** durably reduce conspiracy beliefs, can we extend this framework to personalized persuasion with chatbots?
- **Transgender Rights** remain a contentious issue; personalized communication may increase support.

## Research Design

- **Experimental Design:**
  - Random assignment to *Chatbot treatment* or *Pure Control group*.
  - Pre- and post-intervention surveys measuring attitudes.
- **Chatbot Personalization:**
  - Used OpenAI's GPT-4o via a custom chatbot interface for Qualtrics.
  - Used participants' Moral Foundations Profile.
  - Tailored conversations based on individual attitudes.
- **Dependent Variables:**
  - Transgender rights support (e.g., bathroom access, military service, etc.)
  - Feeling thermometer toward transgender people.
  - Likelihood of taking action (attending a rally, contacting a legislator, etc.)
- **Estimation Strategy:**
  - **Average Treatment Effect (ATE)** of the chatbot treatment on the dependent variables.
  - **Heterogeneous Treatment Effects (HTE)** by moral foundations (e.g., does the chatbot work better for those with high Care/Harm scores?)

## Moral Foundations Questionnaire

**Overview:** The Moral Foundations Questionnaire (MFQ) assesses five fundamental moral foundations that influence human moral reasoning and behavior (Haidt 2008).

### Moral Foundations:

1. **Harm/Care:** Concerns about the suffering of others and the desire to protect the vulnerable.
2. **Fairness/Reciprocity:** Emphasis on justice, rights, and fairness in social interactions.
3. **Ingroup/Loyalty:** Loyalty to one's group, family, or nation, and valuing group cohesion.
4. **Authority/Respect:** Respect for tradition, authority, and social hierarchies.
5. **Purity/Sanctity:** Concerns about cleanliness, sanctity, and avoiding contamination.

## Chatbot Prompt Design

### System Prompt

The chatbot engages voters on transgender rights, aiming to increase support and encourage action. It understands the voter's stance, presents strong arguments, and tailors the conversation based on moral foundations.

#### Guiding Principles:

- Approach with empathy, openness, and respect.
- Use evidence and personal stories.
- Adapt explanations to the user's knowledge level.
- Encourage meaningful action as allies.
- Communicate at a 9th-grade reading level.

### User Profile Summary

- **Moral Foundations Responses:** User's reported relevance and agreement levels on moral dimensions (e.g., Care/Harm, Fairness/Cheating, Loyalty/Betrayal, Authority/Subversion, Sanctity/Degradation).
- **Feeling Thermometer towards Transgender People:** User's warmth rating (0-100).
- **Agreement with Transgender Rights Statements:** User's level of agreement on key issues.
- **Self-Rated Knowledge:** User's perceived understanding of transgender issues.

### Chatbot Operational Steps

1. **Introduction:** Establish contact and introduce the topic.
2. **Ask Opinion:** Inquire about the user's views on transgender people.
3. **Identify Moral Foundations:** Determine key values based on the user's profile.
4. **Align Arguments:** Tailor persuasive arguments to the user's moral foundations.
5. **Strengthen Support or Address Disagreement:**
  - **Reinforce Support:** If the user agrees, provide additional support and encourage action.
  - **Address Disagreement:** If the user disagrees, provide respectful counterarguments and adapt strategies.
6. **Wrap Up:** Summarize the conversation and express appreciation.

## Example Conversation with Respondent

Hi there! I'd love to chat with you about transgender rights. Where do you stand on issues related to transgender people's rights and privileges?

I do not have a say on transgender because I do not recognize people who see themselves as trans. But I can live with them, though.

It's good to hear that you're open to living alongside transgender people. I'm curious, do you know anyone who is transgender personally? If so, how did interacting with that person impact your perspective?

I know a few. I just was not comfortable talking with them at first, but as time went on, I became comfortable with them.

It's understandable to feel uneasy at first, but it's great to hear you've grown more comfortable. I think it's really important to treat everyone with respect and care, especially since compassion for those who are suffering is a crucial virtue you hold.

Given that you value fairness and minimizing suffering, do you think transgender individuals should be treated with the same respect and fairness as everyone else?

Yes, I think everyone is human and should be treated equally and respectfully.

Absolutely, and that aligns perfectly with your value of ensuring that everyone is treated fairly. When we respect transgender people and ensure they have the same rights, we help reduce the emotional suffering they might experience from unfair treatment or discrimination.

For example, allowing transgender individuals to use bathrooms that match their gender identity helps them live more comfortably and safely. How do you feel about supporting such policies to ensure fairness and respect for everyone?

I think such policies should be approved so that we can all live peacefully.

That's wonderful to hear. By supporting these policies, we're not only promoting peace but also affirming the dignity and safety of transgender individuals.

For those of us who believe in compassion and fairness, actions like contacting lawmakers about supportive bills or joining advocacy groups can make a real difference. Do you think you might be interested in taking such actions to help promote equality?

Yes, I am.

That's fantastic! Your support can truly help make a positive impact. To sum up our conversation, we've agreed that treating everyone with respect and fairness is crucial, and supporting policies that protect transgender rights is a step toward ensuring a just society for all. If you're interested in getting involved, you might consider contacting your local representatives or supporting groups that advocate for transgender rights. Thank you for taking the time to discuss this important issue!

Thank you, too.

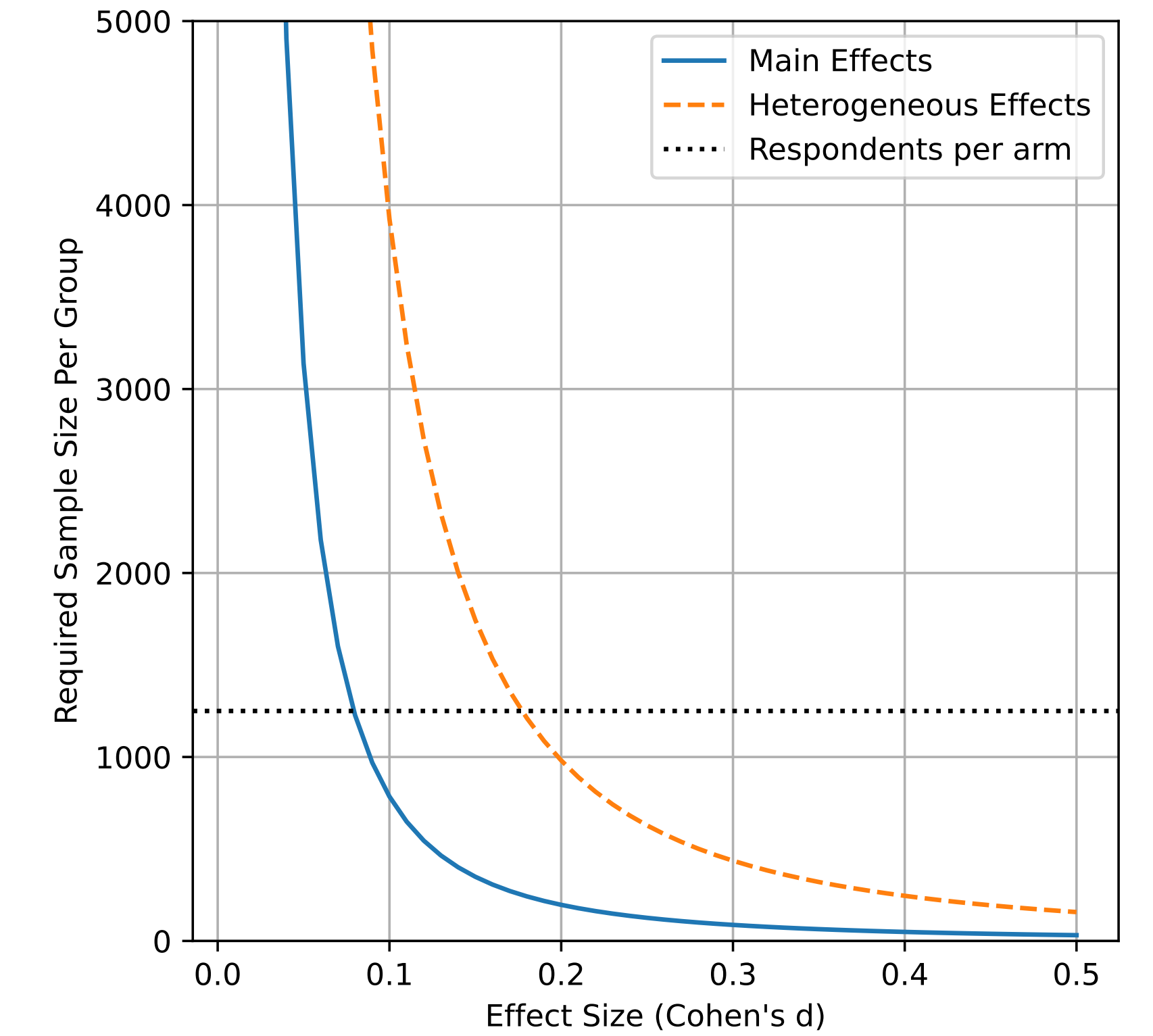
You're very welcome! I'm so glad we had this conversation about promoting fairness and respect for everyone, including transgender individuals. Your willingness to support these policies is truly commendable.

That's all the time we have for today. Thanks for taking the time to talk with me; you can now proceed forward by clicking the arrow below.

Type your response here...

Submit

## Power Analysis



- **Main Effects:** 2500 participants (1250 per arm) needed to detect a small effect size ( $d = 0.08$ ) with 80% power.
- **Interaction Effects:** Can detect heterogeneous treatment effects of roughly 0.18 with 80% power.
- Study well powered for main effects and moderate interaction effects.

## Next Steps

1. **Pilot Study:**
  - 500 respondents from Prolific, sampled to be representative of the US population, currently underway.
2. **Full Study:**
  - 2500 respondents from Prolific, with random assignment to treatment and control groups.
  - Pre- and post-intervention surveys measuring attitudes.
3. **Analysis:**
  - Estimate ATE and HTEs of the chatbot treatment on the dependent variables.
  - Assess the effectiveness of personalized persuasion

## Try the Chatbot!

