ESSAYS IN BEHAVIORAL ECONOMICS

by

NATHAN R BIEMILLER

A DISSERTATION

Presented to the Department of Economics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2019

DISSERTATION APPROVAL PAGE

Student: Nathan R Biemiller

Title: Essays in Behavioral Economics

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Economics by:

Michael A. Kuhn	Chair
Benjamin Hansen	Core Member
Jiabin Wu	Core Member
Aaron Gullickson	Institutional Representative

and

Janet Woodruff-Borden	Vice Provost and Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2019

DISSERTATION ABSTRACT

Nathan R Biemiller

Doctor of Philosophy

Department of Economics

June 2019

Title: Essays in Behavioral Economics

I use applied and experimental methods to empirically investigate individual decisions in a number of settings where behavior may be more consistent with models from behavioral economic theory than from traditional economic theory. These situations provide us with insights about real-world situations in which behavioral models may be more applicable than traditional models. In Chapter II, I find that losing candidates in U.S. House of Representatives elections are more likely to run in the subsequent election if they outperform their expectations relative to opponents. Chapter III estimates the effect of opening recreational marijuana markets on domestic violence; I find that opening markets increases reported domestic violence incidents in treated states. Finally, in Chapter IV, which is co-authored with Michael Kuhn and Jeffrey Naecker, we construct an experiment to test individual commitment demand in a group framework.

This dissertation includes previously unpublished co-authored material.

CURRICULUM VITAE

NAME OF AUTHOR: Nathan R Biemiller

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
Franklin & Marshall College, Lancaster, PA

DEGREES AWARDED:

Doctor of Philosophy, Economics, 2019, University of Oregon
Master of Science, Economics, 2015, University of Oregon
Bachelor of Arts, Mathematics and Economics, 2012, Franklin & Marshall College

AREAS OF SPECIAL INTEREST:

Behavioral Economics
Applied Microeconomics
Econometrics

GRANTS, AWARDS AND HONORS:

Kleinsorge Summer Research Fellowship, University of Oregon, 2017

Kleinsorge Fellowship Award, University of Oregon, 2014-2015

John Marshall Scholarship, Franklin & Marshall College, 2008-2012

ACKNOWLEDGEMENTS

I would like to thank my advisor, Mike Kuhn, for his invaluable support; my committee, for their advice and insights; my friends, for making graduate school infinitely more enjoyable; and the University of Oregon Department of Economics.

To my parents, for giving me the best childhood anyone could ask for.

TABLE OF CONTENTS

LIST OF FIGURES

Figure Page

LIST OF TABLES

Table Page

Table Page

CHAPTER I

INTRODUCTION

Behavioral economics, generally, is concerned with predictable deviations from traditional economic theory. In this dissertation, I investigate a number of situations in which individuals make decisions consistent with predictions from models in the behavioral economics literature.

In Chapter II, I examine how past experiences influence present decisions. I model U.S. House of Representatives candidates expectations of their vote share outcome based on their mean poll share. Assuming that these expectations are unbiased, candidates receive a signal of unexpected value when election results are realized. I find that candidates are unresponsive to the strength of the signal, but that candidates who lose their elections exhibit a discontinuity in the probability of running in the subsequent election around the zero-signal value. Losing candidates either believe that their probability of winning the next election is larger following a positive signal, or these candidates receive some amount of utility from the positive signal and account for that utility in future decisions.

In Chapter III, I find that reported domestic violence incidents increase when recreational marijuana markets open in Colorado, Washington, and Oregon. Since 2014, ten states and the District of Columbia have legalized the recreational use and sales of marijuana. Past literature makes contradictory predictions about the expansion of marijuana use on crime. I use a difference-in-differences model and high-frequency crime data to investigate the causal effect the opening of the recreational marijuana market has on domestic violence in Colorado, Washington, and Oregon, which were the first three states to begin recreational sales. I find that the beginning of recreational sales leads to a 2.9-6.2% increase in domestic violence incidents in these states; there is no effect of legalization, which occurs prior to the start of sales. The least severe category–intimidation, which does not involve physical harm to the victim–increases by the largest percentage. The closest counterfactual crime, non-intimate-partner in-home assaults, remains unchanged when recreational sales begin. While there are multiple mechanisms through which this effect could act, additional evidence is most consistent with an increase in within-household conflict.

In Chapter IV, I investigate group commitment demand in an unpublished work co-authored with Michael Kuhn and Jeffrey Naecker. When individuals are aware that their future preferences differ from their current preferences, they may seek commitment devices to ensure that their future selves take the proper action. There is evidence in the literature of real-world situations in which individuals seek commitment, from decisions about smoking to choices about retirement savings contributions. In this paper, we investigate commitment in group settings by conducting an experiment in which subjects perform a task, indicate that they would like to perform the same task again in the future, and then have the opportunity to commit to that future task by decreasing the future payout in the event that the task does not get completed. We randomly assign participants into either an individual commitment choice or one of a number of group commitment frameworks. We find that commitment demand is much larger in the individual treatment (42% of the budget) than in any group treatment (10% of the budget), and we hypothesize that this may be due to individuals perceiving higher follow-through rates in group treatment settings.

CHAPTER II

THE POWER OF THE POLLS: THE ROLE OF EXPECTATIONS IN INTERPRETING ELECTION RESULTS

Introduction

Consider two political candidates, Alice and Bob, who run for the same office in different districts in the same election. Both candidates are underdogs in polls conducted before the election. Specifically, suppose that Alice polls at 40% of the projected vote share, while her main opponent sits at 60%. Bob expects to receive a higher portion of the vote than Alice: he polls at 44%, while his main opponent sits at 56%. Prior to the election, these candidates have different expectations about their relative performance. Alice should expect to lose her election by 20 percentage points, while Bob should expect to lose his election by 12 percentage points.

On election night, the local television stations broadcast the results: unsurprisingly, both Alice and Bob lose their elections. Each candidate's main opponent receives exactly 58% of the total vote share in the respective districts. Alice slightly outperforms her expectations relative to her opponent by receiving 42% of the vote share, while Bob's 42% of the vote share comes in just below his expectations about his relative performance.

Assuming that the election results provide a true measure of candidate quality, these two candidates are identical, as they each received 42% of the vote. As described, they differ in just one important way: Alice received a positive signal about her performance relative to her expectations of the outcome, while Bob received a negative signal about his performance relative to expectations. In the future, each of these candidates must decide whether to run for this office again in the next election cycle. This paper attempts to determine whether the valence of this signal received by a candidate affects the probability that the candidate runs for the same office in the subsequent election. This is closely related to research on biased information processing and updating.

Eil and Rao (2011)–henceforth referred to as "ER"–conduct a lab experiment to test the following hypothesis about individual updating behavior: when faced with signals regarding attributes about themselves, subjects will respond more to favorable news than to unfavorable news. This type of updating behavior is intuitive. In many facets of life, individuals receive both positive and negative signals about themselves, and it may be the case that individuals process and update differentially based on whether the signal received represents "good news" or "bad news." A compliment from a stranger could cause an individual to think, "That person doesn't even know me, but she thinks highly of me. It's very likely that I truly possess this positive quality." Conversely, following an insult, an individual might think, "That person doesn't even know me, and she made this horrible comment. Why would I believe that what she said has any merit at all?"

ER find that when facing signals about their own attributes, individuals differentially update their beliefs about these attributes depending on the valence of the signal received. In tasks related to individuals' attractiveness and intelligence, individuals who received positive signals updated their beliefs more strongly than individuals who received negative signals. The authors call this the "good news-bad news effect."

This paper is motivated by ER's experiment: does the good news-bad news effect generalize outside the lab? There are a variety of important contexts in which individuals could display this type of behavior: high school students updating the distribution of colleges to which they apply after receiving an SAT or ACT score; homeowners updating their valuations of their houses following a signal about a change in the home's risk of catastrophe; or ill patients deciding whether to receive treatment following a diagnosis about a particular health condition. In each case, the signal informs the individual about the state of the world relative to the individual's expectations.

I investigate how the valence of the signal received by United States House of Representatives candidates affects candidate decisions to run for the same office in the next election. This situation provides a context in which individuals plausibly have an observable, unbiased expectation, subsequently receive an unpredictable signal relative to that expectation, and finally make a relevant, related decision in the future.

At the time the election takes place, each candidate has an expectation of his or her vote share margin. I assume that candidates formulate their expectations of their vote shares as the difference between their mean poll share prior to the election and their main opponent's mean poll share prior to the election. When the election occurs, each candidate experiences a signal about his or her performance in the form of the actual vote share received. The signal received by the candidate is given by

$$(VS_C - VS_O) - (PS_C - PS_O),$$

where VS_C is the candidate's realized vote share, VS_O is the opponent's realized vote share, PS_C is the candidate's mean poll share, and PS_O is the opponent's mean poll share. While this signal is not truly random, since it is determined by the behavior of the electorate, the candidate is unable to predict the value or the valence of the signal. After experiencing this signal in the current election, candidates must decide whether to run for the same office in subsequent elections.

The hypothesis derived from the good news-bad news effect is that candidates update their beliefs about the probability of winning the next election in a specific way: those candidates receiving positive signals would "properly" update their beliefs about winning a future election, while those candidates receiving negative signals would fail to update their beliefs about the probability of winning as far down as they should. Since there are two components—beliefs about the future probability of winning or losing an election and the predicted utilities derived from winning or losing the future election—of the expected utility calculation, the good news-bad news effect is not the only potential theoretical underpinning of candidates' decisions to run in a subsequent election, however. Haggag et al. (2016)–henceforth HP–describe a bias caused by selective memory that could affect candidates' projected utility of winning or losing: attribution bias occurs when the state of the world in which an individual originally experienced a certain type of consumption affects the individual's prediction about the utility derived from future consumption of the same type. This is similar to projection bias, which is described by Loewenstein et al. (2003), in which the current state of the world influences individuals' expectations about the utility they will derive from future consumption of the same type. In both of these models, individuals may be aware that the future state of the world could differ and still fail to fully account for this factor. This type of behavior is consistent with the model of misattribution of reference-dependence presented by Bushong and Gagnon-Bartsch (2016), in which individuals fail to account for the extent to which their own reference-dependence affected their memories of an experience.

How attribution bias occurs depends on the shape of the experienced utility function. Candidates could have some variant of reference-dependent preferences, where the utility function exhibits a kink or slope change at a specific reference point. If candidates derive some utility from the value of the signal received, there could be changes in the effect of the signal on the probability of running again or discontinuities in the probability of running in the future at a reference point, especially if candidates experience attribution bias. A candidate who loses while outperforming expectations must predict her utility of running in the next election. If her utility is higher in the outperforming-expectations state of the world than in the state in which she underperforms, she may fail to account for the difference in the utility she would receive in each state of the world. In this way, not experiencing a loss in the underperforming-expectations state of the world could cause her to overestimate her utility of running in the future. Similarly, if she mistakenly fails to account for the reference-dependent nature of her preferences, she may overestimate the future utility of running after previously experiencing a positive signal. In both cases, candidates who outperform expectations fail to properly predict their utilities of running in the future, which causes these candidates to re-run more frequently than candidates who underperform relative to expectations.

Using polling and outcome data for U.S. House elections from 2002, 2006, 2008, 2010, 2012, and 2014, I estimate the probability that a candidate runs for office in the subsequent election as a function of the signal received relative to the expectation using a regression discontinuity approach, where the signal received is the running variable.

For candidates who won their elections, I estimate a small (5.2 percentage points) but statistically insignificant positive discontinuity in the probability of running in the subsequent election at a signal value of zero. For winners, the estimated slope of the relationship between the signal received and the probability that the candidate runs in the next election is statistically insignificant from zero. This is true both for winners who receive a positive signal and for winners who receive a negative signal. Together, these results imply that when candidates win an election, they pay little attention to the value of the signal received relative to expectations. Once a candidate has won an election, it is likely that other factors, such as job satisfaction or approval rating, play a much larger role in the decision to run in the subsequent election. Unsurprisingly, the mean probability of running in the subsequent election for winners (82.9%) is over 60 percentage points higher than the mean probability of running in the subsequent election for losers (18.0%).

Candidates who lost their elections, on the other hand, display a larger and statistically significant discontinuity in the probability of running in the next election at a signal value of zero. For these candidates, receiving a positive signal instead of a negative signal increases the probability of running in the subsequent election by 17.7 percentage points (in the full sample). While this estimated effect may seem quite large from an intuitive standpoint, much of the effect seems to result from underperforming candidates failing to run again. As shown in Figure 8, the estimated probability of running in the next election for losers who received negative signals is very close to zero for all negative signal values. It appears that receiving any negative signal relative to expectations causes candidates to decide not to run in the subsequent election.

The size of this estimated discontinuity is larger in subsamples in which the signal bandwidth is restricted to be closer to the hypothesized reference point of a zero signal. As is the case for the winning candidates, I do not estimate a positive slope of the relationship between the signal value and the probability to run in the next election. Losers with a negative signal exhibit a slope that is statistically indistinguishable from zero, while losers with a positive signal have a negative slope. This estimated negative slope is unexpected; one potential explanation is that those candidates who outperform their expectations by large amounts decide to run for more prestigious offices in the future, such as Senate or gubernatorial seats, rather than again attempting to win a House seat.

In the following section of the paper, I discuss the literature and theory; section 3 presents the model and the data sources; I discuss the estimation results in section 4; and section 5 concludes the paper.

Literature and Hypotheses

The literature that is most relevant to this question comes from two distinct categories: behavioral economics' models of reference-dependent preferences and political science's investigation of elections using regression discontinuity techniques. While this paper is motivated by testing the good news-bad news effect described by ER, various other models of utility and beliefs provide different predictions about the behavior of individuals around a given reference point.

Behavioral economics

In a simple expected utility setup, the candidate has beliefs about the probability of winning an election and a utility function governing how she feels about winning or losing. Together, these provide an expectation of the amount of utility she would receive from running in the next election. If this expected utility is greater than the utility she would receive from not running in the election, she will choose to run. Performance relative to expectations could affect this expected utility through the candidate's beliefs or through the candidate's utility of winning or losing. An individual could derive utility from winning with a larger signal value or losing with a larger signal value; the individual could also experience some utility irrespective of the outcome of the election. If candidates derive utility from their performance relative to expectations, traditional economic theory predicts that this utility function is smooth and upward-sloping, with no discontinuities or kinks at any reference points, like their expectations.

Reference-dependent preferences

Politicians clearly possess expectations about election outcomes. Behavioral economics has produced a variety of formulations of utility functions that depend on individuals' reference points. Kőszegi and Rabin (2006) develop a model of reference-dependent preferences that includes gain-loss utility, where an individual's overall utility is comprised both of consumption utility and of a utility term capturing the effect of a "gain" or a "loss" relative to a reference point. In their framework, individual utility is given by

$$u(c|r) = m(c) + n(c|r),$$

where c is the consumption bundle, r is the reference point, $m(c)$ is the standard consumption utility, and $n(c|r)$ is the gain-loss utility derived from the position relative to the reference point. The gain-loss utility is modeled as $n(c|r) = \mu(m(c) - m(r))$, where $\mu(\cdot)$ satisfies the properties of the value function from Kahneman and Tversky (1979).

If candidates derive utility from their performance relative to expectations and exhibit value functions like the one hypothesized by Kőszegi and Rabin, their expected utility of running for office in the next election would incorporate an additive gain-loss term capturing the additional value of the signal received. In the data, this would manifest as a positive slope of the relationship between signal value and probability of re-running for those in the positive-signal domain and a positive and steeper slope for those in the negative-signal domain. Additionally, if the slopes are decreasing as the magnitude of the signal increases, this would be evidence of the concavity in gains and convexity in losses described in the standard Kahneman-Tversky loss-aversion model.

Authors have shown that reference-dependent preferences matter in various empirical settings. Some recent examples include Allen et al. (2016), who show that marathon runners exhibit reference-dependence around even finishing times by bunching just under these hypothesized reference points. DellaVigna et al. (2017) write a job search model that incorporates reference-dependent preferences relative to an individual's most recent income level and present evidence from Hungary to argue that the reference-dependent model fits observed hazard rates better than alternative models. Rees-Jones (2017) shows that there is excess mass of tax returns at a value of 0, and he argues that this is evidence that tax filers have a reference point at zero tax liability. Tax filers who owe additional payments are in the loss domain, and these individuals reduce their tax liability more than individuals who are owed a refund.

Another form of reference-dependent preferences comes from Diecidue and Van De Ven (2008), who write a model of utility derived from attaining an aspiration level. They hypothesize that individuals receive some utility from achieving a specified level of wealth, which would cause a discontinuous jump in the utility function at the given aspiration level. The aspiration-level expected utility function is similar to the expected utility function from Kőszegi and Rabin. However, in a subsequent paper (Diecidue et al., 2015), the authors conduct a lab experiment and fail to find evidence of this type of behavior, although they mention that this lack of evidence could be a function of individuals having heterogeneous aspiration levels. If I estimate a discontinuity at the reference point, it would provide evidence that candidates possess this sort of utility function.

While candidates may truly derive gain-loss utility from their performance relative to expectations, they may also misattribute the utility of the election when making a decision about the next election. Haggag et al. (2016) show evidence of attribution bias, which occurs when individuals' predicted utility of future consumption depends upon the state in which the individual previously experienced that type of consumption. This is similar to the model of projection bias introduced by Loewenstein et al. (2003). Attribution bias exists if an individual's predicted utility of consumption in a given state lies between the utility derived from the consumption in the initial state and the true utility derived from consumption in the new state. Formally, an individual's predicted utility of consuming c in state s_t is given by

$$\tilde{u}(c, s_t) = (1 - \gamma)u(c, s_t) + \gamma u(c, s_{t-1})$$

for $\gamma \in [0, 1]$.

When making an expected utility calculation to decide whether to run for office again, candidates must predict the utility they will receive for each potential election outcome. Candidates could exhibit attribution bias: in the current election, candidates experience the election outcome only in one state of the world. No candidate experiences the election outcome while receiving both a positive and a negative signal. Losing candidates who outperform their expectations may predict that the utility they would derive from losing in time $t + 1$ is exactly the same as the utility they actually experienced from losing in time t without realizing that the positive signal they received mitigated their unhappiness about losing. If this is the case, losers receiving positive signals would predict higher levels of expected utility of running in the future, which would lead to this group of candidates running more frequently in the subsequent election than the losing candidates who underperformed expectations.

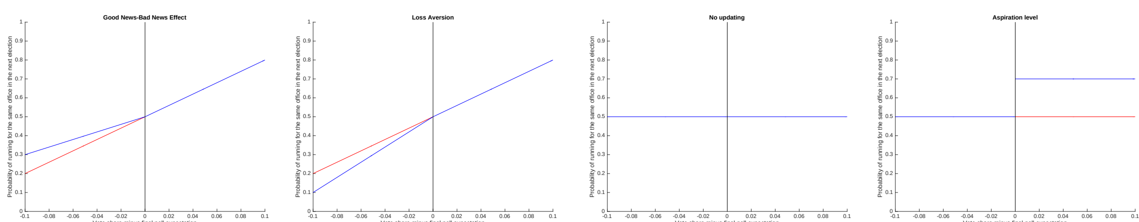
While theoretical models discuss behavior around reference points, a key empirical challenge is determining the proper reference point used by the individuals making the decision of interest. As Barberis (2013) discusses in his overview of prospect theory, determining this reference point is one of the major difficulties in investigating reference-dependent preferences. Though there is no guarantee that the signal value of zero is the correct reference point for candidates, this makes more intuitive sense than other candidate reference points. When evaluating performance, candidates must compare outcomes to some type of expectations. I assume that these expectations are derived from public polls, which play a significant role in elections by informing both candidates' campaigns and voters about the electorate's sentiments about candidates and issues. While campaigns conduct private polls, I am unable to obtain these to use as candidates' expectations in the model. As such, I assume that public polls provide an unbiased estimate of a candidate's expected vote share prior to the election occurring.

Good news-bad news effect

ER conduct a lab experiment in which subjects are ranked on attractiveness and intelligence among a group of ten participants. For attractiveness, these ranks are determined by a speed-dating exercise and subsequent questionnaire; for intelligence, these ranks are determined by an IQ test. The individuals do not know their true ranks. For each attribute, subjects are then asked to specify prior beliefs about what ranking they occupy in the distribution. After specifying prior beliefs, subjects receive a signal telling them either that they are ranked higher or lower in the distribution than another randomly selected participant. Finally, the subjects again specify their beliefs about where they rank in the distribution of participants. This framework allows ER to assess how participants differentially update beliefs as a result of receiving "good news" or "bad news." In the attractiveness task, the authors find that subjects "properly" update in a Bayesian sense when receiving positive signals and fail to update as much as they should when receiving negative signals. A similar effect occurs in the intelligence task, although the difference in slopes is not statistically significant. In a control task with no relevance to personal attributes, the way in which subjects receiving positive and negative signals updated was indistinguishable. The authors call this sort of behavior the "good news-bad news effect."

If the good news-bad news effect holds in the context of elections, those candidates who receive positive shocks—vote shares above expectation—properly update their expectations about the decision to run for the same office in the future. However, those who receive negative shocks do not fully update downward. Graphically, when plotting the probability of running for the same office against the valence of the shock received, this effect would appear as a kink at 0 in the best-fit lines, with the negative-shock best-fit line having a flatter slope than the positive-shock best-fit line. Intuitively, this would mean that while candidates properly internalize positive information about their chances of getting elected, they fail to properly internalize negative signals. Figure 1 depicts the relationship between the signal received and the probability of re-running for each of the hypotheses.

FIGURE 1.
Hypothesized potential effects



(a) Good news-bad news effect (b) Loss aversion (c) No updating (d) Aspiration level

Political science

In political science, the application of the regression discontinuity approach to elections was pioneered by Lee (2008), who uses a regression discontinuity design to investigate U.S. congressional elections; he argues that in close elections, candidates are unable to sort into the positive side of the election-winning vote share threshold in order to win the election. He implements this discontinuity design to investigate the incumbency effect: Democratic candidates who barely won an election at time t are much more likely to win an election at time $t+1$ than Democratic candidates who barely lose an election at time t. Unsurprisingly, he also finds a significant discontinuity in the probability a candidate runs in an election at time $t+1$ at the zero margin-of-victory cutoff for the election at time t. Candidates who barely win an election in time t are much more likely to run in time $t+1$.

The literature is fairly clear that an incumbency advantage exists. Levitt and Wolfram (1997) investigate the incumbency advantage in congressional elections, which increased from approximately 3.4% in the 1950s to approximately 8.0% in the 1980s, conclude that most of the incumbency advantage comes from candidates' abilities to deter high-quality challengers, instead of from direct benefits associated with holding office. Ashworth and Bueno de Mesquita (2008) model the incumbency advantage as a result of the fact that incumbents have greater ability than challengers, which stems from two factors: high-ability candidates are more likely to get elected, and high-quality incumbents are more likely to deter challengers. While I do not specifically investigate the incumbency advantage, I do find that candidates who win elections in time t are significantly more likely to run in elections in time $t+1$, which is consistent with the literature.

Caughey and Sekhon (2011) dispute Lee's use of the regression discontinuity design by claiming that those candidates who barely win elections are significantly different in certain covariates than those candidates who barely lose elections; they use this evidence to claim that candidates are able to sort across the victory threshold in elections, which would invalidate the use of a regression discontinuity design in the context of elections. However, Eggers et al. (2015) analyze a larger sample of elections and show that this type of sorting occurs only in U.S. House elections after 1946; all other elections in their sample do not exhibit this feature. As such, they discuss the possibility that this perceived sorting occurs only due to random variation. The regression discontinuity design in this paper does not require that candidates not sort across the election-winning threshold; instead, we assume that candidates do not sort across the zero-signal threshold.

Another implementation of the regression discontinuity design in elections comes from Anagol and Fujiwara (2016), who investigate the effect of being the runner-up in an election instead of finishing third. Using election data from Brazil, India, and Canada, the authors find that being the runner-up confers a significant positive effect on both the probability to run in future elections and the probability to win future elections. As in their paper, I investigate the effect of crossing a threshold on the probability of running in the subsequent election. In both papers, the threshold across which candidates may not sort occurs away from the election-winning threshold disputed in the literature: Anagol and Fujiwara assume that candidates do not sort across the second-third threshold, while I assume that candidates do not sort across the zero-signal threshold. Diermeier et al. (2005) write a political economy model of congressional careers, where they attempt to quantify an individual's returns to a congressional career. In the model, they estimate the probability of winning an election, and include this estimated probability in the candidate's decision to run for office or pursue other opportunities. However, in the model of the probability of winning the election, the authors do not include information about vote share or vote share relative to polling data. In this paper, I model candidates' decisions to run again as a function of their performance relative to expectations, and I investigate how these decisions differ relative to a meets-expectations reference point.

Methodology and Data

Data

I obtain data on the true vote share received by candidates in elections from the Federal Election Commission. I also use this FEC data to generate the variable indicating whether the candidate ran in the subsequent election. Following the approach in Downey (2017) to gathering polling data, I collect data for each published poll available from www.realclearpolitics.com. I also enter demographic data, such as gender, race, party affiliation, and incumbent status from the polling data webpages. These data include the name and date of the poll, and, most importantly, the poll share received by each candidate.

Unfortunately, the relative dearth of polling data is the limiting factor in the data analysis. While there are 435 U.S. House seats up for election every two years, the majority of the races are won by such a large margin that polls are not conducted. According to ballotpedia.org, the average margin of victory in House races was 35.8% in 2014 and 37.1% in 2016. Only 16 of 2016's 435 House races were decided by a margin smaller than five percentage points. Generally, polls are conducted only for races that could plausibly be considered close prior to the election. I use all available polling data from realclearpolitics.com in my analysis. The sample includes data from 321 polled elections spanning the years 2002, 2006, 2008, 2010, 2012, and 2014. Of these 321 elections, all but two include poll share only for a Democrat and a Republican. In the other two races, a third party candidate received a poll share. This means that the sample consists of approximately 54 polled elections per election year, or that I have polling data from, on average, 12.3% of U.S. House elections in this time period. As a result, the margins in the elections used in the sample for this paper are necessarily closer than the margins in an average U.S. House election.

Summary statistics for both groups of candidates are displayed in Table 1. Winners—those candidates who become incumbents heading into the subsequent election—run in the next election approximately four-and-a-half times as frequently as losers. Each party comprises approximately half of each group.

TABLE 1.
Summary statistics

	Winners				**Losers**			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Ran in next election	.829	.377	0	1	.180	.384	0	1
Signal	.0319	.0646	-.199	.201	-.0326	.0636	-.201	.199
Female	.153	.360	0	1	.238	.427	0	1
Black	.0156	.124	0	1	.0186	.135	0	1
Democrat	.486	.501	0	1	.508	.501	0	1
Observations	321				323			

In absolute value, the largest signals received are close to .20: no candidate in our dataset outperformed or underperformed expectations by more than 20 percentage points. While 20 percentage points represents a sizable deviation from the poll, the majority of election results fall relatively close to the poll shares.

Approximately 85% of observations (544 out of 641) receive a signal within 10 percentage points of zero, which means that the true election outcome was within 10 percentage points of the polled outcome. As expected, the mean signal for winners is positive, while the mean signal for losers is negative. The slight difference in absolute value of the mean of the signal for the groups is a result of the few third-party candidates who are additional losers in specific races. For third-party candidates, the "main opponent" is defined as the opponent who received the highest mean poll share prior to the election; I assume that these candidates use the leader of the race as the proverbial measuring stick.

Model

I conceptualize a simple model of a candidate's decision to run for a U.S. House seat in time period $t + 1$ after having run for office in time period t. I assume that candidates take their mean poll share relative to their main opponent's mean poll share as their expectation of the vote share they will receive in the election; I assume also that candidates process their true vote share relative to their opponent's vote share as an objective signal about their desirability as a candidate. With these two assumptions in mind, the candidate's signal in time period t is:

$$(VS_C - VS_O) - (PS_C - PS_O),$$

where VS_C is the candidate's realized vote share, VS_O is the opponent's realized vote share, PS_C is the candidate's mean poll share, and PS_O is the opponent's mean poll share. After receiving the signal in time period t, the candidate must make a decision to run in time period $t+1$.

Empirical model

I use a regression discontinuity approach to estimate the effect of receiving a positive signal relative to receiving a negative signal for both winners and losers. One key facet of a regression discontinuity design is that individuals must not be able to sort across the treatment threshold. As previously discussed, this assumption has been contested in the context of election results; Caughey and Sekhon (2011) argue that candidates are able to sort across the election-winning threshold, while Lee (2008) and Eggers et al. (2015) argue that this type of behavior is not possible.

In this framework, it is not important if candidates sort across the election-winning threshold, but candidates must not sort across the zero-signal threshold. Given how difficult it would be to sort across the election-winning threshold, it is highly unlikely that candidates would be able to sort across the zero-signal threshold. Even if this type of sorting were possible, candidates should be much more interested in sorting across the election-winning threshold than they are in sorting into a region where they receive a positive signal.

To formally test that no sorting occurs near the treatment, I implement both a McCrary (2008) test and a similar test of the form described by Cattaneo et al. (2018) designed to detect discontinuities in the density of the running variable. The McCrary test discontinuity estimate of the log difference in the density function at the zero-signal value is $-7.64\mathrm{e}^{-6}$, and the p-value for the Cattaneo, Jansson, and Ma test is 0.986. Both of these tests suggest that no sorting occurs at the zero-signal threshold. Figure 2 and Figure 3 present the kernel density of the signal received for winners and losers, respectively. With no sorting present, this necessary regression discontinuity assumption is satisfied.

FIGURE 2.
Signal density for winners

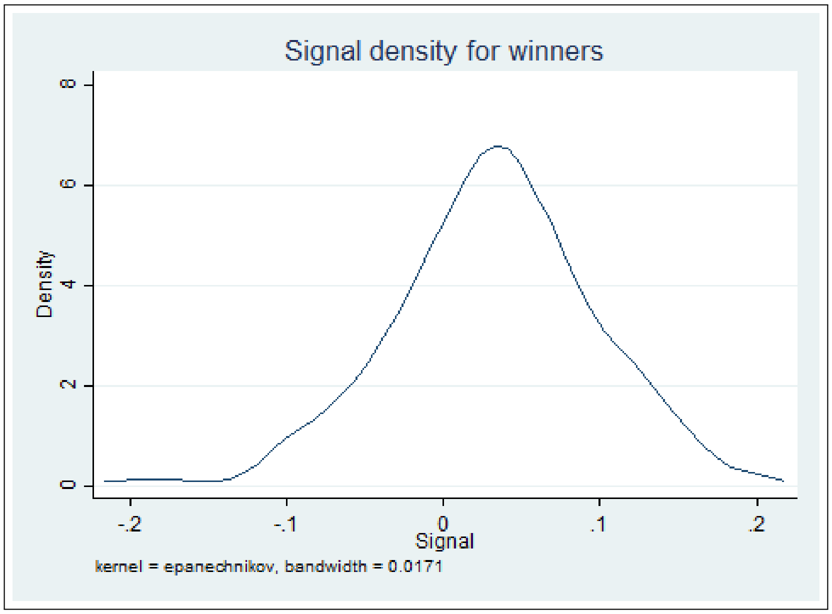
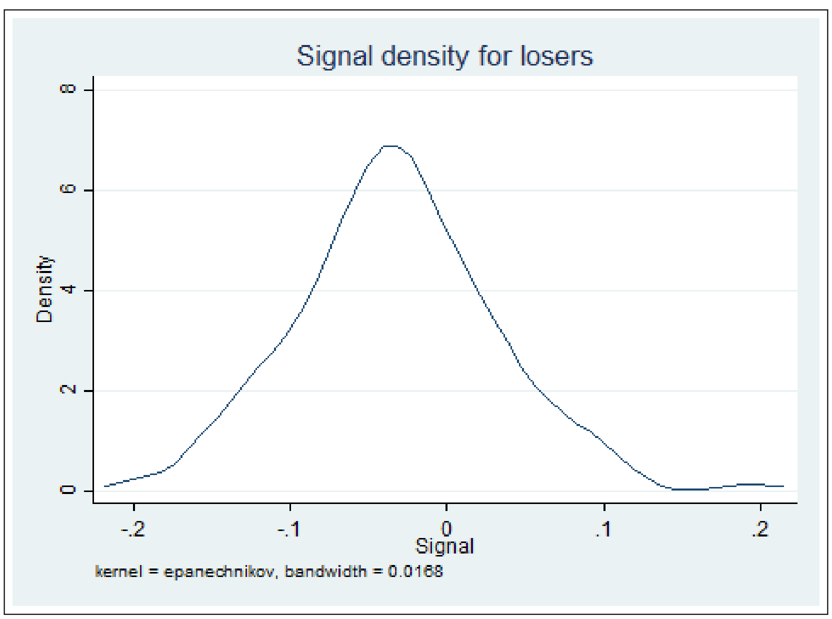


FIGURE 3.
Signal density for losers



- The base model estimates the linear probability model:

$$
\begin{aligned}
Pr(Y_i = 1) =& \alpha + \beta_1 S_i + \beta_2 W_i + \beta_3 P_i + \beta_4 W_i S_i \\
&+ \beta_5 P_i W_i + \beta_6 P_i S_i + \beta_7 P_i W_i S_i + X_i \gamma + \epsilon_i,
\end{aligned}
$$

where

- Y_i is a dummy variable equal to one if the candidate runs in the subsequent election and equal to zero otherwise.

- S_i is the value of the signal received by the candidate.

- W_i is a dummy variable equal to one if the candidate wins the election and equal to zero otherwise.

- P_i is a dummy variable equal to one if $S_i > 0$ and equal to zero otherwise.

- X_i are controls, such as party, race, gender, and incumbency status.

The form of the estimation allows for easy interpretation of coefficients. The interaction terms allow for the estimation of differentiated behavior between groups. The coefficients can be interpreted as follows: the coefficient α is the predicted probability of running again for losers who receive a negative signal just to the left of zero. The sum $\alpha + \beta_3$ is the predicted probability of running again for losers who receive a positive signal just to the right of zero. The sum $\alpha + \beta_2$ is the predicted probability of running again for winners who receive a negative signal just to the left of zero. The sum $\alpha + \beta_2 + \beta_3 + \beta_5$ is the predicted probability of running again for winners who receive a positive signal just to the right of zero. The sum β_1 is the effect of the strength of the signal on the probability of running again for losers who receive a negative signal. The sum $\beta_1 + \beta_6$ is the effect of the strength of the signal on the probability of running again for losers who receive a positive signal. The sum $\beta_1 + \beta_4$ is effect of the strength of the signal on the probability of running again for winners who receive a negative signal. The sum $\beta_1 + \beta_4 + \beta_6 + \beta_7$ is effect of the strength of the signal on the probability of running again for winners who receive a positive signal. The sum β_6 represents the difference between winners and losers in the effect of signal strength on the probability of running again. This coefficient is relevant to various hypotheses: if the probability of running again increases as signal strength becomes more positive, β_6 can provide evidence of the good news-bad news effect (when $\beta_6 > 0$) or of loss aversion (when $\beta_6 < 0$).

Results

Main estimates

Table 2 presents estimates of the model by signal bandwidth; I restrict the sample to those candidates receiving a signal within 15, 10, and 5 percentage points of zero. Slightly more than half of the observations (53.6%) have a signal within 5 percentage points of zero. Due to the nature of which elections get polled, the dataset is already a subset of elections that are closer than the average election; these sample restrictions further concentrate on specifically close elections. Approximately 34.6% of observations are included in the within-five-percentage-points subsample. In all estimates presented, I cluster the error term at the candidate level (519 clusters), although clustering at the state level does not produce substantively different results. All estimates include year fixed effects with a base year of 2002.

TABLE 2.
Mean poll share by signal bandwidth, clustered by candidate

	Full sample	Within 15%	Within 10%	Within 5%
Winner	0.632***	0.613***	0.591***	0.629***
	(7.15)	(6.56)	(5.93)	(4.68)
Signal	-0.0786	-0.218	-1.106	-2.568
	(-0.16)	(-0.38)	(-1.13)	(-0.85)
Winner * Signal	-0.276	-0.713	-0.761	0.110
	(-0.22)	(-0.48)	(-0.46)	(0.02)
Positive	0.177**	0.189**	0.258***	0.200
	(2.29)	(2.16)	(2.80)	(1.46)
Winner * Positive	-0.125	-0.112	-0.181	-0.121
	(-1.15)	(-0.95)	(-1.47)	(-0.68)
Positive * Signal	-2.095**	-2.150	-2.713*	4.168
	(-2.07)	(-1.45)	(-1.72)	(0.83)
Winner * Positive * Signal	1.786	2.239	4.474*	-1.242
	(1.07)	(1.05)	(1.94)	(-0.19)
Vote differential	0.621**	0.693**	0.731**	0.878***
	(2.07)	(2.22)	(2.04)	(2.82)
Female	0.0579	0.0585	0.0377	0.0638
	(1.48)	(1.48)	(0.87)	(1.11)
Asian	-0.168	-0.185	-0.199	-0.0707
	(-1.04)	(-1.03)	(-0.97)	(-0.24)
Black	0.108	0.118	0.126	0.228
	(0.80)	(0.80)	(0.84)	(0.71)
Democrat	-0.0325	-0.0289	-0.0171	-0.0366
	(-1.10)	(-0.95)	(-0.50)	(-0.84)
Incumbent winner	-0.178***	-0.186***	-0.153***	-0.159***
	(-3.86)	(-4.02)	(-3.23)	(-2.70)
Incumbent loser	-0.0162	-0.0160	-0.0353	0.0184
	(-0.33)	(-0.32)	(-0.64)	(0.25)
Constant	0.218***	0.212***	0.198***	0.146
	(3.69)	(3.35)	(2.74)	(1.30)
Observations	644	623	541	345

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As expected, winners are much more likely to run in the subsequent election than losers. However, winners appear to be almost completely unresponsive to the signal received: the effect of the strength of the signal received on the probability of running in the next election is statistically insignificant from zero both for winners who receive a positive signal and winners who receive a negative signal. I estimate a positive but statistically insignificant discontinuity for winners at a zero signal value. All told, it appears that the signal received relative to expectations plays almost no role in winners' decisions to run in the next election. This result is consistent with the idea that once candidates become incumbents, other factors—job satisfaction, job performance, approval rating, etc.—determine whether the candidate decides to run again.

In contrast to the winners, I do find evidence that the signal received has an effect on the losing candidates' chances of running in the subsequent election. In the full sample, losing candidates who receive a positive signal are 17.7 percentage points more likely to run in the subsequent election than losing candidates who receive a negative signal. This effect is significant at the five percent level, and the magnitude of the effect does not diminish in narrower signal bandwidths or vote differential bandwidths. In fact, the estimated effect is larger in close elections than in the full sample. To assuage concerns that the estimated effect may just be an artifact of candidates preferring to run again if the election was close, I control for vote differential.

For losers with a negative signal, the effect of the strength of the signal on the probability of running again is not significantly different from zero across almost all subsamples, and for losers with a positive signal, this estimated effect is negative across almost all subsamples. This estimated negative relationship is perplexing.

Additionally, the other groups in the sample also appear to be unresponsive to the strength of the signal received. It is possible that those losers who outperform their expectations by large margins run for "better" positions or move on to other career opportunities based on the signal they received. The estimated slope for losers with a positive signal within a 5% signal bandwidth is insignificantly different from zero, which would support the hypothesis that this negative slope of the signal is driven by those losers who are greatly outperforming expectations.

Graphically, the binary nature of the outcome variable makes it difficult to visualize the estimated discontinuity. One way I depict the estimated discontinuity is by estimating the model and predicting the probability of running in the subsequent election for each candidate. I then plot these predicted probabilities against the signal received. Figure 4 presents predicted probabilities for winners within a 5% signal bandwidth, which looks extremely similar to the plot for winners within a 15% signal bandwidth. Figure 5 presents the predicted probabilities for losers within a 15% signal bandwidth, where the negative effect of the signal strength for overperforming candidates is evident, and Figure 6 presents the predicted probabilities for losers within a 5% signal bandwidth.

FIGURE 4.
Predicted probabilities of winners

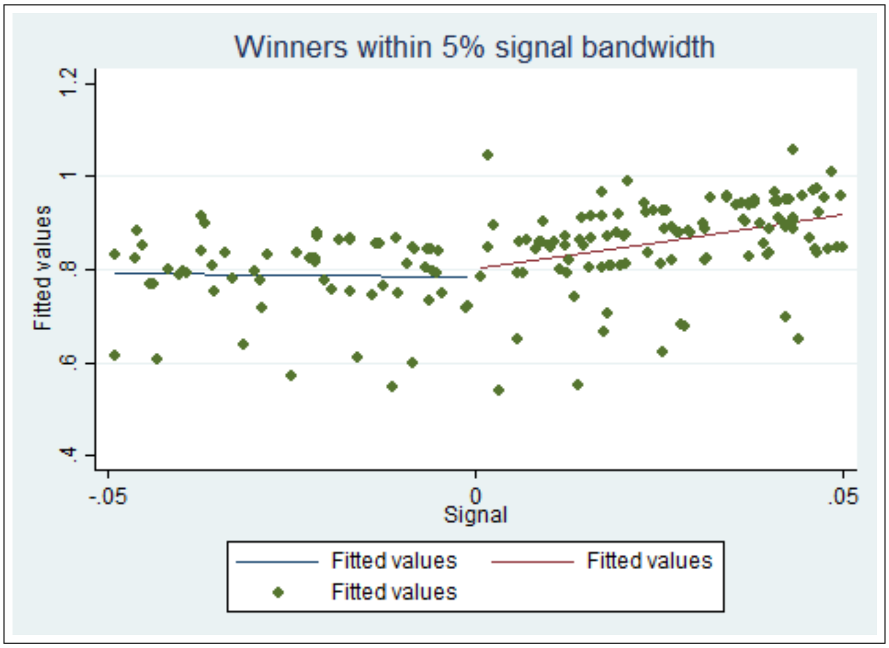


FIGURE 5.
Predicted probabilities of losers

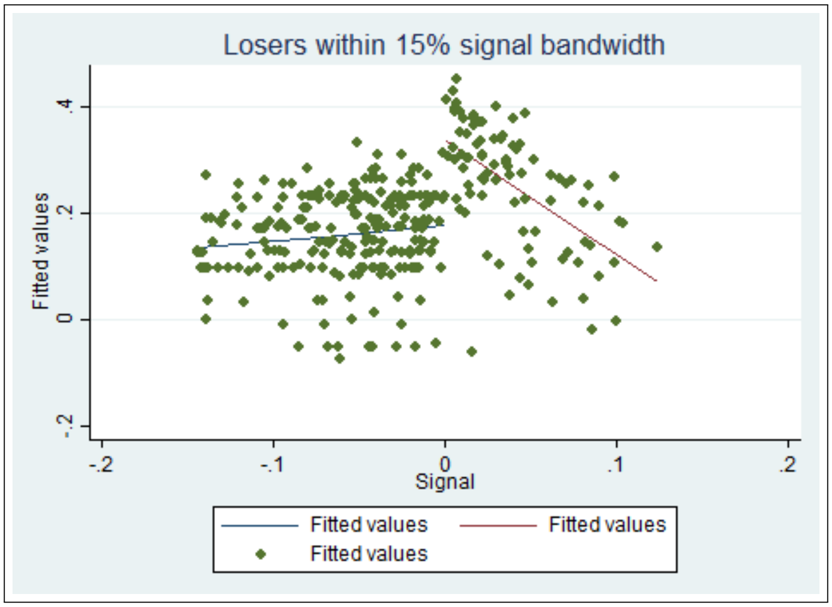
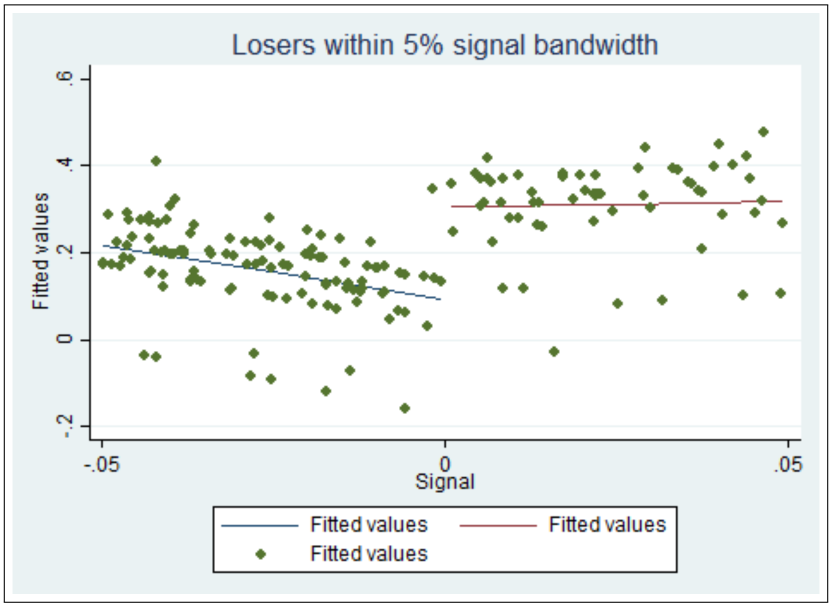


FIGURE 6.
Predicted probabilities of losers



A second way to visualize the discontinuity and test its robustness is through nonlinear estimation. Using fractional polynomial estimation, I plot nonlinear fits of the data for winners in Figure 7 and for losers in Figure 8. These estimations replace the linear signal term from the main specification (S_i) with two nonlinear polynomial terms as specified. The polynomials are chosen using the fractional polynomial optimization routine in STATA from the set of possible powers $\{.5, 1, 2, 3, 4, 5\}$.

- Winners with a positive signal: $S_i^{(5,5)'}\beta = \beta_1 S_i^5 + \beta_2 S_i^5 * ln(S_i)$

- Winners with a negative signal: $S_i^{(2,2)'}\beta = \beta_1 S_i^2 + \beta_2 S_i^2 * ln(S_i)$

- Losers with a positive signal: $S_i^{(.5,.5)'}\beta = \beta_1 S_i^{.5} + \beta_2 S_i^{.5} * ln(S_i)$

- Losers with a negative signal: $S_i^{(2,2)'}\beta = \beta_1 S_i^2 + \beta_2 S_i^2 * ln(S_i)$

Each nonlinear estimation includes all covariates from the main specification, with the exception of terms including the indicator variables for winning the election and for receiving a positive signal. The nonlinear estimated discontinuity for losers is 0.40, or 40 percentage points. The estimated discontinuity for winners is 0.26, or 26 percentage points. Both of these estimates of the discontinuity at the signal value are larger than the estimated discontinuities from the linear model.

FIGURE 7.
Fractional polynomial estimation for winners

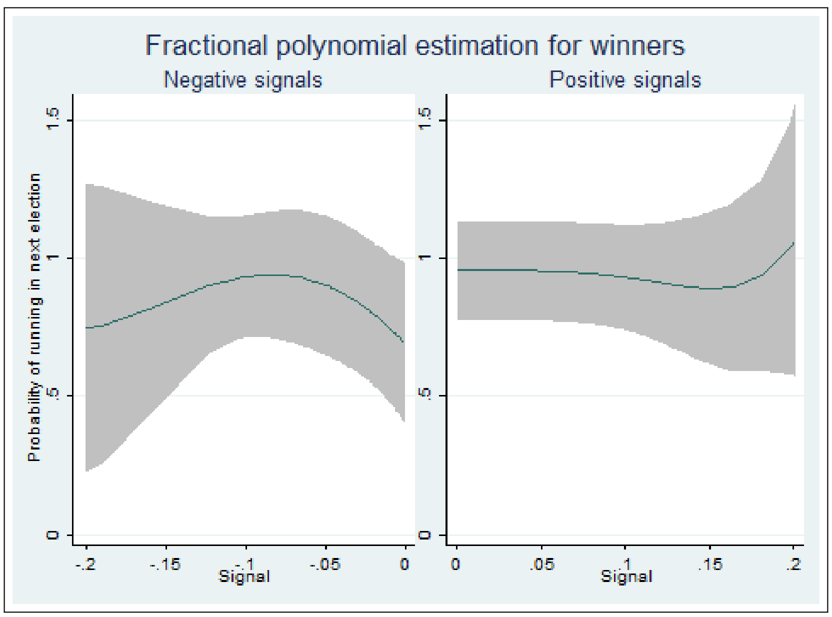
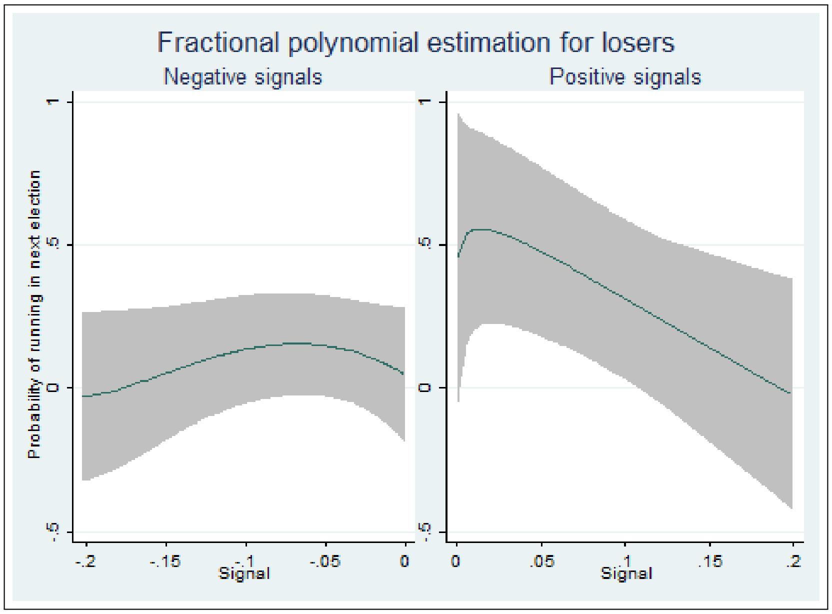


FIGURE 8.
Fractional polynomial estimation for losers



The lack of any estimated positive effects of the strength of the signal on the probability of running again for any of the groups of candidates fails to conform to predictions about candidates either deriving utility from the strength of the signal or updating their beliefs about the probability of winning based on the strength of the signal, both of which predict positive slopes. Furthermore, there is no evidence of the good news-bad news effect in this context. However, if it were the case that candidates were entirely oblivious to the signal, estimates of both the effect of the signal and of any discontinuity would be zero.

There are two potential explanations for the losers' estimated discontinuity at a zero signal value. Either losers update their beliefs about the probability of winning the next election only as a result of receiving a positive signal, or losers receive some utility boost from the positive signal, which then factors into their expected utility calculations in the next time period. Updating beliefs would affect the expected probabilities in the expected utility calculation, while a utility term would be additive. This utility discontinuity would support Diecidue and van de Ven's model of an aspirational utility level in which losing candidates then exhibit attribution bias, as described by Haggag and Pope, with respect to the aspirational utility term.

Once candidates lose while experiencing a positive signal, they form new predictions about their utility of losing in a future election. Suppose that losing candidates' experienced utility of losing is given by

$$u(lose_t) = w(lose_t) + \mu * 1(signal_t > 0),$$

where $\mu > 0$ is the effect of receiving the positive signal. If a candidate loses with a positive signal, the experienced utility of that loss is given by

$$u^+(lose_t) = w(lose_t) + \mu.$$

If the candidate does not realize that this experienced utility is larger as a result of outperforming expectations, the candidate's prediction of the utility of losing in the next period becomes

$$E[u^+(lose_{t+1})] = u^+(lose_t) = w(lose_t) + \mu.$$

Candidates who lose while underperforming expectations do not experience the additional utility from the positive signal, which means their experienced utility of losing is given by

$$u^-(lose_t) = w(lose_t).$$

As a result, their predicted utility of running in the next election is

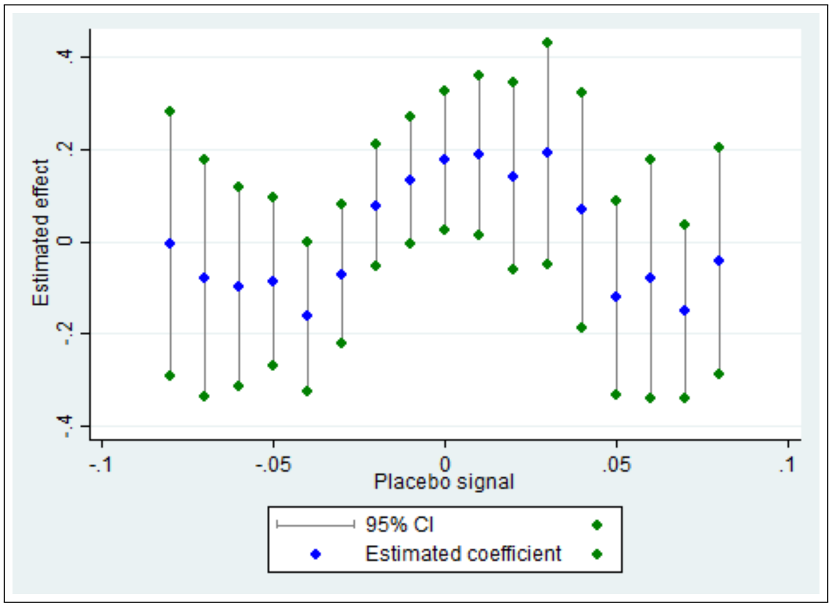
$$E[u^{-}(lose_{t+1})] = u^{-}(lose_t) = w(lose_t).$$

The losing candidates outperforming expectations predict a higher utility level from losing in the next election, which leads to a higher predicted utility level of running in the next election. These candidates are then more likely to run in the next election than their losing counterparts who underperformed expectations.

Robustness

As a check on the validity of the estimated discontinuity, I implement a placebo treatment test by moving the treatment value of the signal and estimating the resulting effect. This placebo treatment is for the full sample of the main linear specification discussed in the previous section. Figure 9 plots estimated treatment effects for each placebo treatment value from -.08 to .08. The estimated effect is slightly larger at treatment values of .01 and .04, but only the estimated effects at 0 and .01 are statistically significant from zero. While the treatment effect is not perfectly maximized at a treatment value of zero, the shape of the plot of the estimated treatment effects suggests that the discontinuity truly occurs at zero.

FIGURE 9.
Placebo estimated treatment effects



Heterogeneity

Given that candidates display this discontinuity in the probability of running again at a zero signal value, a natural follow-up question is whether different groups of candidates exhibit the effect more strongly than others. Is this estimated effect only the result of behavior by candidates of a certain party or incumbency status?

I investigate heterogeneity by estimating the model separately by party and by incumbency status. I also estimate the model for smaller vote differential bandwidths, which allows for investigation of the effect in close elections. A candidate is an incumbent if the candidate was in office in time period $t-1$, prior to the election in time t in which the signal occurs. Table 3 presents estimates of the model by vote differential bandwidth; Table 4 presents estimates by party by signal bandwidth; and Table 5 presents estimates by incumbency status by signal bandwidth.

The estimated discontinuity for losers is larger in close elections than in the full sample, as the effect size increases from 17.7 percentage points in the full sample to 43.4 percentage points in the sample of elections in which the realized vote differential is five percentage points or fewer. This means that performance relative to expectations is even more important to candidate decisions to re-run if the election is relatively close.

The losers' estimated discontinuity remains positive for both Democrats and Republicans across all signal and vote differential bandwidths. Splitting the sample reduces the number of observations, but the estimated discontinuity remains significant at the 1% level for Republicans in both the 5% and 10% vote differential bandwidths and significant at the 10% level for Democrats in both vote differential bandwidths. In each of the four bandwidths presented, the estimated effect is larger for Republican candidates than for Democratic candidates. When I split the sample by entrants and incumbents, the estimated discontinuity remains positive across all groups and bandwidths, and remains statistically significant in most bandwidths. Neither the incumbents nor the entrants exhibit a consistently larger effect across all bandwidths. The heterogeneity analysis is limited by the sample size, as splitting the data further decreases the power of the estimates.

TABLE 3.
Mean poll share by vote differential bandwidth, clustered by candidate

	Full sample	Within 15%	Within 10%	Within 5%
Winner	0.632***	0.523***	0.479***	0.500***
	(7.15)	(5.32)	(4.51)	(3.11)
Signal	-0.0786	-0.123	-0.988	-3.630**
	(-0.16)	(-0.21)	(-1.25)	(-2.00)
Winner * Signal	-0.276	-0.124	-0.479	1.476
	(-0.22)	(-0.09)	(-0.38)	(0.66)
Positive	0.177**	0.187**	0.266***	0.434***
	(2.29)	(2.28)	(2.76)	(3.02)
Winner * Positive	-0.125	-0.162	-0.180	-0.172
	(-1.15)	(-1.43)	(-1.42)	(-0.94)
Positive * Signal	-2.095**	-2.310**	-1.888	0.965
	(-2.07)	(-2.13)	(-1.36)	(0.41)
Winner * Positive * Signal	1.786	2.337	3.141*	0.462
	(1.07)	(1.32)	(1.69)	(0.16)
Vote differential	0.621**	1.427***	2.107***	4.012**
	(2.07)	(3.41)	(2.94)	(2.21)
Female	0.0579	0.0774*	0.0729	0.106
	(1.48)	(1.75)	(1.32)	(1.37)
Asian	-0.168	-0.173	-0.298	-1.131***
	(-1.04)	(-0.94)	(-1.54)	(-10.35)
Black	0.108	0.122	0.0145	0.229
	(0.80)	(0.73)	(0.10)	(1.02)
Democrat	-0.0325	-0.0211	-0.0336	-0.104*
	(-1.10)	(-0.66)	(-0.88)	(-1.93)
Incumbent winner	-0.178***	-0.150***	-0.166***	-0.269***
	(-3.86)	(-3.01)	(-2.78)	(-3.44)
Incumbent loser	-0.0162	-0.0361	-0.0137	0.00942
	(-0.33)	(-0.70)	(-0.23)	(0.10)
Constant	0.218***	0.264***	0.224***	0.113
	(3.69)	(3.48)	(2.71)	(0.90)
Observations	644	557	419	223

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 4.
Mean poll share by signal bandwidth, clustered by candidate

	Rep., 10%	Dem., 10%	Rep., 5%	Dem., 5%
Winner	0.621***	0.384**	0.749***	0.471**
	(4.81)	(2.43)	(4.30)	(1.98)
Signal	-1.224	-0.588	-6.417	1.765
	(-0.90)	(-0.41)	(-1.32)	(0.45)
Winner * Signal	-0.0604	-2.816	4.175	-5.131
	(-0.03)	(-0.90)	(0.68)	(-0.71)
Positive	0.337**	0.223*	0.329	0.129
	(2.42)	(1.79)	(1.52)	(0.73)
Winner * Positive=1	-0.310*	-0.00578	-0.233	-0.0195
	(-1.82)	(-0.03)	(-0.90)	(-0.07)
Positive * Signal	-2.897	-2.973	8.696	-0.175
	(-1.15)	(-1.44)	(1.02)	(-0.03)
Winner * Positive * Signal	3.257	6.848*	-8.094	9.276
	(1.01)	(1.78)	(-0.84)	(0.98)
Vote differential	1.264***	0.984***	0.895	0.438
	(2.75)	(2.60)	(1.64)	(0.91)
Female	0.0586	0.0384	0.0936	0.0228
	(1.02)	(0.56)	(1.08)	(0.28)
Asian	0	-0.212	0	-0.0543
	(.)	(-0.95)	(.)	(-0.17)
Black	0.202	0.0555	0.555	0.129
	(0.98)	(0.27)	(1.50)	(0.33)
Incumbent winner	-0.163**	-0.140*	-0.0734	-0.218**
	(-2.12)	(-1.84)	(-0.80)	(-2.20)
Incumbent loser	0.0321	-0.0890	0.113	-0.0433
	(0.36)	(-1.00)	(0.93)	(-0.37)
Constant	0.202**	0.200*	0.0900	0.0646
	(2.40)	(1.79)	(0.66)	(0.35)
Observations	269	269	172	172

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 5.
Mean poll share by signal bandwidth, clustered by candidate

	Inc., 10%	Entrants, 10%	Inc., 5%	Entrants, 5%
Winner	0.348**	0.669***	0.554**	0.651***
	(2.01)	(5.70)	(2.53)	(4.32)
Signal	0.241	-2.028	-6.602	-0.795
	(0.18)	(-1.54)	(-1.37)	(-0.21)
Winner * Signal	-2.445	0.175	2.167	1.182
	(-0.82)	(0.10)	(0.30)	(0.24)
Positive	0.214	0.283***	0.510**	0.106
	(1.19)	(2.59)	(2.17)	(0.67)
Winner * Positive	-0.162	-0.164	-0.432	0.0228
	(-0.70)	(-1.16)	(-1.33)	(0.12)
Positive * Signal	-5.702**	-0.945	-3.558	5.337
	(-2.14)	(-0.47)	(-0.42)	(0.95)
Winner * Positive * Signal	8.566**	1.946	10.13	-8.175
	(2.02)	(0.79)	(0.94)	(-1.16)
Vote differential	1.085	0.511	0.765	0.963***
	(1.64)	(1.21)	(1.11)	(2.66)
Democrat	-0.0630	-0.0150	-0.140	0.0266
	(-0.79)	(-0.32)	(-1.33)	(0.46)
Female	0.0795	0.0294	0.159	0.0629
	(0.98)	(0.58)	(1.57)	(0.95)
Asian	-0.811***	-0.0751	-0.752***	0.210
	(-10.79)	(-0.37)	(-9.06)	(1.08)
Black	-0.0699	0.288	-0.267*	0.645**
	(-0.50)	(1.41)	(-1.91)	(2.01)
Constant	0.236*	0.141	-0.0327	0.193
	(1.94)	(1.55)	(-0.18)	(1.48)
Observations	199	342	132	213

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Conclusion

To test ER's lab results regarding the good news-bad news effect, I examine U.S. House candidates' decisions to run for the same office in the following election. In the current election period, candidates hold expectations about their outcomes in the form of the mean poll share. The election provides an unpredictable signal of varying magnitude and valence, which informs the candidate about the outcome relative to expectations. After receiving the signal, the candidate must decide whether to run for the same office in the next period.

There is no evidence of a positive relationship between the value of the signal and the probability that a candidate runs in the subsequent election, which means a larger signal value does not necessarily correspond to a higher probability of running again. These relationships are not as strongly identified as the estimates of the discontinuity at the zero-signal value. Winning candidates appear not to account for the signal value at all, as both estimated slopes and the estimated discontinuity are insignificant from zero. Losing candidates display a full-sample discontinuity of 17.7 percentage points, which means that losing candidates who receive a slightly positive signal are significantly more likely to run for the same office again than those losing candidates who receive a slightly negative signal, and the estimated discontinuity is larger in narrower signal and vote differential bandwidths. The model estimates that losing candidates who underperform expectations almost never run in the subsequent election.

This discontinuity could result from an updating of beliefs about winning the next election or from some utility derived from the signal itself. A utility interpretation would mean that candidates receive some amount of utility from obtaining a positive signal, which is consistent with Diecidue and van de Ven's model of an aspirational level utility and HP's model of attribution bias. If losing candidates who outperform their expectations do not realize that the utility they experienced was larger as a result of outperforming expectations, these candidates will predict higher utility levels when losing in the future than those candidates who failed to meet their expectations. This difference in predicted utility levels between groups of candidates is consistent with the estimated discontinuity: losing candidates who receive slightly positive signals are estimated to be significantly more likely to run in the next election than losing candidates who receive slightly negative signals.

While I have assumed throughout this paper that the decision to run in the subsequent election is the candidate's, it is likely that the true decision is made jointly between the candidate and the candidate's party. If the candidate has most of the power in making this decision, the utility interpretation could be correct. If the party gets to make the final decision, this effect is likely the result of the party believing that a candidate has a better chance of winning the next election if the candidate received a positive signal in this election. Alternatively, the party could observe that a candidate underperforms expectations and immediately decide to replace the candidate in the subsequent election. It could also be the case that the estimated discontinuity is a combination of effects: the candidate receives some utility and becomes more likely to want to run again, and the party observes the positive signal and believes that the candidate remains a good choice to run in the next election. However, whether the decision is the candidate's or the party's, the estimated discontinuity implies that a candidate with a lower vote share and a positive signal may be more likely to run than a candidate with a higher vote share a negative signal, despite the fact that the candidate with the higher vote share is of higher quality.

CHAPTER III

THE EFFECT OF OPENING RECREATIONAL MARIJUANA MARKETS ON DOMESTIC VIOLENCE

Introduction

The landscape of marijuana use in the United States has shifted considerably over the past fifty years. In 1970, the federal government passed the Controlled Substances Act[1], which defined marijuana as a Schedule I drug and prohibited its use. Since then, there has been a slow and steady liberalization of policy at the state level. For example, in 1973, Oregon reduced the penalty for possession of up to one ounce of marijuana to a maximum fine of $100, which began a trend of states decriminalizing possession.[2]

The next major shift in marijuana law occurred in 1996, when California passed Proposition 215, which legalized the use of marijuana for medical purposed with a doctor's recommendation[3]. Over the next fifteen years, fifteen additional states, including Washington, Colorado, and Oregon, passed legislation legalizing medical marijuana use.[4] In 2012, Washington and Colorado became the first states to pass laws allowing consumers to use marijuana recreationally, and the first recreational marijuana stores in the United States opened on Jan. 1, 2014, in

[1] Government Publishing Office (1970)

[2] Single (1981)

[3] California Department of Public Health (2018)

[4] Anderson et al. (2015)

Colorado.[5] Recreational sales in California, the nation's most populous state, began on Jan. 1, 2018.[6]

Despite these widespread policy changes, the Drug Enforcement Administration maintains a Schedule I classification for marijuana, which is "defined as drugs with no currently accepted medical use and a high potential for abuse."[7] However, a 2013 memorandum from the Department of Justice states, "enforcement of state law by state and local law enforcement and regulatory bodies should remain the primary means of addressing marijuana-related activity." However, the memo also states that the federal government may seek enforcement if state enforcement efforts are not preventing certain harms, including distribution to minors, diversion to states with no legal possession, violence, and revenue to gangs.[8]

While the potential effects of recreational legalization have been widely debated, the majority have not been studied. Understanding how these changes in the availability of recreational marijuana affect social and public health outcomes is crucial for voters and policymakers considering future laws and regulations. This paper investigates the impact these recreational marijuana laws have on domestic violence, an outcome that imposes high costs to society through economic, physical health, and mental health channels (Max et al., 2004; Black, 2011; Sabia et al., 2013; Breiding et al., 2008; Silverman et al., 2001; Aizer, 2011).

Using daily data on reports of crime incidents at the reporting agency level from the FBI's National Incident-Based Reporting System (NIBRS) from 2012 to

---

[5] McGhee and Ingold (2013)

[6] Nicolewski (2016)

[7] Drug Enforcement Agency (2019)

[8] Department of Justice (2013)

2015, I employ a difference-in-differences model to estimate the effect of recreational marijuana sales in Washington, Colorado, and Oregon on intimate partner violence (IPV). Using a variety of state-specific time trends and year, month, day-of-week, and agency fixed-effects, I find that IPV incidents increase by 2.9-6.2% when recreational marijuana sales begin. I consider treatment at the state level, the county level, and a distance-weighted measure of treatment. The results are qualitatively similar across these definitions of treatment. While the estimated effect is positive for each of the three reported severity levels of IPV (aggravated assault, simple assault, intimidation), the effect is largest for intimidation incidents, which are the least severe type.

The literature suggests that domestic violence responds to economic, social, and psychological cues: Aizer (2010) uses administrative data on female assault hospitalization to show that decreases in the gender wage gap lead to a drop in the level of domestic violence against women; Card and Dahl (2011) use the NIBRS data on intimate partner violence to show that domestic violence increases as a result of unexpected losses in NFL games. Cesur and Sabia (2016) show that assignment to combat status leads to an increased probability of domestic violence committed by members of the military. These papers highlight mechanisms that lead to domestic violence occurence: the relative bargaining power of women in relationships and emotional cues that lead to anger affect the incidence rates of intimate partner violence.

In light of these findings, it would not be surprising if domestic violence also responded to health policy and changes in marijuana spending and use in states that begin selling recreational marijuana. The introduction of a recreational marijuana market likely has effects on the nature of intimate partners' relationships. This could be a new source of conflict or a shared bonding experience; couples could spend more or less time with each other; individuals' behavior likely changes when consuming (or even planning to consume) marijuana. Understanding both the effect of marijuana legalization on IPV and the mechanism by which it occurs is central to the development of effective regulation in this new market.

Recent literature (Crost and Guerrero, 2012; Crost and Rees, 2013; Anderson et al., 2013) suggests that marijuana alcohol function as substitutes. This mechanism would provide a clear pathway if the estimated effect of recreational sales on IPV were negative. However, as the estimated effect is positive, there are multiple potential mechanisms by which the beginning of recreational marijuana sales could cause the incidence of domestic violence to increase. The introduction of legal marijuana could introduce conflict within couples, especially if the couples disagree about whether to use the drug. These conflicts could also exacerbate arguments about money: if one partner wants to spend a portion of the budget on marijuana, tension could rise, leading to more reported incidents. A second mechanism is a simple exposure story: if recreational marijuana causes more couples to spend time together at home instead of participating in separate or public activities, domestic violence could increase as hours spent at home together increase. This is especially true if couples move alcohol consumption from bars to their homes. A third option is a direct effect through complementarity with alcohol: if individuals are more likely to engage in heavy alcohol use as a result of their new purchases of marijuana or if marijuana physically exacerbates the effects of alcohol, this could lead to an increase in alcohol-related incidents.

Other mechanisms could occur through reporting behavior, and not through additional actual assaults. One channel is individual reporting: if individuals were previously afraid or unwilling to report domestic violence because they had marijuana in the house, they may become more willing to report these incidents when they know that recreational marijuana is legal and less stigmatized. Another channel is a change in police behavior. If law enforcement agencies no longer allocate resources to policing marijuana possession, they may be able to respond to more reports of domestic violence, which would show up as an estimated positive effect. Even if agencies did not change enforcement behavior, it could be the case that they change their reporting behavior. Using a variety of additional datasets and alternative specifications, I attempt to disentangle there potential mechanisms. The mechanism most consistent with the evidence is an increase in relationship conflict, but it is possible that multiple mechanisms simultaneously affect the reported incidence of intimate partner violence.

The paper proceeds as follows: Section 2 lays out relevant literature regarding effects of marijuana legalization, domestic violence, and the complicated relationship between marijuana and alcohol, as well as various potential hypotheses about mechanisms through which recreational marijuana legalization could affect reported intimate partner violence. Section 3 discusses the data used and describes the model, including a variety of ways to define treatment for a given agency in the sample. Section 4 presents results from the main specification. Section 5 incorporates additional datasets to investigate the likelihood of each proposed mechanism; Section 6 concludes.

Literature

Marijuana

There is a significant body of literature investigating the relationship between marijuana legalization, marijuana use, and important social, criminal, and public health outcomes. In general, these studies have focused on the effects of medical marijuana legalization laws. Medical marijuana legalization in the United States began in 1996, when California passed Proposition 215, which allowed consumers to legally purchase marijuana for medical purposes with a doctor's recommendation.[9] By 2018, 31 states and the District of Columbia had laws in place that allow for the legal use of medical marijuana.[10]

In contrast, recreational marijuana legalization has been a recent development in the United States. In November 2012, both Colorado and Washington passed legislation that provided a framework for future legal recreational marijuana sales. Consumers in Colorado could begin purchasing recreational marijuana on Jan. 1, 2014, when stores opened in Breckenridge, Central City, Denver, Edgewater, Idaho Springs, Northglenn, Pueblo West, and Telluride (McGhee and Ingold, 2013). In Washington, recreational shops opened on July 8, 2014 in over a dozen cities (Baca, 2014). Oregon's Measure 91, which passed in November 2014, legalized the recreational use of marijuana, and on Oct. 1, 2015, medical marijuana dispensaries began recreational sales to consumers.[11]

While Anderson et al. (2014a) point out that "the legalization of marijuana for medicinal purposes approaches de facto legalization of marijuana for

[9] California Department of Public Health (2018)

[10] National Conference of State Legislatures (2019)

[11] Oregon Liquor Control Commission (2019)

recreational purposes," there are almost assuredly many consumers who purchase recreational marijuana who were not purchasing medical marijuana prior to these law changes. These consumers may previously not have purchased marijuana from a medical dispensary due to lack of a prescription from a doctor, an unwillingness to lie to a doctor to obtain a prescription, or the social stigma associated with using a previously illegal drug; this population of individuals who previously did not purchase medical marijuana is now treated by the policy change. Other individuals who previously purchased marijuana through black market channels may substitute away from illegal purchases to legal purchases of the drug.

Previous studies have investigated many social and public health effects related to the institution of these medical marijuana laws. Marijuana legalization could be significantly detrimental to social welfare if these laws lead to increases in crime or negative public health outcomes. However, much of the literature regarding medical marijuana laws has found positive effects for society. Some of the social benefits attributed to medical marijuana laws include a decrease in alcohol-related traffic fatalities (Anderson et al., 2013), a reduction in the suicide rate among men between the ages of 20 and 39 (Anderson et al., 2014b), and a lowered probability of obesity (Sabia et al., 2017). Additionally, Anderson et al. (2015) do not find evidence that medical marijuana laws lead to increased marijuana use among teenagers, and Ullman (2017) finds that medical marijuana legalization leads to a decrease in absences from work due to sickness.

Since recreational marijuana sales have begun very recently, far less literature exists documenting the effects of the opening of recreational marijuana sales markets. Anderson et al. (2014a) discuss a potential "worst-case scenario" of recreational marijuana legalization through the lens of research documenting effects on marijuana price, marijuana use, alcohol use, and crime; they conclude that they expect recreational marijuana legalization to lead to increased marijuana usage and decreased alcohol usage. Finally, they predict that the net public health benefits resulting from such a policy will be positive.

Domestic violence

According to the National Coalition Against Domestic Violence, more than 10 million men and women are physically abused by an intimate partner each year; one in three women and one in four men have been victims of some form of physical violence by an intimate partner.[12] Domestic violence has both direct and indirect welfare consequences for victims. Max et al. (2004) use data from the National Violence Against Women Survey to estimate that the total cost of intimate partner violence in the United States in 1995 was $5.8 billion. These costs include physical and mental health care services, as well as lost productivity resulting from injury or death. This estimate increases to $9.7 billion when the costs are considered in 2018 dollars. The CDC notes that these are likely underestimates, as the total estimated cost does not include costs incurred by the criminal justice system.[13] Black (2011) provides an overview of the literature linking intimate partner violence to adverse mental and physical health consequences. Victims of domestic violence experience a litany of additional health effects, including additional mental health, cardiovascular, reproductive, and immune problems. Sabia et al. (2013) show that women who are victims of sexual violence have decreased labor market participation rates and wages.

---

[12] National Coalition Against Domestic Violence (2019)

[13] Centers for Disease Control and Prevention (2019)

Aside from direct welfare, productivity, and medical costs, indirect costs for victims of intimate partner violence exist as well. Breiding et al. (2008) use data from the Behavioral Risk Factor Surveillance System to conclude that being a victim of intimate partner violence is associated with increases in health risk behaviors, such as smoking, binge drinking, and failing to regularly visit a doctor. Silverman et al. (2001) conclude that experiencing dating violence increases the rates of many risky behaviors, including unhealthy weight control behaviors, pregnancy, and suicidality, in adolescent girls. Aizer (2011) estimates that a mother being hospitalized during pregnancy as a result of an assault decreases the birth weight of a child by 163 grams.

A body of psychology literature investigates the relationship between marijuana use and domestic violence. Moore and Stuart (2005) and Testa and Brown (2015) discuss cross-sectional studies that imply a positive correlation between marijuana use and intimate partner violence. Moore and Stuart mention three possible mechanisms: psychopharmological; violent behaviors occurring in the context of obtaining illegal substances; and general deviance, which essentially posits that some individuals are predisposed to deviant behavior, making them more likely to engage in both marijuana usage and violence. Testa and Brown discuss several studies that show a positive association between marijuana usage and intimate partner violence, but they acknowledge that the mechanism through which this might occur is not well understood. They also note that the effects are attenuated when controlling for factors like antisocial behavior, other substance use, and psychopathology. Marijuana use may also affect parenting behavior; Freisthler et al. (2015) survey individuals and find that marijuana use is positively related to child physical abuse.

Crane et al. (2016) study behavioral outcomes, such as satisfaction, anger experience, and relationship quality, relative to couples' marijuana use. The paper finds that when both partners in a couple use or abstain from marijuana, the couple displays better conflict resolution behavior than when only one partner in a relationship uses the drug. This finding is consistent with the idea that marijuana use by a single partner could increase conflict within a relationship. While these studies suggest links between marijuana use and domestic violence, none present causal evidence of the effect of a change in access to marijuana on intimate partner violence for a population.

Alcohol

Alcohol is a potentially crucial actor in the relationship between marijuana and domestic violence. The literature generally agrees that increased use of alcohol leads to increases in violent crimes. Markowitz (2005) uses data from the National Crime Victimization Survey to show that increases in the beer tax rate decrease the probability of victimization as a result of assault. Markowitz et al. (2012) find evidence of a similar negative relationship between alcohol prices and the probability of assault victimization, although they find that other alcohol-related policies, such as drunk driving laws and penalties, do not affect these victimization probabilities. Carpenter and Dobkin (2015) use a regression discontinuity approach at the minimum legal drinking age to estimate the effect of crossing the legal-drinking-age threshold on crime; they find that many types of crime, including violent crime, increase across this threshold. This finding is corroborated by Callaghan et al. (2016) using evidence from the minimum legal drinking age threshold in Canada.

While the literature presents clear evidence that alcohol increases violent crime, the relationship between alcohol use and marijuana use is much more contested. Pacula and Sevigny (2014) and Guttmannova et al. (2016) provide overviews of the literature on whether alcohol and marijuana are complements or substitutes. Pacula (1998) uses National Longitudinal Survey of Youth data to estimate individual demand equations for alcohol and marijuana; she finds that alcohol and marijuana function as complements. Williams et al. (2004) find that alcohol-restriction policies decrease marijuana use, which would imply that alcohol and marijuana are complements.

More recent evidence has generally favored the idea that alcohol and marijuana function as substitutes. Crost and Guerrero (2012) implement a regression discontinuity design at the minimum legal drinking age and find that marijuana consumption decreases across this threshold, while Crost and Rees (2013) use National Longitudinal Survey of Youth 1997 data and a regression discontinuity approach to show that alcohol and marijuana are not complements. Anderson et al. (2013) argue that their finding of medical marijuana legalization decreasing alcohol-related traffic fatalities provides evidence that alcohol and marijuana are substitutes.

It may be the case that alcohol and marijuana interact differently for different segments of the population. Wen et al. (2015) discuss the possibility that marijuana and alcohol, especially high-dose consumption of alcohol, may act as substitutes for individuals seeking mild relaxation while acting as complements for individuals using higher doses while seeking a more intense euphoria. Additionally, they estimate that passage of medical marijuana laws is associated with a 10% increase in binge drinking and an 18-22% increase in simultaneous marijuana and alcohol use, which is consistent with their hypothesis that alcohol and marijuana may function as complements for heavy users of the substances. Additional evidence to support this hypothesis comes from Kerr et al. (2017), who find that after recreational marijuana legalization, Oregon college students' usage of marijuana increased only for those students who had reported recent heavy use of alcohol. Subbaraman and Kerr (2015) analyze National Alcohol Survey data and find that individuals who consume both marijuana and alcohol tend to consume them simultaneously; additionally, they find that simultaneous use of these substances increased the odds of self-harm, social consequences, and drunk driving relative to the use of only alcohol.

One potential pathway is a physiological complementarity between alcohol and marijuana. In an experimental setting, Lukas and Orozco (2001) find that subjects who consume ethanol in addition to tetrahydrocannabinol (THC)–the active ingredient in marijuana– experience greater euphoria and higher levels of plasma THC than subjects who consume placebo ethanol. This suggests a pharmacological complementarity between the two drugs: consuming ethanol may increase the body's absorption of THC.

Hypotheses

Based on the literature, there are competing predictions of the effect of recreational marijuana legalization on domestic violence. Multiple mechanisms could cause domestic violence to decrease when recreational sales begin, while a number of other mechanisms could produce the opposite effect.

Substitution from alcohol

If alcohol use increases the likelihood of domestic violence, and if alcohol and marijuana function as substitutes, then the story is similar to the traffic fatality story told by Anderson et al. (2013): legalizing recreational marijuana causes substitution away from alcohol, which lowers domestic violence incidence. Card and Dahl (2011) find that alcohol-related intimate partner violence incidents are more affected by upset football losses than those incidents not involving alcohol, although they note that the power of the model is limited due to potential underreporting of the involvement of alcohol in domestic disputes.

Direct incapacitation

Another mechanism through which recreational legalization could decrease domestic violence is a direct incapacitation effect. Marijuana is generally used to induce relaxation and mild euphoria; these effects may inhibit a person's desire or ability to commit violent acts on partners. Anecdotally, individuals may be less inclined to commit violent acts if, after using marijuana, they just want to sit on the couch and watch TV. This incapacitation mechanism is distinct from a possible substitution mechanism: users of alcohol who substitute away from marijuana may encounter both the substitution (decreased alcohol usage) and incapacitation (relaxation) effects, but individuals who were not previously using alcohol may experience only an incapacitation effect. This type of incapacitation could affect both offenders and victims: if offenders are incapacitated, they may commit fewer assaults; if victims are incapacitated, they may be less likely to report an incident to a law enforcement agency.

Complementarity with heavy alcohol use

On the other hand, if marijuana and heavy alcohol use function as complements, as described by Wen et al. (2015), Kerr et al. (2017), and Subbaraman and Kerr (2015), then the advent of recreational marijuana sales may spur individuals to simultaneously consume marijuana and large amounts of alcohol, which subsequently leads to an increase in intimate partner violence incidents. These substances could also interact differently across different situations: perhaps alcohol and marijuana function as substitutes in social situations or instances where consumers are outside of the home, while functioning as complements when consumed privately.

Conflict within a relationship

The ability to purchase recreational marijuana may also inject additional conflict into a relationship, especially if there is already conflict about budget allocation. While Crane et al. (2016) do not discuss budgets, they do find that relationships tend to function better when both partners either use marijuana or abstain from marijuana use. In a situation where only one partner wants to spend a portion of the couple's monthly budget on marijuana at the newly opened store, a natural consequence would be increased tension, conflict, and perhaps violence.

Increased exposure

Once a marijuana market opens, it could be the case that many individuals choose to consume marijuana at home instead of participating in other forms of recreation outside the home. Domestic violence could increase if many individuals choose to stay home instead of going out to a bar, especially if those individuals choose to stay home and drink alcohol instead of going to bar. Spending more time at home with a partner would mechanically increase the number of domestic violence incidents.

Methodology and Data

Data

I use the FBI's National Incident-Based Reporting System (NIBRS) data as my source for reported domestic violence incidents.[14] Unlike the FBI's Uniform Crime Reporting (UCR) data, which tracks only arrests, the NIBRS data provides detailed information about reported crime incidents. For reporting purposes, the NIBRS defines an incident as "one or more offenses committed by the same offender, or group of offenders acting in concert, at the same time and place." Importantly, many incidents reported in the NIBRS data do not results in arrests, which means that less-severe incidents are still recorded.

The NIBRS has multiple attractive features for use in this analysis. The most important of these is the level of detail about both the victim and the offender; specifically, the inclusion of a variable labelling the relationship of the victim to the offender allows for the separation of assault incidents into those involving an intimate partner and those that do not. I follow Card and Dahl (2011) in defining intimate partner violence incidents: an "intimate partner" is defined as a victim whose relationship to the offender is spouse, common-law spouse, ex-spouse, boyfriend/girlfriend, or homosexual relationship; a domestic violence incident occurs when a victim of this sort experiences some form of assault. Specifically,

[14] This data was accessed through the University of Michigan's Inter-university Consortium for Political and Social Research's National Archive of Criminal and Justice Data.

I define the incident as intimate partner violence if the reported UCR offense codes for a given victim represent an assault and if the victim's relationship to any offender falls under the intimate partner classification, as defined above.

Assaults are classified into three categories in the data: aggravated assault, simple assault, and intimidation. Aggravated assault occurs when "the offender uses a weapon or displays it in a threatening manner, or the victim suffers obvious severe or aggravated bodily injury." Simple assault is "an unlawful physical attack by one person upon another where neither the offender displays a weapon, nor the victim suffers obvious severe or aggravated bodily injury." Finally, intimidation occurs when an offender "unlawfully place(s) another person in reasonable fear of bodily harm through the use of threatening words and/or other conduct, but without displaying a weapon or subjecting the victim to actual physical attack." These distinct levels of severity allow me to investigate how the effect of the beginning of recreational sales varies by incident type. The vast majority of incidents in the dataset are simple assault; these make up 75.3% of incidents. Aggravated assault and intimidation comprise similar portions of the remainder, as 12.7% of incidents are classified as aggravated assault, while 12.1% of incidents are classified as intimidation.

A second attractive property of the NIBRS data is its frequency: the timing of each incident is reported to the day and hour. As a result of this, I can narrow the bandwidth of the sample around the date of treatment in each treated state. Additionally, this temporal granularity allows me to investigate the heterogeneity of the estimated effect by day of week and time of day.

One drawback of the NIBRS data is that law-enforcement agencies voluntarily report incidents, which means that only populations in specific jurisdictions are included in the data. Some states, notably including California, do not participate in the program at all. However, the FBI estimates that approximately 30% of the US population is covered by law enforcement agencies that report crime incidents to the program. My sample runs from the beginning of 2012, in which 6,115 agencies covering 90,290,162 people reported incidents, to the end of 2015, in which 6,648 agencies covering 96,087,615 people reported incidents. Relative to earlier years of the NIBRS program, there is very little change in the population covered during this sample. Colorado is one of 16 states that reports all of its crime through NIBRS, while the program covers a larger proportion of the Washington (approximately 74%) and Oregon (approximately 42%) populations than the population covered in the average state. Unfortunately, Portland, the largest city in Oregon, does not report crime data to NIBRS. [15] My sample runs from the beginning of 2012 to the end of 2015.

Due to the size of the NIBRS dataset, each year of data is broken up into multiple segments for easier use. Each incident in the dataset is assigned an incident number, which can be used to link incidents across different segments of the data. While the NIBRS codebook maintains that these incident numbers are unique, duplicate observations and incident numbers exist in the data. Some observations with the same incident number are perfectly duplicated, including the victim sequence number; some observations with the same incident number are duplicated across agencies, while all other reported values are the same; and some incident numbers span large numbers of observations across states, times, types of crimes, and victims, even for victims with the same victim sequence number. Duplicated observations comprise approximately 1% of the sample. For

[15] Federal Bureau of Investigation (2012, 2015)

observations that are duplicated perfectly or across agencies, I keep one copy and discard duplicates. I drop incident numbers that span multiple incidents, as shown by duplicate victim sequence numbers within an incident at different places and times.

A second important data cleaning consideration is the presence of a “report date indicator” variable. The NIBRS data contains an incident date variable, which reports when the incident occurred. However, if the report date indicator variable in the administrative segment is turned on, the incident date variable actually represents the date on which the incident was reported to the law enforcement agency, not the date on which the incident occurred. No variable is included to indicate the date on which the incident was reported to have occurred, which means that for each incident, the data includes either the incident date or the date on which the incident was reported, but never both. These “reported date” incidents make up approximately 12% of the sample. While I do not have reason to believe this indicator varies systematically across treatment, I drop these “reported date” incidents from the sample to eliminate a source of measurement error in the dependent variable. In general, including these reported date incidents slightly increases the estimated treatment effect but does not qualitatively change the results.

Model

I specify a difference-in-differences model for the number of intimate partner violence incidents reported by a given law enforcement agency on a given day in the sample period. This model takes the form

$$\mu_{it} = \alpha + \beta_1 T_{it} + \theta_i + \delta_t + X_{it}\gamma + \epsilon_{it},$$

where μ_{it} represents the number of intimate partner violence incidents reported by agency i on day t, θ_i represents an agency fixed-effect, which controls for time-invariant characteristics that vary across agencies, and δ_t represents a variety of time fixed-effects. In the main specification, I include time controls for year fixed-effects, month fixed-effects, day-of-week fixed effects, and the full set of holiday dummies used by Card and Dahl (2011).[19] I also include an indicator variable representing if the law making recreational marijuana legal in a given state has passed. The matrix X_{it} includes controls that vary by time and agency, including demographic controls (state-month-level population, unemployment[20], fraction of the population that is female, fraction of the population that is black, and average age)[16], indicator variables for the date of a Super Bowl played by teams representing the state in which the agency lies, and state-specific time trends. I explore three methods of defining the treatment variable T_{it}: state-level treatment, county-level treatment, and a distance-weighted treatment method. I discuss these treatment options in more detail in the following section.

[19]These include indicators for Christmas Day, Christmas Eve, New Year's Day, New Year's Eve, Halloween, Valentine's Day, St. Patrick's Day, Columbus Day, Memorial Day, Labor Day, Veterans' Day, and Thanksgiving.

[20]State-month-level unemployment data comes from the Federal Reserve Economic Data website, https://fred.stlouisfed.org/

[16]Population and demographic data come from the Current Population Survey via the Integrated Public Use Microdata Series: Sarah Flood, Miriam King, Steven Ruggles, and J. Robert Warren. Integrated Public Use Microdata Series, Current Population Survey: Version 4.0 [dataset]. Minneapolis, MN: University of Minnesota, 2015. http://doi.org/10.18128/D030.V4.0.

Defining treatment

As the goal of this paper is to investigate the effect of the availability of recreational marijuana on domestic violence, treatment should occur when an agency or county is first exposed to legal recreational sales. It is not obvious that there is a "right" way to define treatment in this situation. Defining treatment at the state level imposes treatment on all consumers in a state, even those who live quite far from a recreational store, while defining treatment at the county level treats many consumers who live near a store in an adjacent county as untreated.

I estimate models using three different definitions of treatment. The first of these is treatment at the state level, which means that every reporting agency in a state is defined as being treated on the first day of legal recreational sales occurring anywhere in that state. This method will likely underestimate the effect of the policy, since many agencies that are geographically distant from functioning recreational stores are included in the treatment group.

The second method of defining treatment is at the county level. In this case, an agency is defined as being treated on the first day of the first month in which the county containing the agency reported tax revenue to the state government. While this method allows treatment to vary across time within a given state, counties without operational stores will be part of the control group, even if those counties border counties in which recreational sales have begun. Consumers in these counties can easily drive to the store in the next county, which makes it likely that these "control" counties are actually significantly treated. As is the case with treatment at the state level, this definition of treatment will likely underestimate the true effect of the policy. Agencies that span the entire state, such as state police, are coded as being treated on the first day of recreational sales occurring anywhere in the state.

In an attempt to balance the issues of imposing treatment on agencies that are likely untreated and assigning control status to agencies that are likely treated, I construct a distance-weighted treatment variable. Using population-weighted county centroids from the U.S. Census Bureau[17], I calculate the distance from the centroid of county i to the centroid of each other county j, for all county pairs i, j in the sample. I then calculate each county's distance to the nearest county with an active recreational sales market in a given month. This process results in a treatment variable T_{it} for each county:

$$T_{it} = \begin{cases} 1, \text{ if agency } i \text{ is in a county with an active recreational market in time } t \\ \frac{1}{1+.0009d_{it}^2}, \text{ if agency } i \text{ is not in a county with an active recreational market in time } t, \end{cases}$$

where d_{it} is the minimum distance to the population-weighted centroid of a county with an active recreational market in time t.

This distance-weighting function means that any county with active recreational sales is defined as fully treated. I choose the form of the distance-weighting function so that treatment level decreases non-linearly, and I choose the scale parameter so that the treatment variable takes on a value of .1 for a county that is 100 miles away from the nearest county with an active recreational market. A county is essentially "half-treated" (where $T_{it} = .5$) if it is located 33.3 miles from the nearest treated county. Allowing treatment to be continuous allows each county to be partially treated according to its distance, rather than enforcing

[17]Population-weighted centroid data comes from https://www.census.gov/geo/reference/centersofpop.html

binary treatment status. Additionally, this distance-weighted treatment variable decreases at a decreasing rate, which is consistent with the idea of non-linear travel costs to consumers. Since the choice of distance-weighting function is arbitrary, I also estimate the model using alternative scale parameters in the distance-weighting function. Changes to the distance-weighting function do not qualitatively change the results.

While defining treatment at the state or county level crisply defines boundaries and doesn't allow for spillover treatment effects, using a distance-weighted treatment allows areas across state borders to be considered in the treatment group. This is especially important considering that Hansen et al. (2017) show that there is significant diversion of recreational marijuana sold in Washington across the Oregon border in the months following the opening of the recreational marijuana market in Washington.

Results

Main estimates

In Table 6, I present the difference-in-differences estimates of the pooled treatment effect among Colorado, Washington, and Oregon by the order of the state-specific time trend used in the model. This model uses the full sample of data, which consists of all agency-day counts of domestic violence incidents from January 1, 2012 to December 31, 2015.

With state-specific linear time trends included in the model, the pooled treatment effect of the beginning of recreational sales is 0.0104, which means that being in a state that has begun the sale of recreational marijuana increases reports of domestic violence by 0.0104 incidents per agency per day. The estimated pooled

TABLE 6.
Estimated effect of the beginning of recreational marijuana sales by definition of treatment variable, standard errors adjusted for clustering at the level of treatment

	(1)	(2)	(3)	(4)	(5)
Treatment:	Pooled State	County	Distance1	Distance2	Separate State
Legalization	-0.0009	0.001	0.0009	0.001	-0.002
	(-0.22)	(0.16)	(0.11)	(0.16)	(-0.51)
Market opens	0.0104***	0.0213	0.0177*	0.0205	
	(4.99)	(1.54)	(1.72)	(1.59)	
Percentage change	2.9%	6.2%	5.2%	6.0%	
CO market opens					0.0120***
					(3.49)
Percentage change					3.0%
WA market opens					0.0088***
					(6.02)
Percentage change					2.5%
OR market opens					0.0229***
					(5.46)
Percentage change					11.0%
Observations	3434302	3429594	3429594	3429594	3434302
Number of clusters	36	1678	6623	6623	36

t statistics in parentheses

All models include year, month, day-of-week, and agency fixed effects

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

treatment effect is statistically significant at the 1% level when standard errors are clustered at the state level and statistically significant at the 10% level when standard errors are clustered at the agency level. The average law enforcement agency in the sample reports 0.338 incidents of domestic violence per day; this means that the estimated treatment effect from the model with a linear time trend represents a 2.9% increase in the number of reported incidents. The estimated treatment effect remains positive, statistically significant, and of similar magnitude when the model is instead estimated with quadratic, cubic, or quartic state-specific time trends. These estimates can be found in Appendix Table 1. Notably, while the estimated effect of the opening of the recreational market is positive and statistically significant, the estimated effect of legalization–when the law passes making recreational use acceptable, but without active legal recreational markets–is virtually zero in every specification.

In the main specification, treatment occurs at the state-day level on the first day of recreational sales anywhere in the treated state. The table also presents estimates of the treatment effect using the previously discussed alternative methods of treatment, county-level and distance-weighted; the standard errors are adjusted for clustering at the level of treatment, which is the county level. The estimated effect size for all three specifications is larger than the estimated effect size when treatment occurs at the state. These effects are now less statistically significant; the first distance-weighted treatment estimate is statistically significant at the 10% level, while the estimates for county-level treatment and the alternative distance-weighted treatment are close to statistical significance (t-statistics of 1.54 and 1.59, respectively). The first distance-weighted treatment is defined by the function $\frac{1}{1+.0009d_{it}^2}$, where d_{jt} is the minimum distance to the population-weighted centroid of a county with an active recreational market in time t, which means that a county 100 miles from the nearest county with an active recreational market is .1 treated. The alternative distance-weighted treatment is defined by the function $\frac{1}{1+.0036d_{it}^2}$, where d_{jt} is the minimum distance to the population-weighted centroid of a county with an active recreational market in time t, which means that a county 50 miles from the nearest county with an active recreational market is .1 treated.

I also present estimates of the treatment effect in each of the three treated states separately using the full sample of data. The estimated treatment effects are 0.0120 in Colorado, 0.00883 in Washington, and 0.0228 in Oregon. These estimates are all statistically significant at the 1% level. Relative to the mean number of reported domestic violence incidents at the agency-day level in each state, these estimates correspond to increases in reported incidents of 3.0% in Colorado, 2.5% in Washington, and 11.5% in Oregon.

Additionally, I estimate a Poisson regression model using the full sample of observations. The incidence-rate ratio, which is the exponentiated coefficient from the Poisson regression, allows for interpretation of effects as changes in relative rates of the event in question caused by a one-unit change in the given independent variable. The estimated treatment effect is the incident-rate ratio minus one, which allows the coefficient to be interpreted as the percentage change in the dependent variable, domestic violence incidents. The estimated treatment effect from the Poisson model incorporating state-level treatment and linear state-specific time trends is 0.0305, which represents a 3.1% increase in domestic violence incidents in treated states after treatment occurs, and this estimated treatment effect is statistically significant at the 10% level. Appendix Table 2 provides Poisson estimates of the treatment effect by definition of treatment and order of state-specific time trend. These estimated treatment effects range from a 2.9% increase in domestic violence incidents (state-level treatment with quadratic state-specific time trends) to a 6.4% increase (county-level treatment with quartic state-specific time trends).

Robustness

One potential concern is that agencies could change their reporting behavior across treatment, so that the effect is caused by small agencies that previously did not report incidents beginning to report. If this changes systematically across treatment, agency fixed-effects, which assume that the level differences are time-invariant, will not account for the issue. Table 7 provides estimates of the effect by sub-samples of agencies by the percentiles of the number of days in the four-year sample on which they report any crime. The estimated effects remain positive, statistically significant, and unchanged across agency reporting percentiles.

TABLE 7.
Estimated effect of the beginning of recreational marijuana sales on IPV, full sample by agency reporting percentile, state-level treatment, standard errors adjusted for clustering at the state level

	(1) All agencies	(2) Top 75	(3) Top 50	(4) Top 25	(5) Report every day
Legalization	-0.0009	-0.00184	-0.00146	0.00483	0.0600
	(-0.22)	(-0.25)	(-0.14)	(0.26)	(1.00)
Market opens	0.0104***	0.0133***	0.0211***	0.0374***	0.0786***
	(4.99)	(4.92)	(4.29)	(4.88)	(5.20)
Percentage change	2.9%	3.1%	4.1%	4.5%	4.8%
Observations	3434302	2576436	1720058	863637	243925

t statistics in parentheses
All models include year, month, day-of-week, and agency fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Another potential concern is that many states may not represent good control groups for Colorado, Washington, and Oregon. For instance, we may not expect Alabama to be similar to the states that have voted to legalize recreational marijuana sales. In Table 8, I present estimated treatment effects by order of time trend when estimating the model using only Colorado, Washington, and Oregon. In this model, the two states whose treatment statuses remain constant through the treatment date in the third state act as the controls for the treated state. These estimated effects are not statistically significant, but they remain positive and of similar magnitudes to the estimated effects when the entire sample of states is included as controls. With linear state-specific time trends, the estimated effect represents a 3.0% increase in incidents, which is almost identical to the estimated effect in the full model.

TABLE 8.
Pooled state-level treatment by order of time trend, clustered by state, sample includes only CO, OR, WA

	(1) Linear	(2) Quadratic	(3) Cubic	(4) Quartic
Market opens	0.0109	0.00719	0.00889	0.00685
	(1.86)	(1.43)	(2.09)	(0.78)
Percentage change	3.0	2.0	2.5	1.9
Observations	331058	331058	331058	331058

t statistics in parentheses
All models include year, month, day-of-week, and agency fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Next, I investigate the treatment effect in each state in a narrower time bandwidth. For each treated state, I estimate the model using data in a 120-day window around the state-level treatment date: the subsample begins 60 days prior to the beginning of recreational sales in the state and ends 60 days after the beginning of recreational sales. In Colorado, this window runs from November 2, 2013 to March 1, 2014; in Washington, the window runs from May 9, 2014 to September 6, 2014; and in Oregon, the window runs from August 2 to November 29. In each bandwidth, recreational sales policy changes only in the treated state at the time of treatment. These estimates, which include state-specific linear time trends, can be found in Table 9.

TABLE 9.
Estimated effect of the beginning of recreational sales on IPV, 60-day bandwidth, state-level treatment, by allowing time trend of treated state to break

	(1) Colorado	(2) Washington	(3) Oregon
No trend break	0.0562***	0.0423***	0.0351***
	(3.63)	(7.41)	(6.63)
Percentage change	14.1	12.0	17.7
Trend break	0.0559***	0.0422***	0.0349***
	(3.61)	(7.47)	(6.57)
Observations	264906	296594	294731

t statistics in parentheses
All models include year, month, day-of-week, and agency fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In each of the three states, the estimated treatment effect is larger in the narrower bandwidth than in the full sample: the estimated effects are 0.0562, 0.0423, and 0.0351 in Colorado, Washington, and Oregon, respectively. These effects represent a 14.1% increase in reported domestic violence incidents in agencies in Colorado, a 12.0% increase in reported incidents in Washington, and a 17.7% increase in reported incidents in Oregon. All of these estimated effects are statistically significant at the 1% level. I also estimate the model for each state in the narrower bandwidth using a Poisson specification. The estimated effects relative to the mean of reported incidents are 16.7% in Colorado, 11.8% in Washington, and 15.2% in Oregon.

Up to this point, I have enforced on this model the assumption that the state-specific linear time trends in the treated states are the same in the pre-treatment and post-treatment periods. Table 9 also displays estimates for each state when this assumption is relaxed, and the treated state is allowed to have separate pre-treatment and post-treatment linear time trends. In each of the three treated states, allowing the time trend to break at the treatment date has little impact on the estimated treatment effect and does not affect the estimated statistical significance of the result.

I provide a visual characterization of the effect in Figure 10 and Figure 11. Here, I estimate the model in the 120-day window around treatment for each state without including a treatment variable. Figure 1 plots the smoothed empirical density of the sum of the residuals from the without-treatment model for each of the three treated states, as well as an identical plot for the pooled residuals from the treated states. While the combined distribution does not perfectly shift to the right, multiple portions of the distribution shift to the right, which corresponds to the positive estimated treatment effect. In Figure 2, I present the pooled empirical CDF of these residuals for the treated states when the model is estimated without treatment; this figure makes it more obvious that the distribution of these residuals has shifted to the right following the treatment date.

FIGURE 10.
Density of residuals of model without treatment

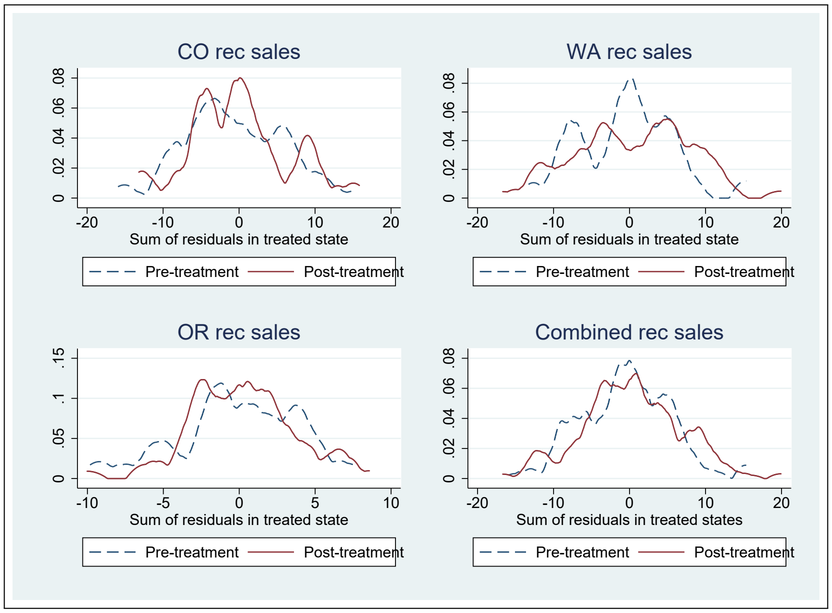
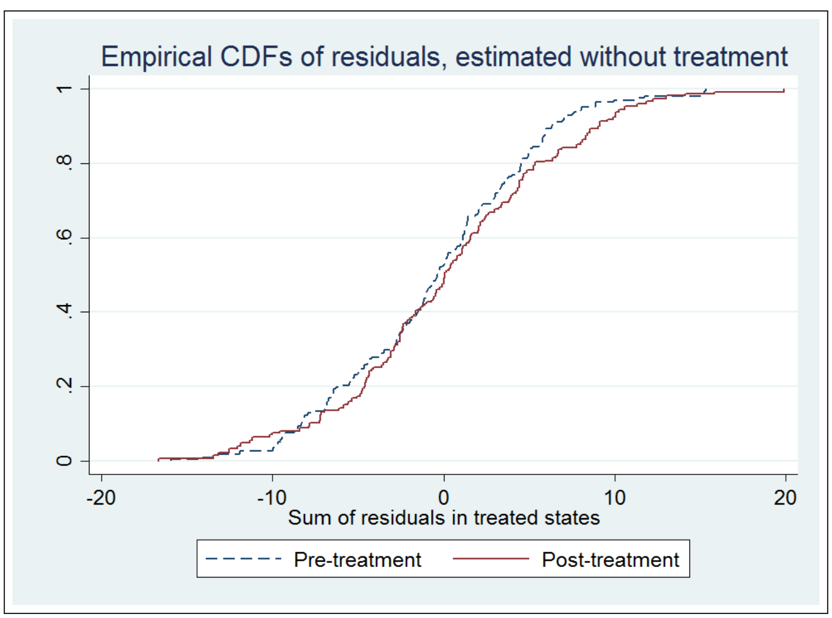


FIGURE 11.
Cumulative density of residuals of model without treatment



Potential Mechanisms

Is the estimated effect merely due to changes in reporting?

Victim reporting

One possibility is that the introduction of recreational marijuana sales does not actually cause more domestic violence incidents. Instead, it could be the case that individual victims are more likely to report incidents once the stigma of marijuana use has diminished. For instance, a victimized individual may previously have refrained from calling the police about a crime due to the presence of marijuana in the home. Because the estimated effect of recreational sales on domestic violence represents the increase in incidents reported by law enforcement agencies, this effect is composed of both the actual increase in incidents that occur and any potential increase in reporting of incidents that would have occurred absent treatment.

If the increase in reported incidents were driven by changes in individual reporting behavior, we would expect to see the largest estimated effects for groups that previously underreported these types of incidents. I use data from the National Crime Victimization Survey[18] from 2007-2015 to determine reporting rates for the types of intimate partner violence incidents taken from the NIBRS data. This survey is a questionnaire conducted by the Bureau of Justice Statistics to measure crime victimization, and it includes a variable specifying whether the incident described by the respondent was reported to the police.

[18]This data was accessed through the University of Michigan's Inter-university Consortium for Political and Social Research's National Archive of Criminal and Justice Data.

I investigate how the treatment effect varies across severity levels of assault. As mentioned earlier, the three categories coded in the data are aggravated assault, simple assault, and intimidation. Table 10 presents estimated effects effects by type of intimate partner violence for the county- and distance-weighted treatment definitions clustered at the county level, which provide the most conservative estimates of statistical significance. While each type of assault increases when recreational sales begin, the largest effect size in term of percentage change from the mean occurs in intimidation, which is the least severe category. Interestingly, the middle category, simple IPV, sees the smallest increase.

TABLE 10.
Estimated effect by severity level of IPV, county and distance-weighted treatment, clustered at treatment level, state-specific linear time trends

	(1) IPV	(2) Agg. IPV	(3) Sim. IPV	(4) Int. IPV
County-level treatment	0.0211	0.00443***	0.0131	0.00355*
	(1.53)	(3.03)	(1.07)	(1.96)
Percentage change	6.2%	9.8%	4.9%	13.3%
Distance-weighted treatment	0.0177*	0.00264	0.0108	0.00430**
	(1.72)	(1.58)	(1.19)	(2.22)
Percentage change	5.2%	5.8%	4.0%	16.2%
Alternative distance-weighted treatment	0.0205	0.00360**	0.0130	0.00399**
	(1.59)	(2.37)	(1.13)	(2.23)
Percentage change	6.0%	7.9%	4.9%	15.0%
Observations	3429594	3429594	3429594	3429594

t statistics in parentheses
All models include year, month, day-of-week, and agency fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

This pattern may indicate that the total effect is a composite of multiple mechanisms. Intimidation, as the least severe type of IPV, is the most marginal category. Individuals who previously would not have reported assaults may have been more likely not to report instances of intimidation, while reporting a higher fraction of the simple assaults that truly occurred and an even higher fraction of the aggravated assaults that occurred. The effects from this mechanism would be largest for intimidation and smallest for aggravated assault.

Another way to investigate individual underreporting is to look at incidents involving drugs. If domestic violence incidents involving drugs were reported less frequently than similar incidents not involving drugs, the estimated increase in intimate partner violence could result from increased reporting by victims of drug-using offenders. However, this does not appear to be the case. The NCVS data includes a question about whether the offender of the crime was using any drugs at the time of the offense. From 2007 to 2015, victims of intimidation, simple assault, or aggravated assault by intimate partners were more likely to report the incident to the police (63.3%) if the offender was using drugs than if the offender was not using drugs (57.7%). There is no evidence that incidents involving drugs are underreported relative to those incidents not involving drugs; in order for the main estimated effect to be driven by changes in victim reporting behavior, victims of drug-using offenders (who already reported crimes at a higher rate) would have to differentially increase their reporting relative to victims of non-drug-using offenders.

Another possibility is that certain demographic groups are more likely to underreport these types of incidents than others. If these underreporting segments see the largest increase in domestic violence incidents, it could provide support for the idea that the estimated effect is truly just a change in individual reporting behavior.

In the NCVS data, black victims are more likely to report incidents to the police than white victims across all types of violent crime. Specifically, for the years 2007-2015, black victims report intimidation, simple assault, or aggravated assault

TABLE 11.
IPV by victim race, full sample, state-specific linear time trends

	(1) All	(2) White	(3) Black	(4) Other
Legalization	0.00155	0.00265	0.000966	0.000168
	(0.43)	(1.24)	(0.45)	(0.45)
Market opens	0.0126***	0.00626*	0.00496***	0.00125
	(4.48)	(1.99)	(3.51)	(1.31)
Percentage change	3.5%	2.6%	4.5%	18.9%
Observations	3834926	3834926	3834926	3834926

t statistics in parentheses
All models include year, month, day-of-week, and agency fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

by an intimate partner to the police 65.4% of the time, while white victims (58.6%) and victims of other races (59.2%) report these crimes at lower rates. However, the estimated effect of the recreational marijuana market opening is larger in percentage terms (4.5%) for black victims than for white victims (2.6%). Table 11 presents the estimated effect of the market opening by race; while the estimated effect is extremely large in percentage terms (18.9%) for victims of other races, the estimated effect is not statistically significant from zero. Additionally, these victims of other races account for under 3.5% of the total pre-treatment incidents.

Based on these data on reporting behavior by type of victim, it does not appear to be the case that the estimated increase in domestic violence incidents is being driven by groups that were previously reporting incidents less frequently than the average level of victim reporting.

Police enforcement

Another potential mechanism for the estimated increase in domestic violence incidents is through police resources: if law enforcement agencies were previously devoting employee-hours to policing drug crimes in a way that kept them from responding to all domestic violence calls, a shift from drug enforcement to enforcement of other crimes could manifest as an increase in domestic violence. While the NIBRS data represent any incident to which a law enforcement agency arrived, I do not have data on what fraction of domestic violence calls to an emergency number receive a response in the form of an officer.

One important aspect of pinning down the mechanism of this positive effect is understanding effect heterogeneity. If the estimated treatment effect were due to a shift in police reporting behavior, we would expect to see similar estimated treatment effects across all days of the week and across all times of day. However, this is not the case. The estimated effect is larger on weekends and in the evenings.

In Table 12, I present estimated treatment effects by day of the week, as well as the percentage change from the mean that each of the effects represents. The estimated positive effects are concentrated on the weekend: the largest estimated effect (12.9%) occurs on Sunday; the next-largest estimated effect, which occurs on Saturday, is barely half as large (6.6%). The next two largest effects occur on Friday and Monday, each of which contain a portion of the weekend. Friday night after individuals return from work represents the beginning of the weekend, and the early hours of Monday morning (after 12 a.m. Sunday night) represent the end of the weekend. Table 13 presents similar estimates for weekdays relative to weekends.

The way in which the estimated treatment effect is distributed throughout the week suggests that the mechanism is not purely a reporting story, either on the part of individuals or on the part of law-enforcement agencies. In order for the effects to be caused only by changes in reporting, either individuals or law-enforcement agencies would need to differentially change their reporting behavior

TABLE 12.
Full sample IPV by day of week, distance-weighted treatment, standard errors adjusted for clustering at the county level

	(1) Estimated effect	(2) Percentage change
Monday	0.0132	4.3%
	(1.00)	
Tuesday	-0.00566	-1.9%
	(-0.55)	
Wednesday	-0.00222	-0.7%
	(-0.21)	
Thursday	0.00813	2.7%
	(0.69)	
Friday	0.0157	4.9%
	(1.51)	
Saturday	0.0271	6.6%
	(1.08)	
Sunday	0.0582*	12.9%
	(1.68)	
Observations	3429594	

t statistics in parentheses
All models include year, month, day-of-week, and agency fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE 13.
Full sample IPV by weekend, distance-weighted treatment

	(1) Clustered by state	(2) Clustered by county	(3) Clustered by agency
Weekday	0.00764	0.00764	0.00764
	(1.46)	(0.97)	(1.01)
Weekend	0.0314	0.0314*	0.0314
	(1.44)	(1.71)	(1.59)
Observations	3429594	3429594	3429594

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

by day of the week after treatment. While this is theoretically possible, it is unlikely. Neither group should be expected to increase reporting on the weekend but not during the week. However, while this is evidence that the mechanism is not purely reporting, it does not explicitly rule out an impact of a change in reporting behavior.

While this pattern does not pin down the mechanism of the effect, it is consistent with multiple other potential mechanisms. The largest increases in reported incidents could occur on the weekends if individuals spend more time at home in order to use marijuana, if conflicts about a partner purchasing marijuana occur more frequently on the weekend, if individuals consume marijuana and large quantities of alcohol together on the weekends, or some combination of these three behaviors. Specifically, it is easy to imagine and individual who, without available recreational marijuana, would have spent some or all of the weekend doing some activity away from home. Once sales begin and this individual has access to marijuana, he or she decides instead to spend that time at home using marijuana and, as suggested by Wen et al. (2015), alcohol.

In Table 14, I present estimates of the effect size by time of day; I divide the day into six parts: early morning (12 a.m.-4 a.m.), morning (4 a.m.-8 a.m.), late morning (8 a.m.-12 p.m.), afternoon (12 p.m.-4 p.m.), evening (4 p.m.-8 p.m.) and night (8 p.m.-12 a.m.). Incidents are assigned to specific day-part categories based on the incident hour variable in the NIBRS dataset. The estimated effect is positive for the late morning, afternoon, evening, and night segments, and the largest effect sizes by percentage change from the pre-treatment mean occur in the afternoon, evening, and pre-midnight hours. The largest estimated effect does not occur after midnight, when we would expect to see individuals using marijuana, alcohol, or both simultaneously; this implies that the effect is not being caused by a direct effect of consumption of marijuana. Instead, it seems as though the largest effect occurs at times when individuals are likely purchasing marijuana: during the day and after work or school.

TABLE 14.
IPV by time of day, full sample, clustered by state

	(1) Estimated effect	(2) Percentage change
12 a.m.-4 a.m.	-0.000416 (-0.24)	-0.6%
4 a.m.-8 a.m.	0.00000657 (0.01)	0.02%
8 a.m.-12 p.m.	0.00152* (2.01)	3.4%
12 p.m.-4 p.m.	0.00269 (1.08)	5.0%
4 p.m.-8 p.m.	0.00293*** (3.57)	4.1%
8 p.m.-12 a.m.	0.00592 (1.63)	6.1%
Observations	3429594	

t statistics in parentheses
All models include year, month, day-of-week, and agency fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

One method to get at the question of police resources is to look at data from traffic stops in the states that began selling recreational marijuana. I use data from the Stanford Open Policing Project[19] to estimate the effect of legalization and recreational sales beginning on traffic stops, searches, and contraband found

[19]https://openpolicing.stanford.edu/

by state police in Washington and Colorado.[20] While state police are likely not responding to the majority of domestic violence calls, a decrease in traffic stops when recreational sales begin could indicate that law enforcement agencies have substituted some fraction of their time from policing traffic violations to enforcing other types of crime, like domestic violence.

In Table 15, I present the estimated effects of both the legalization of marijuana and the beginning of recreational sales in Washington on the total number of traffic stops, the total number of searches, and the total number of times contraband was found. Table 16 presents the same estimated effects for Colorado. In both tables, the model uses data from 2012-2015 and includes year, month, and day-of-week fixed effects.[21]

TABLE 15.
Traffic stops, searches, and contraband found in Washington, 2012-2015

	(1) Stops	(2) Searches	(3) Contraband
Legalization	-30.23	-13.87***	-6.868***
	(-0.39)	(-4.20)	(-7.99)
Percentage change	-0.92%	-15.6%	-45.5%
Market opens	1.242	-0.596	0.271
	(0.03)	(-0.36)	(0.63)
Percentage change	0.04%	-0.72%	2.2%
Observations	1461	1461	1461

t statistics in parentheses
All models include year, month, and day-of-week fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[20]While the website provides a file for Oregon, all of the necessary variables are empty.

[21]The Washington data erroneously duplicates all stops in the month of April in 2015; the results are strikingly different if these are not corrected.

TABLE 16.
Traffic stops, searches, and contraband found in Colorado, 2012-2015

	(1) Stops	(2) Searches	(3) Contraband
Legalization	-58.11	-2.905***	-2.275***
	(-0.96)	(-5.43)	(-5.83)
Percentage change	-4.9%	-50.3%	-66.1%
Market opens	12.79	-0.723	-0.293
	(0.13)	(-0.82)	(-0.46)
Percentage change	1.10%	-14.8%	-10.1%
Observations	1461	1461	1461

t statistics in parentheses
All models include year, month, and day-of-week fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

The pattern of changes seen at the times the laws take effect and the times when recreational sales begin is consistent across Colorado and Washington. In both states, the number of searches and the number of times contraband is found decrease dramatically when the law legalizing marijuana use takes effect, but there is almost no change when recreational sales begin. The number of daily stops decreases by a relatively small fraction in both states when the laws take effect, and there is no evidence that the number of stops changes in either state when recreational sales begin. In terms of police resources, these estimates suggest that while some police time constraint could be slackened when legalization takes effect, it is not the case that fewer traffic stops are happening once recreational sales begin.

Another way to investigate whether the estimated effect could be caused by changes in police enforcement behavior is to look at crimes police respond to in similar ways to domestic violence. Using the NIBRS data, I estimate the model using non-intimate-parter home assaults, which should be the most similar crime to intimate partner assaults.

Both crimes require some sort of call to police, who must then respond and file a report about the incident. If the estimated increase in intimate partner violence incidents is caused by additional available police resources, we should see similar increases in non-intimate-partner home assaults. However, this is not the case. Table 17 presents these estimates, which show that there is virtually no change in non-intimate-partner home assaults when recreational sales begin.

TABLE 17.
IPV and non-IPV home assaults, full sample, clustered by state

	(1) IPV	(2) Non-IPV home assaults
Legalization	0.00155	0.00817
	(0.43)	(1.26)
Market opens	0.0126***	0.000813
	(4.48)	(0.18)
Observations	3834926	3834926

t statistics in parentheses
All models include year, month, day-of-week, and agency fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Neither of these techniques suggest that there is increased police enforcement of crimes that were already happening prior to the advent of the recreational marijuana market. Additionally, Law Enforcement Officers Killed or Assaulted (LEOKA) data, which provide the number of officers hired in each state in each year, do not show a consistent upward trend in hiring in treated states during the years in the sample. In sum, there does not appear to be any evidence that the estimated increase in domestic violence is merely additional enforcement of pre-existing crimes.

Is there increased conflict within relationships?

If the estimated increase in domestic violence incidents does not stem from changes in police or individual reporting behavior, it must be the case that more of these incidents truly occur once recreational sales of marijuana begin. One potential source of new or increased conflict within a household comes from the introduction of a new pressure on the household's budget constraint: the legal availability of recreational marijuana expands the household's choice set, and purchasing marijuana requires using some portion of the budget that was previously devoted to some other expense. If one individual in the household spends money previously used for necessities on marijuana, tension will increase, and in some cases, this will lead to an increase in domestic violence incidents.

To investigate changes in household expenditure in states in which recreational sales begin, I use data from the Consumer Expenditure Survey,[22] in which American consumers fill out a detailed diary of all expenditures over a two-week period. These expenditures are classified by Universal Classification Code, which provides extremely fine categorization of expenses.

Each of the two weeks of an individual's diary are assigned separate identification codes in the data; however, because the data does not identify the specific start date within a month, I collapse an individual's two diary-weeks into one observation per individual. Each observation then consists of an individual diary and is assigned to the month in which the majority of diary days occurred.[28] The finest level of geographic information available is the individual's state. An

---

[22]Data taken from the Bureau of Labor Statistics' public-use microdata data files page at https://www.bls.gov/cex/pumddata.htm.

[28]For example, a diary recording expenditures on 8 days in May, 2013 and on 6 days in June, 2013 is assigned to May, 2013.

observation is then considered treated if the individual comes from a state with an active recreational market. I partition the set of expenditures into broad categories: food, alcohol, housing, apparel, transportation, entertainment.[23]

One important consideration when using this dataset is the size of outliers relative to normal expenditures within categories. For example, one household in the dataset made a $167,500 retirement contribution during its two-week diary period, while another household spent $3,995 on a catered affair at a restaurant. While these are likely genuine data (as opposed to data-entry errors), it is unlikely that these are regular expenses that are truly representative of the household's spending habits. I investigate the data across three methods of dealing with these outliers: using the full sample, in which case the outliers are treated as true data and not amended; winsorizing the data by category, in which case the extreme values of the data above a certain percentile are replaced with the value of the cutoff percentile; and trimming the data by category, in which case observations above a certain percentile are dropped from the sample.

Another important consideration is how "smooth" expenditures are within each category. Many household expenditures are both large and infrequent, like the previously mentioned retirement contribution or purchasing a house in cash during the diary period. Two rough methods of "smoothness" of various categories of expenditures are the standard deviation of expenditure divided by the mean of expenditure and the percentage of zeroes within a given category. For example, savings is the least "smooth" of the categories: the standard deviation of expenditure is 36 times as large as the mean, and 79.7% of the observations

[23] I treat any catered affairs at restaurants as entertainment, not food expenditure, as these are infrequent events that can cost many multiples of an individual's food purchases for the diary period. Two additional categories, housing and savings, exhibit large outliers.

are zeroes. At the other end of the spectrum, food is the “smoothest” category by a large margin. The standard deviation of food is smaller than the mean (in every other category, the standard deviation is at least twice as large as the mean), and only 0.75% of diaries report zero food expenditures (the category with the second-fewest zeroes, housing, sees no expenditures reported 11.2% of the time). As a result, we should expect the estimated effects for the food category to be the least responsive to various methods of dealing with outliers, and this is indeed the case.

More importantly, food is likely to be a frequent expense whose budget is reasonably fungible. It is also a necessity, and changes in food expenditures could certainly cause strife within a household. In Table 18, I present the estimated effects of legalization and the start of recreational sales on expenditures for six categories in the data, excluding housing and savings.[24] This table uses data from which observations above the 99th percentile of each category have been trimmed; the same tables using the full and winsorized datasets are available in the appendix. All models include year, month, and state fixed effects. Treatment is pooled across treated states.

TABLE 18.
Consumer expenditure by category, pooled treatment, trimmed at 99, clustered by state

	(1) Food	(2) Alcohol	(3) Healthcare	(4) Entertainment	(5) Transportation	(6) Apparel
Legalization	9.989	5.284***	-9.029	23.63	61.81***	15.01*
	(0.70)	(3.89)	(-1.61)	(1.35)	(5.26)	(1.91)
Market opens	-34.20***	-4.187***	-6.170***	22.12	-0.221	-10.59*
	(-4.25)	(-3.35)	(-2.99)	(1.37)	(-0.07)	(-1.86)
Observations	20302	20302	20302	20302	20302	20302

t statistics in parentheses
All models include year, month, and state fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

[24]Housing is likely less fungible than the categories presented; the vast majority (79.7% of households) report $0 in savings during the diary period.

While significant effects are estimated for multiple categories, many of these estimates (and their precision) fluctuate significantly depending upon the method of dealing with outliers in the data. However, the estimated effect of beginning recreational sales on food expenditure is extremely stable across these methods. In the table provided, the estimated effect represents an average decrease in food expenditures of $34.20 per two-week diary entry in households treated by the opening of recreational marijuana markets. Appendix Tables A.3 and A.4 show that these estimates are relatively stable across various percentiles of both winsorization and trimming. Appendix Tables A.5 and A.6 present expenditure categories for the full sample and the winsorized sample. Additionally, in each of the estimated models, the estimated negative treatment effect is composed of an estimated negative treatment effect on food expenditures in each of the treated states.

The estimated decrease in food expenditures is larger for households in the bottom tercile of the income distribution. Table 19 presents the estimated effect of recreational markets opening on food expenditures using an interaction between treatment and an indicator variable denoting whether the household is in the bottom third of the income distribution, as calculated from the before-tax income variable in the consumer expenditure data.[25] The estimated differential effect for low-income households relative to households in the top two-thirds of the income distribution is statistically significant and approximately three times as large as the estimated effect for households not in the bottom tercile.

[25] A small fraction (37 observations) of the household are listed with negative before-tax incomes; the second column of Table 19 presents estimates with these negative-income households dropped from the sample. The estimated effects are qualitatively unchanged.

This larger decrease in food expenditures for low-income households is consistent with a conflict mechanism: low-income households' budget constraints likely bind more tightly than the budget constraints of high-income households, and a larger food expenditure effect likely causes more strife in a low-income household than a high-income household. Although there is no way to match incomes to victims in the NIBRS data, this suggests that the estimated increase in intimate partner violence is largely driven by low-income households.

TABLE 19.
Food expenditure by before-tax income tercile, pooled treatment, clustered by state

	(1) Full sample	(2) Non-negative incomes
Legalization	1.264	5.521
	(0.10)	(0.40)
Market opens	-13.08	-13.03
	(-1.38)	(-1.38)
Bottom tercilexLegalization	27.37*	27.91*
	(1.81)	(1.84)
Bottom tercilexMarket opens	-44.82***	-44.80***
	(-2.81)	(-2.82)
Observations	20507	20470

t statistics in parentheses
All models include year, month, and state fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

While the estimated effect may seem large in dollar terms, it is not unreasonable given other estimates of marijuana expenditure. For example, Kilmer (2016) estimates that residents of Vermont (in which recreational sales are not legal) spend \$125 million to \$225 million on marijuana each year. This result implies that the average person spends \$200-\$360 per year on marijuana, or \$17-\$30 per person per month. Once we account for the facts that many households are composed of multiple individuals and that many individuals in treated states begin to purchase marijuana once legal sales begin, the estimated effect size–an average household decrease of $68.40 in food expenditures over a four-week period–seems much more reasonable.

These estimates provide evidence that treated households spend significantly less money on food once recreational marijuana markets open in their states. This reduction in food expenditure is likely to be a source of conflict among budget-constrained households, especially those households in the bottom of the income distribution. This suggests that at least a portion of the estimated increase in domestic violence incidents can be attributed to increased conflict within relationships as a result of financial conflict caused by substitution of expenditure from the household's food budget to marijuana.

Is the effect caused by individuals spending more time at home?

Another possibility is that when recreational sales start, some individuals choose to purchase marijuana and consume it at home rather than doing recreational activities away from home (like going to a bar, for instance). If this were the case, domestic violence could be occurring at the same rate, conditional on victim and perpetrator time spent together, as before the policy change took effect, but the larger population at home with partners would mechanically lead to an increase in the number of incidents. This mechanism is not consistent with the estimated heterogeneity of effect by time of day: there is no estimated increase in IPV from 12 a.m.-4 a.m., when this type of substitution would likely occur.

To investigate this effect, I estimate the treatment effect of the opening of recreational markets on assaults at bars/nightclubs, as determined by the location type variable in the NIBRS dataset. Assaults are defined exactly as they are in the main specification: the variable consists of aggravated assaults, simple assaults, and intimidation. These results are presented in Table 20.

If it were the case that individuals substituted away from bars to staying home, there should be a mechanical reduction in bar assaults that corresponds with the estimated increase in domestic violence incidents. Instead, as seen in the table, the estimated effect of the beginning of recreational sales on bar assaults is slightly positive and statistically insignificant from zero. This implies that there is not a mechanical exposure effect as a result of substitution away from other recreational activities.

TABLE 20.
IPV and bar assaults, full sample, clustered by state

	(1) IPV	(2) Bar assaults
Legalization	0.00155	0.000391
	(0.43)	(1.01)
Market opens	0.0126***	0.00137
	(4.48)	(1.40)
Observations	3834926	3834926

t statistics in parentheses
All models include year, month, day-of-week, and agency fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Is there a direct effect of marijuana consumption?

Another possible mechanism that could cause an increase in intimate partner violence incidents is a complementary relationship between marijuana use and alcohol use: if marijuana use leads to more violence, or if it exacerbates the effect of alcohol use. Past literature (Card and Dahl, 2011) suggests that a complementarity with alcohol could produce an increase in violent behavior. I investigate whether there is evidence of this effect through the consumer expenditure survey data and the uniform crime reporting data.

In the offender segment of the NIBRS, there exist variables documenting whether the officer reported the offender as being under the influence of alcohol or under the influence of drugs. However, these variables are likely subject to significant underreporting by law enforcement officers who respond to incident reports, as discussed in Card and Dahl (2011) and Lindo et al. (2018). In Table 21, I present estimated effects of the opening of the recreational marijuana sales market on IPV incidents marked as associated with alcohol use and drug use. The effect on alcohol-related incidents is small, positive, and statistically insignificant, while the effect on drug-related incidents is small, negative, and statistically insignificant. It is important to note that officer reporting behavior of these variables may not be constant across treatment; once recreational sales of marijuana begin, officers may be more likely not to note the presence of marijuana in their report. While both of these estimated effects would suggest that the estimated increase in domestic violence incidents are not caused by a direct effect of marijuana and alcohol complementarity, these indicator variables are likely not very reliable.

While the consumer expenditure model using data trimmed at the 99th percentile estimates a significant negative effect of the opening of the recreational market on alcohol expenditure, other methods of dealing with outliers (using the full data set and winsorization) do not estimate an effect statistically different from zero. Table 22 depicts the instability of these effects across a various trim percentiles; the estimated effect is also not consistent across various winsorization percentiles. Despite the inconsistency of these estimates, none of the models

TABLE 21.
Estimated effect of the beginning of recreational sales of marijuana on IPV by type, state-level treatment, state-specific linear time trends, standard errors adjusted for clustering in state

	(1)	(2)	(3)
	IPV	Susp. alc.	Susp. drugs
Market opens	0.0104***	0.0007	-0.0001
	(4.99)	(0.35)	(-0.41)
Observations	3429960	3429960	3429960

t statistics in parentheses
All models include year, month, day-of-week, and agency fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

estimate a significant positive effect on alcohol expenditure. This implies that the increase in domestic violence incidents is not driven by households consuming more alcohol and subsequently committing more violent crime.

TABLE 22.
Alcohol expenditure by trim percentile, pooled treatment, clustered by state

	(1)	(2)	(3)	(4)
	99th	95th	90th	80th
Legalization	5.284***	2.275*	1.910***	0.0864
	(3.89)	(1.89)	(7.28)	(0.12)
Sales start	-4.187***	1.708	-1.357	-0.998**
	(-3.35)	(1.62)	(-0.69)	(-2.17)
Observations	20302	19482	18457	16406

t statistics in parentheses
All models include year, month, and state fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

I also estimate the effect of the opening of the recreational marijuana market on DUI arrests using UCR arrest reports[26] as the dependent variable. I present these results in Table 23. These estimated effects of the opening of recreational

[26]This data was accessed through the University of Michigan's Inter-university Consortium for Political and Social Research's National Archive of Criminal and Justice Data.

sales markets are slightly positive, but not statistically significant from zero. These results are not consistent with the story told in Anderson et al. (2013) regarding traffic fatalities: medical marijuana legalization causes individuals to substitute away from alcohol to marijuana, which decreases drunk driving and alcohol-related traffic fatalities. However, it it important to note that the UCR data does not specify a separate category of arrest for driving under the influence of non-alcohol drugs, so any arrests for driving under the influence of marijuana are included in this arrest category. As such, it may be the case that drunk driving decreases while driving under the influence of marijuana increases, but this is inconclusive.

TABLE 23.
Pooled full sample UCR DUI arrests by sample of agencies, clustered by state

	(1)	(2)
	All agencies	Agencies reporting every day
Legalization	0.770**	1.868
	(2.44)	(1.23)
Market opens	0.391	1.340
	(1.05)	(0.92)
Observations	286527	64978

t statistics in parentheses
All models include year, month, day-of-week, and agency fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

As discussed previously, the estimated increase in intimate partner violence incidents is largest in the afternoon, not at night. This provides additional evidence that the result is not working through a direct effect of complementary usage of marijuana and alcohol.

Additionally, if the true mechanism were a direct effect of increased marijuana purchases leading to increased alcohol consumption and domestic violence incidents, we would expect to see the estimated increase in incidents increasing over time.

Instead, as discussed previously, the estimated effect is larger in the 60-day window around the opening of the market than in the full sample, which implies that the effect is diminishing over time, not growing.

The consumer expenditure data does not suggest that individuals spend more money on alcoholic beverages once the recreational market opens, and the uniform crime reporting data does not suggest that individuals are more likely to be arrested for DUI once sales start. The estimated effect is not largest at night, when we expect individuals to be consuming marijuana and alcohol simultaneously. Finally, the estimated effect of sales starting on intimate partner violence incidents does not grow over time. Together, these pieces of evidence suggest that the increase in domestic violence incidents are not caused by a direct effect of increased alcohol use as a result of the beginning of recreational sales of marijuana.

Conclusion

Using the FBI's NIBRS data on reported crime incidents from 2012-2015, I estimate the effect of the opening of the recreational marijuana market on intimate partner violence incidents reported by law enforcement agencies; I find that the beginning of recreational sales of marijuana in a treated state is associated with a 2.9-6.2% increase in the number of domestic violence incidents reported. This increase occurs only when recreational markets open, and there is no estimated effect of the passage of laws that legalized recreational marijuana.

The effect could be caused by a variety of mechanisms, including conflict, exposure, changes in reporting behavior, and the direct effect of complementarity between marijuana use and heavy alcohol use. Using a number of additional datasets, I investigate these mechanisms. The supporting evidence is most consistent with a conflict mechanism: individuals–specifically those in low-income households whose budget constraint binds tightly–spend some of the household budget on food, which causes conflict within relationships. As a result, increased tension between partners causes an increase in intimate partner violence incidents.

The evidence presents a somewhat surprising picture: opening legal recreational marijuana markets leads to an increase in domestic violence, while legalizing recreational marijuana use does not. This suggests that the effect probably has little to do with marijuana at all: instead, partners do not match on preferences for the good sold in the new market, and conflict results. Indeed, future work could explore the question of whether the expansion of markets of other controversial goods or products produces similar effects.

In the longer term, potential partners will likely begin to sort along this preference dimension in ways they had not previously, causing the effect of opening the new market to diminish over time. Couples may also separate or divorce at a higher rate in the short term. In the long run, individuals would learn to incorporate this specific type of preference into their matching behavior, at which point conflict would likely return to its pre-treatment levels.

While this paper is not a comprehensive analysis of the costs and benefits of recreational marijuana legalization, it provides information on one important outcome. This estimated effect may be surprising to some, as the hypothesized substitution between alcohol and marijuana would predict a decrease in domestic violence incidence when recreational sales begin. As such, this result represents a potentially unanticipated and previously unconsidered social cost associated with legalizing recreational marijuana. As policymakers and voters consider the social costs and benefits of future marijuana liberalization legislation, they should take into consideration that the evidence from this study suggests that opening a recreational marijuana sales market in a state leads to increased reported intimate partner violence incidents, and that these additional reports are likely additional actual assaults.

CHAPTER IV

TYING EACH OTHER TO THE MAST: COMMITMENT DEMAND IN GROUP TASKS

This is unpublished co-authored material with Michael Kuhn and Jeffrey Naecker.

Introduction and Literature

Individuals with time-inconsistent preferences may be aware of these inconsistencies. For instance, someone may decide on an exercise plan in which she wakes up early to go the gym three days per week. In the present, the individual has ascertained that the benefits from this future action outweigh the costs. When the time comes to wake up and work out, however, she may want to hit the snooze button and go back to sleep. In this new present, the individual has decided that the costs (getting out of bed) outweigh the benefits. If the individual is aware of potential time inconsistencies in her preferences, she may seek a commitment device: she may be willing to make some payment in order to change the prices of her future choices so that she can follow through on the action her present self prefers. Strotz (1955) pioneered the idea of these time inconsistencies in utility maximization, and O'Donoghue and Rabin (1999) formalize the idea of naive (unaware of these inconsistencies) and sophisticated (aware of these inconsistencies) consumers.

Bryan et al. (2010) provide an overview of the literature on commitment devices; there is anecdotal evidence of time inconsistency of preferences in domains from smoking to New Year's resolutions to weight loss. Laboratory experiments have found that individuals demand commitment in situations like choosing movie titles and avoiding sugary foods. Thaler and Benartzi (2004) formulate a retirement savings plan that increases contributions over time automatically to act as a commitment device for employees to increase their savings rates; Giné et al. (2010) conduct an experiment in which individuals choose to take the opportunity to commit to a savings plan to quit smoking.

Some literature investigates the behavior of individuals in rotating savings and credit associations (ROSCAs), which provide informal savings mechanisms for individuals in settings without access to formal banking services. In these situations, a group of individuals will meet on a regular basis to contribute money into a collective "savings pot." At each meeting, a different individual takes home the entirety of the contributions. Gugerty (2007) uses data from ROSCAs in Kenya to argue that these groups serve as a collective commitment mechanism for individuals with time-inconsistent preferences, and Basu (2011) models these ROSCAs as commitment devices. However, in situations involving saving in underdeveloped nations, there are benefits to participation beyond mere commitment: taking part in a ROSCA decreases the likelihood that those funds will be used by another family member or stolen.

While there has been significant research conducted on individual commitment choices and follow-through, less work has been done on joint commitment decisions. Specifically, there is little laboratory evidence of whether commitment behavior changes when the commitment decision is undertaken bilaterally. Consider, for instance, the example of deciding to commit to a regular exercise plan with a gym buddy instead of committing to the same exercise regimen alone. Do individuals demand these type of arrangements? Does this type of group commitment lead to greater follow-through of the desired action (in this case, getting to the gym) than individual commitment?

In this paper, we conduct an experiment to estimate the effect of group commitment settings on commitment demand relative to individual commitment demand. We recruit individuals from Amazon's Mechanical Turk to complete an initial task and verify that the individuals would like to complete the same task in the future. Once subjects have indicated that they would like to complete the future task, they are randomized into one of a number of treatments and presented with the opportunity to demand commitment.

When considering the effect a partner may have on an individual's commitment choice, there are a number of possible motivations in play. Subjects may strategically attempt to manipulate partners' follow-through rates using the subject's own choice of commitment; subjects may not want to let down the "team", as described in Babcock et al. (2015); subjects may not want to demand commitment due to the risk of a partner not following through on the future task; and subjects may want to increase commitment demand if the choice is publicly observable, as described in Exley and Naecker (2016). We design four different group treatments, which we call joint, asymmetric, separate-with-knowledge, and separate-without-knowledge, to try to disentangle the various motivations driving any overall group effect.

Subjects demand significantly less commitment in all group treatments than they do in individual treatments. Of a possible \$2.25 budget available to them, subjects in the individual treatment demand an average of \$0.95 of commitment (42%), while individuals in group treatments demand an average of \$0.25 of commitment (10%); none of the average commitment demands for the various group treatments are statistically significant from each other. Similarly, subjects in the individual treatment demand some positive amount of commitment 52% of the time, while subjects in group treatments demand positive commitment only 21% of the time. Given the construction of the treatments, the large negative effect can be attributed to the team motivation: how a subject's commitment choice affects the partner's payoffs.

Experimental Design

In this project, we conduct an experiment in which we manipulate the type of commitment decisions available to participants. We randomly assign individuals to treatments in which the payoffs for completing a future task depend upon the actions of randomly assigned partners in various ways. Using a number of treatments, we attempt to determine the effects of a number of motivations for commitment demand: the risk of a partner defaulting on the future task; the public knowledge of the individual's commitment choice; the strategic motivation to manipulate a partner's behavior; and the desire not to let down the team.

Our subject pool consists of individuals recruited from Amazon's Mechanical Turk web service, which allows requesters to post human intelligence tasks (HITs) for workers ("Turkers") to complete in exchange for payment. These tasks can include work like completing surveys, image classification, and data processing. For our purposes, this method of recruiting participants allows us to find a large pool of subjects whom we can easily pay for their time. Additionally, we can use the Mechanical Turk API to contact these workers with links to the future task.[1]

---

[1] All interactions with Mechanical Turk and the subjects occurred through the Mechanical Turk API using the mTurkR package in R, including publication of HITs, contacting of workers, and granting of bonuses.

The initial HIT on Mechanical Turk tells participants that it will take approximately 15 minutes to complete the task, and that they will receive $2 for their participation.[2] Guidelines for being a good requester suggest paying workers at least the federal minimum wage ($7.25/hr) for their time, although many requesters pay significantly less.[3] We decide on this payment both to comply with good requester practice and to incentivize workers to do the future task: in order for subjects to make any decisions about committing to a future task, they must first have decided that they would like to complete the same task again.

Once a worker has accepted the HIT, the page redirects to a Qualtrics survey in the participant's browser window. After a few demographic questions[4], we provide subjects with a description of the main task. This task consists of counting the number of zeros in 10 10x15 tables[5], where each cell contains either a zero or a one. Participants may not move on to the next table until they have entered the correct answer. We intend the task to fulfill two requirements: it should be tedious, so that completing an identical future task is not costless to participants; but it should not be so challenging or unpleasant that subjects choose not to complete it again in the future.[6] The actual nature of the task has no bearing on our question: we merely want subjects to choose to complete the future task and subsequently to decide on a level of commitment.

---

[2]We restrict our HIT to be available to Turkers in the United States with at least an 80% HIT approval rate using the qualification requirements feature. Additionally, participants who completed the task in a pilot version or earlier batch of the survey are prevented from accepting the HIT again.

[3]http://turkrequesters.blogspot.ca/2012/09/tips-for-academic-requesters-on-mturk.html; http://wiki.wearedynamo.org/index.php/GuidelinesforAcademicRequesters

[4]Participants are asked for their age, sex, and highest completed level of education.

[5]Each page of the survey shows only one table.

[6]The initial version of this project included simple arithmetic problems, but we decided to exchange these for a more onerous task.

When an individual has correctly completed all of the tables, we ask if the subject would like to complete an identical task three weeks in the future for $2.25 and an additional $1 bonus to the Make-a-Wish Foundation. We select the value of $2.25 for two reasons: first, we think that individuals are more likely to want to do the future task if the payment is larger than the payment for the original point, since subjects' references points are likely to be $2; second, when we ask subjects to choose a level of commitment, the heuristic of choosing an even number ($2) will not lead to a zero-commitment choice as it would if the zero-commitment payment choice were $2. In general, subjects are quite willing to agree to do the future task. Of 809 participants who completed the survey in the main body of the experiment, 767 chose to do the future task (94.8%). This is important, as it provides a sizable sample while paying relatively few workers who do not provide commitment choices. Individuals who choose not to complete the future task are thanked for their time, and their surveys are redirected to the submission page on Mechanical Turk.[7]

Once subjects have indicated that they would like to complete the future task, they are randomized into one of five treatments, which we call individual, joint, asymmetric, separate-with-knowledge, and separate-without-knowledge. In each treatment, participants are told that they will receive an email within 48 hours that contains a link to the Qualtrics survey containing the future task, which will not become open for three weeks. Subjects must correctly answer two simple comprehension questions about the timing of the follow-up task to progress past the instructions page. Each subject subsequently receives an additional email 24-48 hours before the follow-up task becomes active with a simple reminder that the

[7]Once surveys are completed and HITs are submitted to Mechanical Turk, HITs are auto-approved after one hour.

window for the follow-up task will begin shortly. However, this reminder email does not contain a new link to the survey, so participants must either save the previous link or return to the original email to access the follow-up task.

Treatments

There are a number of motivations that could affect subjects' decisions about commitment in individual versus group scenarios, and we design the treatments to try to capture these different motivations. We label these motivations as follows:

1. Team: when a partner's payoff depends up on the subject's follow-through, we expect that individuals will not want to let down the "team."[8]

2. Public: when an individual's commitment choice is made observable to a partner, we expect that subjects will increase commitment demand.[9]

3. Strategic: when an individual's payoff depends upon the follow-through of an anonymous partner, individuals may be able to manipulate the partner's follow-through rate by selecting a higher level of commitment.

4. Risk: when an individual's payoff depends upon a partner completing a follow-up task, we expect the subject to demand less commitment to mitigate exposure to the risk that the partner fails to complete the task.

In the individual treatment, subjects are told that if they complete the future task, they will receive a payment of \$2.25, and that we will donate \$1 to the Make-a-Wish Foundation on their behalf. They also learn that if they do not complete the task, they will receive \$X, and the Make-a-Wish Foundation will not receive

[8] As seen in Babcock et al. (2015).

[9] As seen in Exley and Naecker (2016).

the donation ($0). For each of the treatments, the instructions page specifies that subjects will make a choice of $X on the following page, and the payoffs for all possible outcomes are shown in both table form and descriptive text. Once the individual has advanced to the next page, a price list is shown: subjects must select $X to be any value from $0 to $2.25 in $0.25 increments.[10] We include timing questions for both the instructions page and the choice page in an attempt to determine which subjects sped through questions to finish the survey as quickly as possible, without regard for the choice they are asked to make. Commitment demand in the individual treatment is the baseline from which we can estimate the effects of various group treatments on commitment demand.

We call the second treatment the joint treatment. In this case, individuals are told that they will be randomly assigned a partner who has also agreed to complete the future task. Each partner makes a commitment choice of $X, and one of the choices is randomly selected as the commitment level for the pair. If both partners complete the follow-up task, each partner receives $2.25 and the Make-a-Wish Foundation receives $1. If at least one of the partners fails to complete the follow-up task, each partner receives $X and the Make-a-Wish Foundation receives $0. Subjects are told that they will learn their partner's selected value of $X in the email they receive, and that their partners will also learn their selected value of $X.[11] In this treatment, partners' payoffs are dependent upon subjects' follow-through (team); subjects can attempt to influence partners' follow-through through choice of $X (strategic); partners learn subjects' choice (public); and subjects are

[10]When we discuss "commitment," we mean the value of $(2.25-X), which is the dollar amount the partner has "committed."

[11]The emails sent to participants specify the individual's chosen value of $X, the partner's chosen value of $X, and the randomly selected $X that will hold for the pair.

exposed to the risk that their partners may not complete the future task, which affects the subject's payoffs (risk). This treatment is designed to subsume all of the motivations in the group scenario.

The third treatment is asymmetric: individuals are randomly assigned a partner who has chosen to do the future task, but only one partner is chosen to have the responsibility of completing the follow-up task. Each partner selects a value of $X, and one partner is randomly selected to receive the email containing the link to the follow-up task. If the randomly selected partner completes the task, each partner receives $2.25 and the Make-a-Wish Foundation receives $1. If the randomly selected partner fails to complete the task, each partner receives the selected partner's choice of $X and the Make-a-Wish Foundation receives $0. Subjects are told that they will learn their partner's selected value of $X in the email they receive, and that their partners will also learn their selected value of $X.[12] In this treatment, subjects' partners learn the choice of $X (public) and subjects who are chosen can let down the team if they fail to complete the task (team), but a subject's choice of $X cannot influence the partner's payoffs (no strategic) or expose the subject to additional risk of lower payoffs if the partner fails to complete the task (no risk).

In the separate-without-knowledge treatment, individuals are again randomly assigned a partner. Here, each subject selects a value of $X that does not apply to the subject's partner. If both partners complete the future task, each partner receives $2.25 and the Make-a-Wish Foundation receives $1. If at least one of the partners does not complete the follow-up task, each partner receives the his/her

[12]The emails sent to participants specify whether the participants has been selected as the partner to complete the future task, the individual's chosen value of $X, the partner's chosen value of $X, and the $X that will hold for the pair.

own chosen value of \$X. In this way, each subject's payoffs depend only on the partner's follow-through, not on the partner's commitment choice. Subjects are explicitly told that they will not learn their partner's chosen value of \$X, and that their partners will not learn their chosen values. In this treatment, subjects may care about letting down the team by not following through (team) and are exposed to risk that their partners will not complete the task (risk), but their commitment choices are not divulged to their partners (no public) and their commitment choices do not strategically affect their partners' payoffs (no strategic).

Finally, the separate-with-knowledge treatment is identical to the separate-without-knowledge treatment, with the exception of the fact that subjects are explicitly told that both partners will learn both partners' commitment choices. Here, subjects may care about letting down the team by not following through (team); they are exposed to risk that their partners will not complete the task (risk); and their commitment choices become observable (public); however, these commitment choices still do not strategically affect their partners' payoffs (no strategic).

Results

We estimate the basic model

$$C_i = \alpha + \beta_J J_i + \beta_A A_i + \beta_{SK} SK_i + \beta_S S_i + \epsilon_i,$$

where:

- C_i is the commitment demanded in dollars by individual i;

- J_i is equal to 1 if individual i is in the joint treatment and equal to 0 otherwise;
- A_i is equal to 1 if individual i is in the asymmetric treatment and equal to 0 otherwise;
- SK_i is equal to 1 if individual i is in the separate-with-knowledge treatment and equal to 0 otherwise;
- S_i is equal to 1 if individual i is in the separate-without-knowledge treatment and equal to 0 otherwise.

The omitted category is the individual treatment, which allows us to interpret the intercept as the mean commitment demand for subjects in the individual treatment. Given these estimated coefficients and the design of the experiment, we can calculate the effects on commitment of the different possible motivations as follows:

- Strategic $= \beta_J - \beta_{SK}$
- Risk $= \beta_{SK} - \beta_A$
- Public $= \beta_{SK} - \beta_S$
- Team $= \beta_S -$ Risk $= \beta_S - (\beta_{SK} - \beta_A) = \beta_S - \beta_{SK} + \beta_A$

In this section, we present our main findings. Table 24 contains simple summary statistics of the sample.[13] Mean commitment across all treatments is \$0.39, or about 17% of the \$2.25 budget available for commitment; additionally,

[13] The survey data comes from two batches: one batch began on March 19-20, 2019; the second batch began on March 25, 2019.

across all treatments, 27.5% of subjects demand some non-zero level of commitment. In terms of demographics, 58.4% of the subjects are male, and the average age is 35.6 years old. We also ask about education level: the most commonly reported education level is "College (bachelor's degree)," which comprises 45.5% of individuals who agreed to complete the future task.

TABLE 24.
Summary statistics

Statistic	N	Mean	St. Dev.	Min	Max
Commitment	767	0.390	0.748	0	2.25
Non-zero commitment	767	0.275	0.447	0	1
Male	767	0.584	0.493	0	1
Age	767	35.641	10.840	18	75

In Table 25, we present the main results. Column 1 includes only treatment assignment as explanatory variables; mean commitment for subjects in the individual treatment is \$0.95, which represents 42% of the available \$2.25 budget. Commitment demand is significantly lower for each of the group treatments: for the joint treatment, mean commitment is \$0.29, which is 13% of the budget; for the asymmetric treatment, mean commitment is \$0.24, which is 11% of the budget; for the separate with knowledge treatment, mean commitment is \$0.28, which is 12% of the budget; for the separate without knowledge treatment, mean commitment is \$0.19, which is 9% of the budget. In columns 2, 3, and 4, we include sex, education level, and age, respectively. Figure 12 presents a bar graph of mean commitment levels chosen by treatment, while Figure 13 presents empirical CDFs of the commitment choices by treatment.

The inclusion of these additional demographic controls does not affect our results: the individual commitment level remains much larger than any of the

TABLE 25.
Commitment demand by treatment

	Dependent variable:			
	Commitment (in dollars)			
	(1)	(2)	(3)	(4)
Constant (Individual)	0.953***	0.916***	0.968***	1.068***
	(0.056)	(0.063)	(0.097)	(0.131)
Joint	−0.664***	−0.667***	−0.671***	−0.667***
	(0.079)	(0.079)	(0.080)	(0.080)
Sep. with knowledge	−0.673***	−0.673***	−0.679***	−0.673***
	(0.079)	(0.079)	(0.080)	(0.080)
Sep. without knowledge	−0.760***	−0.765***	−0.765***	−0.766***
	(0.079)	(0.079)	(0.080)	(0.080)
Asymmetric	−0.718***	−0.715***	−0.719***	−0.717***
	(0.079)	(0.079)	(0.079)	(0.079)
Male		0.065	0.064	0.058
		(0.051)	(0.051)	(0.051)
College			−0.078	−0.084
			(0.086)	(0.086)
Masters			−0.035	−0.034
			(0.107)	(0.107)
Some college			−0.009	−0.012
			(0.090)	(0.090)
PhD			−0.366	−0.367
			(0.233)	(0.233)
Professional			−0.201	−0.207
			(0.274)	(0.274)
Age				−0.003
				(0.002)
Observations	767	767	767	767
Note:				*p<0.1; **p<0.05; ***p<0.01

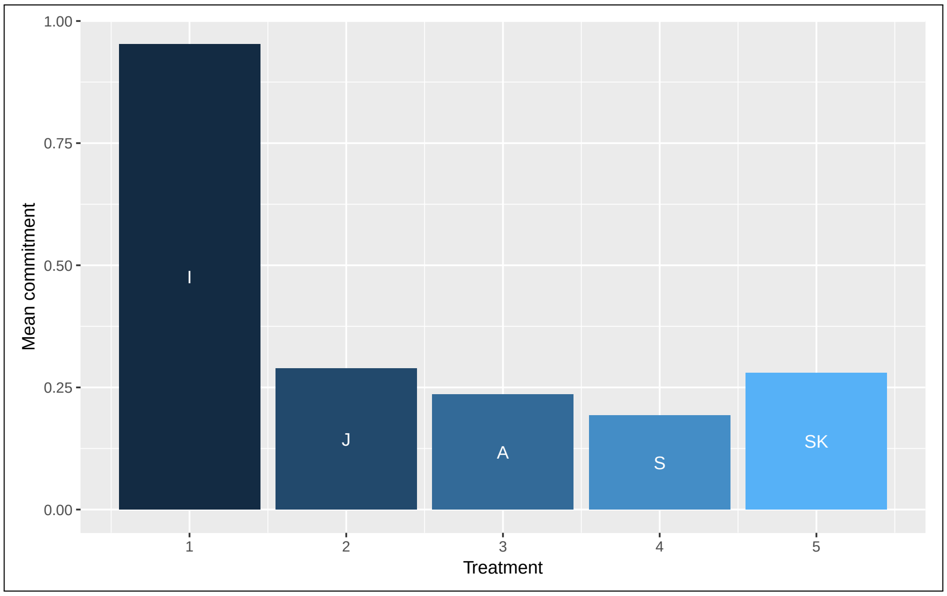
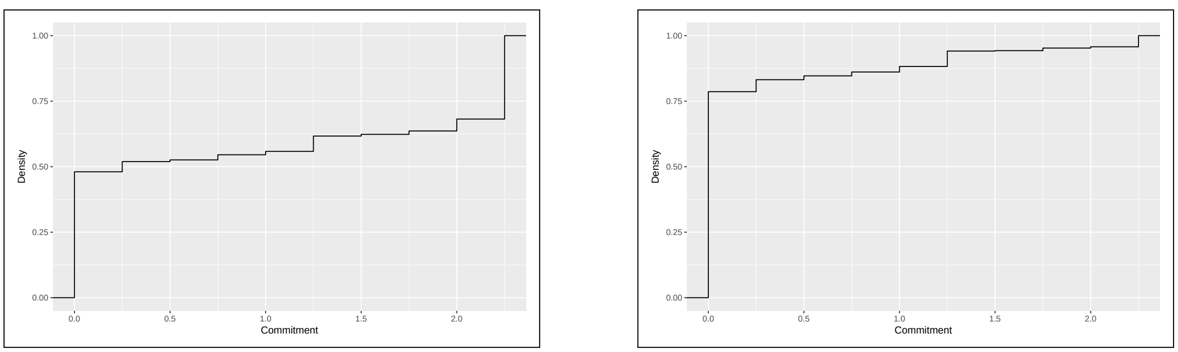


FIGURE 12.
Mean commitment by treatment



(a) Individual commitment

(b) Any partner commitment

FIGURE 13.
CDFs of commitment by treatment

group treatment effects. While it appears that higher levels of education may have small negative effects on commitment demand, none of these estimated effects are statistically significant from zero[14]; similarly, while it appears age may have a small

[14]Interestingly, all 10 subjects with a PhD selected no commitment, but there are too few in the sample to precisely estimate any potential effect.

negative effect on commitment demand, the effect is not statistically significant from zero.

Given our estimated effects on commitment demand, we can calculate each of the motivational factors discussed previously. In Table 26, we present these estimated factors for the main model that includes only treatment variables. Given the large negative coefficients on each of the group treatment variables, the large negative team estimate is not surprising. Additionally, this is the only estimated effect that is statistically significant from zero. This team effect represents how the subject's propensity to let down a partner by failing to complete the follow-up task affects the subject's commitment demand. One potential explanation for this large negative effect is that subjects may anticipate higher follow-through rates in joint treatments than in individual treatments. If this were the case, there would be little need for commitment, since the subject believes it extremely unlikely that she fails to complete the future task.

TABLE 26.
Calculated factors, only treatment variables

	Factor	Effect	p-value
1	Risk	0.009	0.5771
2	Strategic	0.044	0.9061
3	Team	-0.804^{***}	$7.0e^{-13}$
4	Public	0.087	0.2745

While we are most interested in the effect of treatment on commitment demand, we can also investigate how the chances that a subject selects a positive level of commitment vary with treatment. Table 27 presents estimates of the effect of treatment on the probability that a participant selects a non-zero level of commitment. Consistent with the observed effects on levels of commitment selected, participants in the individual treatment select a positive amount of commitment much more frequently (52% of the time) than individuals in any of the group treatments. In all models, any of the group treatment variables have large, negative, statistically significant effects on the probability of non-zero commitment. Figure 14 presents a bar graph of the fraction of subjects who select a positive amount of commitment by treatment.

TABLE 27.
Probability of non-zero commitment by treatment

	Dependent variable:		
	Non-zero commitment		
	Linear	*Logit*	*Probit*
	(1)	(2)	(3)
Constant (Individual)	0.519^{***}	0.078	0.049
	(0.035)	(0.161)	(0.101)
Joint	-0.286^{***}	-1.265^{***}	-0.775^{***}
	(0.049)	(0.250)	(0.150)
Sep. with knowledge	-0.296^{***}	-1.322^{***}	-0.809^{***}
	(0.049)	(0.253)	(0.152)
Sep. without knowledge	-0.323^{***}	-1.489^{***}	-0.905^{***}
	(0.049)	(0.260)	(0.154)
Asymmetric	-0.318^{***}	-1.456^{***}	-0.886^{***}
	(0.049)	(0.258)	(0.153)
Observations	767	767	767

Note: *p<0.1; **p<0.05; ***p<0.01

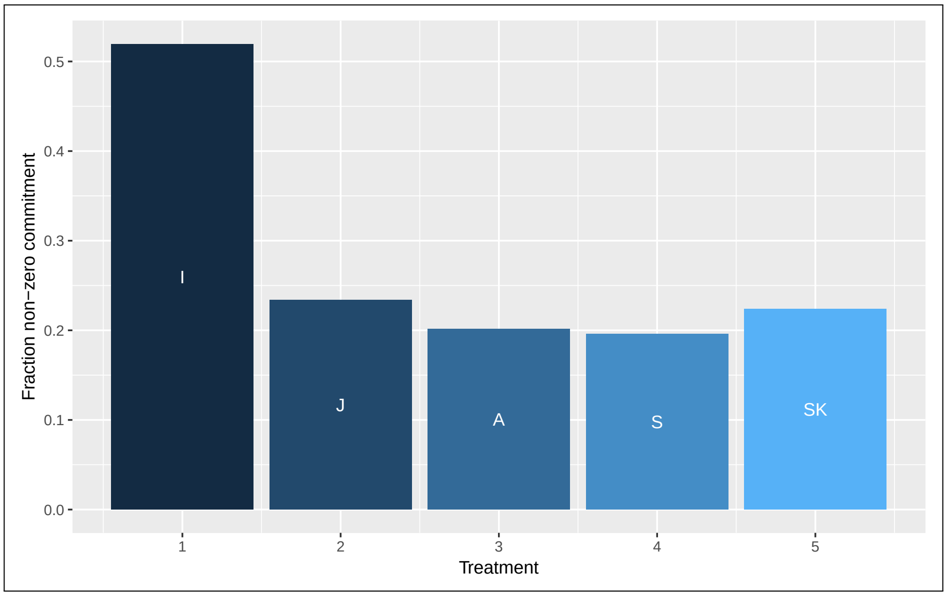


FIGURE 14.
Non-zero commitment by treatment

One concern we have is that some subjects may not pay attention to the instructions and attempt to click through the survey as quickly as possible. We included timing questions on the instructions page in the survey, which allows us to split the sample by minimum time thresholds. We estimate the model for various minimum time-on-instructions cutoffs and present these results in Table 28, where it appears that the main results are stable across these various truncated samples. In general, our main findings of large negative effects of any group treatment on commitment demand are robust across all models and sample subsets.

Commitment demand in the individual treatment increases as the minimum time threshold increases; we estimate this effect directly in Table 29, which includes the time spent on the instructions page as an independent variable. The estimated effect is positive, nears statistical significance at the 10% level, and can be interpreted to say that taking an additional 6 seconds on the instructions page increases commitment by $0.01.

TABLE 28.
Commitment by minimum time on instructions

	Exclude fastest			
	1%	5%	10%	20%
Constant	0.957***	0.979***	1.011***	1.066***
	(0.056)	(0.057)	(0.059)	(0.063)
Joint	−0.666***	−0.682***	−0.722***	−0.760***
	(0.079)	(0.081)	(0.083)	(0.087)
SepWith	−0.684***	−0.724***	−0.761***	−0.812***
	(0.080)	(0.081)	(0.083)	(0.088)
SepWithout	−0.767***	−0.793***	−0.839***	−0.879***
	(0.080)	(0.082)	(0.083)	(0.088)
Asymmetric	−0.727***	−0.764***	−0.812***	−0.916***
	(0.079)	(0.081)	(0.084)	(0.091)
Observations	759	728	690	613

Note: *p<0.1; **p<0.05; ***p<0.01

TABLE 29.
Commitment by time on instructions and choice

	Dependent variable:	
	Commitment	
	(1)	(2)
Constant	0.920^{***}	0.969^{***}
	(0.060)	(0.060)
Joint	-0.680^{***}	-0.654^{***}
	(0.080)	(0.080)
SepWith	-0.687^{***}	-0.660^{***}
	(0.080)	(0.081)
SepWithout	-0.773^{***}	-0.750^{***}
	(0.080)	(0.080)
Asymmetric	-0.725^{***}	-0.707^{***}
	(0.079)	(0.080)
TimeOnInstructions	0.002	
	(0.001)	
TimeInstructionsChoice		-0.0004
		(0.001)
Observations	767	767
Note:	*p<0.1; **p<0.05; ***p<0.01	

We also investigate the follow-through of subjects on the future task that they previously agreed to complete. In the individual treatment, 26.6% of subjects completed the future task; the effect of each group treatment is small, positive, and not statistically significant from zero. In Table 30, we present estimates of the effect on follow-through of being in any group treatment relative to being in the individual treatment. The estimated effect is positive (5.0 percentage points in the full sample, 5.9 percentage points for subjects who took at least 10 seconds on the instructions page), but it is not statistically significant from zero. In general, the results suggest that being assigned to any group treatment increases follow-through rates.

TABLE 30.
Follow-through by individual versus group

	Dependent variable:	
	Completed future task	
	Full sample	Time > 10
Constant	0.266***	0.296***
	(0.037)	(0.044)
AnyPartner	0.050	0.059
	(0.042)	(0.050)
Observations	691	544
Note:	*p<0.1; **p<0.05; ***p<0.01	

The formulation of the joint treatment allows us, for a small subset of the sample, to estimate the causal impact of commitment on follow-through. For individuals who are assigned the joint treatment and whose commitment choice differs from the group's randomly selected commitment level, both the treatment and the level of commitment have been exogenously assigned. Table 31 presents the estimated causal effect of commitment on follow-through for these subjects; the estimated effect is positive, but not statistically significant. The coefficient can be interpreted to say that an additional dollar of commitment increases a subject's probability of completing the future task by 4.9 percentage points.

TABLE 31.
Effect of commitment on follow-through, joint treatment

	Dependent variable:
	Completed future task
Constant	0.327^{**}
	(0.122)
Commitment	0.049
	(0.115)
Observations	33
Note:	*p<0.1; **p<0.05; ***p<0.01

Conclusion

We design and implement an experiment in which subjects complete a task and have the opportunity to demand commitment across an individual treatment and a variety of group treatments. The treatments are designed to tease out the separate effects of a number of possible motivations the subjects could have for commitment: strategic incentives, risk, public observability of the choice, and the desire not to let down the team by failing to follow through on the future task. We find that both commitment demand and the fraction of subjects who demand some level of commitment are significantly larger in the individual treatment than in any of the joint treatments. There is some evidence that being in a joint treatment leads to a higher follow-through rate on the future task, but this estimated effect is not significant.

Given the experimental design, we find that the risk, strategic, and public factors are not statistically significant from zero, although we do estimate a positive public factor, which is consistent with the findings of Exley and Naecker (2016). However, the large negative effect of being assigned to any group treatment poses new questions. Evidently, the effect on a partner's payoffs as a result of a subject's potential failure to follow through on the future causes subjects to demand significantly less commitment. This could be the result of subjects anticipating that follow-through rates will be higher in these treatments, which would lead to little need for commitment devices. If subjects expect that they will always complete the future task, we would expect commitment demand to be zero.

Moving forward, we should elicit information about subjects' beliefs about both their own follow-through rates and the follow-through rates of partners. We discussed including these types of questions in the initial study, but we decided to omit them for fear of priming subjects about the types of choices they were making. If we could determine that subjects' beliefs about their own follow-through rates (and, potentially, partners' follow-through rates) differed significantly across individual and group treatments, that could further speak to the mechanism of the negative group treatment effect on commitment demand.

CHAPTER V

CONCLUSION

I empirically investigate individual behavior in a number of real-world settings to determine whether these decisions are consistent with models from behavioral economic theory.

In Chapter II, I estimate the effect of outperforming or underperforming expectations on U.S. House of Representatives' decisions to run again in the next election. While candidates likely try to sort across election-winning thresholds of vote share, they are unlikely to intentionally sort across the zero-signal threshold. Although there appears to be no effect of outperforming expectations for winning candidates, losing candidates who outperform expectations exhibit a significant positive discontinuity in the probability of running in the next election; this type of behavior is consistent with theories of attribution bias and aspiration-level utility.

Chapter III uses granular data on reported crimes to examine the effect of opening recreational marijuana markets on reported domestic violence incidents in Colorado, Washington, and Oregon. I estimate positive effects in all three states, and I employ a number of different datasets to provide additional evidence for the most likely mechanism. The estimated effects are concentrated on times of day and days of week in which individuals are likely to purchase marijuana, and estimates from Consumer Expenditure Survey data suggest that household food expenditures decrease when marijuana sales start. These, combined with additional evidence, suggest that household budget conflict is the most likely channel for the observed results.

Finally, in Chapter IV, which is co-authored with Michael Kuhn and Jeffrey Naecker, we design an experiment to elicit commitment demand from subjects in group commitment settings. Using Amazon's Mechanical Turk and Qualtrics, we have subjects perform a task, agree to perform the task again, and then select a level of commitment for a variety of treatments. We design the experiment to separate possible motivations described in the behavioral economics literature; we find that there is a large negative effect of being assigned to any partner on commitment demand, and we suggest that this is evidence that individuals do not want to let down the "team." Additionally, there is weak evidence that higher commitment increases follow-through on the follow-up task.

APPENDIX

SUPPLEMENTAL TABLES

TABLE A1.
Estimated effect of the beginning of recreational marijuana sales, full sample by order of time trend, state-level treatment, standard errors adjusted for clustering by state

	(1) Linear	(2) Quadratic	(3) Cubic	(4) Quartic
Market opens	0.0104***	0.0129***	0.0134***	0.0126**
	(4.99)	(3.58)	(3.41)	(2.54)
Percentage change	2.9%	3.5%	3.7%	3.5%
Observations	3434302	3434302	3434302	3434302

t statistics in parentheses
All models include year, month, day-of-week, and agency fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A2.
Pooled Poisson estimates of the effect of the recreational sales market opening by order of state time trend by level of treatment

	(1) Linear	(2) Quadratic	(3) Cubic	(4) Quartic
State-level treatment	0.0305*	0.0290*	0.0326*	0.0308
	(1.72)	(1.70)	(1.71)	(1.32)
County-level treatment	0.0478*	0.0526*	0.0558**	0.0640*
	(1.74)	(1.78)	(1.98)	(1.85)
First distance-weighted treatment	0.0418*	0.0442*	0.0511**	0.0555**
	(1.79)	(1.86)	(2.41)	(2.24)
Second distance-weighted	0.0471*	0.0508*	0.0561**	0.0634**
	(1.77)	(1.81)	(2.14)	(1.99)
Observations	3097018	3097018	3097018	3097018

t statistics in parentheses
All models include year, month, day-of-week, and agency fixed effects
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A3.
Food expenditure by winsorization percentile, pooled treatment, clustered by state

	(1) Full sample	(2) 99th	(3) 95th	(4) 90th	(5) 80th
Legalization	11.39	10.74	10.86	10.77	11.26
	(0.69)	(0.68)	(0.77)	(0.87)	(1.21)
SalesStart	-27.36*	-28.81**	-28.27**	-27.15**	-25.09**
	(-1.92)	(-2.26)	(-2.59)	(-2.39)	(-2.56)
Observations	20507	20507	20507	20507	20507

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A4.
Food expenditure by trim percentile, pooled treatment, clustered by state

	(1) Full sample	(2) 99th	(3) 95th	(4) 90th	(5) 80th
Legalization	11.39	9.989	5.933	12.89**	11.99**
	(0.69)	(0.70)	(0.48)	(2.07)	(2.45)
SalesStart	-27.36*	-34.20***	-27.84*	-18.11	-26.23**
	(-1.92)	(-4.25)	(-1.94)	(-1.45)	(-2.60)
Observations	20507	20302	19482	18457	16406

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A5.
Consumer expenditure by category, pooled treatment, clustered by state

	(1) Food	(2) Alcohol	(3) Healthcare	(4) Entertainment	(5) Transportation	(6) Apparel
Legalization	11.39	-3.102	-27.15*	51.49	126.8	29.50***
	(0.69)	(-1.09)	(-1.74)	(1.38)	(1.63)	(2.97)
SalesStart	-27.36*	1.197	0.181	81.25	-36.36	-7.297
	(-1.92)	(0.30)	(0.01)	(0.70)	(-0.58)	(-0.71)
Observations	20507	20507	20507	20507	20507	20507

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

TABLE A6.
Consumer expenditure by category, pooled treatment, winsorized at 99, clustered by state

	(1) Food	(2) Alcohol	(3) Healthcare	(4) Entertainment	(5) Transportation	(6) Apparel
Legalization	10.74	0.140	0.104	41.88	56.67***	25.62***
	(0.68)	(0.08)	(0.01)	(1.59)	(4.28)	(3.07)
SalesStart	-28.81**	-1.722	-11.14	30.15	15.03***	-13.76***
	(-2.26)	(-0.74)	(-1.17)	(0.52)	(3.19)	(-8.81)
Observations	20507	20507	20507	20507	20507	20507

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

REFERENCES CITED

Aizer, A. (2010). The gender wage gap and domestic violence. *American Economic Review*, 100(4):1847–59.

Aizer, A. (2011). Poverty, violence, and health the impact of domestic violence during pregnancy on newborn health. *Journal of Human resources*, 46(3):518–538.

Allen, E. J., Dechow, P. M., Pope, D. G., and Wu, G. (2016). Reference-dependent preferences: Evidence from marathon runners. *Management Science*, 63(6):1657–1672.

Anagol, S. and Fujiwara, T. (2016). The runner-up effect. *Journal of Political Economy*, 124(4):927–991.

Anderson, D. M., Hansen, B., and Rees, D. I. (2013). Medical marijuana laws, traffic fatalities, and alcohol consumption. *The Journal of Law and Economics*, 56(2):333–369.

Anderson, D. M., Hansen, B., and Rees, D. I. (2015). Medical marijuana laws and teen marijuana use. *American Law and Economics Review*, 17(2):495–528.

Anderson, D. M., Rees, D. I., et al. (2014a). The legalization of recreational marijuana: how likely is the worst-case scenario. *Journal of Policy Analysis and Management*, 33(1):221–232.

Anderson, D. M., Rees, D. I., and Sabia, J. J. (2014b). Medical marijuana laws and suicides by gender and age. *American journal of public health*, 104(12):2369–2376.

Ashworth, S. and Bueno de Mesquita, E. (2008). Electoral selection, strategic challenger entry, and the incumbency advantage. *The Journal of Politics*, 70(4):1006–1025.

Babcock, P., Bedard, K., Charness, G., Hartman, J., and Royer, H. (2015). Letting down the team? social effects of team incentives. *Journal of the European Economic Association*, 13(5):841–870.

Baca, R. (2014). These 24 washington marijuana shops are licensed to open july 8. *The Cannabist*.

Barberis, N. C. (2013). Thirty years of prospect theory in economics: A review and assessment. *Journal of Economic Perspectives*, 27(1):173–96.

Basu, K. (2011). Hyperbolic discounting and the sustainability of rotational savings arrangements. *American Economic Journal: Microeconomics*, 3(4):143–71.

Black, M. C. (2011). Intimate partner violence and adverse health consequences: implications for clinicians. *American journal of lifestyle medicine*, 5(5):428–439.

Breiding, M. J., Black, M. C., and Ryan, G. W. (2008). Chronic disease and health risk behaviors associated with intimate partner violence18 us states/territories, 2005. *Annals of epidemiology*, 18(7):538–544.

Bryan, G., Karlan, D., and Nelson, S. (2010). Commitment devices. *Annu. Rev. Econ.*, 2(1):671–698.

Bushong, B. and Gagnon-Bartsch, T. (2016). Learning with misattribution of reference dependence. Technical report, working paper.

California Department of Public Health (2018). Laws and regulations. https://www.cdph.ca.gov/Programs/CEH/DFDCS/MCSB/Pages/Legislation.aspx.

Callaghan, R. C., Gatley, J. M., Sanches, M., and Benny, C. (2016). Do drinking-age laws have an impact on crime? evidence from canada, 2009–2013. *Drug and alcohol dependence*, 167:67–74.

Card, D. and Dahl, G. B. (2011). Family violence and football: The effect of unexpected emotional cues on violent behavior. *The Quarterly Journal of Economics*, 126(1):103–143.

Carpenter, C. and Dobkin, C. (2015). The minimum legal drinking age and crime. *Review of economics and statistics*, 97(2):521–524.

Cattaneo, M. D., Jansson, M., and Ma, X. (2018). Manipulation testing based on density discontinuity. *The Stata Journal*, 18(1):234–261.

Caughey, D. and Sekhon, J. S. (2011). Elections and the regression discontinuity design: Lessons from close us house races, 1942–2008. *Political Analysis*, 19(4):385–408.

Centers for Disease Control and Prevention (2019). Preventing intimate partner violence. https://www.cdc.gov/violenceprevention/intimatepartnerviolence/consequences.html.

Cesur, R. and Sabia, J. J. (2016). When war comes home: The effect of combat service on domestic violence. *Review of Economics and Statistics*, 98(2):209–225.

Crane, C. A., Testa, M., Schlauch, R. C., and Leonard, K. E. (2016). The couple that smokes together: Dyadic marijuana use and relationship functioning during conflict. *Psychology of addictive behaviors*, 30(6):686.

Crost, B. and Guerrero, S. (2012). The effect of alcohol availability on marijuana use: Evidence from the minimum legal drinking age. *Journal of health economics*, 31(1):112–121.

Crost, B. and Rees, D. I. (2013). The minimum legal drinking age and marijuana use: New estimates from the nlsy97. *Journal of health economics*, 32(2):474–476.

DellaVigna, S., Lindner, A., Reizer, B., and Schmieder, J. F. (2017). Reference-dependent job search: Evidence from hungary. *The Quarterly Journal of Economics*, 132(4):1969–2018.

Department of Justice (2013). Memorandum. https://www.justice.gov/iso/opa/resources/3052013829132756857467.pdf.

Diecidue, E., Levy, M., and van de Ven, J. (2015). No aspiration to win? an experimental test of the aspiration level model. *Journal of Risk and Uncertainty*, 51(3):245–266.

Diecidue, E. and Van De Ven, J. (2008). Aspiration level, probability of success and failure, and expected utility. *International Economic Review*, 49(2):683–700.

Diermeier, D., Keane, M., and Merlo, A. (2005). A political economy model of congressional careers. *American Economic Review*, 95(1):347–373.

Downey, M. (2017). Losers go to jail: Congressional elections and union officer prosecutions. Technical report, UCSD Working paper.

Drug Enforcement Agency (2019). Drug scheduling. https://www.dea.gov/drug-scheduling.

Eggers, A. C., Fowler, A., Hainmueller, J., Hall, A. B., and Snyder Jr, J. M. (2015). On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races. *American Journal of Political Science*, 59(1):259–274.

Eil, D. and Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. *American Economic Journal: Microeconomics*, 3(2):114–38.

Exley, C. L. and Naecker, J. K. (2016). Observability increases the demand for commitment devices. *Management Science*, 63(10):3262–3267.

Federal Bureau of Investigation (2012). Nibrs participation by state. https://ucr.fbi.gov/nibrs/2012/resources/nibrs-participation-by-state.

Federal Bureau of Investigation (2015). Nibrs participation by state. https://ucr.fbi.gov/nibrs/2015/tables/pdfs/participation-by-state-2015.pdf.

Freisthler, B., Gruenewald, P. J., and Wolf, J. P. (2015). Examining the relationship between marijuana use, medical marijuana dispensaries, and abusive and neglectful parenting. *Child abuse & neglect*, 48:170–178.

Giné, X., Karlan, D., and Zinman, J. (2010). Put your money where your butt is: a commitment contract for smoking cessation. *American Economic Journal: Applied Economics*, 2(4):213–35.

Government Publishing Office (1970). Statute 84. https://www.govinfo.gov/content/pkg/STATUTE-84/pdf/STATUTE-84-Pg1236.pdf.

Gugerty, M. K. (2007). You cant save alone: Commitment in rotating savings and credit associations in kenya. *Economic Development and cultural change*, 55(2):251–282.

Guttmannova, K., Lee, C. M., Kilmer, J. R., Fleming, C. B., Rhew, I. C., Kosterman, R., and Larimer, M. E. (2016). Impacts of changing marijuana policies on alcohol use in the united states. *Alcoholism: Clinical and Experimental Research*, 40(1):33–46.

Haggag, K., Pope, D. G., Bryant-Lees, K. B., and Bos, M. W. (2016). Attribution bias in consumer choice. *The Review of Economic Studies*.

Hansen, B., Miller, K., and Weber, C. (2017). The grass is greener on the other side: How extensive is the interstate trafficking of recreational marijuana? Technical report, National Bureau of Economic Research.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Kerr, D. C., Bae, H., Phibbs, S., and Kern, A. C. (2017). Changes in undergraduates marijuana, heavy alcohol and cigarette use following legalization of recreational marijuana use in oregon. *Addiction*, 112(11):1992–2001.

Kilmer, B. (2016). Marijuana legalization, government revenues, and public budgets. *RAND*.

Kőszegi, B. and Rabin, M. (2006). A model of reference-dependent preferences. *The Quarterly Journal of Economics*, 121(4):1133–1165.

Lee, D. S. (2008). Randomized experiments from non-random selection in us house elections. *Journal of Econometrics*, 142(2):675–697.

Levitt, S. D. and Wolfram, C. D. (1997). Decomposing the sources of incumbency advantage in the us house. *Legislative Studies Quarterly*, pages 45–60.

Lindo, J. M., Siminski, P., and Swensen, I. D. (2018). College party culture and sexual assault. *American Economic Journal: Applied Economics*, 10(1):236–65.

Loewenstein, G., O'Donoghue, T., and Rabin, M. (2003). Projection bias in predicting future utility. *the Quarterly Journal of economics*, 118(4):1209–1248.

Lukas, S. E. and Orozco, S. (2001). Ethanol increases plasma δ9-tetrahydrocannabinol (thc) levels and subjective effects after marihuana smoking in human volunteers. *Drug and alcohol dependence*, 64(2):143–149.

Markowitz, S. (2005). Alcohol, drugs and violent crime. *International Review of Law and economics*, 25(1):20–44.

Markowitz, S., Nesson, E., Poe-Yamagata, E., Florence, C., Deb, P., Andrews, T., and Barnett, S. B. L. (2012). Estimating the relationship between alcohol policies and criminal violence and victimization. *German economic review*, 13(4):416–435.

Max, W., Rice, D. P., Finkelstein, E., Bardwell, R. A., and Leadbetter, S. (2004). The economic toll of intimate partner violence against women in the united states. *Violence and victims*, 19(3):259.

McCrary, J. (2008). Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2):698–714.

McGhee, T. and Ingold, J. (2013). A list of colorado recreational marijuana stores open on jan. 1. *The Denver Post*.

Moore, T. M. and Stuart, G. L. (2005). A review of the literature on marijuana and interpersonal violence. *Aggression and Violent Behavior*, 10(2):171–192.

National Coalition Against Domestic Violence (2019). Statistics. http://ncadv.org/learn-more/statistics.

National Conference of State Legislatures (2019). State medical marijuana laws. http://www.ncsl.org/research/health/state-medical-marijuana-laws.aspx.

Nicolewski, R. (2016). 7 things to know now that prop 64 is law. *San Diego Union-Tribune*.

O'Donoghue, T. and Rabin, M. (1999). Doing it now or later. *American Economic Review*, 89(1):103–124.

Oregon Liquor Control Commission (2019). Faqs: Personal use. https://www.oregon.gov/olcc/marijuana/Pages/FAQs-Personal-Use.aspx.

Pacula, R. L. (1998). Does increasing the beer tax reduce marijuana consumption? *Journal of health economics*, 17(5):557–585.

Pacula, R. L. and Sevigny, E. L. (2014). Marijuana liberalizations policies: why we cant learn much from policy still in motion. *Journal of policy analysis and management:[the journal of the Association for Public Policy Analysis and Management]*, 33(1):212.

Rees-Jones, A. (2017). Quantifying loss-averse tax manipulation. *The Review of Economic Studies*, 85(2):1251–1278.

Sabia, J. J., Dills, A. K., and DeSimone, J. (2013). Sexual violence against women and labor market outcomes. *American Economic Review*, 103(3):274–78.

Sabia, J. J., Swigert, J., and Young, T. (2017). The effect of medical marijuana laws on body weight. *Health economics*, 26(1):6–34.

Silverman, J. G., Raj, A., Mucci, L. A., and Hathaway, J. E. (2001). Dating violence against adolescent girls and associated substance use, unhealthy weight control, sexual risk behavior, pregnancy, and suicidality. *Jama*, 286(5):572–579.

Single, E. W. (1981). The impact of marijuana decriminalization. In *Research advances in alcohol and drug problems*, pages 405–424. Springer.

Strotz, R. H. (1955). Myopia and inconsistency in dynamic utility maximization. *The Review of Economic Studies*, 23(3):165–180.

Subbaraman, M. S. and Kerr, W. C. (2015). Simultaneous versus concurrent use of alcohol and cannabis in the national alcohol survey. *Alcoholism: Clinical and Experimental Research*, 39(5):872–879.

Testa, M. and Brown, W. C. (2015). Does marijuana use contribute to intimate partner aggression? a brief review and directions for future research. *Current opinion in psychology*, 5:6–12.

Thaler, R. H. and Benartzi, S. (2004). Save more tomorrow: Using behavioral economics to increase employee saving. *Journal of political Economy*, 112(S1):S164–S187.

Ullman, D. F. (2017). The effect of medical marijuana on sickness absence. *Health economics*, 26(10):1322–1327.

Wen, H., Hockenberry, J. M., and Cummings, J. R. (2015). The effect of medical marijuana laws on adolescent and adult use of marijuana, alcohol, and other substances. *Journal of health economics*, 42:64–80.

Williams, J., Liccardo Pacula, R., Chaloupka, F. J., and Wechsler, H. (2004). Alcohol and marijuana use among college students: economic complements or substitutes? *Health economics*, 13(9):825–843.