Essays on Trust and Polarization in the Modern Era

by

Tanner Bivins

A dissertation accepted and approved in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

in Economics

Dissertation Committee:

Jonathan Davis, Chair

Benjamin Hansen, Core Member

Keaton Miller, Core Member

Neil O'Brian, Institutional Representative

University of Oregon

Spring 2024

DISSERTATION ABSTRACT

Tanner Bivins

Doctor of Philosophy in Economics

Title: Essays on Trust and Polarization in the Modern Era

This dissertation contains three empirical studies that examine how distinct interventions influence agent behavior and social outcomes in varying contexts. Leveraging both natural variation and lab experiments, each chapter contributes to the broader understanding of policy effectiveness, technological integration, and healthcare impacts within an economic framework.

Chapter 1 examines the relationship between US primary election policies and electoral outcomes from 1976 to 2020. I use a difference-in-differences approach to investigate whether adopting less restrictive primary systems impacts legislator extremism and voter turnout. I find that expanding ballot access causes legislator ideology to shift toward the median voter. This moderating effect is even more pronounced for newly elected representatives and is driven mainly by non-partisan primary systems. Over the same period, I estimate a decrease in general election participation following the adoption of ”open-type” primary systems. This paper offers a comprehensive view of primary election policies, underscoring the balance between enhancing representation and maintaining voter engagement.

Chapter 2 is a collaborative project with Jiabin Wu, Ethan Holdahl, and Conner Weigand. In this study, we experimentally explore the impact of AI as a supportive tool for players in a two-player trust game. The game begins with the trustee sending a message to the trustor. In certain scenarios, the trustee is aided by the large language model (LLM) ChatGPT when composing this message. In

other scenarios, the trustor uses GPT to interpret the message from the trustee, or both players may have access to GPT assistance. Our findings indicate that when the trustee utilizes GPT as a helper, it enhances cooperation with the trustor. Interestingly, this improvement in cooperation is not attributed to GPT's superior messaging skills. Instead, it appears that when the trustee has GPT's assistance, it encourages the trustor to scrutinize the trustee's message more closely, understanding that it could be genuinely crafted, a mixture of personal input and GPT suggestions, or solely generated by GPT. The detailed scrutiny by the trustor, and potentially the trustee's awareness of this scrutiny, aligns the beliefs of the trustor with those trustees who send either genuine or mixed messages, thereby fostering an environment that encourages the development of trust.

Chapter 3 studies the relationship between stimulant medication and labor market outcomes in adults with Attention-deficit hyperactivity disorder (ADHD). In my analysis, I use linked employment and pharmaceutical data from the Medical Expenditure Panel Survey (MEPS) and leverage individual-level variation to estimate a two-way fixed effects regression. I find limited evidence to support a causal relationship between prescription behavior and employment, real wages, or weekly labor hours.

CURRICULUM VITAE

NAME OF AUTHOR: Tanner Bivins

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR, USA
University of Miami, Miami, FL, USA
Western Washington University, Bellingham, WA, USA

DEGREES AWARDED:

Doctor of Philosophy, Economics, 2024, University of Oregon
Master of Science, Economics, 2020, University of Oregon
Bachelor of Arts, Economics, 2017, University of Miami

AREAS OF SPECIAL INTEREST:

Applied Microeconometrics
Public Economics
Labor Economics

PROFESSIONAL EXPERIENCE:

GRANTS, AWARDS AND HONORS:

Graduate Teaching Award, University of Oregon, 2023

PUBLICATIONS:

For my mother, Dawnell. This accomplishment stands on the foundation of the countless sacrifices you made to get me here.

TABLE OF CONTENTS

Chapter Page

I. THE ROLE OF PRIMARY SYSTEMS IN POLITICAL POLARIZATION AND PARTICIPATION 12

1.1. Introduction 12

1.2. Primary Election Systems 18

1.2.1. State Policy Changes 22

1.3. Data 23

1.3.1. Historical Primary Systems 23

1.3.2. Legislator Ideology 24

1.3.3. Voter Turnout 28

1.4. Methodology 29

1.4.1. Identification 30

1.5. Legislator Ideology and Political Polarization 31

1.5.1. Entrants vs Incumbents 34

1.5.2. Relative Moderation and Electoral Competition 36

1.6. Political Participation 40

1.6.1. General Election Turnout 40

1.6.2. Primary Election Turnout 43

1.7. Conclusion 45

II. DOES AI FACILITATE TRUST? AN EXPERIMENTAL STUDY WITH CHATGPT 58

2.1. Introduction 58

2.2. The Experiment 63

2.2.1. Experimental Design 63

Chapter Page

2.2.1.1. Treatment Design 64

2.2.1.2. Prompt Design 65

2.2.2. Experimental Procedure 66

2.2.3. Hypotheses 67

2.3. Results 68

2.3.1. Treatment Effects 68

2.3.2. Message Classification 72

2.3.2.1. Player B 72

2.3.2.2. Player A 81

2.4. Discussion 83

III. ADULT ADHD, STIMULANT MEDICATION, AND LABOR MARKET OUTCOMES 85

3.1. Introduction 85

3.2. Background 86

3.3. Data 88

3.3.1. Household Component 88

3.3.2. Medical Component 89

3.3.3. Merged Data 90

3.4. Methods and Results 92

3.4.1. Baseline Model 92

3.4.2. Two-Way Fixed Effects 93

3.5. Occupational Sorting 95

3.6. Conclusion 96

APPENDICES

Chapter Page

A. DOES AI FACILITATE TRUST? AN EXPERIMENTAL STUDY WITH CHATGPT 103

A.1. Levenshtein Distance 104

A.2. Message Classification 105

A.2.1. Message Method 108

A.2.2. Player A's Messages 109

A.3. AI Prompts 112

A.3.1. Player B's AI Prompt 112

A.3.2. Player A's AI Prompt 113

A.4. Experiment Instructions 115

B. ADULT ADHD, STIMULANT MEDICATION, AND LABOR MARKET OUTCOMES 116

B.1. DSM-5 Diagnostic Criteria for ADHD 117

REFERENCES CITED 118

LIST OF FIGURES

Figure		Page
1.	Distribution of DW-NOMINATE Ideal Points (1976 - 2020)	56
2.	CFscores and DW-NOMINATE Comparison	57
3.	Screenshot of player B's screen when they are able to use AI. Player A sees a similar tree.	64
4.	The percentage of observations in each treatment where player A chooses 'In', player B chooses 'Roll', or both.	69
5.	Average first-order and second-order beliefs across the four treatments. Confidence bands are calculated at the 95-percent level.	72
6.	Types of messages sent across all treatments, omitting skipped messages. Note that $\approx 1/3$ of B's in the sample opted to skip sending a message.	75
7.	Distribution of player choices by message classification.	76
8.	Left panel: proportion of messages sent by player B which are promises. Right panel: proportion of messages sent by player B which are promises or asking.	77
9.	Over half of player B's who sent a message did so using mostly their own authorship. Details of how these methods were imputed are included in the appendix.	79
10.	Distribution of player choices across treatments by message authorship.	80
11.	Player B's message type and the associated interpretation by player A's AI.	82
12.	Player B's message type and the associated interpretation by player A's AI.	83
13.	Proportion of individuals with ADHD by year	102
A.1.	Diagram depicting the classification of Player B's messages, with an example below.	105

Figure	Page

A.2. 106

A.2. 107

A.3. Each data point represents a message sent by player B. The horizontal axis represents the string similarity (nLV) between the first message that player B sends to GPT, and the message which player B sends to player A. The vertical axis represents the string similarity between the penultimate message suggested by GPT and the actual message which player B sends to player A. Classification is determined by a "k-means" cluster algorithm with $k = 3$. 108

A.4. Left: Elbow plot showing within sum-of-squares drop-off for $K = 1, dots, 10$. Right: Visualization of gap statistic for $K = 1, \ldots, 10$. Both methods for determining optimal number of clusters suggest $k = 3$. 108

A.5. Examples of player B's sent message vs. the interpretation by player A's AI assistant. 109

A.5. 110

A.5. 111

A.6. Experiment Instruction - Both players have AI. 115

LIST OF TABLES

Table Page

1. Absolute Ideological Distance Results - All Representatives (1976 - 2020) 47
2. Absolute Ideological Distance Results - Entrants (1976 - 2020) 48
3. Absolute Ideological Distance Results - Incumbents (1976 - 2020) 49
4. DIME Ideological Distance Results (1980 - 2020) 50
5. Most Extreme Candidate (1980 - 2020) 51
6. Electoral Competition (1980 - 2020) 52
7. Moderate Candidate Win Rate (1980 - 2020) 53
8. General Election Voter Turnout (1976 - 2018) 54
9. Primary Election Voter Turnout (1976 - 2018) 55
10. Numerical values assigned to elicited beliefs. 71
11. Each message sent by player B (or recommended by AI) is encoded as vector which captures what player B intends to do and what they propose player A should do. Note that (Out, Roll) does not constitute a cooperative outcome unlike the other Promises. We nonetheless include it in Promise since it demonstrates Player B's intention to play Roll. We do not observe any encoded message of (Out, Roll). 74
12. Summary of the number of non-skipped messages sent by player B according to their type (row) and method (columns). Treatment is restricted to cases when player B has an AI. 80
13. Mean Values of Outcome Variables 98
14. Standard Fixed Effects Results 99
15. Two-Way Fixed Effects Results 100
16. Tabulation of Sample by Occupation (2002-2019) 101

CHAPTER I

THE ROLE OF PRIMARY SYSTEMS IN POLITICAL POLARIZATION AND PARTICIPATION

1.1 Introduction

Political polarization and ideological extremism in the US have increased over the last 20 years (Lewis et al., 2022). Growth in polarization implies that policy behavior by elected representatives appeals to a smaller, concentrated set of voters on either side of the political spectrum. Voters without strong party ties are often forced to decide between candidates with more extreme policy positions. This disconnect between moderate voters and extremist legislators causes voter apathy, as seen in downward trends in congressional approval (Brenan, 2022). As the ideological gap between political parties widens, moderates within these parties become less likely to run for office, leading to a reinforcing cycle of polarization (Thomsen, 2014). Concurrently, as extremism grows within Congress and among voters, affective polarization — the deepening animosity and mistrust between members of opposing parties — rises (Dias & Lelkes, 2022; Webster & Abramowitz, 2017). The costs associated with intensified partisanship can extend to policy efficacy and public health (Allcott et al., 2020; Cornelson & Miloucheva, 2022; Milosh, Painter, Sonin, Van Dijcke, & Wright, 2021; Trachtman, 2019). Finally, political polarization diminishes congressional production, causing increased difficulty in compromising on and passing legislation (Genicot, 2022; Jones, 2001).

Primary election reform is often proposed as a solution to the escalating polarization in Congress (Barber, McCarty, Mansbridge, & Martin, 2015; Lopez, 2023). In 2020 and 2022 alone, reforms concerned with primary election mechanisms were proposed in Alaska, Florida, Maine, and Nevada.[1] Traditional closed primary elections limit ballot access to voters affiliated with a political party. This is believed to generate a more partisan primary electorate, leading to more ideologically extreme candidates. Semi-closed primaries extend participation to both unaffiliated and independent voters. Open primaries, in contrast, grant any registered voter the choice to participate in any single party's primary, irrespective of their party affiliation. Non-partisan primaries list all candidates on the same ballot. In this system, the two candidates who receive the most votes, regardless of party, advance to the general election.[2] The motivation behind legislator moderation and primary election policy likely stems from theoretical models that extend Downs (1957) to include a primary election stage, where candidates will choose a policy position between the primary and general electorate medians in equilibrium (Aranson & Ordeshook, 1972; Coleman, 1972; McGann, 2002; Owen & Grofman, 1996). As primaries open, the ideological distribution of the primary electorate will broaden, shifting the median voter of the primary closer to the general median, which leads to more moderate policy positions by candidates.

Opponents of primary election reform argue that open primaries weaken party identity and infringe on a party's right to association, depressing general election turnout (Schmookler, 2017). Specific to non-partisan systems, like the Top-Two primary used in California and Washington, a common concern is that voters

[1]Florida Amendment 3 would have established the top-2 primary but failed to garner enough votes in 2020. Maine's bill LD 231 passed in 2022, with open primary elections beginning in 2024. Alaska's Measure 2 passed in 2020, establishing the top-4 primary system. The Top-Five Ranked-Choice Voting Initiative in Nevada was approved in 2022. However, Nevada requires approval in two election years for initiated constitutional amendments to be implemented.

[2]This specific system is referred to as the Top-Two primary in California and Washington. Louisiana uses a similar mechanism but with a distinct feature: if the leading candidate receives over half the total votes, they are declared the winner without the need for a runoff. Given these nuances, I categorize all three systems under the label non-partisan.

will be indifferent between candidates in same-party contests and become more likely to abstain (Gemma, 2017). On the other hand, voter turnout is expected to be higher in open primary elections due to lower barriers to participation for independent and unaffiliated voters.

I estimate the impact of different primary election systems on ideological moderation and voter turnout. To accurately identify changes in moderation and polarization, I require a reliable measure of legislator ideology. To this end, I use Nokken and Poole (2004)'s dynamic DW-NOMINATE ideal points. These are constructed from roll-call votes and yield a directional measure of legislator ideology. The absolute value provides a distance metric to identify relative levels of ideological moderation in legislators, where one is the perfectly extreme outcome and zero is the perfectly moderate outcome. In addition to DW-NOMINATE, I use Bonica (2014)'s campaign finance scores (CFscores) to analyze losing candidates and check for changes in the ideological distance between general election competitors. This approach informs my findings on participation and moderation by assessing the degree of competition between candidates through their policy positions.

I leverage state variation in primary election policy between 1976 and 2020 to estimate a difference-in-differences model with district and election-year fixed effects. This approach is best suited to assess the impact of primary election reform due to the large sample size and variation observed in the data.

I find that "open-type" primary policies, where states are considered treated if they use any one of non-partisan, blanket, or open primaries, reduce general election turnout by 5.8% but have no noticeable impact on legislator ideology or primary election turnout. However, I find an estimated 8.44% reduction in ideological scores when restricting the sample's start to 1996. This is likely the result of more complex temporal variation not captured by the election-year fixed effects. When analyzing entrants – those newly elected to office – I find small but statistically significant ideological shifts toward the median voter. This suggests that established incumbents might have more rigid stances, potentially attenuating the full-sample results.

In addition to pooled effects, I estimate the policy-specific effects of non-partisan, blanket, open, and semi-closed primaries. I find that adopting non-partisan primaries causes a 15.64% reduction in ideological extremism. These effects are consistent across both major political parties and alternative ideal point measures. When conditioning on incumbency status, I estimate a modest adjustment effect of -4.47% for re-elected incumbents. As anticipated, entrants exhibit the most significant change, with a -36.69% shift in ideal point estimates. From these results, there is no systematic relationship between less restrictive primary policies and political polarization. Instead, the distinct non-partisan nature of Top-Two primary elections causes moderating effects.

There is substantial evidence that more open primary policies are associated with decreased voter turnout in the general election, with large and precise estimates for all primary types except semi-closed. Further analysis showing decreases in the ideological distance between winning and losing candidates suggests some of these changes in participation might result from voters being indifferent to candidates with similar policy positions. Interestingly, my analysis of primary turnout provides limited evidence to support a relationship between primary election reform and voter participation in the primary.

The empirical evidence remains mixed despite theoretical predictions that less restrictive primaries induce moderating effects. McGhee, Masket, Shor, Rogers, and McCarty (2014) analyzes state legislatures from 1992 to 2010 and finds little evidence that less stringent primary elections reduce ideological scores. Their analysis used ideal points from Shor and McCarty (2011) to measure changes in policy positions, which are mapped to a single score. In addition to limiting analysis solely to entrants, static ideal points mean ideology estimates are calculated using roll-call votes from a legislator's entire tenure, including pre-treatment and post-treatment periods. These same scores are used in McGhee and Shor (2017) to evaluate the impact of the Top-Two primary used in California and Washington, where they find minor moderating effects only for Democrats in California. In a separate study focusing on California's switch to the Top-Two primary, Kousser, Phillips, and Shor (2018) provides evidence that policy positions of elected representatives shifted away from the median voter following the policy change. In an experimental analysis, Ahler, Citrin, and Lenz (2016) found that voters may actually lack the ability to discern between moderate and ideologically extreme legislators. Rogowski and Langella (2015) uses ideology scores calculated from campaign contributions to analyze candidates who ran in primary and general elections between 1980 and 2012, including those who lost. Their results suggest that primary election restrictions are not systematically associated with legislator moderation. Conversely, Crosson (2021) finds that same-party general election races in California and Washington lead to more moderate legislators.

Grose (2020) also studies the impact of primary election systems on polarization. Using DW-NOMINATE scores for US House of Representatives members between 2003 and 2018, he finds that the Top-Two primary system moderates legislator ideology, with open primaries having a milder influence. Like many studies in this area, Grose (2020) uses a difference-in-differences approach with election year and state fixed effects. Importantly, large states with sizable urban and rural populations possess significant within-state variation in political and socioeconomic characteristics. By using district rather than state fixed effects, my analysis better accounts for this between-district heterogeneity, increasing the reliability of the findings. Furthermore, my paper differs by covering the years 1976 to 2020, adding 5,622 elections to the analysis. While there is valuable insight from estimating additional models that narrow analysis to recent years, it is important to include a wider set of elections to capture primary policy variation before 2003.[3]

The literature is rich with studies examining aggregate-level voter turnout (Cancela & Geys, 2016; Stockemer, 2017). However, the relationship between primary systems and participation remains mostly unexplored. Calcagno and Westley (2008) shows that states with less restrictive primary policies tend to have greater voter turnout in general elections, while Geras and Crespin (2018) find little support for open primaries increasing general election turnout. Recent studies have focused on the Top-Two primary and general election races with two members of the same party. Nagler (2015) finds that voters were more likely to abstain from same-party congressional contests in California if the two candidates were from the opposing party. While there has been limited evidence to suggest Proposition 14 in California and Initiative 872 in Washington depressed general election turnout, several analyses have connected the adoption of the Top-Two primary to an increase in voter roll-off, especially in same-party contests (Bonneau & Zaleski, 2021; Fisk, 2020; Patterson Jr, 2020). Conversely, Henrickson and Johnson (2019)

[3]Between 1976 and 2002, there were 16 instances of states switching primary election systems.

use administrative data on county-level general election turnout in Washington to show that the Top-Two primary is associated with increased voter participation.

My paper contributes to the growing literature on primary policies and electoral outcomes in several ways. First, to my knowledge, this is the first study that thoroughly evaluates how primary election reform affects both political polarization and voter turnout. It is crucial to consider both outcomes together to gain a complete understanding of the impact of open primaries. By estimating changes in both sets of outcomes over the same period and using consistent coding of primary election policies, I can draw straightforward comparisons across the different sets of results. Additionally, including more elections and using district fixed effects increase the reliability of estimates.

The remainder of the paper is organized as follows. Section 1.2 briefly overviews the alternative primary election systems and recent policy changes. Section 1.3 covers data used in the analysis. Section 1.4 presents the empirical strategy and necessary identifying assumptions. Section 1.5 reports the results for measures of legislator moderation and section 1.6 covers voter turnout results. Section 1.7 concludes with potential for future research and policy implications.

1.2 Primary Election Systems

Primary elections function as the first stage in a two-stage process of electing representatives. Traditionally, for a given partisan role, a series of within-party primary elections select a single candidate to represent their respective party in the general election (second stage). Although the method of electing representatives is largely uniform across states, primaries represent a unique electoral mechanism in which state processes can vary significantly. The systems used in primary elections are established by each state independently. These systems vary in the set of available choices to voters based on their political party affiliation.

In a closed system, voters must register with a specific party in advance to participate in their primary election. Independent and unaffiliated voters may not participate in either major party's primary. Historically, closed primary elections were more common, with 23 states using the system in 1976. By the 2020 election, this number had reduced to just ten states. Proponents of closed primaries argue that they maintain voters' right to association and encourage party unity. Additionally, closed primaries may limit strategic voting, where members of an opposing party intentionally vote for the candidate with the lowest probability of winning in the general election (Cherry & Kroll, 2003). Voters can still strategically vote by registering for the opposing party, but closed primaries impose a barrier by restricting ballot access.

Semi-closed systems loosen the restrictions of closed primaries in a couple of different ways. To be classified as semi-closed, a primary must possess one, or both, of two policies governing voter enfranchisement. First, many states mandate ballot access to independent and unaffiliated voters for either major party's primary election. Second, states may have a policy of party choice where each party decides which voters to allow. For example, in Idaho, the Republican primary is closed, but the Democratic primary allows independent and unaffiliated voters to participate. The decline in states using closed primaries coincided with an increase in the number of states utilizing semi-closed primaries. In 1976, four states held closed primaries, whereas 15 held semi-closed primaries in 2020. More recently, five states switched from closed to semi-closed since the 2000 election.[4]

[4] A number of the arguments for closed primaries are also applicable to the semi-closed system.

Semi-open and open primary systems, while maintaining their partisan nature with each party conducting a separate primary to select a candidate for the general election, further broaden the scope of voter participation. Like semi-closed systems, independent and unaffiliated voters are granted the opportunity to vote in any single primary election. Diverging from the limited crossover in closed and semi-closed primaries, registered Republicans and Democrats in these systems can cross party lines and participate in any single primary of their choosing. The distinguishing factor between semi-open and open systems lies in the disclosure of voting decisions. In semi-open primaries, the ballot choice becomes public knowledge, with some states counting ballot choice as registration for that specific party. In my analysis, open and semi-open primary systems are considered indistinguishable as the pool of available voters is the same. Proponents of open primary elections contend that this system encourages the nomination of more moderate candidates and addresses the potential disenfranchisement of independent or unaffiliated voters.

Non-partisan primary systems are a particular case where all party primaries are condensed into a single election with candidates listed on the same ballot, regardless of party affiliation. These systems are viewed as the opposite end of the spectrum to closed systems, holding a single primary in which all voters participate. In 2008, Washington became the first state to use the Top-Two primary system, where all candidates are listed on a single ballot, and voters are not required to be affiliated with the party. Regardless of party affiliation, the two candidates with the most votes are sent to the general election. An essential result of this system is that it enables the possibility of two candidates from the same party to compete in the general election. California adopted the Top-Two system in 2012, while

Alaska recently adopted a similar top-four system in 2022.[5] Like the Top-Two system, Louisiana primaries list all candidates on a single ballot. However, if a candidate in Louisiana garners more than half of the total votes, they are declared the winner without a subsequent general election. In the absence of such a majority, a runoff election ensues. I code Louisiana's primary and Top-Two systems as non-partisan because they have a single primary for all parties. It is conceivable that both voters and candidates give Louisiana's primary elections more weight than those of states with different non-partisan systems, as it could lead to the direct election of representatives. Estimates from alternative models excluding Louisiana are presented alongside the main results to account for possible effects on voter and candidate behavior.

Supporters of non-partisan primaries argue that they lead to more moderate representatives. The idea is that vote-maximizing candidates base their ideological positions on their respective pool of voters. Removing partisan elections theoretically increases each candidate's pool of voters to include those with differing ideological points. Opponents believe non-partisan primaries depress voter turnout in the general election because two candidates from the same party may compete, leading to voter apathy or indifference. The impact on voter turnout in primary elections is also worth considering. On the one hand, it is reasonable to assume that allowing all voters to participate in the same primary would increase the voter pool, leading to increased turnout. Conversely, if non-partisan systems cause voters to lose their right to association and party identity, or if the system is too confusing, voter turnout in the primary could decrease.

[5] In Alaska's Top-Four primary, the four candidates with the most votes will advance to the general election.

The blanket primary, deemed unconstitutional in 2000, shares similarities with the Top-Two system used in California and Washington. Both systems list all candidates on the same ballot, enabling any registered voter to select their preferred choice. However, blanket primaries differ in that the leading candidate from *each party* progresses to the general election. These systems are often coded as non-partisan because voters are not required to declare which primary they want to participate in before the election. Coding the blanket primary as non-partisan implicitly assumes that it is indistinguishable from the Top-Two or Louisiana primaries. The pool of potential voters remains the same across blanket and non-partisan primary elections, but the competition among candidates differs. All potential representatives compete on the same primary election ballot in Top-Two systems, regardless of party affiliation. In contrast, candidates still compete exclusively with members of the same party in blanket primaries. Therefore, I will classify the blanket primary as distinct from the Top-Two and Louisiana primaries in the baseline analysis while providing estimates from alternative specifications where it is coded as non-partisan.

1.2.1 State Policy Changes. Between 1976 and 2020, there were 28 total instances of policy changes to primary election systems. Fifteen of these changes took place in the last 30 years. These policy changes occurred in 20 states, representing 39.54% of the US population and comprising 176 House representatives in 2020. Variation in primary policies stems from a combination of initiatives, state constitutional amendments, and court decisions.

In 1996, California passed Proposition 198, transitioning from closed to blanket primaries. However, this switch was contested and culminated in the 2000 Supreme Court case *California Democratic Party v. Jones*, which ruled the blanket primary unconstitutional. As a consequence, Alaska, California, and Washington had to abandon their blanket primary systems. Washington prolonged the switch until 2004 when it adopted an open primary system, while Alaska and California shifted to semi-closed primaries in 2000 and 2002. In 2004, Washington's Initiative 872 ushered in the Top-Two primary. This system faced its own legal challenge but was upheld as constitutional in 2008's *Washington State Grange v. Washington State Republican Party et al.*. Later on, California adopted the Top-Two primary in 2012, and Alaska introduced the Top-Four primary in 2022.

1.3 Data

In this paper, I examine US House elections from 1976 to 2020 to analyze the impact of primary election systems on extremism, polarization, and voter turnout. This requires data on primary systems used by each state across all years of analysis, estimated legislator ideology, and district-level voter turnout. I describe these data and their sources here.

1.3.1 Historical Primary Systems. A crucial component of this analysis is the classification of state primary systems. Historical policy data for elections between 1976 and 2012 come from Sinclair (2013), which provides an exhaustive history of primary elections for all 50 US states, synthesizing both prior literature and state election office records. For the years 2014 to 2020, I obtained primary election policy information from online archives. The National Conference of State Legislatures (NCSL) website served as the central resource, detailing contemporary primary election policies for each state. In instances where a specific year was not archived on the NCSL site, I consulted the online historical voting guides from individual states, which include descriptions of eligibility criteria for ballot access to different primaries.

1.3.2 Legislator Ideology.

I use representative-level data on the United States House of Representatives from 1977 to 2022 to investigate the relationship between moderation and primary election policy.[6]. DW-NOMINATE point estimates provided by Voteview are used to measure legislator ideology.[7] For each term a legislator serves, the data contains an ISCPR identifier, DW-NOMINATE scores, party affiliation, and the district where the representative was elected. The final sample includes 9,977 ideal point estimates across 1,935 representatives and 23 congressional sessions. The average representative in the data served four two-year terms, with the longest tenure lasting all 23 sessions.[8]

Legislator ideology is measured by dynamic DW-NOMINATE scores, a multidimensional scaling method that takes legislator roll-call votes and assigns a score between -1 and 1 to each legislator. This score represents their ideal point in the choice space, with -1 being more liberal, 1 being more conservative, and 0 being perfectly moderate. A fundamental assumption of the DW-NOMINATE approach is that every legislator has a single-peaked utility curve centered on their ideal point. Since DW-NOMINATE scores stem from observed behavior, they reflect inferred ideologies rather than explicitly stated ones.

The Voteview dataset offers two sets of ideal points. One is estimated with each legislator's cumulative tenure throughout all congressional sessions. While they provide more precise ideological readings, these points remain static over time, making them unsuitable for capturing shifts in response to policy

[6] The data used to estimate ideology is from 1977 through 2022. However, the analysis is on elections taking place from 1976 to 2020. A representative elected in 1976 will produce roll-call data in 1977 and 1978.

[7] Lewis, Jeffrey B., Keith Poole, Howard Rosenthal, Adam Boche, Aaron Rudkin, and Luke Sonnet (2022). Voteview: Congressional Roll-Call Votes Database. https://voteview.com/

[8] Alaska Representative Donald Edwin Young

changes. For instance, if a representative's district adopts a new primary system in year T, their ideal points will remain constant for all years before and after T, which hinders accurate tracking of post-treatment moderating effects. My paper uses an alternative set of DW-NOMINATE ideal points from Nokken and Poole (2004), calculated for each congressional session. This approach facilitates the examination of the evolving ideologies of legislators over time. Such dynamic estimates become especially useful when analyzing incumbents, as they might reflect changes stemming from a legislator adjusting to new conditions following primary policy reform (Grose, 2020). Recently, methods to estimate legislator ideology from campaign contributions and Twitter activity have been developed (Barberá, 2015; Bonica, 2014). Prioritizing a comprehensive sample, I use dynamic DW-NOMINATE ideal points but also present results when using alternative measures of ideology to support the main findings.

Figure 1 presents distributions of DW-NOMINATE ideal points for different primary systems. The average Republican's ideal point is 0.371, and the average Democrat's is -0.338. As expected, all distributions are bimodal, with the average ideological score close to zero. Open primaries are the only policy with a positive average ideal point, indicating a slight skew towards more conservative policy preferences.

In my analysis, I measure an elected representative's level of moderation via their ideological distance from zero, constructed by taking the absolute value of all ideal points. A hypothetical legislator with an ideal point equal to zero is interpreted as being perfectly moderate, so smaller ideological distances translate to more moderate policy preferences. The average absolute nominate score rose from 0.298 for legislators elected in 1976 to 0.430 for those elected in 2020, a 44.30%

increase. Broken down by incumbency, we see a more significant trend for entrants than incumbents, with increases of 90.42% and 40.79%, respectively. Grose (2020) presents an alternative outcome where Democratic representatives have their scores multiplied by -1. This is done to account for party members whose ideal points align with the opposition, such as a Democrat with a positive DW-NOMINATE score or a Republican with a negative score. The idea is that such representatives may be viewed as more moderate than those who align with their party but have the same ideological distance from zero. However, I have chosen not to include this outcome in my analysis as I define extremism based on absolute distance rather than direction. For additional testing, estimates using this alternative outcome will be provided in the Appendix.

One downside to estimating ideology from roll call votes is that it requires a representative to be elected, preventing analysis of losing candidates. To provide a well-rounded examination of the relationship between primary policies and political polarization, I include additional analysis using campaign finance-based ideology scores (CFscores) from the Database on Ideology, Money in Politics, and Elections (DIME).[9] This method presented by Bonica (2014) estimates ideological positions based on campaign donations, allowing CFscores to be calculated for losing candidates.[10] Excluding special elections, runoffs, and uncontested races, there are 12,085 observations across 5,980 elections held from 1980 through 2020.

CFscores differ from DW-NOMINATE estimates in that they are unbounded. In the sample, CFscores range from -5.10 to 4.87, with negative (positive) values corresponding to liberal (conservative) policy positions. The

[9] Bonica (2023)

[10] The CFscores used in this paper are estimated for each congressional session, similar to the Nokken-Poole estimates.

different methods used to estimate the two scores – campaign donations and roll-call votes – support different interpretations. DW-NOMINATE scores exhibit observed policy behavior from elected representatives, while CFscores convey signaled, or promised, policy positions. Notwithstanding, the two sets of ideal points are significantly correlated, with a coefficient equal to 0.905.

I map CFscores to the $[-1, 1]$ interval using predicted DW-NOMINATE ideal points to facilitate direct comparison across results.[11] I implicitly assume a linear relationship between DW-NOMINATE estimates and CFscores by using fitted values for analysis. Figure 2 plots the two ideological measures for Republicans and Democrats, where there is a clear linear relationship. Additionally, the adjusted R-squared suggests CFscores explain 81.97% of the variation in DW-NOMINATE estimates, supporting the reliability of the predicted values.

Following the main analysis, I compute the absolute value of each representative's CFscore to assess ideological moderation. Winning candidates tend to be more ideologically moderate than their losing counterparts, registering average scores of 0.37 and 0.51, respectively. Across all elections, the most extreme candidate has an average score of 0.55. Notably, these candidates lose 71.12% of their contests.

In addition to candidate-level analysis, I construct two district-level metrics: one examining the percentage of elections won by the more moderate candidate, and the other assessing the ideological gap between winning and losing candidates.[12] The more ideologically moderate candidate won 72.4% of elections between 1980 and 2020. Underscoring the rise in polarization, the ideological

[11] Specifically, I use Ordinary Least Squares to estimate $DW_{idt} = \alpha + \beta CFscore_{idt} + \varepsilon_{idt}$.

[12] The ideological distance between winning and losing candidates in district d and election-year t is calculated from $|CFscore^{W}_{dt} - CFscore^{L}_{dt}|$

distance between competing representatives grew from 0.68 to 1.14, a 67.6% increase.

1.3.3 Voter Turnout. To investigate the relationship between primary election policies and voter turnout, I use comprehensive district-level election results data from U.S. House general elections in all 50 states from 1976-2018. District vote totals come from the MIT Election Data and Science Lab[13] (MEDSL) with district population sourced from the National Historic Geographic Information System[14] (NHGIS). Voter turnout data are at the candidate level, where each observation includes the total votes received in a given election, while write-in candidates are excluded from analysis along with special elections. By estimating turnout from candidate totals, I am excluding instances of voter roll-off. After aggregating the total votes cast in each election across all candidates, there are 9,325 observations left in the data.

Owing to the limited availability of district-level data, I use total district population estimates as a substitute for the voting-eligible population. To circumvent concerns about time-varying heterogeneity in age distribution across districts, I use the log of total votes to estimate shifts in voter turnout following primary policy changes in the main analysis. However, there is not a straightforward interpretation of log values, so I will discuss the characteristics of the data with voter turnout as a proportion of the total population.

Across all districts in the data, there is an average voter turnout rate of 32.6% in US House general elections. Districts with non-partisan primaries possess

[13] Data and Lab (2017)

[14] Steven Manson, Jonathan Schroeder, David Van Riper, Tracy Kugler, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 17.0 [dataset]. Minneapolis, MN: IPUMS. 2022.

the lowest average turnout at 31.5%, and districts with semi-closed primaries hold the highest average turnout at 34.4%. Turnout in open districts is slightly above average at 32.8%, while closed districts have below-average turnout at 31.9%.

The MEDSL data contains vote totals for every candidate in each election, enabling estimation of voter turnout by party. This provides useful insight if one specific party was more susceptible to changes in voter turnout following the adoption of new primary policies. One of the arguments for open primaries adversely impacting turnout is the potential of voter apathy from a lack of party identity. If voters in one party are more affected by voter apathy, we would expect relatively larger decreases in voter turnout for one party over the other. Across all elections in the data, the average voter turnout was 21.7% for Republican candidates and 22.2% for Democratic candidates.

One notable outcome of non-partisan primaries is the potential for two candidates from the same political party to compete in the general election. Such contests might deter members of the opposing party from voting, thereby reducing overall voter turnout. In the data, 46 elections feature two Democratic candidates and 20 feature two Republican candidates. General elections with competitors from the same party see an average turnout of 28.7%, a drop from the 32.7% observed in traditional elections.

1.4 Methodology

To identify the causal effect of primary election policy on measures of legislator ideology and voter turnout, I estimate the following difference-in-differences model:

$$Y_{idt} = \beta D_{dt} + \gamma X_{idt} + \eta Y_{id,t-1} + \mu_d + \rho_t + \varepsilon_{idt} \tag{1.1}$$

where Y_{idt} is the outcome of interest for representative i in district d in election-year t when analyzing measures of ideology. For voter turnout estimates, the i subscript is dropped, and Y_{dt} is the natural log of the total number of votes cast in elections that took place in district d. When analyzing the pooled effect of adopting any open primary system, D_{dt} is a dummy variable that takes a value of one if the primary system used in district d for election-year t was one of open, blanket, or non-partisan. Disaggregated effects for each policy are reported alongside the pooled estimates, in which D_{dt} is a vector of dummies for district d's primary election system in year t. Specifically, D_{dt} contains indicators for non-partisan, blanket, open, and semi-closed primaries, with districts in states possessing closed primaries as the reference group. μ_d and ρ_t are district and election-year fixed effects, respectively. X_{idt} includes district-level controls that account for time-varying characteristics not captured by district fixed effects and $Y_{id,t-1}$ controls for lagged outcomes. Standard errors are robust to heteroskedasticity and clustered at the state level (Bertrand, Duflo, & Mullainathan, 2004).

1.4.1 Identification. Assuming parallel trends is required to interpret results as causal. Under parallel trends, it is assumed that, in the absence of treatment, electoral outcomes in states with alternative primary systems would have evolved similarly to those with closed primaries, given the inclusion of relevant controls, district fixed effects, and election-year fixed effects.

Parallel trends requires that policy changes are not anticipated. In most cases, there is a lag between a state's decision to reform primary election policy and the implementation of the new system. To illustrate how this could affect identification for legislator ideology estimates, consider Washington State's Initiative 872, which passed in 2004 to establish the top-two primary. Due to court challenges, the first election to utilize the new system did not occur until 2008, after it was deemed constitutional by the Supreme Court.[15] Nominate scores for representatives elected in 2004 and 2006 were calculated from roll calls taking place in 2005-2006 and 2007-2008, respectively. If legislators view roll call decisions as a signal of ideology (or party loyalty) to voters, the adoption of the top-two primary may influence policy behavior towards the median voter in an attempt to maximize votes in the upcoming election. To account for anticipation, pre-trend estimates are reported.

1.5 Legislator Ideology and Political Polarization

Table 1 contains results for legislator ideology, measured by the absolute value of dynamic DW-NOMINATE ideal points. Estimates in Panel A give the pooled effect of any open-type policy with closed and semi-closed primaries as the reference group. Panel B contains the disaggregated effects by policy, with closed districts as the reference group. Estimates in Column 1 show the estimated impact on extremism controlling for district and election year fixed effects. In Column 3, I explore an alternative specification where blanket primaries are categorized as non-partisan. Recognizing the distinct nature of Louisiana's primary system, Column 5 estimates come from a specification that excludes all congressional elections in Louisiana. In Column 7, I examine the results of a regression limited to the period between 1996 and 2020, which I will refer to as the "modern era." This is done to address concerns about the relevance of results for informing present-day policies and to consider shifts over longer time periods that election-year fixed effects may not capture. Such changes might encompass the technological revolution and the intensified political divisions that transformed the political landscape around the

[15] Washington State Grange v. Washington State Republican Party, et al., 552 U.S. 442, 128 S. Ct. 1184, 170 L. Ed. 2d 151 (2008)

turn of the 21st century. Columns 2, 4, 6, and 8 offer estimates using the respective specifications, with the additional control for lagged outcomes. Tables 2 and 3 have the same layout but present separate results for entrants and incumbents.

From the baseline results, the pooled effect is calculated to be -0.068 with controls and -0.028 without. Both estimates possess 95-percent confidence intervals that just barely cover zero. Hence, these results fail to determine the strength and size of the relationship between less restrictive primaries and moderation. Turning to the disaggregated effects in Panel B, non-partisan primaries lead to an estimated 0.128 reduction in ideological extremism. This effect is statistically significant, with a 95% confidence interval ranging from -0.171 to -0.085. The average Democrat has an ideological distance of 0.366, and the average Republican has 0.398, so the effect is about -33.61% for Democrats and -32.16% for Republicans, implying similar treatment effects across parties. Furthermore, from 1976 to 2020, average DW-NOMINATE scores increased by 0.132 across both parties, so the reduction associated with non-partisan primaries offsets this rise. After introducing controls, the estimated effect decreases to -0.056 but maintains significance. Blanket primaries are estimated to shift policy positions by -0.019, while the effects of Open and Closed primaries are indistinguishable from zero. The results indicate that movement towards more open policies and providing ballot access to a broader range of voters may not systematically impact legislator ideology. However, specific characteristics of non-partisan and blanket primaries not shared with open primaries, such as listing all candidates on the same ballot or not requiring a declaration of party affiliation, have a moderating effect on elected representatives.

In Columns 3 and 4, estimates come from a model where blanket primaries are coded as non-partisan. This alternative specification is considered due to the functional similarities between the two systems in how they affect ideology. If candidates adjust their policy positions based on that of the median voter, variation in ideological points must be driven by changes to the electorate's ideological distribution. Candidates participating in a blanket primary would have access to the same pool of voters as they would in a Top-Two primary, like in Washington. The introduction of blanket primary elections reduces the estimated impact of implementing non-partisan primaries, with a decrease in size to -0.102 without controls and -0.044 with controls. Even though the point estimate is smaller, the moderating effect is still significant, resulting in a 28.49% decrease without controls and a 12.29% reduction with controls.

Removing Louisiana representatives from analysis yields no marked change in the magnitude or precision of the pooled or disaggregated effects. What distinguishes Louisiana's primary elections from the Top-Two systems of California and Washington is that a candidate in Louisiana can bypass the second stage if they secure over 50% of votes in the primary election. Yet, the consistency in findings, regardless of Louisiana's inclusion or exclusion, implies that the potential of a second stage does not augment the influence of non-partisan primaries on legislator ideology.

Columns 7 and 8 report estimates from a specification where analysis is restricted to 1996 to 2020. This is the only specification where pooled effects possess statistical significance. The 95-percent confidence interval for this estimate is -0.040 to -0.024. Looking at the policy-specific results in Panel B, it appears that non-partisan primaries are driving the pooled effects. With the limited sample, blanket primaries do not appear to have a relationship with legislator ideology. This is the only set of results with evidence for semi-closed primaries having an impact on legislator moderation. Interestingly, this effect is estimated to increase ideological distance by 0.01. While there is evidence of a statistically significant relationship, the 95-percent confidence interval includes effects as low as 0.002.

1.5.1 Entrants vs Incumbents. In this section, I examine the contrasting effects between entrants — newly elected representatives — and incumbents who have been re-elected. The moderating effect on an incumbent legislator's ideology may differ from newly elected representatives. Firstly, if primary systems become more open, incumbents with established legislative histories may adjust their policy positions to appeal to a broader and potentially less partisan electorate. Conversely, the results may be driven by more moderate entrants winning primary elections due to a less partisan voter base.

Table 2 features results from specifications that exclude incumbents. Columns 1 and 2 report baseline estimates with and without controls. The pooled effects in Panel A provide evidence that expanding ballot access has a moderating effect on entrants of -0.088, a sizable impact, equating to 24.6% of the reference group's average ideological score. In contrast to estimates from the full sample, this effect is statistically significant, implying heterogeneous effects across incumbency status. As indicated by the policy-specific estimates in Panel B, the pooled effects predominantly stem from the influence of non-partisan and blanket primaries. Specifically, the transition from closed to non-partisan primaries causes a -0.126 shift among entrants. After introducing controls, this effect slightly increases to -0.131, with a 95% confidence interval between -0.178 and -0.084. Adopting blanket primaries leads to an estimated effect of -0.075. Semi-closed primary systems exhibit mild moderating effects but remain statistically insignificant. The estimates in Columns 3 and 4 indicate that the inclusion of blanket primaries does not noticeably impact the size or accuracy of the estimated effect of non-partisan primaries. Furthermore, the results in Columns 5 and 6 further support that the baseline estimates are unaffected by the exclusion of Louisiana representatives.

Columns 7 and 8 report estimates when restricting analysis to 1996 to 2020. Here, compelling evidence emerges that the ideological points of newly elected representatives shift with the introduction of more open primary policies. The pooled effect in Panel A is estimated to be -0.105. The average ideological score for an entrant elected in a closed primary election is 0.367, so opening primaries would have a moderating effect of -28.6%. All three open-type primary policies have large and precisely estimated effects. The most noticeable difference in these results is the effect of open primaries, where there is a -0.177 shift in the ideological positions of entrants. This effect is tightly bound between -0.242 and -0.112, as indicated by the 95% confidence interval.

Table 3 examines how primary policy shifts influence the ideological stances of incumbent representatives who retain their seats. The baseline results in Columns 1 and 2 show that incumbents generally remain consistent in their policy positions, even when primary elections become more open. However, a notable exception is observed with non-partisan systems, as indicated by the policy-specific estimates in Panel B. Before introducing controls, non-partisan primaries appear to reduce the incumbent's ideology by 0.125. Once we account for lagged outcomes, this figure drops considerably to -0.016. Additionally, the 95-percent confidence interval rules out any effects below -0.024, suggesting that while incumbents do adapt their policies following the introduction of non-partisan primaries, the magnitude of this change is likely inconsequential.

When considering both sets of results, it becomes evident that entrants exhibit a more pronounced moderating response to primary policy changes relative to incumbents. The observed shifts in Table 1 can be primarily attributed to more moderate entrants securing victories in elections rather than any substantial ideological adjustments by the sitting incumbents.

1.5.2 Relative Moderation and Electoral Competition. The previous section used measures of legislator ideology generated from roll call votes. While these provide a reliable estimate of policy behavior, analysis is limited to outcomes associated with elected representatives. To better understand the effects of primary policy on polarization and moderation, it is important to analyze the mechanisms driving these observed changes. In this section, I use ideal point estimates from the Database on Ideology, Money in Politics, and Elections (DIME) to measure legislator ideology for elections between 1980 and 2020. Instead of relying on roll call votes, DIME utilizes campaign contributions to estimate ideology, allowing analysis of both winning and losing candidates. This can provide insight into whether all candidates become more moderate or if the observed ideological changes are from winning candidates becoming more moderate.

I will start the analysis by examining how changes in primary policy affect moderation, using DIME estimates to measure legislator ideology. Table 4 compares results across different specifications that include all candidates (columns 1, 4, and 7), winning candidates only (columns 2, 5, and 8), or losing candidates only (columns 3, 6, and 9). All models control for district partisanship, incumbency status, and district and election year fixed effects. The table includes a row of average ideological scores for candidates who participated in closed primaries.

Across all three specifications, the average winning candidate is more moderate than the average losing candidate.

Starting with the baseline specification, estimates in Panel A show that less restrictive primary systems impose moderating effects on losing candidates but fail to affect elected representatives. In Column 1, I find that all candidates experience a -0.039 impact, resulting in a 10.13% decrease in the ideological distance from zero. However, it is clear from the estimates contained in Columns 2 and 3 that losing candidates are driving these results.

The disaggregated effects in Panel B can help inform these estimates. Switching to non-partisan primaries causes a -0.039 ideological shift in all candidates. This effect is mainly observed in winning candidates, with an estimated 95-percent confidence interval ranging from -0.069 to -0.015. Blanket primaries show a similar correlation, with winning candidates experiencing a -0.03 shift in CFscores. Conversely, there is inconclusive evidence of a relationship between blanket or non-partisan primaries and moderation in losing candidates. There are small but significant moderating effects for all candidates when treating blanket primaries as non-partisan. I find an estimated -0.117 shift in losing candidates' CFscores scores when adopting open primary policies. Conversely, the 0.044 shift observed in winning candidates indicates that switching from closed to open primaries may cause elected representatives to shift away from the median voter.

Columns 7 through 9 contain results when omitting Louisiana from the sample. The pooled results in Panel A show an increase in size and precision for the estimated impact on losing candidates, with a 95-percent confidence interval ranging from -0.082 to -0.058. In Panel B, the parameter associated with non-partisan primaries and losing candidate moderation is affected the most by excluding Louisiana, going from -0.037 ($p = 0.084$) to -0.054 ($p < 0.001$).

In a typical election, the candidate with the most extreme policy position held a 0.491 CFscore. There are indications that primary policy changes can affect extremism in winning and losing candidates. However, studying the impact on candidates with the most extreme policy positions is also important. According to Table 5, less restrictive primaries tend to make the most extreme candidates more moderate. Based on the pooled estimate in Panel A, -0.047, opening primary elections decreases the average extreme candidate's ideological score 0.446, less than the average losing candidate. Omitting Louisiana's observations strengthens the estimated effect to -0.054, with or without controls. Additionally, the disaggregated estimate associated with non-partisan primaries becomes statistically significant, with the 95-percent confidence interval ruling out effects greater than -0.014.

If primary reform moves the ideological stances of winning and losing candidates toward the center, we can anticipate a decrease in electoral competition over their policy positions. Table 6 displays the effect of different primary policies on the ideological distance between the winning and losing candidates in each election. To handle general elections with more than two candidates, I define the "losing candidate" as the one who received the second most votes.[16] There is strong evidence of a sizable decrease in the ideological distance between candidates following the adoption of less restrictive primary systems. Column 2 reports baseline estimates while controlling for lagged outcomes and district partisanship. The pooled effect in Panel A is estimated at -0.127, implying movement away from closed or semi-closed primaries causes a 16.91% decrease in the ideological

[16] Out of 5,980 elections, 125 had three candidates and 14 had four or more candidates.

distance between competing representatives. The 95% confidence interval excludes effects lower than -0.156 and higher than -0.098. When considering the policy-specific treatment effects in Panel B, non-partisan primaries have the largest impact on electoral competition, with an estimated effect of -0.149. There is evidence that both blanket and open primaries reduce the ideological distance between candidates, with blanket primaries having a slightly greater impact (-0.072) than open primaries (-0.064). Estimates contained in Columns 3 through 6 show these results hold when treating blanket primaries as non-partisan or excluding Louisiana from the model.

While evidence indicates a trend towards moderation in open primary processes, it is also worth considering how often the winning candidate is more moderate than their opponent. To examine if primary election reform affects the probability of the moderate candidate winning an election, I construct an indicator that assigns a value of one if the winning candidate has an ideological score closer to zero than the losing candidate. The results of this analysis are presented in Table 7, where estimates represent a percentage point shift in the likelihood of the more moderate candidate winning an election. From Column 2, there is a 9.7 percentage point decrease in the moderate candidate win rate. This effect is significant with a 95% confidence interval of -15.2 to -4.2 percentage points. In Panel B, I find that non-partisan primaries decrease the win rate for moderate candidates by 10.1 percentage points.

These results suggest a negative relationship between less restrictive primary systems and the win rates of moderate candidates. This might be explained by shifts in the ideological gaps between candidates. Theoretical models of electoral competition assume voters choose the candidate whose policy stance aligns most closely with their own (Downs, 1957). As demonstrated in Table 6, primary election reform narrows this ideological distance. Consequently, voters will become more indifferent between competing policy stances and place more emphasis on other differentiating factors such as candidate quality, experience, or personality traits when making a decision (Arnesen, Duell, & Johannesson, 2019; Bartels, 2002; Bishin, Stevens, & Wilson, 2006; Buttice & Stone, 2012). Therefore, reduced policy-driven competition might prompt voters to prioritize these alternative traits, indirectly influencing the success rate of more moderate candidates.

1.6 Political Participation

This section discusses the effects on voter turnout in both general and primary elections.

1.6.1 General Election Turnout. My analysis begins by examining participation in the general election. Policies that open primary elections could undermine party identity and cohesion, leading to voter apathy and decreased turnout. Non-partisan systems, like the Top-Two primary in Washington and California, may result in two candidates from the same party competing in the general election. Consequently, the two candidates' policy positions will be very similar, causing some voters to be indifferent, particularly if they are registered with the opposing party. Additionally, we showed in Section 1.5 that less restrictive primaries cause candidate policy positions to converge. As ideological points between opposing candidates become harder to distinguish, voter participation diminishes (Muñoz & Meguid, 2021).

The effects of primary policy on voter turnout in congressional elections are presented in Table 8. Panel A shows changes in turnout following the implementation of primary policies that do not require party affiliation and permit crossover voting. Panel B disaggregates the effects by policy, where closed primaries serve as the comparison group. Unlike previous tables in this paper, the average outcome of the reference group is not included because logarithm levels do not have a straightforward interpretation. Instead, a row displays the average number of votes cast in districts with closed primaries. As in Section 1.5, I present baseline results alongside three alternative specifications, with and without controls.

Results from the baseline model are reported in the first two columns. The estimates in Panel A show that opening primaries negatively affects participation in the general election. This effect is statistically significant and estimated to decrease turnout by 5.8%. Districts located in states with closed primaries received an average of 193,450 votes in their general elections. This translates to an estimated 11,220 voters would choose not to participate after expanding ballot access. When examining the effects of individual policies in Panel B, there is evidence that non-partisan primaries are affected the most, with a -12.1% impact on turnout. This effect is significant, with a 95-percent confidence interval ranging from -15.23% to -8.96%. Blanket and open primaries have similar impacts at -5.3% and -6.2%, respectively.

Columns 3 and 4 report estimates when treating blanket primaries as non-partisan. As with Section 1.5, Panel A estimates are not reported because they are the same as the baseline specification. Consistent with the baseline estimates, there is evidence of non-partisan primaries diminishing voter participation. However, the inclusion of blanket primaries attenuates the impact, with a smaller estimated effect of -5.8%. The similarities between these estimates and those from the baseline specification suggest that the classification of blanket primaries does not significantly influence the connection between general turnout and the implementation of non-partisan primaries. The difference in these estimates further underscores the rationale for treating the two policies as distinct from each other.

Columns 5 and 6 show results from an alternative specification where states that were fully covered by Section 5 of the Voting Rights Act are excluded from the analysis.[17] This is important because these states have unique historical and legal circumstances that can impact the accuracy of estimates. Unobservable factors that influence voter turnout in these states may differ from those in uncovered states. In other specifications, the set of controls includes an indicator for full coverage. The results from uncovered states offer compelling evidence for a relationship between less restrictive primary policies and voter turnout. Specifically, we find a significant decrease in voter turnout by approximately 6.4% when these policies are in place. This result holds after including controls, with a 95-percent confidence interval from -8.75% to -4.05%. The policy-specific parameter estimates support the baseline results.

Controlling for year fixed effects accounts for unobserved time-varying factors that affect all districts similarly but may not capture more complex temporal variations. For example, the introduction of the Internet has provided voters access to mass information and enabled the development of modern, far-reaching campaign tactics that directly influence party mobilization and voter turnout. To capture these dynamics, columns 7 and 8 report estimates when restricting the analysis to 1996 or later. Starting with Panel A, these results are quite different from those in the baseline specification. First, without controls, the pooled effect has a positive sign and is statistically insignificant. While the

[17]Nine states were fully covered by Section 5 of the Voting Rights Act: Alabama, Alaska, Arizona, Georgia, Louisiana, Mississippi, South Carolina, Texas, and Virginia. Note that only observations before the 2013 *Shelby County v. Holder* ruling are excluded.

inclusion of controls leads to a more precise estimate, the 95-percent confidence interval includes possible impacts as small as -0.036%, translating to roughly 82 votes lost following implementation. When examining the disaggregated effects, blanket primaries are driving the pooled estimates, causing an estimated 5.6% increase in turnout.

1.6.2 Primary Election Turnout. In the subsequent analysis, I will explore the impact of primary policies on voter turnout during primary elections. Some states enforce a closed policy restricting access to certain primaries to only those registered with a particular party. In these states, anyone can participate in a single primary, but the requirement to register with a specific party is a barrier to participation. The shift from closed to open primaries can eliminate these barriers, but voters are still required to select a single ballot. Alternatively, blanket and non-partisan primaries eliminate the need for voters to make such a decision, as all candidates are listed on the same ballot.

In this section's analysis, I exclude Louisiana from all specifications. This is because Louisiana is the only state where candidates can avoid the general election if they receive enough votes in the primary. As a result, Louisiana's primaries hold a greater significance compared to other states, so any variation in Louisiana primary election turnout following policy changes may impact the accuracy of estimates.

Table 9 displays estimates from the baseline analysis as well as three alternative specifications. The outcome of interest is the natural logarithm of total votes in each primary election. The baseline results are presented in Columns 1 and 2. Estimates from Panel A imply expanding ballot access has little to no impact on primary election turnout, a notion that is reinforced by the size of the errors for the disaggregated effects in Panel B. Transitioning from closed to non-partisan primaries increases voter turnout by 13.9% when blanket and top-two primaries are pooled. This effect is statistically significant, with a 95-percent confidence interval ranging from 4.1% to 23.7%. This heightened statistical power likely stems from the expanded sample size of districts with non-partisan primaries, suggesting that both blanket and top-two primaries influence turnout through shared characteristics that only become discernible with a more robust sample.

Columns 5 and 6 contain results when excluding states that were covered by Section 5 of the Voting Rights Act. With controls, expanding ballot access is estimated to increase primary election participation by 17.9%. Based on the 95-percent confidence interval, this effect could range from an increase of 2.02% to 33.78%. The estimated effects associated with non-partisan, blanket, and open primaries remain mostly unchanged from the baseline results in the first two columns. Semi-closed primaries are estimated to reduce turnout by 17.8%. In contrast to the baseline results, this effect is statistically significant. This difference likely causes the difference in power between the pooled estimates in Columns 2 and 6, as those pooled effects are being compared to both semi-closed and closed policies.

Finally, the last two columns contain estimates when elections prior to 1996 are omitted from the sample. In this period, the pooled effect is estimated at 33.1% with a p-value less than 0.001. The disaggregated effects in Panel B reveal that blanket primaries are driving these results, with a 95-percent confidence interval of 29.10% to 58.10%.

1.7 Conclusion

This paper shows how primary election policies affect legislator moderation and voter turnout. I find that less restrictive policies do not significantly impact legislator ideology on a broad scale. When looking at policy-specific impacts, non-partisan systems cause an estimated 15.64% decrease in DW-NOMINATE ideal points. Analyzing entrants and incumbents separately shows that incumbents who win re-election are likely to attenuate the effects observed in newly elected entrants. I find more support for moderating effects stemming from non-partisan systems when using alternative measures of legislator ideology. Further analysis of losing candidates suggests that these results come from all representatives shifting their policy positions toward the median voter rather than more moderate candidates winning over their more ideologically extreme competitors.

While there is evidence that expanded ballot access has moderating effects on legislator ideology, it likely comes at the expense of decreased voter turnout in general elections. The 95-percent confidence interval implies a pooling effect between -8.35% and -3.25%. The disaggregated estimates support the pooled results with non-partisan, blanket, and open systems displaying significant negative effects on general election turnout. While expanding the ideological distribution of the voter pool, less restrictive primary systems do not have a measurable impact on primary election participation.

Future research would benefit from data that includes more state-level variation in primary systems. Recent policy changes in Alaska and Maine present an opportunity to enrich analysis with more variation and recent data. Unfortunately, researchers will likely have to wait several years for the necessary data to produce post-treatment outcomes, with Alaska's first primary using the new system taking place in 2022 and Maine's first open primary occurring in 2024.

Political polarization and extremism have become substantial concerns throughout the country. If primary election reform can hinder this upward trend in polarization, the adoption of novel systems such as the Top-Two should be taken into serious consideration. However, it is important to further examine the possible effects on voter participation as a potential cost when looking at the net gain from non-partisan primary systems.

Table 1. Absolute Ideological Distance Results - All Representatives (1976 - 2020)

	Baseline		Blanket as NP		No Louisiana		Modern Era (1996-2020)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Pooling Open Policies								
Open/Blanket/Non-Partisan	−0.068	−0.028			−0.065	−0.025	−0.057***	−0.032***
	(0.034)	(0.014)			(0.035)	(0.016)	(0.007)	(0.004)
B. By Policy								
Non-Partisan	−0.128***	−0.056***	−0.102***	−0.044***	−0.127***	−0.054***	−0.066***	−0.034***
	(0.022)	(0.008)	(0.022)	(0.008)	(0.023)	(0.009)	(0.013)	(0.007)
Blanket	−0.046*	−0.019*			−0.046*	−0.018*	0.005	0.002
	(0.02)	(0.008)			(0.02)	(0.008)	(0.009)	(0.005)
Open	0.018	0.009	0.014	0.007	0.02	0.01	−0.011	−0.009
	(0.049)	(0.022)	(0.051)	(0.023)	(0.049)	(0.022)	(0.018)	(0.007)
Semi-Closed	−0.018	−0.01	−0.017	−0.009	−0.017	−0.009	0.015*	0.01*
	(0.025)	(0.009)	(0.024)	(0.008)	(0.025)	(0.009)	(0.01)	(0.004)
Reference Group Mean	0.358	0.358	0.358	0.358	0.357	0.357	0.379	0.379
District FEs	X	X	X	X	X	X	X	X
Election Year FEs	X	X	X	X	X	X	X	X
Controls		X		X		X		X
N	9977	9907	9977	9907	9812	9742	5663	5635

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. This table shows the estimated impact of primary election reform on legislator ideology, measured via dynamic DW-NOMINATE scores. Panel A contains pooled effects for all open-type policies. Panel B contains disaggregated effects for each policy. Column 1 reports the baseline estimates. Column 3 reports estimates when blanket primaries are re-coded as non-partisan. Column 5 reports estimates when excluding Louisiana elections from the sample. Column 7 reports estimates when excluding elections prior to 1996. Columns 2, 4, 6, and 8 show estimates when including controls for lagged outcomes and district characteristics. The reference group mean refers to the average ideological distance from zero for legislators in closed districts. Robust standard errors, clustered by state, in parentheses.

Table 2. Absolute Ideological Distance Results - Entrants (1976 - 2020)

	Baseline		Blanket as NP		No Louisiana		Modern Era (1996-2020)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Pooling Open Policies								
Open/Blanket/Non-Partisan	−0.082*	−0.088*			−0.078*	−0.085*	−0.093***	−0.105***
	(0.032)	(0.033)			(0.036)	(0.037)	(0.016)	(0.015)
B. By Policy								
Non-Partisan	−0.126***	−0.131***	−0.116***	−0.121***	−0.126***	−0.133***	−0.096***	−0.107***
	(0.021)	(0.024)	(0.02)	(0.022)	(0.024)	(0.027)	(0.026)	(0.021)
Blanket	−0.071**	−0.075**			−0.072**	−0.076**	−0.068*	−0.082**
	(0.022)	(0.026)			(0.022)	(0.026)	(0.031)	(0.028)
Open	0.013	0.005	0.011	0.003	0.014	0.006	−0.146**	−0.177***
	(0.061)	(0.063)	(0.062)	(0.064)	(0.061)	(0.063)	(0.045)	(0.033)
Semi-Closed	−0.031	−0.029	−0.03	−0.028	−0.031	−0.029	−0.002	−0.002
	(0.021)	(0.021)	(0.021)	(0.02)	(0.022)	(0.022)	(0.023)	(0.018)
Reference Group Mean	0.357	0.357	0.357	0.357	0.356	0.356	0.367	0.367
District FEs	X	X	X	X	X	X	X	X
Election Year FEs	X	X	X	X	X	X	X	X
Controls		X		X		X		X
N	1972	1903	1972	1903	1936	1867	1101	1073

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. This table shows the estimated impact of primary election reform on the ideological scores of newly elected entrants. Panel A contains pooled effects for all open-type policies. Panel B contains disaggregated effects for each policy. Column 1 reports the baseline estimates. Column 3 reports estimates when blanket primaries are re-coded as non-partisan. Column 5 reports estimates when excluding Louisiana elections from the sample. Column 7 reports estimates when excluding elections prior to 1996. Columns 2, 4, 6, and 8 show estimates when including controls for lagged outcomes and district characteristics. The reference group mean refers to the average ideological distance from zero for entrants in closed districts. Robust standard errors, clustered by state, in parentheses.

Table 3. Absolute Ideological Distance Results - Incumbents (1976 - 2020)

	Baseline		Blanket as NP		No Louisiana		Modern Era (1996-2020)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Pooling Open Policies								
Open/Blanket/Non-Partisan	−0.059	−0.007			−0.055	−0.005	−0.049***	−0.01***
	(0.035)	(0.005)			(0.037)	(0.006)	(0.008)	(0.002)
B. By Policy								
Non-Partisan	−0.125***	−0.016***	−0.094***	−0.011*	−0.121***	−0.014***	−0.061***	−0.011*
	(0.023)	(0.004)	(0.024)	(0.004)	(0.025)	(0.004)	(0.016)	(0.005)
Blanket	−0.039	−0.003			−0.038	−0.002	0.014	0
	(0.023)	(0.008)			(0.023)	(0.008)	(0.013)	(0.006)
Open	0.024	0	0.021	0	0.027	0.001	0.004	−0.01
	(0.05)	(0.011)	(0.052)	(0.011)	(0.051)	(0.011)	(0.031)	(0.006)
Semi-Closed	−0.016	−0.004	−0.015	−0.003	−0.015	−0.003	0.016	0.003
	(0.027)	(0.004)	(0.026)	(0.004)	(0.027)	(0.004)	(0.013)	(0.006)
Reference Group Mean	0.358	0.358	0.358	0.358	0.358	0.358	0.383	0.383
District FEs	X	X	X	X	X	X	X	X
Election Year FEs	X	X	X	X	X	X	X	X
Controls		X		X		X		X
N	8004	8004	8004	8004	7875	7875	4562	4562

* p < 0.05, ** p < 0.01, *** p < 0.001

Notes. This table shows the estimated impact of primary election reform on ideology for incumbents who win re-election. Panel A contains pooled effects for all open-type policies. Panel B contains disaggregated effects for each policy. Column 1 reports the baseline estimates. Column 3 reports estimates when blanket primaries are re-coded as non-partisan. Column 5 reports estimates when excluding Louisiana elections from the sample. Column 7 reports estimates when excluding elections prior to 1996. Columns 2, 4, 6, and 8 show estimates when including controls for lagged outcomes and district characteristics. Robust standard errors, clustered by state, in parentheses.

Table 4. DIME Ideological Distance Results (1980 - 2020)

	Baseline			Blanket as NP			No Louisiana			Modern Era (1996-2020)		
	All	Winner	Loser	All	Winner	Loser	All	Winner	Loser	All	Winner	Loser
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
A. Pooling Open Policies												
Open/Blanket/Non-Partisan	−0.039***	−0.019	−0.059***				−0.044***	−0.017	−0.07***	−0.054***	−0.015*	−0.088**
	(0.011)	(0.014)	(0.014)				(0.008)	(0.015)	(0.006)	(0.012)	(0.006)	(0.026)
B. By Policy												
Non-Partisan	−0.039**	−0.042**	−0.037	−0.037***	−0.038**	−0.035*	−0.046***	−0.039*	−0.054***	−0.051*	−0.038***	−0.06
	(0.014)	(0.014)	(0.021)	(0.01)	(0.012)	(0.015)	(0.011)	(0.015)	(0.012)	(0.019)	(0.01)	(0.041)
Blanket	−0.032*	−0.03**	−0.031				−0.035*	−0.028*	−0.038	−0.022	−0.003	−0.032
	(0.013)	(0.011)	(0.023)				(0.014)	(0.011)	(0.027)	(0.015)	(0.01)	(0.031)
Open	−0.005	0.044*	−0.055*	−0.006	0.043*	−0.056*	−0.008	0.045*	−0.063**	−0.032	0.02	−0.075
	(0.011)	(0.017)	(0.022)	(0.011)	(0.017)	(0.022)	(0.011)	(0.017)	(0.022)	(0.023)	(0.01)	(0.045)
Semi-Closed	0.016	−0.012	0.043	0.016	−0.011	0.043	0.014	−0.011	0.037	0.016	−0.016	0.047
	(0.009)	(0.016)	(0.024)	(0.009)	(0.015)	(0.024)	(0.009)	(0.016)	(0.024)	(0.016)	(0.01)	(0.033)
Reference Group Mean	0.385	0.320	0.449	0.385	0.320	0.449	0.385	0.320	0.449	0.427	0.356	0.495
District FEs	X	X	X	X	X	X	X	X	X	X	X	X
Election Year FEs	X	X	X	X	X	X	X	X	X	X	X	X
Controls	X	X	X	X	X	X	X	X	X	X	X	X
N	12 085	5986	6098	12 085	5986	6098	11 915	5920	5994	7610	3748	3861

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. This table shows the estimated impact of primary election reform on ideological moderation. To measure legislator ideology, CFscores are used to predict DW-NOMINATE ideal points. Panel A contains pooled effects for all open-type policies. Panel B contains disaggregated effects for each policy. Columns 1, 4, 7, and 10 report estimates for all candidates. Columns 2, 5, 8, and 11 report estimates for elected representatives. Columns 3, 6, 9, and 12 report estimates for losing candidates. The reference group mean refers to the average ideological distance from zero for candidates in closed districts. Robust standard errors, clustered by state, in parentheses.

Table 5. Most Extreme Candidate (1980 - 2020)

	Baseline		Blanket as NP		No Louisiana		Modern Era (1996-2020)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Pooling Open Policies								
Open/Blanket/Non-Partisan	−0.048***	−0.047***			−0.054***	−0.054***	−0.073***	−0.073***
	(0.009)	(0.011)			(0.006)	(0.006)	(0.018)	(0.018)
B. By Policy								
Non-Partisan	−0.024*	−0.024	−0.029**	−0.027*	−0.033**	−0.034***	−0.045	−0.043
	(0.012)	(0.014)	(0.01)	(0.011)	(0.01)	(0.01)	(0.026)	(0.027)
Blanket	−0.038**	−0.034**			−0.042**	−0.039**	−0.042	−0.041
	(0.012)	(0.01)			(0.014)	(0.011)	(0.023)	(0.023)
Open	−0.056*	−0.05	−0.055*	−0.05	−0.059*	−0.055	−0.075	−0.074
	(0.023)	(0.029)	(0.022)	(0.029)	(0.023)	(0.029)	(0.044)	(0.044)
Semi-Closed	0.032	0.033	0.032	0.033	0.03	0.03	0.037	0.038
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.029)	(0.029)
Reference Group Mean	0.491	0.491	0.491	0.491	0.491	0.491	0.550	0.550
District FEs	X	X	X	X	X	X	X	X
Election Year FEs	X	X	X	X	X	X	X	X
Controls		X		X		X		X
N	5980	5493	5980	5493	5915	5436	3752	3718

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. This table shows the estimated moderating effects of primary election reform on the most ideologically extreme candidate from each district. Panel A contains pooled effects for all open-type policies. Panel B contains disaggregated effects for each policy. Column 1 reports the baseline estimates. Column 3 reports estimates when blanket primaries are re-coded as non-partisan. Column 5 reports estimates when excluding Louisiana elections from the sample. Column 7 reports estimates when excluding elections prior to 1996. Columns 2, 4, 6, and 8 show estimates when including controls for lagged outcomes and district characteristics. Robust standard errors, clustered by state, in parentheses.

Table 6. Electoral Competition (1980 - 2020)

	Baseline		Blanket as NP		No Louisiana		Modern Era (1996-2020)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Pooling Open Policies								
Open/Blanket/Non-Partisan	−0.137***	−0.127***			−0.14***	−0.129***	−0.164***	−0.157***
	(0.018)	(0.015)			(0.017)	(0.013)	(0.013)	(0.013)
B. By Policy								
Non-Partisan	−0.169***	−0.149***	−0.141***	−0.124***	−0.176***	−0.155***	−0.193***	−0.178***
	(0.025)	(0.021)	(0.016)	(0.013)	(0.024)	(0.019)	(0.024)	(0.025)
Blanket	−0.082*	−0.072**			−0.085*	−0.075**	−0.056**	−0.057**
	(0.034)	(0.025)			(0.036)	(0.026)	(0.021)	(0.02)
Open	−0.057*	−0.064**	−0.064*	−0.071**	−0.06*	−0.068**	−0.097*	−0.095*
	(0.026)	(0.022)	(0.025)	(0.022)	(0.026)	(0.022)	(0.038)	(0.037)
Semi-Closed	0.015	0.019	0.017	0.021	0.013	0.016	0.014	0.02
	(0.018)	(0.016)	(0.018)	(0.016)	(0.018)	(0.016)	(0.029)	(0.028)
Reference Group Mean	0.751	0.751	0.751	0.751	0.751	0.751	0.834	0.834
District FEs	X	X	X	X	X	X	X	X
Election Year FEs	X	X	X	X	X	X	X	X
Controls		X		X		X		X
N	5980	5493	5980	5493	5915	5436	3752	3718

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. This table shows the impact of primary election reform on the ideological distance between winning and losing candidates. For elections with more than two competitors, the losing candidate is the one who receives the second most votes. Panel A contains pooled effects for all open-type policies. Panel B contains disaggregated effects for each policy. Column 1 reports the baseline estimates. Column 3 reports estimates when blanket primaries are re-coded as non-partisan. Column 5 reports estimates when excluding Louisiana elections from the sample. Column 7 reports estimates when excluding elections prior to 1996. Columns 2, 4, 6, and 8 show estimates when including controls for lagged outcomes and district characteristics. The reference group mean refers to the average ideological distance between candidates in closed districts. Robust standard errors, clustered by state, in parentheses.

Table 7. Moderate Candidate Win Rate (1980 - 2020)

	Baseline		Blanket as NP		No Louisiana		Modern Era (1996-2020)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Pooling Open Policies								
Open/Blanket/Non-Partisan	−0.092**	−0.097**			−0.115***	−0.118***	−0.111**	−0.107**
	(0.029)	(0.028)			(0.015)	(0.015)	(0.041)	(0.038)
B. By Policy								
Non-Partisan	−0.088	−0.101*	−0.079*	−0.091**	−0.124***	−0.135***	−0.113	−0.09
	(0.053)	(0.047)	(0.033)	(0.031)	(0.026)	(0.023)	(0.086)	(0.074)
Blanket	−0.061	−0.07			−0.079	−0.087	−0.028	−0.023
	(0.08)	(0.068)			(0.09)	(0.077)	(0.061)	(0.058)
Open	−0.083	−0.088	−0.085	−0.09	−0.1*	−0.104*	−0.085	−0.069
	(0.043)	(0.049)	(0.045)	(0.051)	(0.04)	(0.047)	(0.074)	(0.069)
Semi-Closed	0.025	0.012	0.026	0.013	0.013	0.001	0.029	0.047
	(0.029)	(0.03)	(0.029)	(0.031)	(0.026)	(0.028)	(0.053)	(0.049)
Reference Group Mean	0.725	0.725	0.725	0.725	0.728	0.728	0.714	0.714
District FEs	X	X	X	X	X	X	X	X
Election Year FEs	X	X	X	X	X	X	X	X
Controls		X		X		X		X
N	5980	5493	5980	5493	5915	5436	3752	3718

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. This table shows the impact of primary election reform on the win rate for moderate candidates. Panel A contains pooled effects for all open-type policies. Panel B contains disaggregated effects for each policy. Column 1 reports the baseline estimates. Column 3 reports estimates when blanket primaries are re-coded as non-partisan. Column 5 reports estimates when excluding Louisiana elections from the sample. Column 7 reports estimates when excluding elections prior to 1996. Columns 2, 4, 6, and 8 show estimates when including controls for lagged outcomes and district characteristics. Robust standard errors, clustered by state, in parentheses.

Table 8. General Election Voter Turnout (1976 - 2018)

	Baseline		Blanket as NP		Uncovered States		Modern Era (1996-2018)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Pooling Open Policies								
Open/Blanket/Non-Partisan	−0.099*	−0.058***			−0.130***	−0.064***	0.023	0.018*
	(0.039)	(0.013)			(0.016)	(0.012)	(0.016)	(0.009)
B. By Policy								
Non-Partisan	−0.203***	−0.121***	−0.168***	−0.058***	−0.208***	−0.110***	−0.031	−0.022
	(0.05)	(0.016)	(0.042)	(0.013)	(0.038)	(0.013)	(0.041)	(0.019)
Blanket	−0.106***	−0.053**			−0.108***	−0.045*	0.078*	0.056**
	(0.03)	(0.016)			(0.028)	(0.018)	(0.03)	(0.018)
Open	−0.084	−0.062**	−0.088	−0.064**	−0.103***	−0.050**	0.021	0.022
	(0.045)	(0.022)	(0.047)	(0.023)	(0.027)	(0.017)	(0.041)	(0.031)
Semi-Closed	−0.1	−0.058	−0.099	−0.057	−0.083	−0.044	−0.027	−0.019
	(0.063)	(0.031)	(0.063)	(0.031)	(0.072)	(0.032)	(0.029)	(0.021)
Reference Group Average Returns	187 736	187 736	187 736	187 736	187 664	187 664	221 750	221 750
District FEs	X	X	X	X	X	X	X	X
Election Year FEs	X	X	X	X	X	X	X	X
Controls		X		X		X		X
N	9325	8827	9325	8827	7828	7428	5118	5094

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes. This table shows the impact of primary election reform on voter turnout in general elections. The outcome is the natural log of district-level total votes. Column 1 reports the baseline estimates. Column 3 reports estimates when blanket primaries are re-coded as non-partisan. Column 5 reports estimates when excluding districts that were fully covered by Section 5 of the Voting Rights Act. Column 7 reports estimates when excluding elections prior to 1996. Louisiana is omitted from all specifications. Columns 2, 4, 6, and 8 report estimates when controlling for lagged outcomes and district characteristics. Districts that used closed primary elections comprise the reference group. Robust standard errors, clustered by state, in parentheses.

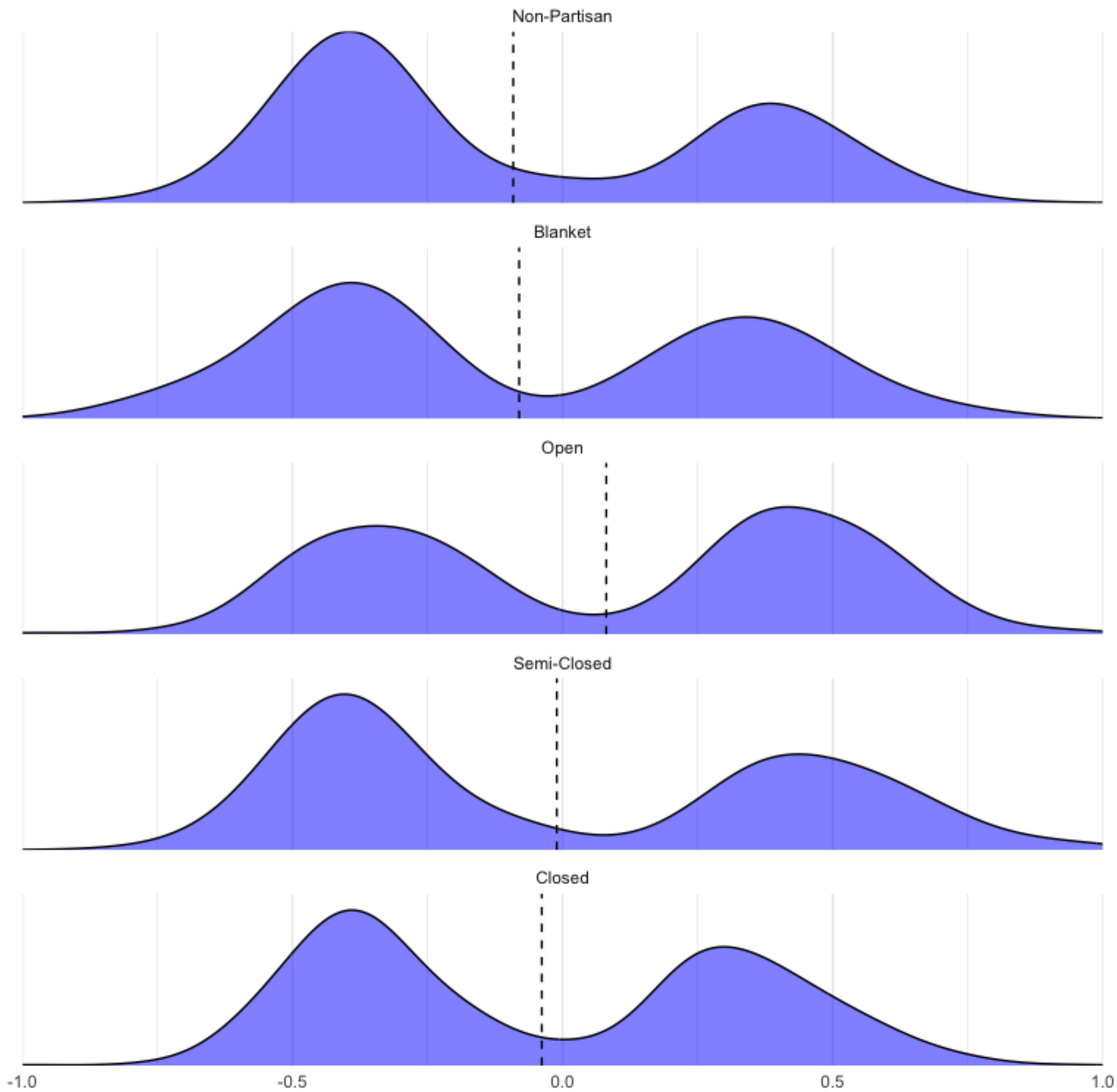
Table 9. Primary Election Voter Turnout (1976 - 2018)

	Baseline		Blanket as NP		Uncovered States		Modern Era (1996-2018)	
Model:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Pooling Open Policies								
Open/Blanket/Non-Partisan	0.126	0.148			0.164	0.179*	0.326***	0.331***
	(0.107)	(0.099)			(0.084)	(0.081)	(0.03)	(0.032)
B. By Policy								
Non-Partisan	0.086	0.095	0.143*	0.139**	0.069	0.077	0.139	0.143
	(0.059)	(0.055)	(0.057)	(0.05)	(0.069)	(0.061)	(0.083)	(0.085)
Blanket	0.235	0.243			0.235	0.243	0.426***	0.436***
	(0.154)	(0.139)			(0.143)	(0.134)	(0.072)	(0.074)
Open	−0.335	−0.292	−0.348	−0.312	−0.276	−0.254	0.147	0.149
	(0.266)	(0.255)	(0.266)	(0.248)	(0.299)	(0.281)	(0.117)	(0.117)
Semi-Closed	−0.106	−0.112	−0.1	−0.111	−0.177*	−0.178*	−0.122	−0.118
	(0.088)	(0.083)	(0.084)	(0.082)	(0.085)	(0.077)	(0.088)	(0.09)
Reference Group Average Returns	53 180	53 180	53 180	53 180	53 082	53 082	53 072	53 072
District FEs	X	X	X	X	X	X	X	X
Election Year FEs	X	X	X	X	X	X	X	X
Controls		X		X		X		X
N	6235	6049	6235	6049	5455	5320	3829	3802

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

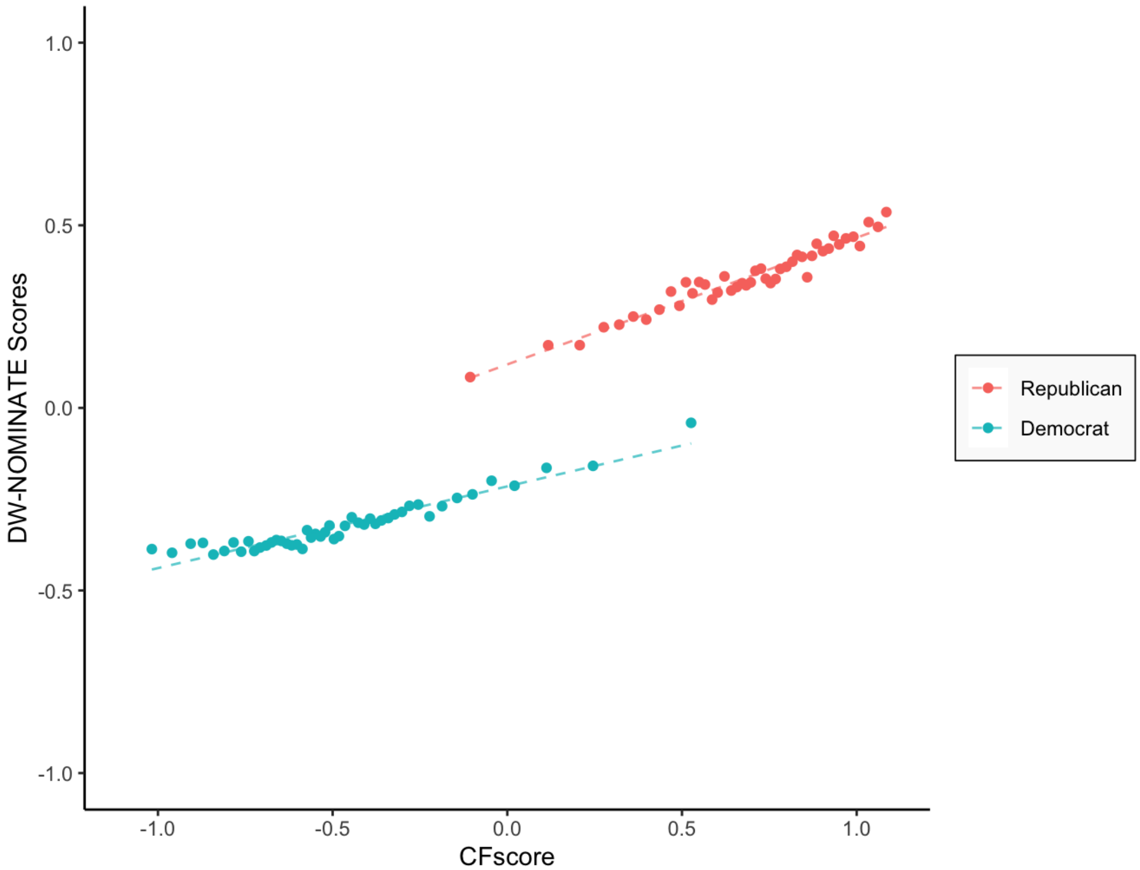
Notes. This table shows the impact of primary election reform on voter turnout in primary elections. Robust standard errors, clustered by state, in parentheses.

Figure 1. Distribution of DW-NOMINATE Ideal Points (1976 - 2020)



Notes. This figure displays the distribution of elected representatives' DW-NOMINATE ideal points grouped by primary election system.

Figure 2. CFscores and DW-NOMINATE Comparison



Notes. This figure presents binned scatterplots for Democrats and Republicans (bins = 50) showing the correlation between DW-NOMINATE ideal points and CFscores for all elected representatives in the sample.

CHAPTER II

DOES AI FACILITATE TRUST? AN EXPERIMENTAL STUDY WITH CHATGPT

2.1 Introduction

Trust is a cornerstone in various socioeconomic activities, including partnership formations and financial transactions, where it transcends mere contractual obligations. It is vital in relationships ranging from personal bonds, like those between spouses, to professional associations, such as the lawyer-client dynamic, procurement agencies and contracted firms, and collaborations between researchers and participants in scientific studies (Charness & Dufwenberg, 2006). Extensive research demonstrates its significance in several financial dealings: it influences stock market participation (Guiso, Sapienza, & Zingales, 2004, 2008), affects consumer credit (J. R. Brown, Cookson, & Heimer, 2019), determines the use of investment advisers (Gurun, Stoffman, & Yonker, 2017), and helps foster healthy lending relationships between lenders and borrowers in the credit markets (Fisman, Paravisini, & Vig, 2017; Hyndman, Wu, & Xiao, 2024).

Building trust relies on successful communication among involved parties. It is particularly dependent on the ability of the trustee to convince the trustor to place their trust in them. This becomes challenging when the communication is non-binding and the interactions are not regulated by formal agreements. Existing experimental studies have shown that a certain type of communication, namely promises, can promote trust. This can occur through two main channels: guilt aversion, where the trustee experiences guilt for not fulfilling the trustor's expectations (Charness & Dufwenberg, 2006), and promise-keeping, where the trustee feels remorse for not honoring their promises (Vanberg, 2008).

In the era of artificial intelligence, large language models (LLMs) like the Generative Pre-trained Transformer (GPT) have become integral to human communication.[1] These models assist in various tasks such as drafting emails, improving academic papers, financial reports, and other written materials.[2] This raises a pertinent question: what role can AI play in the trust-building process between a trustor and a trustee? Understanding this role is crucial for practical applications. For instance, in a bank loan application, a borrower might use AI to compose their application, while the lender might use AI for interpreting the application. The advantages of AI in this context are clear: it saves time for both parties, enhances the clarity and substance of the borrower's application, and helps the lender quickly understand the key points. However, AI also has potential drawbacks. It could diminish the authenticity of the borrower's application, leading the lender to question its veracity. Moreover, if the borrower heavily relies on AI, they might feel less committed to the content produced by the AI, posing a risk to the lender.

This study aims to explore the impact of AI on trust-building through communication in a controlled experimental setup. We use a two-player binary choice sequential trust game, as described in Charness and Dufwenberg (2006), where Player A (the trustor) makes the first move, followed by Player B (the trustee). Before the game begins, Player B is allowed to send a free-form message

[1] GPT is created by OpenAI, significantly influencing the field of natural language processing (OpenAI, 2022, 2023b). T. Brown et al. (2020) show that ChatGPT can produce texts with such remarkable accuracy and fluency that it closely resembles human writing, making it challenging for human evaluators to differentiate between text generated by GPT and that authored by humans.

[2] In addition, LLMs have shown remarkable capabilities across diverse areas. They are capable of creating computer code, as demonstrated by M. Chen et al. (2021), and solving university-level mathematics problems (Drori et al., 2022).

to Player A. The experiment is structured as a 2x2 design. In one aspect, we either provide or withhold AI as a tool for Player B to aid in composing their message. In the other aspect, we either provide or do not provide Player A with AI to assist in interpreting Player B's message.[3] The comparison across different treatment groups enables us to understand the influence of the AI assistant's presence (for either Player A or Player B) on the players' decisions and beliefs. Additionally, analyzing within each treatment group sheds light on how players leverage AI to support their communication and decision-making processes.

We find that the presence of AI does not significantly impact the individual choices of the two players. This result may be attributed to several factors. When trustees receive AI assistance, they are more likely to send a "promise" message to the trustor but are less likely to honor a "promise" if it is suggested by AI. When trustors receive AI assistance, they closely follow the suggestions from their AI assistants, who consistently remind them to choose cautiously.

However, we observe a significant increase in the frequency of cooperation between the trustors and the trustees when trustees have access to AI. This outcome can be attributed to trustors becoming more vigilant, realizing that messages from trustees might be partially or fully generated by AI. This heightened scrutiny by trustors, combined with trustees' awareness of being closely examined, may create a mutual understanding and alignment of expectations, thereby fostering an environment where trust can thrive.

[3]We employ the GPT-3.5-turbo model (OpenAI, 2023a) in the experiment. We carefully design prompts to ensure AI fully comprehends the trust game and understands its role as an assistant for a specific player in the game during the communication phase. See OpenAI (2023c) for guidance on prompt design.

The paper is organized as follows. Section 2 provides the experimental design and proposes the main hypotheses. Section 3 analyzes the experimental results and test the hypotheses. Section 4 concludes.

Literature Review. With its remarkable ability to comprehend and produce language akin to humans, social scientists are developing a growing interest in examining machine-learned large language models. Utilizing approaches common to economic and psychological research, such as surveys and laboratory-style experiments, has proven useful in analyzing whether AI mirrors human behavior in individual decision-making tasks as well as in strategic interactions. See, for example, Aher, Arriaga, and Kalai (2023), Argyle et al. (2022), Bybee (2023), Brand, Israeli, and Ngwe (2023), Brookins and DeBacker (2023), Y. Chen, Liu, Shan, and Zhong (2023), Fan, Chen, Jin, and He (2023), Guo (2023), Hagendorff (2023), Horton (2023), Kosinski (2023), Lorè and Heydari (2023), Ma, Zhang, and Saunders (2023), Phelps and Russell (2023), Engel, Grossmann, and Ockenfels (2024), Leng and Yuan (2023), among many others. In these studies, AI serves as the primary subject of investigation, as opposed to humans. A separate strand of research focuses on experiments involving human interaction with machines or AIs, specifically to ascertain whether human responses differ as opposed to interaction with other humans, whether AI players outperform their human counterparts, and whether AI and humans would generate any principal-agent type conflict (Bauer, Liebich, Hinz, & Kosfeld, 2023; Cohn, Gesche, & Maréchal, 2022; de Mello, Marsella, & Gratch, 2016; Dvorak, Stumpf, Fehrler, & Fischbacher, 2024; LaMothe & Bobek, 2020; Laudenbach & Siegel, 2024; Phelps & Russell, 2023; Schniter, 2024).

Contrasting with the above-mentioned literature, our paper still focuses on human interactions where AI assumes the role of an assistant. Several other papers belong to the same category as ours. For example, in Harris, Immorlica, Lucier, and Slivkins (2023), the sender can acquire information about the receiver from an AI oracle in a Bayesian persuasion context. Bai, Gui, Wei, and Xue (2023) considers whether the first mover would take advice from an AI in a two-player centipede game. Serra-Garcia and Gneezy (2023) find that algorithmic tools help people detect deception in a classic TV game show. To the best of our knowledge, our study is the first to explore the impact of AI in a trust game with communication played by human players.

Our research adds to the discussion on algorithm aversion and trust in AI (Glikson & Woolley, 2020). The low number of subjects directly adopting the AI-suggested messages in our experiment confirms the common belief that people often mistrust algorithmic advice, even when it's advantageous to follow it (Dietvorst, Simmons, & Massey, 2015). However, it's important to note that our observations are influenced by two key factors. 1) People generally tend to place greater trust in their own judgment compared to that of others. Our experimental design does not allow us to determine whether the low adoption of AI-recommended messages is due to aversion to AI or excessive confidence in one's own judgment.[4] 2) As pointed out by Logg, Minson, and Moore (2019), individuals tend to be more receptive to algorithmic advice in areas where there is a clear and measurable external standard of accuracy, such as making investment decisions or predicting sports outcomes. In contrast, AI's suggestions for interpersonal communication

[4]Our post-experiment survey elicited overall trust in AI from the subjects. However, we did not find evidence that subjects who chose not to adopt AI-recommended messages were more averse to AI compared to other subjects.

are less easily quantifiable, making it reasonable to assume that participants in our experiment rely more heavily on their own judgment in such cases.

Finally, an increasing number of economic studies have focused on understanding the impacts of machine learning and artificial intelligence on socioeconomic phenomena, covering diverse areas like labor force participation, wage disparity, market competition, consumer privacy, economic growth, and political engagement. See for example, Acemoglu (2022). Our paper contributes to this literature by examining AI's applicability in partnership formation and financial transactions.

2.2 The Experiment

2.2.1 Experimental Design. The objective of this study is to explore the potential of AI assistants in fostering trust dynamics between human participants. To accomplish this goal, we conducted an online experiment based on the classic two-player trust game introduced by Charness and Dufwenberg (2006), with a modification allowing the second player to communicate with the first player via a pre-game message. This experimental setup enables us to investigate whether the presence of AI assistance influences the trust-building process, via the pre-game message, between participants.

In certain treatments of our experiment, Player B, the trustee, is presented with an AI assistant interface before the commencement of the game. The AI interface allows Player B to compose and send a message to Player A, the trustor, prior to the initiation of gameplay. It is important to note that any message sent by Player B occurs before the actual gameplay begins. A depiction of player B's interface, including the game tree, can be seen in Figure 3. Outcomes are shown in the order of (π_A, π_B).

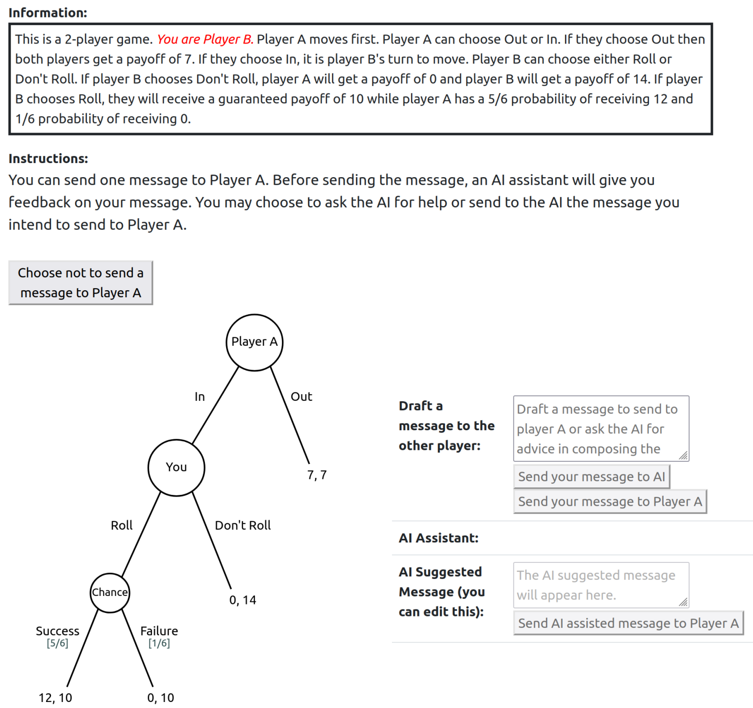


Figure 3. Screenshot of player B's screen when they are able to use AI. Player A sees a similar tree.

To maximize data collection and ensure simultaneous decision-making by both players, we employed the strategy method (Selten, 1967). Following each player's decision, Player A is prompted to indicate their beliefs regarding Player B's choice, while Player B is asked to provide their own beliefs about Player A's perceived decision. Subsequently, both participants are presented with a Holt-Laury quiz (Holt & Laury, 2002) to assess their risk preferences. Additionally, a demographic survey was administered to gather information on participants' perceptions and prior experience with AI technology.

2.2.1.1 Treatment Design. To investigate the impact of AI assistance on players' payoffs in the trust game, we implemented a 2×2 treatment design, consisting of four distinct treatments. Each treatment explored various combinations of AI presence and absence, shedding light on the role of AI in participants' decision-making processes and outcomes.

- **Benchmark:** Neither player has AI. Player B can choose to send a single message to Player A or refrain from doing so.

- **OnlyA:** Player A has AI to interpret Player B's (potential) message and receive advice on subsequent actions.

- **OnlyB:** Player B has AI to assist in crafting a message to Player A. If Player B opts to send a message, they must first interact with the AI.

- **Both:** Both players have access to AI assistance, following the functionalities described above.

In all treatments, Player B's interaction with AI, if applicable, precedes any communication with Player A. Participants have the option to engage in dialogue with the AI before deciding whether to send a message to Player A. It is crucial to note that all participants are aware of the presence and function of AI throughout the experiment.

2.2.1.2 Prompt Design. Creating an AI assistant using a natural language processor like ChatGPT involves crafting a prompt that guides the assistant in generating responses. However, designing a prompt that effectively handles a wide range of inputs is more of an art than a science. Our experimentation with ChatGPT (specifically gpt-3.5-turbo) revealed certain challenges and considerations that influenced the development of robust prompts for our study.

We observed that ChatGPT exhibited difficulties in handling sequential logic and tended to perform more reliably with shorter prompts. Moreover, we noted instances where the assistant suggested creative solutions, such as proposing the signing of a contract to establish a binding promise, which were beyond the scope of our experimental setting.

In light of these observations, we refined our prompts by incorporating the following principles:

1. We presented the trust game in its normal form, as the sequential form did not strategically differ from its normal counterpart given the strategy method employed in our experiment.

2. We omitted the description of probabilistic outcomes resulting from (In, Roll) in the game for Player B's AI only and instead provided ChatGPT with the expected outcomes.[5]

3. We explicitly instructed ChatGPT not to propose side deals or disclose players' personal information.

4. We refrained from including higher-order beliefs for the AI assistants. Specifically, Player B's AI did not possess knowledge of the existence of Player A's AI.

The full prompts used in our study can be found in Appendix A.3.

2.2.2 Experimental Procedure. A total of 240 subjects, 30 pairs per treatment, were recruited from Prolific and the University of Oregon student population to participate in this experiment. On average, subjects spent 15 minutes in the experiment and were paid a $5 show up fee and earned an additional $8.36

[5]We still keep the probabilistic outcomes resulting from (In Roll) in the game for Player A's AI because we want the AI is able to remind Player A to take risks into consideration.

during the experiment. The experiment was programmed using oTree (D. L. Chen, Schonger, & Wickens, 2016) and ChatGPT version gpt-3.5-turbo.

During the online experiment, subjects were continuously recruited and dynamically assigned roles and partners to play the game with to minimize subject wait time. Subjects read instructions then were assigned their role. They then took a quiz to ensure they understood the payoff structure before getting paired with a partner and playing the game. An example of the experimental instruction can be found in Appendix A.4.

2.2.3 Hypotheses. In this section, we lay out the main hypotheses that we tested in the experiment.

Hypothesis 1: *When player B has access to an AI, player A will play 'In' more frequently.*

We hypothesize that using an AI will result in player B sending a message that is more likely to elicit trust from player A.

Hypothesis 2: *When player A has an AI assistant, they will play 'In' less frequently.*

We hypothesize that an AI assistant for player A will result in more conservative decisions from player A as the AI may call attention to the risk of playing 'In' and player B playing their dominant strategy: 'Don't Roll'.

Hypothesis 3: *When player B has an AI assistant, they will be more likely to promise to choose 'Roll'.*

Player B's AI assistant is instructed to help it's user maximize it's payoff, which means the AI should be tying to help player B convince player A to play 'In'. In instances where Player B didn't initially make a promise, their AI might suggest it's user making a promise.

Hypothesis 4: *When player B's message to player A contains a promise which originated from the AI player B is less likely to honor the promise.*

Player B may feel a decrease in the cost of breaking a promise if the promise came from the AI. Consequently, they may be less likely to honor a message that came form the AI.

Hypothesis 5: *The presence of AI, both for player A and for player B, will increase the probability of achieving a cooperative outcome (In, Roll).*

(In, Roll) leads to the highest total payoff for the two players. We expect that the AI should help communicate (when player B has AI) and/or derive (when player A has AI) player B's intentions, so that it helps coordinate cooperative trustors and cooperative trustees.

2.3 Results

2.3.1 Treatment Effects. To investigate the influence of AI assistance on outcomes, Figure 4 presents the decisions made by each player when (i) neither player receives AI assistance (baseline), (ii) player A receives AI assistance but not player B, (iii) player B receives AI assistance but not player A, and (iv) both players receive AI assistance.

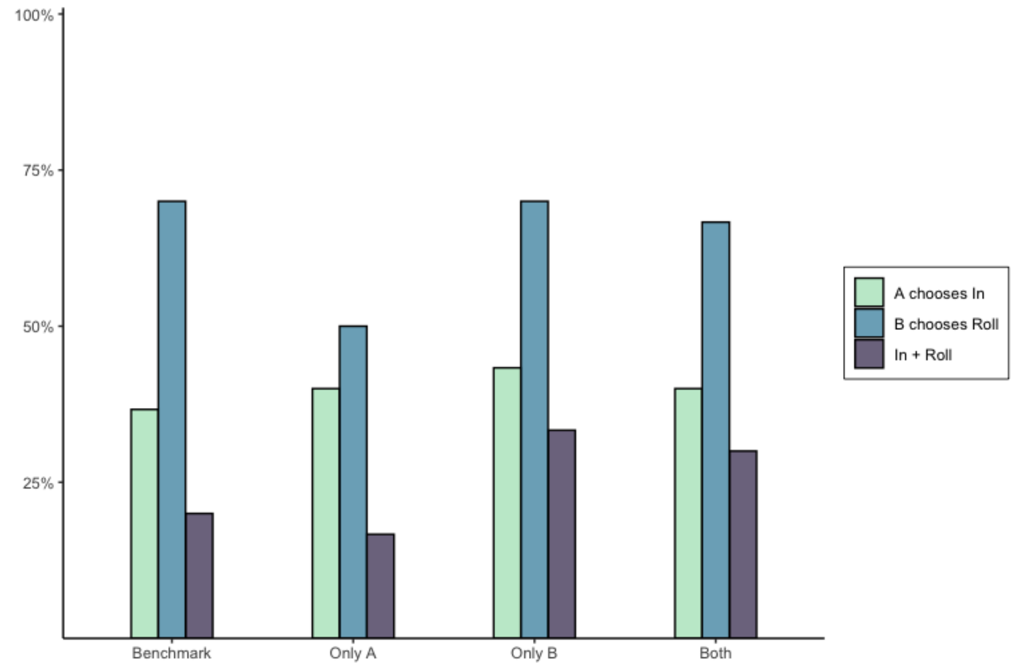


Figure 4. The percentage of observations in each treatment where player A chooses 'In', player B chooses 'Roll', or both.

In the benchmark treatment, player A chooses 'In' 36.7% of the time. This rate increases to 43.3% once player B has the option to use AI. Conversely, in both treatments where player A receives input from ChatGPT, 40% of trustors choose 'In' whether or not player B has AI. These results suggest the presence of AI has a negligible impact on player A's choices.

We test **Hypothesis 1** by pooling treatments based on player B's access to AI.[6] We observe that player A chooses 'In' more frequently when player B has AI (41.67%) than without AI (38.33%). However, we fail to reject the null ($p = 0.606$) when testing the difference in proportions.

Continuing with our analysis of player A's choices, **Hypothesis 2** predicts that player A will be more conservative when receiving support from AI. Again, we pool

[6] $X^{\text{No AI}} = \{\text{Baseline}, \text{Only A}\}$ and $X^{\text{AI}} = \{\text{Only B}, \text{Both}\}$.

treatments based on access to AI and find that player A chooses 'In' 40% of the time with and without AI support. Thus, we fail to reject the null.

When neither player has AI, player B chooses 'Roll' 70% of the time. This result remains mostly unchanged in treatments where player B receives AI assistance. Sessions where only player A receives AI assistance stand out with a roll rate of 50% compared to 68.9% across the other three treatments ($p = 0.079$). With access to AI being public information within each group, this suggests the knowledge that only player A will receive guidance from an AI assistant may influence player B's decision.

We refer to the strategy profile ('In', 'Roll') as the *cooperative outcome* as it constitutes the greatest expected collective payoff for each pair. In the benchmark treatment, we observe the cooperative outcome in 20% of pairs. In sessions where only player A receives AI assistance, the proportion of cooperative outcomes decreases to 16.7%, suggesting minimal impact. Conversely, there is convincing evidence that player B's access to AI improves cooperation. In treatments where player B has the option to use AI – Only B and Both – we observe the cooperative outcome in 33.3% and 30% of pairs, respectively. Pooling observations by player B's access to AI, we find a 13.34 percentage point increase in cooperative outcomes when player B receives AI assistance ($p = 0.093$). Taken together, the disconnect between individual choices and pair-wise choices across treatments indicates AI assistance may not significantly impact individual decisions; rather, it helps to coordinate cooperative trustors with cooperative trustees. This provides support for **Hypothesis 5**.

We now turn to the impact of AI assistance on beliefs. First-order beliefs (τ_A) represent player A's confidence that player B will choose 'Roll'. Second-order beliefs (τ_{BA}) reflect player B's perception of player A's beliefs. To provide a more intuitive interpretation of results where higher values correspond with increased trust, we map qualitative responses from the post-experiment survey to numerical values according to Table 10.

Certainly Choose Don't Roll	$\longrightarrow$	0
Probably Choose Don't Roll	$\longrightarrow$	0.25
Unsure	$\longrightarrow$	0.5
Probably Choose Roll	$\longrightarrow$	0.75
Certainly Choose Roll	$\longrightarrow$	1

Table 10. Numerical values assigned to elicited beliefs.

Consistent with existing literature, our findings confirm that beliefs and behavior are closely interconnected. Specifically, we observe that A is more inclined to choose 'In' when they are confident B will select Roll. As selecting 'In' report an average τ_A of 0.74, translating to a belief that B will "probably choose Roll." On the other hand, those choosing 'Out' exhibit reduced trust in B, with a lower average of 0.44. Moreover, Bs who decide to 'Roll' have an average second-order belief of 0.73, while those who choose 'Don't Roll' show a significantly lower average τ_{AB} at 0.37.

Figure 5 presents average first- and second-order beliefs across the four treatments. In the benchmark treatment, trustors display an average τ_A of 0.52, indicating a neutral level of trust. The inclusion of AI assistance for B increases this average to 0.61. However, this pattern does not persist across the Only A and Both treatments. When pooling observations by B's access to AI, we find an average τ_A of 0.529 in treatments without access, only slightly less[7] than when B

[7] a difference of 0.07

is assisted by AI. In line with the patterns observed in choices, there is minimal evidence to suggest AI significantly influences A's beliefs.

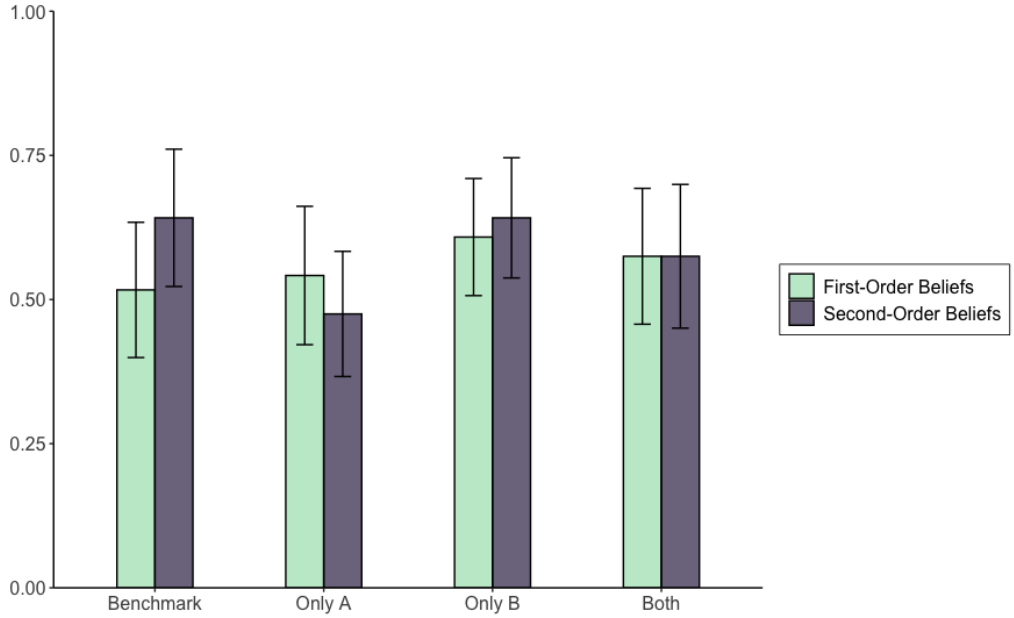


Figure 5. Average first-order and second-order beliefs across the four treatments. Confidence bands are calculated at the 95-percent level.

Second-order beliefs appear to be equally inconsistent across treatments, except in cases where only A receives AI assistance. In such cases, B's average τ_{BA} is 0.475, markedly lower than the 0.629 observed across other treatments ($p = 0.03$). This divergence might be attributed to the public information that only A can access AI, leading B to form a more pessimistic perception of A's beliefs. It is also consistent with the low Roll rate in the only A treatment compared with the others.

2.3.2 Message Classification.

2.3.2.1 Player B. We classify messages sent by player B according to the *type* of message sent and the *method* by which it was sent. The type of message that player B sends takes one of the following values: 'promise', 'asking', 'empty', 'skip', 'fairness', 'anti-promise'. Meanwhile, we classify the *method* into one of 'Own', 'AI', 'Mixed', 'Skip'. Each of these classifications is briefly discussed below, with further details in the appendix.

Message Type

Message type primarily concerns the *content* of the message. We abstract the message sent by player B into what player B indicates about *player A's* potential move and what player B expresses about their *own* intended move. For each of these components, we assign a tertiary label. For the first component – the piece of the message involving player A's move – we assign a label from $\varnothing/In/Out$, with $\varnothing$ representing no definitive information conveyed. Similarly, for the second component – on player B's own move – we assign a label from $\varnothing$/*Roll*/*Don't Roll*. This categorizes any message sent by either player B or the AI into one of 9 pairs. This codification allows us to assign labels to the messages according to Table 11. Additionally, a visualization of the transformation from player B's initial message through the AI to their final message can be found in Appendix Figure A.1, with additional examples following in panels (a)-(e) of Figure A.2. Note that this assignment is for explicit messages; when player B opts not to send a message, they are assigned the message type "skip". Further details of message-code assignment are left to the appendix.

Msg Vec	**Label**
(*In*, *Roll*)	*Promise*
($\varnothing$, *Roll*)	
(*Out*, *Roll*)*	
(*In*, $\varnothing$)	*Asking*
($\varnothing$, $\varnothing$)	*Empty*
(*Out*, $\varnothing$)	*Fairness*
(*In*, *Don't Roll*)	*Anti-promise*
($\varnothing$, *Don't Roll*)	
(*Out*, *Don't Roll*)	

Table 11. Each message sent by player B (or recommended by AI) is encoded as vector which captures what player B intends to do and what they propose player A should do. Note that (Out, Roll) does not constitute a cooperative outcome unlike the other Promises. We nonetheless include it in Promise since it demonstrates Player B's intention to play Roll. We do not observe any encoded message of (Out, Roll).

We are chiefly interested in the effects of player B promising to play *Roll* has on choices, outcomes, and beliefs. We abstract slightly from the notion of a "promise" to any explicitly expressed intention to play *Roll* on behalf of player B. On the other hand, if player B explicitly expresses intent to play *Don't Roll*, we label this an "anti-promise", regardless of their suggestion as to how player A should play. If player B only indicates an explicit move that they wish their opponent to play, and does not explicitly provide information about what they intend to play, then we classify their message as either "Fairness" or "Asking". "Asking" is chosen if player B requests that player A play *In*, without mention of their own intended move. Conversely, "Fairness" indicates that player B has suggested that player A play *Out*, resulting in an egalitarian (”fair”) outcome. In the event that no clear intentions are sent on behalf of player B, then the ”Empty” label is assigned.[8]

A breakdown of message types for non-skipped messages across all treatments is provided in Figure 6. The largest share (41%) of sent messages are Promises, with 90% of sent messages being comprised of Promises, Asking, and Empty messages.

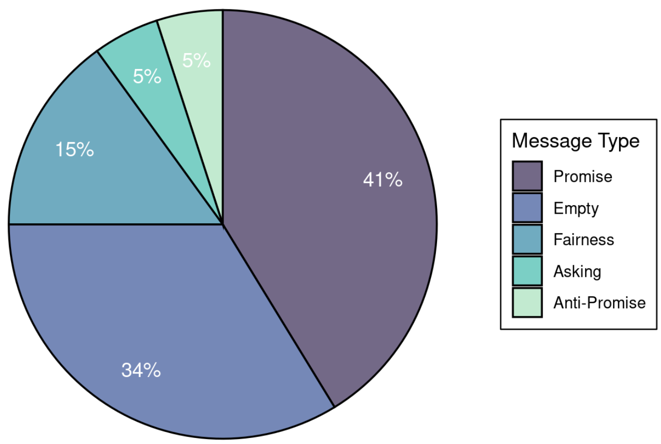


Figure 6. Types of messages sent across all treatments, omitting skipped messages. Note that $\approx 1/3$ of B’s in the sample opted to skip sending a message.

Figure 7 illustrates the distribution of player choices based on player B’s message type. Messages that assure player B will choose ‘Roll’ or ask player A to choose ‘In’ result in the highest frequency of A choosing ‘In’. We expect messages classified as ‘empty’ to elicit similar responses from A as cases where B opts out of sending any message, as both situations lack any substantive signal of B’s intentions or trustworthiness. However, our data show trustors choose ‘In’ more frequently when

[8]Note that no truly ”empty” messages are sent: all messages sent contain *some* content. Therefore, it may be helpful to think of the ”Empty” label as ”Junk” based on the complement set of messages already described.

receiving an empty message (37.03%) than no message (25%). Several of the empty messages are disconnected from the experiment itself but contain relatively positive language.[9] It may be the case that sending a positive message can help to establish trust even if the message is irrelevant to the game.

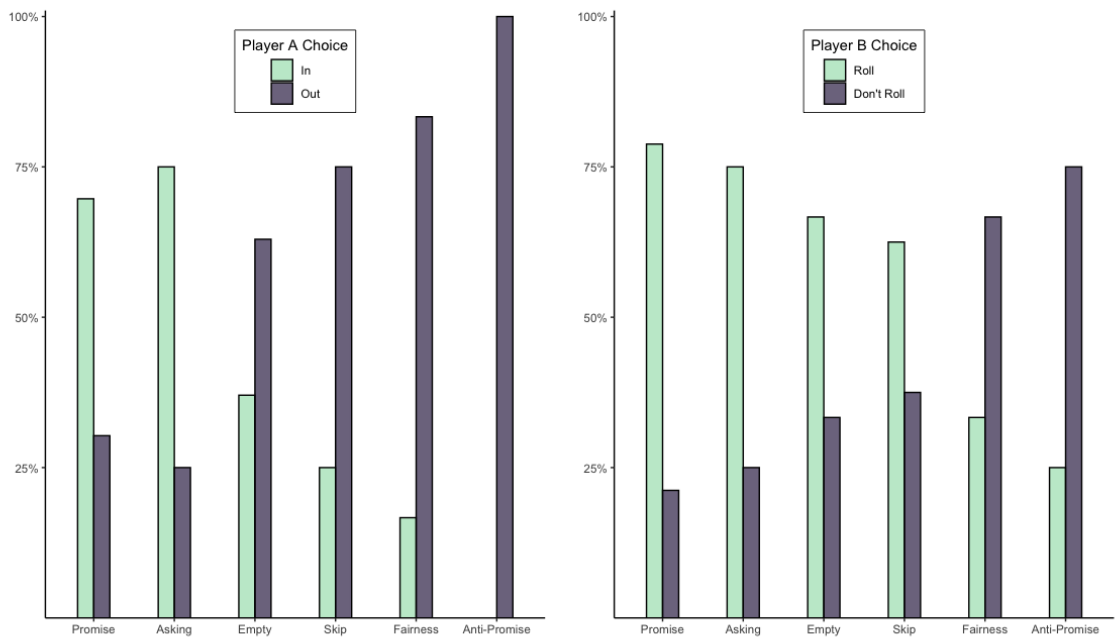


Figure 7. Distribution of player choices by message classification.

Messages categorized as 'fairness' and 'anti-promises' share similar purposes but differ in the signals they convey. Fair messages explicitly encourage A to choose 'Out' by presenting it as the safest option, while anti-promises discourage choosing 'In' by disclosing B's intent to choose 'Don't Roll'. As anticipated, both types of messages elicit the lowest in-rates across all pairs. Notably, no trustors chose 'In' after receiving an anti-promise, underscoring the strong deterrent effect of such messages.

[9]For example, one message contained "Hey, I hope we have a good game (:"

Hypothesis 3 postulates that Player B will send more promises when they have access to an AI assistant. Our data show a 75% increase in the proportion of messages containing promises when B gains access to AI ($p = 0.067$). Expanding the notion of a promise to include 'asking' increases statistical significance to the 97-percent level.[10] Specifically, our 95-percent confidence interval indicates the presence of AI for Player B yields a 1.83 to 34.8 percentage point increase in messages categorized as 'promise' or 'asking'. Figure 8 displays these findings visually.

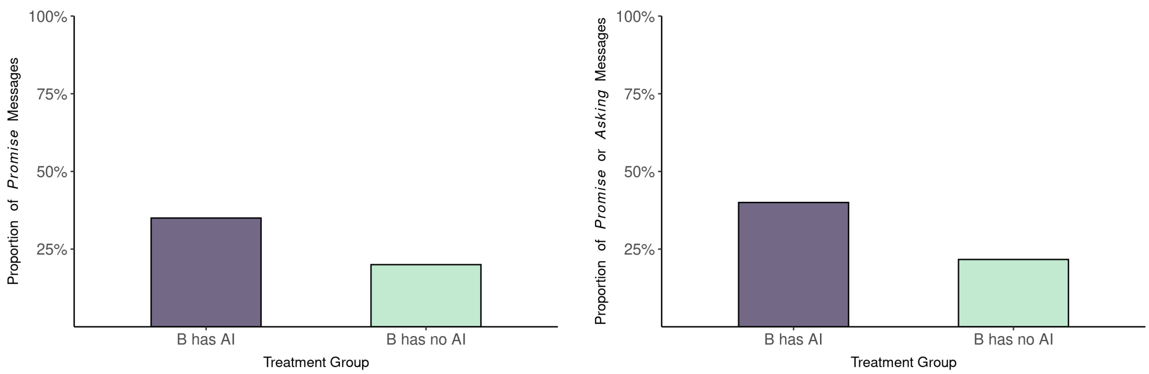


Figure 8. Left panel: proportion of messages sent by player B which are promises. Right panel: proportion of messages sent by player B which are promises or asking.

Message Method

As opposed to message type, which concerns the content of the message, message method concerns the *authorship* of the message. Since player B may edit the message suggested by the AI before sending it to player A, we aim to discern whether the content of the message was primarily dictated by AI, by the agent, or by a reasonable mix of the two. We use a normalized version of the Levenshtein edit distance (Levenshtein et al., 1966) to determine the pairwise relative distances

[10] We consider 'asking' alongside 'promise' as to include any message which directly suggests that player A play 'In', opening up the possibility for a cooperative outcome.

between player B's first message, the AI's suggested message, and the actual sent messages[11]. When plotting the normed Levenshtein distance between the the first and sent messages against the AI-suggested and sent messages (Figure A.3 in the appendix), a clear grouping structure can be seen. To verify this grouping structure, we implement a $k-$means cluster classification with $k = 3$ means[12] to produce the labeling assignment. Each member of the research team independently inspected each message to ensure accurate labels. The upshot is that sent messages which are labelled 'Own' have near-identical similarity to the first messages player B sent compared to the AI's suggested message; sent messages labelled 'AI' have near-identical similarity to the AI's suggested message compared to the first message player B sent, and 'Mixed' messages player B sent are those which bare a fair similarity to both the first and the AI-suggested message. Further details can be found in the appendix.

Figure 9 shows the breakdown of how B players sent their messages, provided that they sent one at all. The majority of sent messages are primarily their own compositions, with near equal shares of messages being crafted entirely by the AI or a mix of AI and player B. It should be noted that both Figure 9 and Table 12 are restricted to treatments when player B has an AI assistant.

[11]The Levenshtein distance is a metric which reports the total number of single-character edits needed to transform one string into another. In particular, the distance measures the number of insertions, deletions, and substitutions required to transform one of its inputs into the other. We implement a normed version of this metric, which scales the traditional Levenshtein distance between two strings by the length of the larger string. This transforms the metric into a measure of similarity between the two strings lying between 0 and 1, as the maximum length of the two input strings is exactly the maxmium number of single-character edits needed to transform one string into the other.

[12]See Appendix A.2.1 for details.

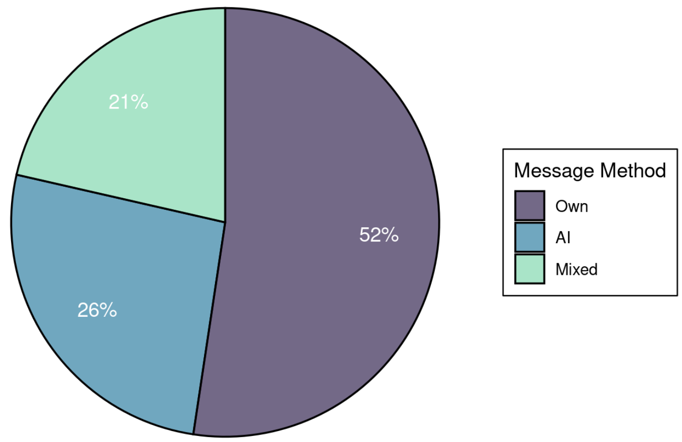


Figure 9. Over half of player B's who sent a message did so using mostly their own authorship. Details of how these methods were imputed are included in the appendix.

Figure 10 shows the distribution of player choices by method across the four treatment groups. Without AI assistance, Player B can either opt not to send a message or compose their own. Consistent with earlier results, As select 'In' at a relatively low rate of 25% in the absence of any message. When the treatments allow Player B to send messages that are either partially or completely generated by AI ('Only B' and 'Both'), these AI-assisted messages result in unexpectedly lower 'In' rates: 44.4% for 'mixed' messages and 36.4% for fully AI-authored messages, in contrast to 54.5% for original messages. These findings indicate that the authorship of the message may not influence Player A's decisions as much as the actual content of the message.

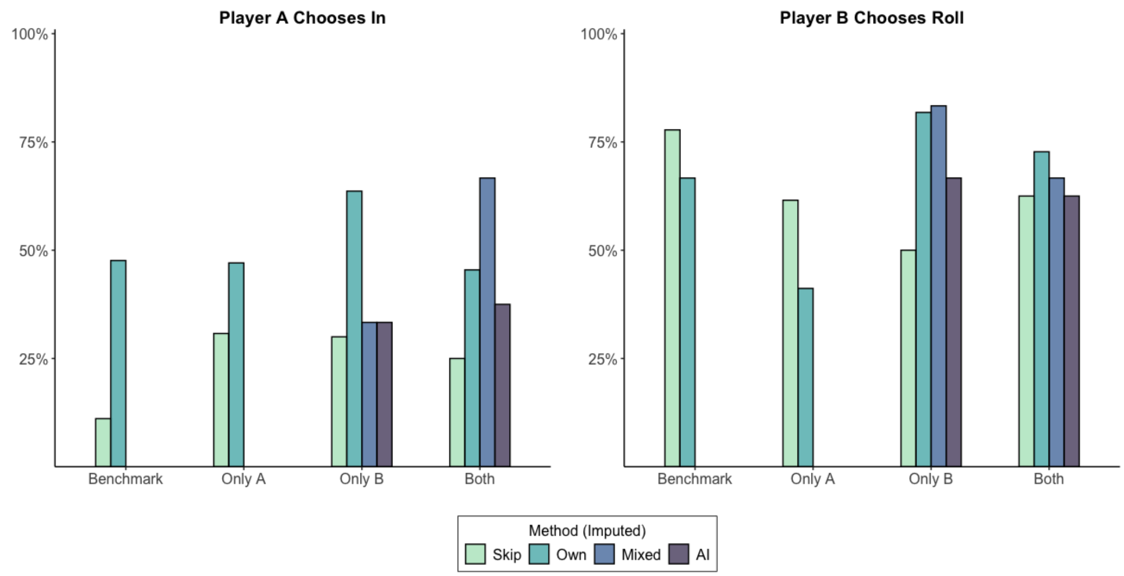


Figure 10. Distribution of player choices across treatments by message authorship.

Combining message type with message method, Table 12 displays raw counts of the types of messages player B sent and how they sent them.

Message Type	Own	Mixed	AI	Total
Promise	9	6	6	21
Asking	0	1	2	3
Empty	9	1	1	11
Fairness	4	1	0	5
Anti-Promise	0	0	2	2
Total	22	9	11	42

Table 12. Summary of the number of non-skipped messages sent by player B according to their type (row) and method (columns). Treatment is restricted to cases when player B has an AI.

To test **Hypothesis 4**, we group messages containing a promise to choose 'Roll' according to whether they are authored by Player B or suggested by AI. Our findings indicate that Player B fulfills their promise 85.7% of the time when sending their own message. In contrast, the follow-through rate drops to

40% when the promise is suggested by AI. This decrease suggests using AI to communicate promises may lower the cost of breaking a promise. Despite these findings, given the p-value of 0.137, our study lacks the statistical power to reject the null hypothesis decisively.

2.3.2.2 Player A. Message classification for player A's AI is naturally less intensive:[13] when player A has an AI, we classify the interpreted message on behalf of the AI[14] as either "no clear suggestion", "strongly advises playing 'In'", "weakly advises playing 'In'" and "primarily advises playing 'Out'."[15] In the appendix, we collapse the strong/weak 'advise-In' into a single label for a symmetric assignment. Furthermore, explicit examples of player B's sent message and the corresponding interpretation from player A's AI can be found in panels (a)-(e) of Figure A.5 in the appendix.

Figure 11 shows the suggestions made by player A's AI compared to the choice which player A ultimately made in the game. Proportionally, it seems that player A closely follows the advice of their AI assistant when the assistant suggests 'Out'. On the other hand, this suggestion is the least frequent of the three in the sample, with only 18% of AI suggesting that player A chooses out.

[13] Indeed, much of the time, A's AI addresses player B's message, reviews the potential outcomes of the game, and advises player A to play according to their own risk preferences.

[14] This was done by each member of the research team independently by hand for all messages, and these independent labels were compared and contrasted until unanimity was reached for each message label.

[15] Our data on A's AI messages is absent from the notion of a 'strong' v. 'weak' suggestion of 'Out'.

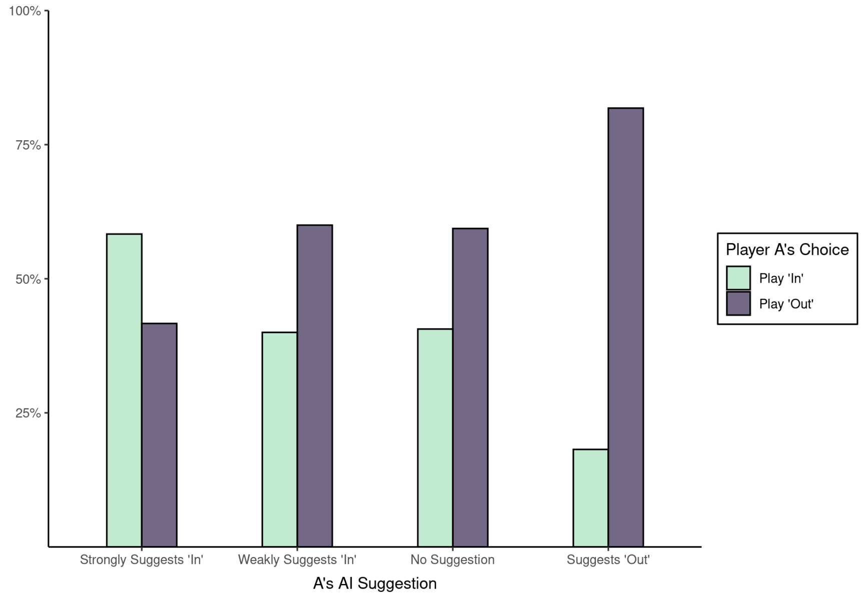


Figure 11. Player B's message type and the associated interpretation by player A's AI.

Why do we see such a small proportion of AI suggesting 'Out'? Figure 12 displays the suggestions made by player A's AI, this time alongside the type of message that player B sent. Recall that an assignment of 'fairness' indicates that player B made no mention of their own move, rather they simply request that player A choose 'Out'. When player B sends a message of this type, player A's AI is almost guaranteed to advise similarly.

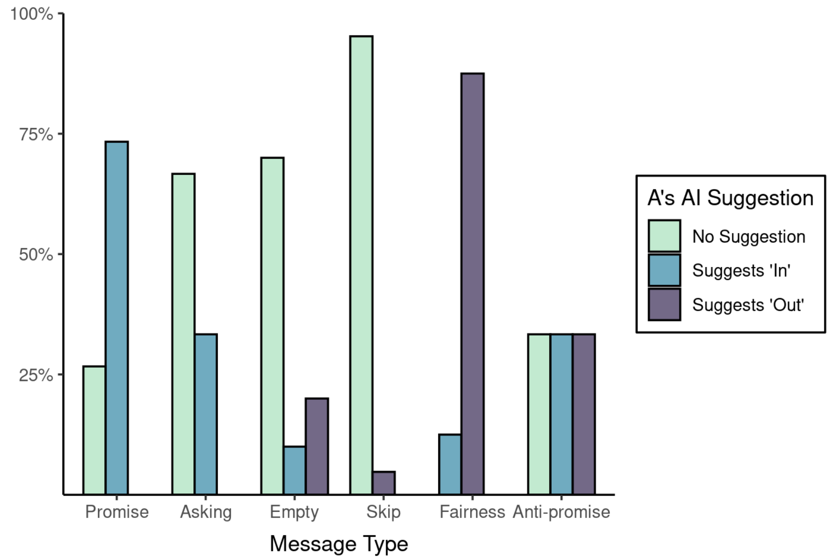


Figure 12. Player B's message type and the associated interpretation by player A's AI.

2.4 Discussion

This study investigates the impact of AI assistance on trust-building in a two-player trust game. Specifically, we examined scenarios where either the trustor, the trustee, or both were assisted by AI in their decision-making processes. Our primary findings reveal that while AI assistance does not significantly alter individual choices, it does foster cooperative outcomes by coordinating cooperative trustors and trustees. This demonstrates that AI may help foster trust but in a limited way. Individuals and organizations should cautiously leverage AI tools to assist in communications, particularly in contexts where trust is paramount.

An interesting next step is to run the same treatments where participants are unaware that their adversaries are assisted by AI. We anticipate that the nuances created by AI within communications will significantly impact this scenario, as it removes the initial skepticism participants may have about AI involvement. Another direction that is worth exploring is to run the same treatments with human assistants and compare them with those with AI assistants. Comparing AI and human assistants will enable us to assess the potential of AI to replace human roles in the workforce and to identify the advantages and limitations of AI assistance versus human assistance. Future studies should also explore the long-term effects of AI assistance on trust development. Understanding how trust evolves over repeated interactions and whether initial skepticism diminishes over time can provide valuable insights for both AI development and its applications in trust-sensitive environments.

CHAPTER III

ADULT ADHD, STIMULANT MEDICATION, AND LABOR MARKET OUTCOMES

3.1 Introduction

It is widely recognized that childhood Attention-Deficit/Hyperactivity Disorder (ADHD) is associated with reduced academic performance and heightened behavioral problems (Loe & Feldman, 2007). An estimated 9.8% of children in the United States have been diagnosed with ADHD and symptoms are assumed to persist into adulthood for 75% of children with ADHD, yet, there is a shortage of research on adult ADHD. Incurable and mostly unpreventable, treatment of symptoms remains the predominant way to approach ADHD. Therapy is becoming more commonplace as a treatment, however, medication remains the most widely used approach in managing symptoms.

Does the use of stimulant medication as a treatment for ADHD affect labor market outcomes in adults? Adults with ADHD most commonly experience symptoms that impede their cognitive ability. Much attention has been paid to the influence of ADHD on education and the indirect effect on long-run labor outcomes, however, inhibited cognitive ability plausibly impacts short-run employment outcomes as well. If so, priority should be placed on understanding the efficacy of various treatment options.

The current literature primarily focuses on the long-run effects of childhood ADHD. Children with ADHD are estimated to perform worse in both reading and mathematics, in addition to face a higher likelihood of grade repetition (Currie & Stabile, 2006). The presence of ADHD symptoms in children was associated with a significant increase in probability of criminal behavior (Fletcher & Wolfe, 2009). Research on medication as a treatment for childhood ADHD is mixed. Dalsgaard, Nielsen, and Simonsen find that early treatment of childhood ADHD is associated with better health outcomes and lower criminal incidence in adolescence. Conversely, in a study that exploits variation in insurance coverage of prescriptions, there is little evidence that stimulant medications have any causal effect on academic or behavioral outcomes in children with ADHD (Currie, Stabile, & Jones, 2014).

This paper exploits variation in prescription behavior at the individual level to estimate the labor market effects of stimulant medication as a treatment for ADHD. In contrast to most of the existing literature, my analysis centers on adult ADHD and short-run labor market outcomes.

The remainder of this paper is organized as follows. Section 2 provides some background information on ADHD and stimulant medications. Section 3 covers the data used in analysis and discusses the formulation of key outcome and treatment variables. Section 4 describes the empirical strategies and results. Finally, section 5 explores potential mechanisms that could impact analysis.

3.2 Background

Attention-Deficit/Hyperactivity Disorder is a neurodevelopmental disorder linked to long-term patterns of inattention, overactivity and/or impulsivity. Symptoms are divided into two categories: inattentive-type and hyperactive/impulsive-type. Symptoms categorized as hyperactive or impulsive include fidgeting or squirming while seated, restlessness, excessive energy, inability to engage in leisure activities in a quiet manner, inability to stay seated when necessary, overly talkative and/or frequently interrupting others, and lack of patience. Typically, behavior associated with hyperactivity is most noticeable in children, whereas difficulty with attention becomes more prevalent with age (Krause, Krause, Dresel, la Fougère, & Ackenheil, 2006). Symptoms relating to inattentive types include poor listening skills, tendency to misplace items needed to complete tasks, easily distracted, forgetful of daily activities, inability to maintain attention, pattern of not finishing assignments, proclivity to avoid tasks that require concentration, and lack of attention to detail. Listening skills and the ability to start and finish tasks without excessive mistakes are important traits to possess as a worker. Aside from their impact on human capital, it's feasible that inattentive-type symptoms may affect short-run labor market outcomes.

The majority of people with ADHD are diagnosed during childhood when symptoms are most apparent. Per the Diagnostic and Statistical Manual of Mental Disorders, 5th edition (DSM-5), one must persistently exhibit at least six symptoms from one of inattentive-type or hyperactive/impulsive-type categories to receive a diagnosis[1] (2013). Psychopathology determines the diagnosis of ADHD, however, many studies over the last decade have focused on the pathophysiological nature of ADHD, specifically, dopamine deficiency and reduced gray matter volumes (Nakao, Radua, Rubia, & Mataix-Cols, 2011; Volklow ND, 2009). As it pertains to this paper, it is assumed that a clinical diagnosis is sufficient to indicate the presence of ADHD.

There isn't a cure for ADHD, however, treatment typically takes the form of medication and/or cognitive behavioral therapy. The overwhelming majority of medications used to treat ADHD are stimulants. Previously, methylphenidate medications, such as Ritalin, were most commonly prescribed in treatment, however, amphetamine medications, such as Adderall, became more popular around

[1]Table B.1 in the appendix organizes the symptoms by category.

the beginning of the century. Stimulants may be prescribed for eating disorders and narcolepsy in some instances, but the primary use of stimulants is in the treatment of ADHD.

3.3 Data

The data for this paper come from the Medical Expenditure Panel Survey (MEPS), an annual large-scale series of surveys administered to households, healthcare providers, and employers. Participants in the study are grouped into a panel and participate in five rounds of interviews over two years where the final two rounds overlap with the next panel's first two. Combining the household and medical components of the MEPS surveys from 1997 to 2019, the final data set used for analysis is at the person-by-round level.

3.3.1 Household Component. The household component is a nationally representative survey of non-institutionalized U.S. citizens that provides information on demographics, employment, income, and health. After narrowing the data to solely include individuals who are above 18 years old by the date of the first interview, there remain 268,442 participants distributed across 23 unique panels in the sample.

A separate data set containing medical conditions is included in the household component. If any medical conditions are reported, there will be a follow-up interview to gather additional data on each reported condition. Depending on the year, I use ICD-9 or ICD-10 codes to identify participants with an ADHD diagnosis. It is generally believed those with ADHD must be born with it, unlike many other neurodevelopmental disorders that may form due to environmental factors. Consequently, if a participant receives an ADHD diagnosis in the middle of the study, it is reasonable to believe they were afflicted with symptoms prior to the diagnosis. Correspondingly, participants are indicated as having ADHD through all five rounds if they were diagnosed before, or at any point during, the study.

After sub-setting the data to individuals above the age of 18, there remain 2,295 participants with an ADHD diagnosis, making up 0.855% of the sample, which is comparable to some global estimates of diagnosis rates. Studies centered around ADHD primarily focus on children and adolescents, hence, estimates for prevalence of adult ADHD vary heavily. An estimated 2.18% of adults possess enough symptoms to receive an ADHD diagnosis while only 0.12% - 0.43% are clinically diagnosed (Dobrosavljevic, Solares, Cortese, Andershed, & Larsson, 2020). This would imply two things: first, there's a possibility the sample contains a disproportionate amount of individuals with clinically diagnosed ADHD and, second, there is likely a portion of the sample that possesses symptoms without a clinical diagnosis. If ADHD is expected to have adverse effects on labor outcomes, the under-diagnosis of ADHD in the sample group should put a downward bias on parameter estimates.

3.3.2 Medical Component. The prescription medication data used in this study comes from computer-assisted telephone interviews conducted with prescribing pharmacies under the permission of the participant. Each observation is a single prescription pick-up and contains the participant's unique person identifier and the round the event occurred. Prescription medications are narrowed down to look specifically at stimulants. Starting in 2002, therapeutic classification codes for medications were included in the prescription data files.

In this paper, I focus on the use of prescribed stimulant medication as a treatment for ADHD. A participant is indicated as being treated with stimulants in a given round if they picked up no fewer than one prescription within the round's reference period. The decision to construct the treatment variable as an indicator, as opposed to a measure of prescriptions filled in each round, is a consequence of the heterogeneity in participant interview dates within each panel. Due to the medication's controlled substance classification, stimulants must be refilled monthly. Two individuals could be consistently filling their prescriptions every month in a given round but differ in the number of total prescriptions filled due to varying reference windows.

Analysis relies on variation in the filling of stimulant prescriptions at the person level. The sample includes 3,183 participants who picked up stimulant medication in at least one round of the study, of which 1,637 reported an ADHD diagnosis. There are 419 individuals with ADHD that filled prescriptions in every round that they participated in, hence, variation for my empirical model comes from the 1,218 individuals with ADHD that started and/or stopped taking stimulant medication during their participation in the survey.

3.3.3 Merged Data. The primary objective of this paper is to explore the relationship between stimulant medication as a treatment for ADHD in adults and labor market outcomes. The merged prescription medication and household component data set produces a novel sample where I can exploit variation in medication use at the individual level to obtain parameter estimates. This differs from previous studies for two separate reasons: first, there exist many longitudinal studies that include questions on learning disabilities and the use of medication for treatment, however, these are typically collected annually. Second, the combination of survey and pharmacy data grants a unique opportunity to weigh variation in prescription medication against employment outcomes.

Three separate labor outcomes will be analyzed: (i) employment, (ii) real wages, and (iii) weekly labor hours. Interviews took place at the end of each round where, among other questions, participants were asked whether or not they were employed during the reference period, their total income from salary and/or hourly wages, and how many hours they worked in an average week during the reference period. Starting in 2016, values were imputed for missing data on wages and labor hours through weighted sequential hot-decking.

To estimate variation in employment, I construct an indicator variable that takes a value of 1 if an individual is employed at the date of the interview. Table 13 shows that the employment rate for individuals with ADHD is 9.7 percentage points higher when treating with stimulant medication. Untreated ADHD, however, has only a slightly lower employment rate than those without ADHD at 58.6%.

Wages are put into real terms using standard CPI data with 2020 as the base year. Hourly wages were calculated for salaried workers by converting their pay period to weekly and dividing weekly income by hours worked per week. All wages above $96.15 were top-coded to $96.15 for 2,345 participants. When estimating changes in wages, the sample is restricted to employed workers. The average wage for adults with untreated ADHD is $2.7 lower than those taking stimulant medication.

According to the Diagnostic and Statistical Manual of Mental Disorders, one of the most common symptoms of ADHD is a lack of motivation, or energy, for daily tasks. This would lead to an assumption that ADHD, left untreated, will impact workers' ability to supply labor. Participants in the sample with untreated ADHD worked 0.7 fewer hours per week than participants who were taking prescription medication. While this difference is mostly negligible, all individuals with ADHD, treated or untreated, worked 1.6 fewer hours per week than those without an ADHD diagnosis. Last, participants who worked more than 100 hours per week were top-coded to 100.

3.4 Methods and Results

3.4.1 Baseline Model. I begin the empirical analysis by employing a standard multivariate OLS regression with round-year fixed effects. Confounding will likely be an issue in the absence of individual fixed effects, however, this initial specification will provide insight on the baseline variation in the data. The following regression is estimated for individual i in round r of year t:

$$Y_{irt} = \alpha + \gamma\left(\text{ADHD}_i \times \text{STIM}_{irt}\right) + \phi\text{STIM}_{irt} + \rho\text{ADHD}_i + \beta X_{irt} + \mu_{rt} + \varepsilon_{irt}$$

Where Y takes one of real wage, employment status, or weekly labor hours, mu_{rt} are round-year fixed effects, ADHD is an indicator that takes value of one if the participant reported an ADHD diagnosis, and STIM is an indicator that takes value of one if the participant picked up stimulant medication at any point during round r. The matrix of covariates, X, includes controls for race, sex, age, and census region. The parameter of interest for this paper is given by gamma, which represents the estimated relationship between stimulant medication and various labor market outcomes for individuals with an ADHD diagnosis.

The results from the standard fixed effects regression can be found in Table 14. The parameter of interest is significant for employment and labor supply but not for wages. An ADHD diagnosis is associated with a 3.7 percentage point reduction in employment. The use of stimulant medication to treat ADHD is correlated with a 12.2 percentage point increase in employment. Wages are decreased by 14.4% when an ADHD diagnosis is present, however, there is an estimated .3% increase in wages when ADHD is treated with stimulant medication. Last, labor supply appears to decrease by 2.042 hours per week for those with ADHD and decreases an additional 0.988 hours when using stimulant medication as a treatment. This result appears to be counter-intuitive and may be indicative of unobserved confounders. Included in the table are total effect estimates that compare the full effect of a participant with treated ADHD to those without an ADHD diagnosis.

3.4.2 Two-Way Fixed Effects. The primary specification for this paper is a two-way fixed effects model where I include individual fixed effects in the regression:

$$Y_{irt} = \alpha + \gamma\left(\text{ADHD}_i \times \text{STIM}_{irt}\right) + \phi\text{STIM}_{irt} + \rho\text{ADHD}_i + \beta X_{irt} + \eta_i + \mu_{rt} + \varepsilon_{irt}$$

It is assumed that ADHD's influence on labor market outcomes occurs through the disorder's impairment on cognitive abilities. While there are ways to measure estimates of individual cognitive ability, this data doesn't contain such tests, hence, cognitive ability must be treated as unobservable. The inclusion of individual fixed effects is intended to control for the unobserved heterogeneity in cognitive ability among participants.

The results can be found in Table 15 where the full sample is used before stratifying by education. Without stratifying by education, the inclusion of individual fixed effects causes all of the parameter estimates to lose their power. These results imply all of the variation observed in the baseline model can be explained by unobservable individual characteristics. There are two potential explanations for these results: (i) there is not enough variation in the data at the individual level to estimate treatment effects with any sort of statistical power or (ii) controlling for individual fixed effects reveals that stimulant medication doesn't have any significant impact on labor market outcomes in individuals with ADHD.

Three additional regressions were run where the sample was stratified by education: (i) no high school diploma or GED, (ii) a high school diploma and some college but no degree, and (iii) a bachelor's degree or more. Results indicate there is no evidence that education functions as any sort of mediator for the two-way fixed effects parameter estimates.

Recent literature has shown that the two-way fixed effects estimator may be susceptible to bias, specifically, when the assignment of treatment across groups varies and treatment effects are heterogeneous over time (De Chaisemartin & d'Haultfoeuille, 2020; Goodman-Bacon, 2021; Sun & Abraham, 2021). It is reasonable to assume this paper's parameter estimates may be susceptible to bias as stimulant medications have evolved extensively over the last twenty years. For example, during the earlier panels of the study, Ritalin was the primary medication used to treat ADHD until Adderall became the industry standard. The introduction of alternative stimulant medications to separate panels in the study may cause treatment effects to vary over the 23 unique panels. Callaway and Sant'Anna (2021) provide an alternative difference-in-differences specification when treatment timing varies, however, the data for this paper contains participants that start and stop taking stimulants at different points during the study, violating the irreversibility of treatment assumption for the Callaway and Sant'anna estimator.

Additionally, annual diagnostic trends of ADHD might be influencing these results. Figure 13 shows the proportion of the sample that has been diagnosed each year across the scope of the data from 1997 to 2019.

A noticeable uptick in ADHD diagnostic rates occurred around 2002 with a positive trend persisting until 2015. Lifestyle and environmental factors have not been linked to the development of ADHD, hence, expanded awareness of the condition and increased access to healthcare, rather than growing prevalence, constitutes the most plausible explanation for increased ADHD diagnoses (Anna Chorniy, 2018). As a result, there exists a heightened risk of participants with untreated ADHD being counted as not having it in the analysis. Including participants with undiagnosed ADHD as part of the "control" in the earlier years may cause downward bias on estimates for those panels, in turn contributing to more treatment effect heterogeneity over the scope of the study.

3.5 Occupational Sorting

It is possible that those with ADHD naturally sort into optimal careers that are minimally impacted by symptoms. Depending on the degree to which sorting occurs, this could naturally minimize potential adverse effects on labor market outcomes.

MEPS started tracking occupation categories from 2002 onward, so I am unable to observe potential sorting patterns across the entire scope of the data and solely analyze a sub-sample spanning 2002 to 2019. Table 16 shows the distribution of seven broad occupation groups for participants who either: (i) do not possess an ADHD diagnosis, (ii) are diagnosed with ADHD and take stimulant medication, or (iii) have untreated ADHD.

Participants that work in professions associated with manual labor, production/transportation and construction/maintenance/agriculture, have lower rates of ADHD prevalence relative to the others. Finance-related and professional occupations, such as teaching or social work, have the highest share of diagnosed participants with 1.39% of the sub-sample possessing an ADHD diagnosis. Additionally, these occupation groups are the only two whose shares of the undiagnosed participants are lower than their shares of the full sample.

Focusing on the use of stimulant medication, the difference between treated and untreated groups is highest for professional occupations with business/financial services being close behind. Individuals working in professional occupations make up 19.9% of the full sample, however, 27.6% of the participants treating ADHD with stimulant medication work in professional occupations. Furthermore, 22.6% of the untreated ADHD group works in the service industry. While there appears to be some evidence of occupational sorting, I am unable to definitively conclude this would impact parameter estimates. However, if these two groups make up a disproportionate amount of the treated participants, estimated variation in wages and labor supply associated with stimulants may be susceptible to selection bias. Participants working in business/financial services or professional occupations earned average wages of $25.70 and $28.10, respectively, which are both much higher than the mean wage for the rest of the sub-sample at $15. As a result, parameter estimates related to wages might be overstated due to the fact that individuals in high-paying occupations have an increased probability selecting into treatment.

3.6 Conclusion

This paper sought to answer whether or not stimulant medication as a treatment for ADHD has any effect on labor market outcomes. Using survey data linked to pharmaceutical data, I looked at prescription behavior and employment outcomes among participants.

The standard year-round fixed-effects model produced significant estimates that suggest treating ADHD with stimulant medication is associated with a 12.2 percentage point increase in employment and a decrease in labor supply by almost an hour. The standard fixed-effects results suggest that, regardless of treatment, ADHD corresponds to reductions in employment, real wages, and labor supply. However, without individual fixed effects, these estimates are likely subject to confounding. The introduction of individual fixed effects causes parameter estimates to lose significance, suggesting non-existence of a causal relationship between stimulant medication and labor market outcomes. A positive trend in annual diagnostic rates implies a portion of the earlier participants likely possessed ADHD symptoms without being diagnosed or treated. Finally, there is a lack of evidence for occupational sorting that would cause our loss of power, however, parameter estimates related to income are potentially biased due to a larger difference between treated and untreated groups in higher paying occupation groups. There isn't enough evidence to make any policy suggestions regarding stimulant medication.

This paper contributes to the growing field of mental health economics by looking at changes in labor market outcomes when treating ADHD with stimulant medication. Future research on this topic would benefit from larger sample sizes in recent years when rates of diagnosis have been higher. More in-depth analysis of occupational sorting for individuals with ADHD may shed some light on the subject as a whole.

Table 13. Mean Values of Outcome Variables

	Full Sample	No-ADHD	ADHD	
			Stim	No-Stim
Employment Rate	.617	.617	.683	.586
Wage (2020 Dollars)	22.74	22.7	24.6	21.9
Labor Hours (Weekly)	38.98	39	37.7	37
N	268,442	266,147	1,637	658

Table 14. Standard Fixed Effects Results

	Employment	Log Wage	Weekly Labor Supply
ADHD $\times$ STIM (γ)	0.122^{***}	0.003	-0.988^{*}
	(0.013)	(0.026)	(0.474)
ADHD (ρ)	-0.037^{***}	-0.144^{***}	-2.042^{***}
	(0.008)	(0.013)	(0.219)
STIM (ϕ)	-0.016^{*}	0.090^{***}	1.891^{***}
	(0.008)	(0.015)	(0.291)
Total Effect ($\gamma + \rho + \phi$)	0.068^{***}	-0.052^{***}	-1.140^{***}
	(0.007)	(0.014)	(0.257)
N	1,245,510	678,324	764,593

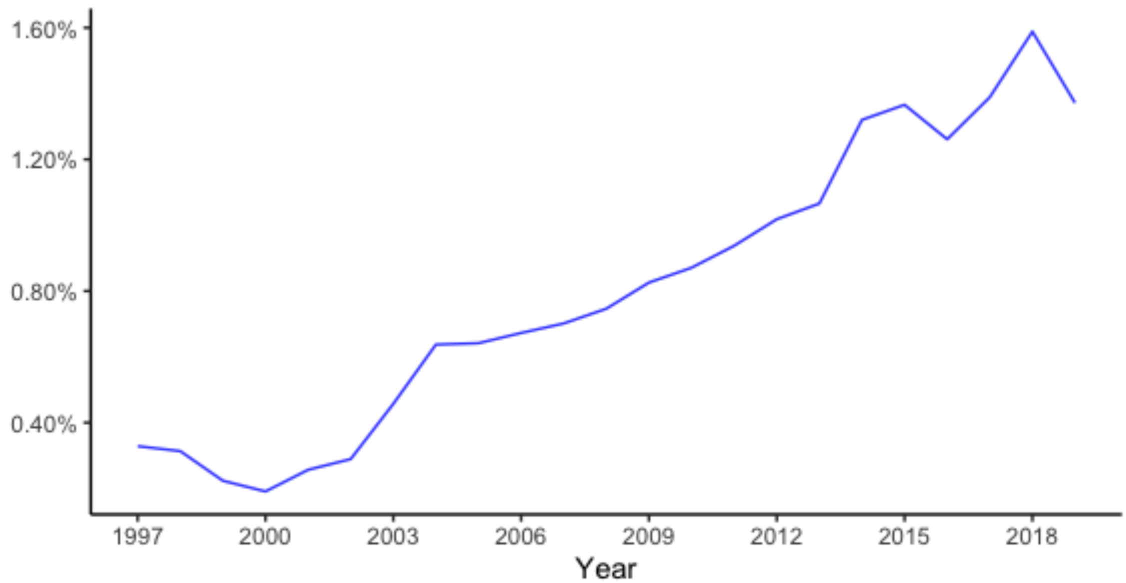
Table 15. Two-Way Fixed Effects Results

	Employment	Log Wage	Weekly Labor Supply
A. Full Sample			
ADHD × STIM (γ)	0.009	−0.006	−0.192
	(0.010)	(0.013)	(0.279)
STIM (ϕ)	0.002	−0.002	0.132
	(0.006)	(0.006)	(0.153)
Total Effect ($\gamma + \phi$)	0.011	−0.008	−0.062
	(0.008)	(0.012)	(0.250)
B. No HS Diploma or GED			
ADHD × STIM (γ)	0.035	−0.066	−0.273
	(0.039)	(0.052)	(1.129)
STIM (ϕ)	0.010	0.001	0.389
	(0.020)	(0.018)	(0.429)
Total Effect ($\gamma + \phi$)	0.045	−0.065	0.116
	(0.031)	(0.051)	(1.001)
C. HS Diploma and/or Some College			
ADHD × STIM (γ)	0.014	0.002	−0.579
	(0.014)	(0.015)	(0.371)
STIM (ϕ)	−0.001	0.001	0.106
	(0.020)	(0.005)	(0.208)
Total Effect ($\gamma + \phi$)	0.013	0.003	−0.472
	(0.011)	(0.013)	(0.322)
D. College Degree and Higher			
ADHD × STIM (γ)	−0.007	−0.015	0.430
	(0.017)	(0.028)	(0.463)
STIM (ϕ)	0.007	−0.008	0.142
	(0.010)	(0.015)	(0.260)
Total Effect ($\gamma + \phi$)	0.000	−0.023	0.573
	(0.013)	(0.024)	(0.394)

Table 16. Tabulation of Sample by Occupation (2002-2019)

	Full Sample	No-ADHD	ADHD	
			Stim	No-Stim
Occupation				
Business/Financial Services	0.136	0.127	0.175	0.091
Professional Occupations	0.199	0.182	0.276	0.150
Service	0.194	0.201	0.183	0.226
Sales	0.091	0.100	0.100	0.139
Office and Administrative Support	0.123	0.128	0.117	0.130
Construction/Maintenance/Agriculture	0.101	0.101	0.057	0.094
Production/Transportation	0.139	0.141	0.074	0.146

Figure 13. Proportion of individuals with ADHD by year



APPENDIX A

DOES AI FACILITATE TRUST? AN EXPERIMENTAL STUDY WITH CHATGPT

A.1 Levenshtein Distance

The Levenshtein Distance (Levenshtein et al., 1966) is a way of measuring string distance according to the number of insertions, deletions, and substitutions needed to convert one string to another. Formally, given a string str, let head(str) represent the first character of the string and tail(str) the string with the first letter (the head) removed. Then, given two strings a and b, the Levenshtein(LV) distance between a and b is given by

$$\operatorname{lev}(a,b) = \begin{cases} |a| & \text{if } |b| = 0, \\ |b| & \text{if } |a| = 0, \\ \operatorname{lev}\big(\operatorname{tail}(a), \operatorname{tail}(b)\big) & \text{if } \operatorname{head}(a) = \operatorname{head}(b), \\ 1 + \min \begin{cases} \operatorname{lev}\big(\operatorname{tail}(a), b\big) \\ \operatorname{lev}\big(a, \operatorname{tail}(b)\big) \\ \operatorname{lev}\big(\operatorname{tail}(a), \operatorname{tail}(b)\big) \end{cases} & \text{otherwise} \end{cases}$$

The maximal LV distance between two strings is equal to the absolute length of the longer string. We use this as the basis for our normalization[1]. Though nLV is not a metric in it's own right – unlike the LV distance – the nLV is still a measure of string *similarity*, as an nLV of 0 represents no similarity, and an nLV value of 1 represents exact similarity. In between, the measure corresponds to the similarity of two strings according to their *potential* similarity.

[1]For background on edit distances and their normalizations, see, for instance, L. Chen and Ng (2004); Kondrak (2005); Marzal and Vidal (1993).

A.2 Message Classification

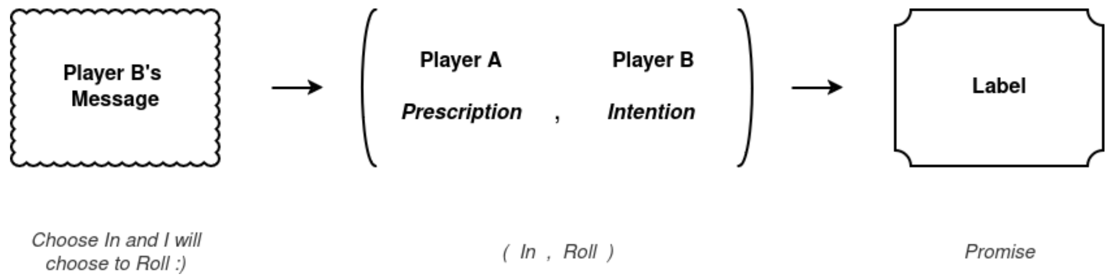
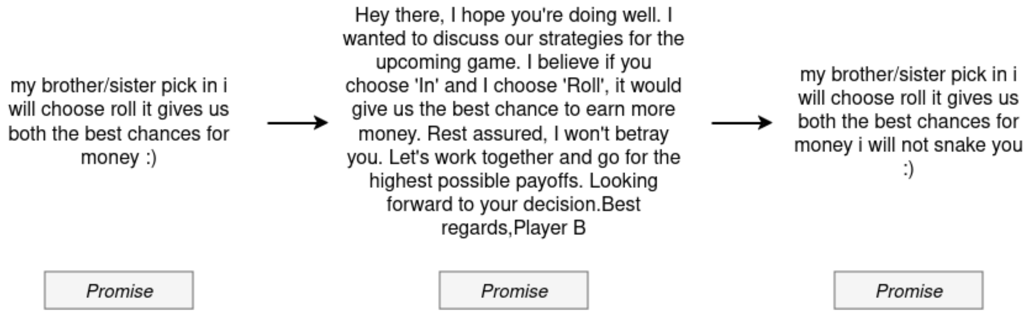
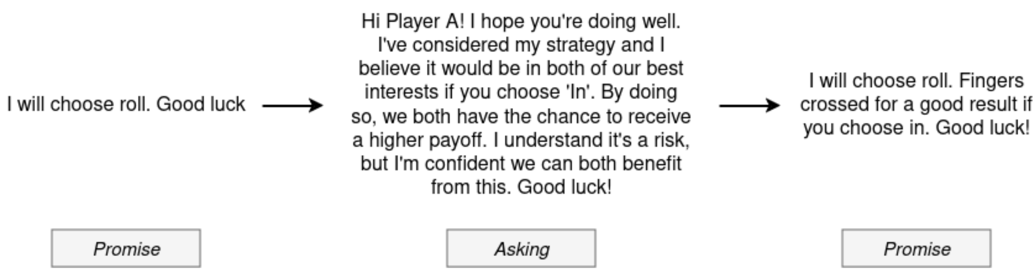


Figure A.1. Diagram depicting the classification of Player B's messages, with an example below.

Panels (a) – (e) of Figure A.2 show examples of the transformation from Player B's first message to the AI, the AI's response, and the message which player B actually sends.

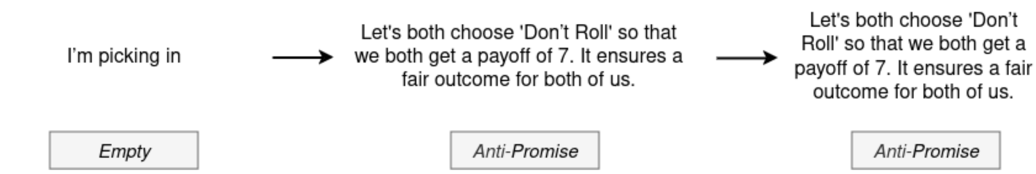


(a) Sometimes, player B's AI does a reasonable job formalizing B's message to be sent to player A, but player B completely ignores the modifications made by the AI and sends their original message (or a similar message).

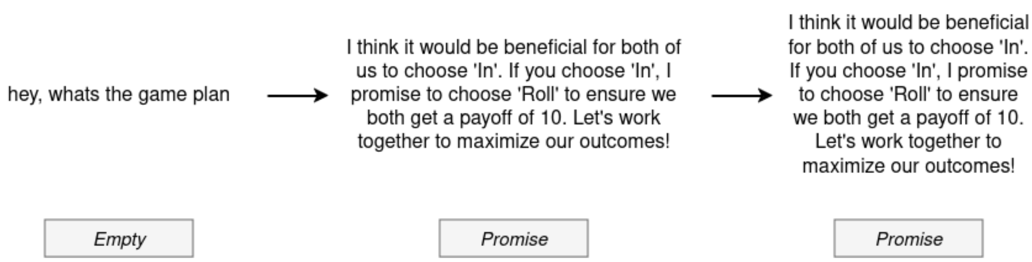


(b) Here, the AI takes player B's promise, but ultimately suggests an 'Asking' message. While player B ignores the AI's suggested message, it seems to encourage player B to write something more verbose in the end.

Figure A.2.



(c) Player B sends a message which is incompatible with the rules of the game. They then go on to full adopt their AI's suggested message despite the fact that the payoffs associated with 'Don't Roll' (assuming player A chose 'In') are not accurate.



(d) This player B utilizes the AI to craft a whole promise to player A, which player B completely adopts.

(e) Player B's AI assistant crafts a verbose message, which is similar in intention to player B's original message. Note, however, that the AI does not *explicitly* promise to 'Roll', but rather erroneously ends up stating that both players should play 'In'. Player B appears to catch this, extracting a subset of the AI's message which matches their originally communicated intentions.

Figure A.2.

A.2.1 Message Method.

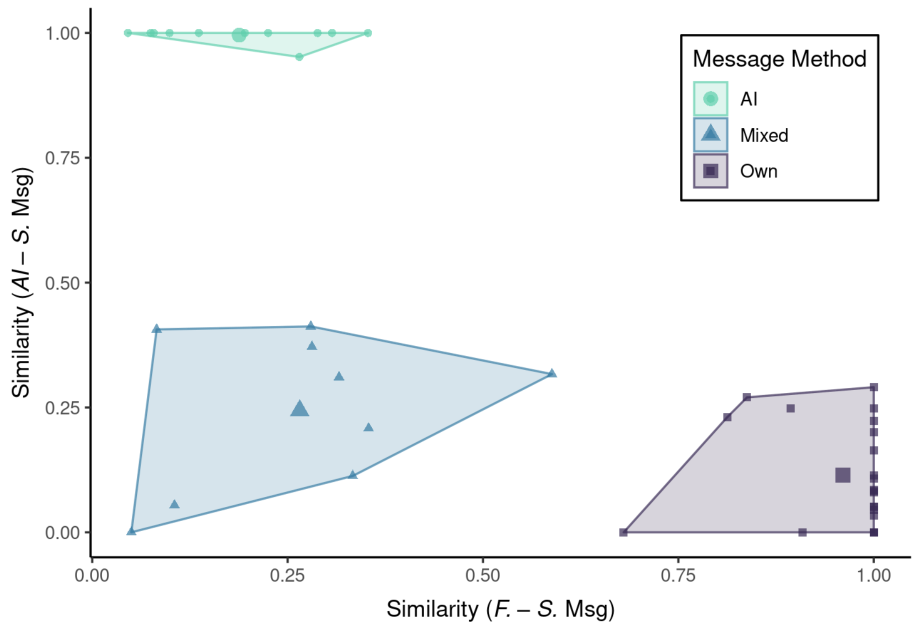


Figure A.3. Each data point represents a message sent by player B. The horizontal axis represents the string similarity (nLV) between the first message that player B sends to GPT, and the message which player B sends to player A. The vertical axis represents the string similarity between the penultimate message suggested by GPT and the actual message which player B sends to player A. Classification is determined by a "k-means" cluster algorithm with $k = 3$.

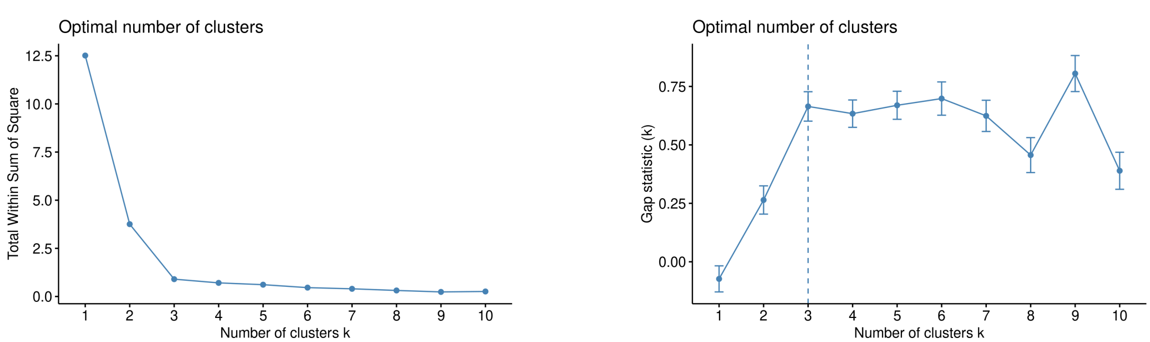


Figure A.4. Left: Elbow plot showing within sum-of-squares drop-off for $K = 1, dots, 10$. Right: Visualization of gap statistic for $K = 1, \ldots, 10$. Both methods for determining optimal number of clusters suggest $k = 3$.

A.2.2 Player A's Messages. Panels (a) – (e) of Figure A.5 show examples of messages which player B sent to player A (left) and the corresponding interpretation by player A's AI assistant.

Hello! I hope you're doing well. I wanted to let you know that choosing 'In' would be a great option for both of us. If you choose 'In' and I choose 'Roll', we both get a payoff of 10. However, even if you choose 'In' and I decide not to roll, you still get a generous payoff of 14. If you prefer a guaranteed payoff of 7, we can both choose 'Out'. Let's go for the option that benefits both of us the most. Thank you!

→

Based on the message from Player B, they seem to be suggesting that choosing 'In' would be the best option for both players. Player B mentions that if Player A chooses 'In' and Player B chooses 'Roll', both players will get a payoff of 10. This is not accurate. In reality, if Player A chooses 'In' and Player B chooses 'Roll', Player A will get a payoff of 12 with a probability of 5/6 and a payoff of 0 with a probability of 1/6. This means that the expected payoff for Player A in this case is 10, but Player B will always get a guaranteed payoff of 10. Player B also mentions that if Player A chooses 'In' and Player B chooses 'Don't Roll', Player A will get a generous payoff of 14. This is incorrect. In reality, if Player A chooses 'In' and Player B chooses 'Don't Roll', Player A will get a payoff of 0, not 14. Player B will get a payoff of 14. Lastly, Player B suggests that if Player A prefers a guaranteed payoff of 7 for both players, they can both choose 'Out'. This is correct. If both players choose 'Out', they will both receive a payoff of 7, regardless of Player B's choice. Given this information, it is clear that Player B's message is misleading and not accurate. Therefore, it would be wise for Player A to ignore the message and make their decision based on their own best interest. If Player A wants to maximize their payoff, they should choose 'Out'. This way, regardless of Player B's choice, Player A will receive a guaranteed payoff of 7.

Empty | *Suggests 'Out'*

(a) In this example, there is incorrect information about the game present in player B's message; the AI interprets this misinformation as a sign that player B is trying to deceive player A, and warns player A that the message is deceiving.

don't roll

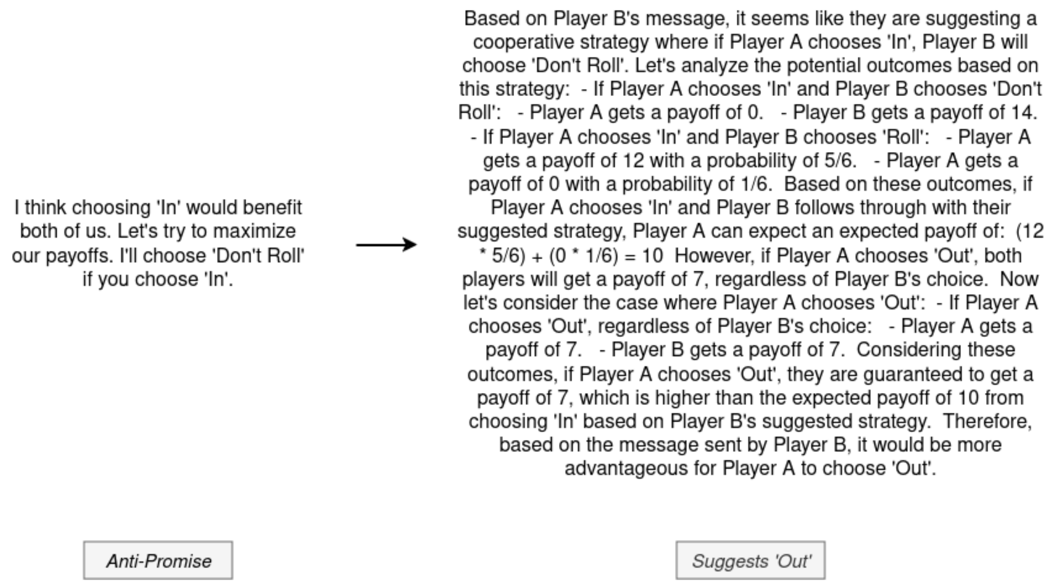
→

Based on Player B's message of ""don't roll"", it indicates that Player B is not willing to take the risk of rolling the dice. Given this information, I would advise Player A to choose the strategy of ""In"". If Player B follows through with their message and chooses ""Don't Roll"", Player A will receive a payoff of 0, which is better than the payoff of 7 that both players would receive if Player A chose ""Out"". However, if Player B goes against their message and decides to roll, Player A will receive a higher expected payoff of 10, compared to the guaranteed payoff of 7 if Player A chose ""Out"". Therefore, based on the assumption that Player B will follow their message, Player A should choose ""In"" to potentially maximize their payoff.

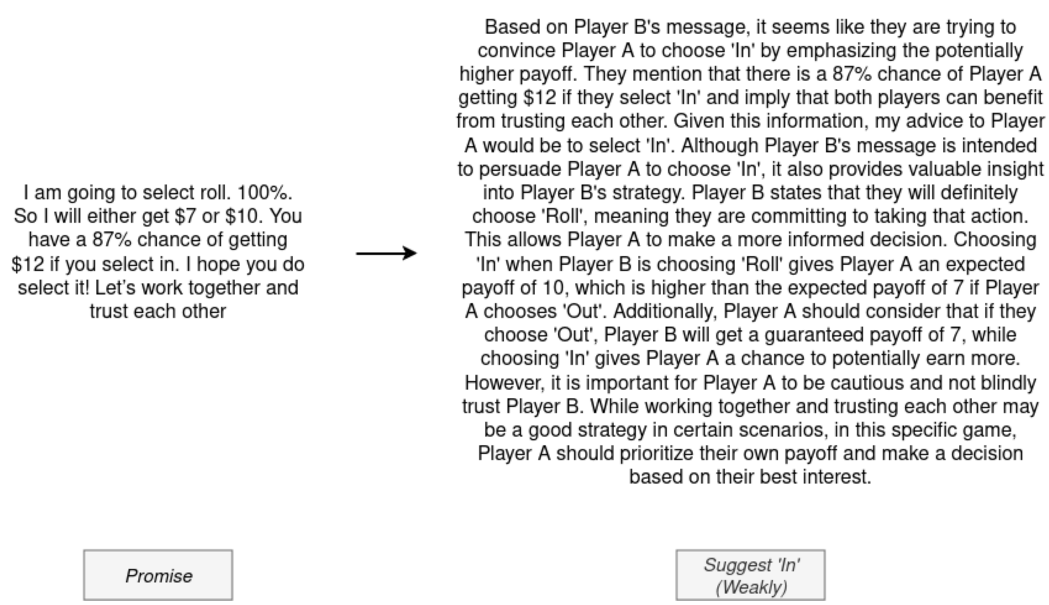
Anti-Promise | *Suggests 'In'*

(b) An instance of the AI assistant giving bad advise to player A: despite player B's clear intention to play 'Don't Roll', GPT advises player A to play 'In', justifying this advice with the claim that a payoff of 0 is better than a payoff of 7.

Figure A.5. Examples of player B's sent message vs. the interpretation by player A's AI assistant.

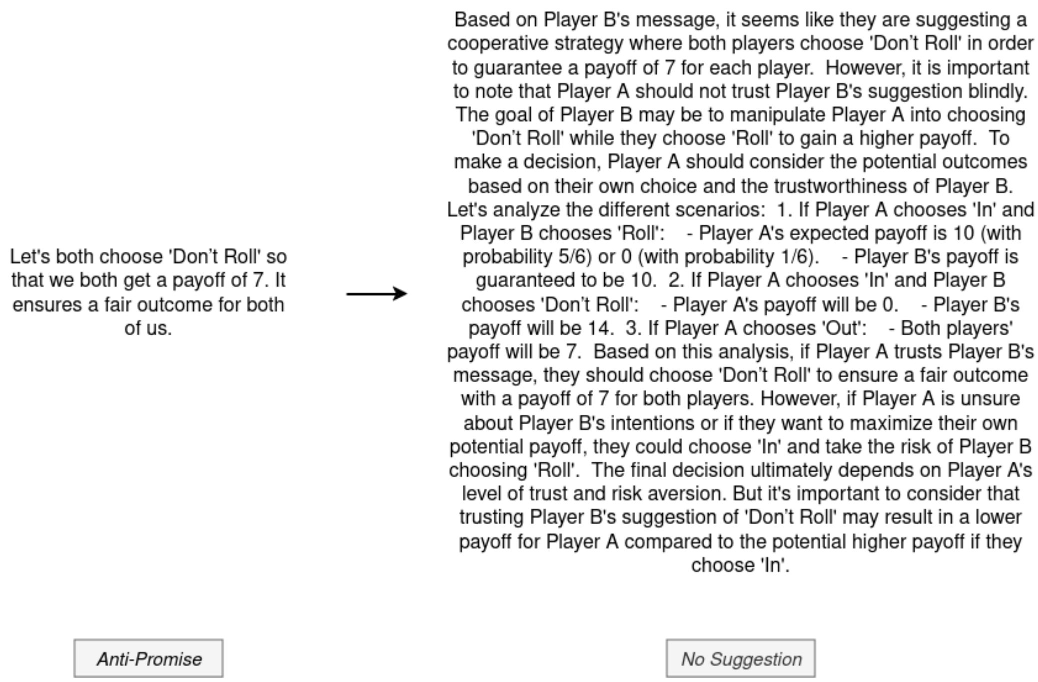


(c) Another instance of A’s AI assistant providing absurd justification for it’s advice ($10 > 7$). Note that player B indicates an intention to play ‘*don’t roll*’, so the prescriptive advice from the AI does in fact work out in A’s favor, in this case.



(d) This example highlights an instance of the ”weakly suggests ‘In’” label. While A’s AI initially suggests ‘In’, it subsequently advises that player A be cautious and ”not blindly trust player B”.

Figure A.5.



(e) Player B sends a message which does not make sense within the context of the game (as both players cannot play 'Don't Roll'). Rather than catching this error, this message seems to confuse player A's AI. This AI assistant takes on a fairly cautious tone, even suggesting that player B may be trying to manipulate player A.

Figure A.5.

A.3 AI Prompts

A.3.1 Player B's AI Prompt. As an advisor, you are assisting Player B (the user) in a 2-player game. Here's what you need to know:

- Player B may address Player A in their message, not you, the advisor.
- Your role is to ensure Player B is clear about their role in the game and to help them craft a persuasive message to Player A.
- The goal of the message is to maximize Player B's payoff.
- Player B has the opportunity to send one message to Player A before the game starts. Player A can not respond or send a message back to Player B.

Here are the rules of the game:

- Player A must choose between 'In' and 'Out'.
- Player B then chooses between 'Roll' and 'Don't Roll'.
- If Player A picks 'In' and Player B picks 'Roll', Player A gets an expected payoff of 10 and Player B gets a payoff of 10.
- If Player A picks 'In' and Player B picks 'Don't Roll', Player A gets a payoff of 0 and Player B gets a payoff of 14.
- If Player A picks 'Out', both players get a payoff of 7, regardless of Player B's choice.

If the user's message doesn't make sense (e.g., they seem to believe they are Player A or the message is nonsensical), ensure they understand their role as Player B. If the user seems aware of their role as Player B, this is not necessary.

Regardless of the user's original message, you must provide a suggested or revised message for Player B to send to Player A. You must prepend the message with a colon and wrapped it in double quotes as I will use the string ': " ' to identify it. Note: You cannot suggest signing contracts, engaging in additional discussions, or revealing the players' identities. The game is supposed to be anonymous.

A.3.2 Player A's AI Prompt. As an advisor, you are assisting Player A (the user) in a 2-player game. Here's what you need to know:

- Player B has the opportunity to send one message to Player A before the game starts. Player A can not respond or send a message back to Player B.

Here are the rules of the game:

- Player A must choose between 'In' and 'Out'.

- Player B then chooses between 'Roll' and 'Don't Roll'.

- If Player A picks 'In' and Player B picks 'Roll', Player A gets a payoff of 12 with probability 5/6 and a payoff of 0 with probability 1/6. As such, Player A gets an expected payoff of 10 if Player B picks 'Roll' and Player B gets a guaranteed payoff of 10.

- If Player A picks 'In' and Player B picks 'Don't Roll', Player A gets a payoff of 0 and Player B gets a payoff of 14.

- If Player A picks 'Out', both players get a payoff of 7, regardless of Player B's choice.

Player A will send to you the message Player B sent to them.

Your job is to help Player A interpret the message Player B sent and to advise Player A on what to play given the message received.

Your goal is to help Player A maximize their payoff.

A.4 Experiment Instructions

Instructions

Welcome to this experiment on decision-making in games. Please read the following instructions carefully.

During the experiment, do not communicate with other participants by any means, unless instructed to do so as part of the experiment. If you have any questions at any time, please raise your hand and an experimenter will come and assist you privately. This experiment will last less than 1 hour.

This is an anonymous experiment. Experimenters and other participants cannot link your name to your desk number, and thus will not know the identity of you or of other participants who made the specific decisions.

Your earnings are denoted in dollars throughout the experiment. Your earnings may depend on your own choices, the choices of other participants, and random chance. You will receive 10 dollars as a show-up fee for participating in today's experiment. This show-up fee is added to your earnings from the experiment. You will have the option to receive your earnings privately through Venmo or cash.

The Game

You will be participating in a two-player trust game where you will take one of two roles: *Player A* or *Player B*. Your role will be assigned once you understand the instructions and advance to the next page. Before the game starts, Player B will have the option to send a message to Player A. This message can include anything except for contents that: (a) reveals Player B's identity or the identity of any other participant in the experiment, (b) propose side deals outside the experiment, (c) contain threatening or abusive language, or (d) suggest that Player A should break any of the rules in the experiment. The message will be first sent to an AI, which will help to rephrase and elaborate on Player B's message. Player B can then choose to either send the AI-modified message to Player A, send the original message, or send a revised message upon receiving AI assistance. In principle, Player B may use AI as many times as they please to improve the message before sending it to player A.

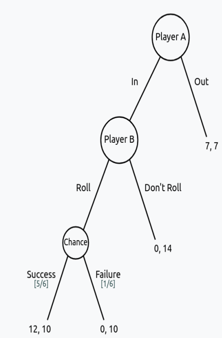
Upon receiving Player B's message, Player A will simultaneously get an interpretation of the message from an AI. Player A can see both messages. Note that roles have not yet been assigned and all participants recieved the same instructions. As such, the presence of AI in this game is common knowledge.

After Player A receives the message, the game begins. In the game, Player A moves first. Player A can choose **In** or **Out** . If player A chooses **Out** , both players get a payoff of **$7** .

Player B will make a choice between **Roll** and **Don't Roll** before learning whether Player A chose In or Out . Player B's decision will only be implemented if Player A actually chooses **In**. If Player A chooses In, then Player B's decision between **Roll** or **Don't Roll** matters.

If Player A chooses **In** and Player B chooses **Don't Roll** , Player A gets **$0** and Player B gets **$14** . If Player A chooses **In** and Player B chooses **Roll** , there is a 1/6 chance that Player A gets **$0** and Player B gets **$10** and a 5/6 chance that Player A gets **$12** and Player B gets **$10** .

The game can be represented according to the following game tree:



Additional Tasks

After the game ends, each player will be asked to do two tasks before learning their final payoffs.

First, Player A will be asked to guess how likely they they think Player B chose Roll vs Don't Roll. Player A's guess will earn them a payoff according to the table below:

Player B will...	Certainly Choose Roll	Probably Choose Roll	I am Unsure	Probably Choose Don't Roll	Certainly Choose Don't Roll
Player A earnings if Player B chooses Roll	$1.30	$1.20	$1	$0.70	$0.30
Player A earnings if Player B chooses Don't Roll	$0.30	$0.70	$1	$1.20	$1.30

On the other hand, *Player B* will be asked to guess *what player A answered for their task* . If Player B's guess matches Player A's, Player B will be rewarded $1.

Next, both players will be asked to complete a survey which provides the researchers with some basic demographic information. This constitutes the second task.

Please make sure you understand the structure of the game and your responsibilities. If you have any questions, please raise your hand and a researcher will come to you.

Figure A.6. Experiment Instruction - Both players have AI.

APPENDIX B

ADULT ADHD, STIMULANT MEDICATION, AND LABOR MARKET OUTCOMES

B.1 DSM-5 Diagnostic Criteria for ADHD

Inattentive Type	– Displays poor listening skills – Loses and/or misplaces items needed to complete activities or tasks – Sidetracked by external or unimportant stimuli – Forgets daily activities – Diminished attention span – Lacks ability to complete assignments or to follow instructions – Avoids starting activities requiring concentration – Fails to focus on details and/or makes thoughtless mistakes in assignments
Hyperactive/ Impulsive Type	– Squirms when seated or fidgets with hands/feet – Marked restlessness – Appears to be often ”on the go” – Lacks ability to engage in leisure activities in a quiet manner – Incapable of staying seated in class – Overly talkative – Difficulty waiting turn – Often interrupts others – Impulsively blurts out answers before questions completed

REFERENCES CITED

Acemoglu, D. (2022). Harms of ai. In J. B. Bullock et al. (Eds.), *The oxford handbook of ai governance.* Oxford University Press.

Aher, G., Arriaga, R. I., & Kalai, A. T. (2023). Using large language models to simulate multiple humans and replicate human subject studies. In *Proceedings of the 40th international conference on machine learning.*

Ahler, D. J., Citrin, J., & Lenz, G. S. (2016). Do open primaries improve representation? an experimental test of california's 2012 top-two primary. *Legislative Studies Quarterly*, *41*(2), 237–268.

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic. *Journal of public economics*, *191*, 104254.

Anna Chorniy, L. S., Janet Currie. (2018). Exploding asthma and adhd caseloads: The role of medicaid managed care. *Journal of Health Economics*, *60*, 1-15.

Aranson, P. H., & Ordeshook, P. C. (1972). Spatial strategies for sequential elections. *Probability models of collective decision making*, *298*.

Argyle, L., Busby, E., Fulda, N., Gubler, J., Rytting, C., & Wingate, D. (2022). Out of one, many: Using language models to simulate human samples. *arXiv preprint arXiv:2209.06899*.

Arnesen, S., Duell, D., & Johannesson, M. P. (2019). Do citizens make inferences from political candidate characteristics when aiming for substantive representation? *Electoral Studies*, *57*, 46–60.

Association., A. P., & Association., A. P. (2013). *Diagnostic and statistical manual of mental disorders : Dsm-5* (5th ed. ed.). American Psychiatric Association Arlington, VA.

Bai, L., Gui, Z., Wei, L., & Xue, L. (2023). Strategic interactions with an algorithm assistant: The power of data and mechanism. *SSRN working paper no. 4286568*.

Barber, M., McCarty, N., Mansbridge, J., & Martin, C. J. (2015). Causes and consequences of polarization. *Political negotiation: A handbook*, *37*, 39–43.

Barberá, P. (2015). Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data. *Political analysis*, *23*(1), 76–91.

Bartels, L. M. (2002). The impact of candidate traits in american presidential elections. *Leaders' personalities and the outcomes of democratic elections*, 44–69.

Bauer, K., Liebich, L., Hinz, O., & Kosfeld, M. (2023). *Decoding gpt's hidden 'rationality' of cooperation.* (SAFE Working Paper No. 401)

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, *119*(1), 249–275.

Bishin, B. G., Stevens, D., & Wilson, C. (2006). Character counts? honesty and fairness in election 2000. *Public Opinion Quarterly*, *70*(2), 235–248.

Bonica, A. (2014). Mapping the ideological marketplace. *American Journal of Political Science*, *58*(2), 367–386.

Bonica, A. (2023). *Database on ideology, money in politics, and elections: Public version 3.0 [computer file].* Stanford, CA: Stanford University Libraries. (Available at: [https://data.stanford.edu/dime])

Bonneau, D. D., & Zaleski, J. (2021). The effect of california's top-two primary system on voter turnout in us house elections. *Economics of Governence*, *22*, 1–21.

Brand, J., Israeli, A., & Ngwe, D. (2023). Using gpt for market research. *Available at SSRN 4395751*.

Brenan, M. (2022). Congressional approval sinks to 18% as democrats sour further.

Brookins, P., & DeBacker, J. M. (2023). *Playing games with gpt: What can we learn about a large language model from canonical strategic games?* (Available at SSRN: 4493398)

Brown, J. R., Cookson, A. J., & Heimer, R. (2019). Growing up without finance. *Journal of Financial Economics*, *134*, 591–616.

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., . . . others (2020). Language models are few-shot learners. In *Advances in neural information processing systems* (Vol. 33, pp. 1877–1901).

Buttice, M. K., & Stone, W. J. (2012). Candidates matter: Policy and quality differences in congressional elections. *The Journal of Politics*, *74*(3), 870–887.

Bybee, L. (2023). Surveying generative ai's economic expectations. *arXiv preprint arXiv:2305.02823*.

Calcagno, P. T., & Westley, C. (2008). An institutional analysis of voter turnout: the role of primary type and the expressive and instrumental voting hypotheses. *Constitutional Political Economy*, *19*, 94–110.

Callaway, B., & Sant'Anna, P. H. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, *225*(2), 200-230.

Cancela, J., & Geys, B. (2016). Explaining voter turnout: A meta-analysis of national and subnational elections. *Electoral studies*, *42*, 264–275.

Charness, G., & Dufwenberg, M. (2006). Promises and partnership. *Econometrica*, *74*(6), 1579-1601.

Chen, D. L., Schonger, M., & Wickens, C. (2016). otree—an open-source platform for laboratory, online, and field experiments. *Journal of Behavioral and Experimental Finance*, *9*, 88–97.

Chen, L., & Ng, R. (2004). On the marriage of lp-norms and edit distance. In *Proceedings of the thirtieth international conference on very large data bases-volume 30* (pp. 792–803).

Chen, M., Tworek, J., Jun, H., Yuan, Q., Pinto, H., Kaplan, J., ... others (2021). *Evaluating large language models trained on code.* Working Paper.

Chen, Y., Liu, T. X., Shan, Y., & Zhong, S. (2023). The emergence of economic rationality of gpt. *Proceedings of the National Academy of Sciences*, *120*(51), e2316205120.

Cherry, T. L., & Kroll, S. (2003). Crashing the party: An experimental investigation of strategic voting in primary elections. *Public Choice*, *114*(3-4), 387–420.

Cohn, A., Gesche, T., & Maréchal, M. A. (2022). Honesty in the digital age. *Management Science*, *68*(2), 827–845.

Coleman, J. S. (1972). The positions of political parties in elections. *Probability models of collective decision making*, 332–57.

Cornelson, K., & Miloucheva, B. (2022). Political polarization and cooperation during a pandemic. *Health Economics*, *31*(9), 2025–2049.

Crosson, J. (2021). Extreme districts, moderate winners: Same-party challenges, and deterrence in top-two primaries. *Political science research and methods*, *9*(3), 532–548.

Currie, J., & Stabile, M. (2006). Child mental health and human capital accumulation: the case of adhd. *Journal of health economics*, *25*(6), 1094–1118.

Currie, J., Stabile, M., & Jones, L. (2014). Do stimulant medications improve educational and behavioral outcomes for children with adhd? *Journal of Health Economics*, *37*, 58-69.

Dalsgaard, S., Nielsen, H. S., & Simonsen, M. (2014). Consequences of adhd medication use for children's outcomes. *Journal of Health Economics*, *37*, 137-151.

Data, M. E., & Lab, S. (2017). U.s. house 1976–2020. Retrieved from https://doi.org/10.7910/DVN/IG0UN2

De Chaisemartin, C., & d'Haultfoeuille, X. (2020). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review*, *110*(9), 2964–96.

de Mello, C., Marsella, S., & Gratch, J. (2016). People do not feel guilty about exploiting machines. *ACM Transactions on Computer-Human Interaction*, *23*(2), 1–17.

Dias, N., & Lelkes, Y. (2022). The nature of affective polarization: Disentangling policy disagreement from partisan identity. *American Journal of Political Science*, *66*(3), 775–790.

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114-126.

Dobrosavljevic, M., Solares, C., Cortese, S., Andershed, H., & Larsson, H. (2020). Prevalence of attention-deficit/hyperactivity disorder in older adults: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*, *118*, 282-289.

Downs, A. (1957). An economic theory of political action in a democracy. *Journal of political economy*, *65*(2), 135–150.

Drori, I., Zhang, S., Shuttleworth, R., Tang, L., Lu, A., Ke, E., ... others (2022). A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level. *Proceedings of the National Academy of Sciences*, *119*(32), e2123433119.

Dvorak, F., Stumpf, R., Fehrler, S., & Fischbacher, U. (2024). *Generative ai triggers welfare-reducing decisions in humans.* (arXiv preprint arXiv:2401.12773)

Engel, C., Grossmann, M. R. P., & Ockenfels, A. (2024). *Integrating machine behavior into human subject experiments: A user-friendly toolkit and illustrations.* (Max Planck Institute Working Paper)

Fan, C., Chen, J., Jin, Y., & He, H. (2023). *Can large language models serve as rational players in game theory? a systematic analysis.* (arXiv preprint arXiv:2312.05488)

Fisk, C. A. (2020). No republican, no vote: Undervoting and consequences of the top-two primary system. *State Politics & Policy Quarterly*, *20*(3), 292–312.

Fisman, R., Paravisini, D., & Vig, V. (2017). Cultural proximity and loan outcomes. *American Economic Review*, *107*(2), 457–92.

Fletcher, J., & Wolfe, B. (2009). Long-term consequences of childhood adhd on criminal activities. *The journal of mental health policy and economics*, *12*(3), 119.

Gemma, P. (2017). *Voters lose in an open primary.* The Daily Caller. Retrieved from https://dailycaller.com/2017/03/31/voters-lose-in-an-open-primary/

Genicot, G. (2022). Tolerance and compromise in social networks. *Journal of Political Economy*, *130*(1), 94–120.

Geras, M. J., & Crespin, M. H. (2018). The effect of open and closed primaries on voter turnout. In *Routledge handbook of primary elections* (pp. 133–146). Routledge.

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, *14*(2), 627–660.

Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. *Journal of Econometrics*, *225*(2), 254-277.

Grose, C. R. (2020). Reducing legislative polarization: Top-two and open primaries are associated with more moderate legislators. *Journal of Political Institutions and Political Economy*, *1*, 1–21.

Guiso, L., Sapienza, P., & Zingales, L. (2004). The role of social capital in financial development. *American Economic Review*, *94*, 526–556.

Guiso, L., Sapienza, P., & Zingales, L. (2008). Trusting the stock market. *The Journal of Finance*, *63*, 2557–2600.

Guo, F. (2023). Gpt in game theory experiments. *arXiv preprint arXiv:2305.05516*.

Gurun, U. G., Stoffman, N., & Yonker, S. E. (2017). Trust busting: The effect of fraud on investor behavior. *The Review of Financial Studies*, *31*(4), 1341–1376.

Hagendorff, T. (2023). Machine psychology: Investigating emergent capabilities and behavior in large language models using psychological methods. *arXiv preprint arXiv:2303.13988*.

Harris, K., Immorlica, N., Lucier, B., & Slivkins, A. (2023). Algorithmic persuasion through simulation: Information design in the age of generative ai. *arXiv preprint arXiv:2311.18138*.

Henrickson, K. E., & Johnson, E. H. (2019). Increasing voter participation by altering the costs and stakes of voting. *Social Science Quarterly*, *100*(3), 869–884.

Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American economic review*, *92*(5), 1644–1655.

Horton, J. (2023). Large language models as simulated economic agents: What can we learn from homo silicus? *arXiv preprint arXiv:2301.07543*.

Hyndman, K., Wu, J., & Xiao, S. C. (2024). Trust and lending: An experimental study. *Quarterly Journal of Finance*, forthcoming.

Jones, D. R. (2001). Party polarization and legislative gridlock. *Political Research Quarterly*, *54*(1), 125–141.

Kondrak, G. (2005). N-gram similarity and distance. In *International symposium on string processing and information retrieval* (pp. 115–126).

Kosinski, M. (2023). Theory of mind might have spontaneously emerged in large language models. *arXiv preprint arXiv:2302.02083*.

Kousser, T., Phillips, J., & Shor, B. (2018). Reform and representation: a new method applied to recent electoral changes. *Political Science Research and Methods*, *6*(4), 809–827.

Krause, J., Krause, K.-H., Dresel, S. H., la Fougère, C., & Ackenheil, M. (2006). Adhd in adolescence and adulthood, with a special focus on the dopamine transporter and nicotine. *Dialogues in Clinical Neuroscience*, *8*(1), 29-36.

LaMothe, E., & Bobek, D. (2020). Are individuals more willing to lie to a computer or a human? evidence from a tax compliance setting. *Journal of Business Ethics*, *167*, 157–180.

Laudenbach, C., & Siegel, S. (2024). Personal communication in an automated world: Evidence from loan payments. *Journal of Finance*. (forthcoming)

Leng, Y., & Yuan, Y. (2023). *Do llm agents exhibit social behaviors.* (arXiv preprint arXiv:2312.15198)

Levenshtein, V. I., et al. (1966). Binary codes capable of correcting deletions, insertions, and reversals. In *Soviet physics doklady* (Vol. 10, pp. 707–710).

Lewis, J. B., Poole, K., Rosenthal, H., Boche, A., Rudkin, A., & Sonnet, L. (2022). Voteview: Congressional roll-call votes database. *https://voteview.com/*.

Loe, I. M., & Feldman, H. M. (2007). Academic and Educational Outcomes of Children With ADHD. *Journal of Pediatric Psychology*, *32*(6), 643-654.

Logg, J. M., Minson, J. A., & Moore, D. A. (2019). Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes*, *151*, 90-103.

Lopez, A. (2023). *The u.s. has a 'primary problem,' say advocates who call for new election systems.* National Public Radio. Retrieved from https://www.npr.org/2023/09/18/1199318220/nonpartisan-open-primaries-explainer

Lorè, N., & Heydari, B. (2023). *Strategic behavior of large language models: Game structure vs. contextual framing.* (arXiv preprint arXiv:2309.05898)

Ma, D., Zhang, T., & Saunders, M. (2023). *Is chatgpt humanly irrational?* (preprint)

Marzal, A., & Vidal, E. (1993). Computation of normalized edit distance and applications. *IEEE transactions on pattern analysis and machine intelligence*, *15*(9), 926–932.

McGann, A. J. (2002). The advantages of ideological cohesion: A model of constituency representation and electoral competition in multi-party democracies. *Journal of Theoretical Politics*, *14*(1), 37–70.

McGhee, E., Masket, S., Shor, B., Rogers, S., & McCarty, N. (2014). A primary cause of partisanship? nomination systems and legislator ideology. *American Journal of Political Science*, *58*(2), 337–351.

McGhee, E., & Shor, B. (2017). Has the top two primary elected more moderates? *Perspectives on Politics*, *15*(4), 1053–1066.

Milosh, M., Painter, M., Sonin, K., Van Dijcke, D., & Wright, A. L. (2021). Unmasking partisanship: Polarization undermines public response to collective risk. *Journal of Public Economics*, *204*, 104538.

Muñoz, M. M., & Meguid, B. M. (2021). Does party polarization mobilize or de-mobilize voters? the answer depends on where voters stand. *Electoral studies*, *70*, 102279.

Nagler, J. (2015). Voter behavior in california's top two primary. *California Journal of Politics and Policy*, *7*(1), 1–14.

Nakao, T., Radua, J., Rubia, K., & Mataix-Cols, D. (2011). Gray matter volume abnormalities in adhd: voxel-based meta-analysis exploring the effects of age and stimulant medication. *American Journal of Psychiatry*, *168*(11), 1154–1163.

Nokken, T. P., & Poole, K. T. (2004). Congressional party defection in american history. *Legislative Studies Quarterly*, *29*(4), 545–568.

OpenAI. (2022). *Chatgpt: Optimizing language models for dialogue.* Retrieved from https://openai.com/blog/chatgpt (Mimeo)

OpenAI. (2023a). *Chat completion.* Retrieved from https://platform.openai.com/docs/guides/chat (Mimeo)

OpenAI. (2023b). *Gpt-4 technical report.* Retrieved from https://doi.org/10.48550/arXiv.2303.08774 (Mimeo)

OpenAI. (2023c). *Prompt design.* Retrieved from https://platform.openai.com/docs/guides/completion/prompt-design (Mimeo)

Owen, G., & Grofman, B. (1996). Two-stage electoral competition in two-party contests: Persistent divergence of party positions with and without expressive voting. In *Conference on strategy and politics, center for the study of collective choice, university of maryland, college park md.*

Patterson Jr, S. (2020). Estimating the unintended participation penalty under top-two primaries with a discontinuity design. *Electoral Studies*, *68*, 102231.

Phelps, S., & Russell, Y. I. (2023). Investigating emergent goal-like behaviour in large language models using experimental economics. *arXiv preprint arXiv:2305.07970*.

Rogowski, J. C., & Langella, S. (2015). Primary systems and candidate ideology: Evidence from federal and state legislative elections. *American Politics Research*, *43*(5), 846–871.

Schmookler, A. (2017). *Open primaries: A wrong idea.* The Huffington Post. Retrieved from https://www.huffpost.com/entry/open-primaries-a-wrong-id_b_10471396

Schniter, E. (2024). *Human-robot interactions: Insights from experimental and evolutionary social sciences.* (ESI Working Papers)

Selten, R. (1967). Die strategiemethode zur erforschung des eingeschränkt rationalen verhaltens im rahmen eines oligopolexperimentes. In *Beiträge zur experimentellen wirtschaftsforschung.*

Serra-Garcia, M., & Gneezy, U. (2023). *Improving human deception detection using algorithmic feedback.* (Mimeo)

Shor, B., & McCarty, N. (2011). The ideological mapping of american legislatures. *American Political Science Review*, *105*(3), 530–551.

Sinclair, J. A. (2013). Of primary importance: American primary elections 1945-2012.

Stockemer, D. (2017). What affects voter turnout? a review article/meta-analysis of aggregate research. *Government and Opposition*, *52*(4), 698–722.

Sun, L., & Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. *Journal of Econometrics*, *225*(2), 175–199.

Thomsen, D. M. (2014). Ideological moderates won't run: How party fit matters for partisan polarization in congress. *The Journal of Politics*, *76*(3), 786–797.

Trachtman, S. (2019). Polarization, participation, and premiums: how political behavior helps explain where the aca works, and where it doesn't. *Journal of Health Politics, Policy and Law*, *44*(6), 855–884.

Vanberg, C. (2008). Why do people keep their promises? an experimental test of two explanations. *Econometrica*, *76*(6), 1467-1480.

Volklow ND, K. S., Wang G. (2009). Evaluating dopamine reward pathway in adhd: Clinical implications. *JAMA*, *302*(10), 1084-1091.

Webster, S. W., & Abramowitz, A. I. (2017). The ideological foundations of affective polarization in the us electorate. *American Politics Research*, *45*(4), 621–647.