From Risk and Financial Markets to Persistence of Racism and Gun Control: Essays in Empirical Microeconomics

Inaugural-Dissertation

zur Erlangung des Grades

Doctor oeconomiae publicae (Dr. oec. publ.)

an der Ludwig-Maximilians-Universität München

2016

vorgelegt von

David Schindler

Referent:	Prof. Dr. Martin G. Kocher
Korreferent:	Prof. Dr. Uwe Sunde
Promotionsabschlussberatung:	10.05.2017

Datum der mündlichen Prüfung: 04. Mai 2017

Namen der Berichterstatter: Martin G. Kocher, Uwe Sunde, Florian Englmaier *to my parents*

Acknowledgements

This dissertation only became possible because of ongoing support and encouragement by advisers, colleagues, friends and family. Acknowledging their contributions to me is not simply courtesy, but an expression of my deep gratitude for each and everyone's individual contributions to this work.

First, I would like to thank my adviser Martin Kocher. It is only because of him welcoming me as a graduate student in his team that I received a chance to prove myself scientifically. His continuing support throughout the years, and his confidence in my abilities have made being his graduate student an incredibly valuable and fantastic experience. Martin's door was always open, and he always took the time to answer my questions and help me advance my projects. I could not have wished for a better mentor.

I am indebted to Uwe Sunde for agreeing to be my second adviser, for supporting my ambitions on the job market, and for serving on my committee. His support from my first year as a graduate student when I sought external funding, to my final year when I needed recommendations for the job market, has been exceptional.

Florian Englmaier established contacts that eventually made my year-long research stay at the University of Pennsylvania's Wharton School possible, something I am extremely grateful for. Apart from his firm commitment to enable my stay abroad, he was always open to discuss research and offered helpful advice, even when the likelihood of success of my ideas was limited to begin with. His support during my job market year was outstanding, and I am also happy to have him as a coauthor in a project that eventually did not make it into this dissertation.

My stay at Wharton would not have been possible without the continuous efforts by Judd Kessler. Judd helped me to prepare for the academic job market and was always open to meet and discuss my job market paper ideas. Additionally, he arranged numerous meetings with scholars working in areas related to my job market paper, without which my final project surely would have failed.

Over the past four years, I had the pleasure to work with many brilliant and dedicated PhD students, junior and senior faculty, all of whom I'd like to thank. Stefan Trautmann has been an important source of inspiration, his comprehensive understanding of complex relationships has left a long-lasting impression with me. I very much enjoyed working with Yilong Xu, who is the most dedicated, polite and likeable researcher I know. Starting grad school as friends, Mark Westcott and I eventually became coauthors, and I still joyfully remember the joint beer garden session in which we developed our project idea. Christoph Koenig agreed to invest time and effort on very short notice to progress our joint project, something I am truly grateful for.

Several students and faculty members in the economics department in Munich and at other institutions have become friends and also given valuable input to all of my projects. Others became co-authors in projects that did not make it into this thesis. Thanks to all of you for the great and exciting four years in such a positive and fun environment. Florentin Krämer deserves special recognition for freeing me from the administrative burden of handing in this dissertation personally.

Over my entire lifetime, my parents have been most influential of my development. It is only because of them that I developed a keen interest in exploring life around me and wanted to understand how the world functioned: I still remember with delight the joy of reading the *How and Why Wonder Books* as a child that they constantly provided me with. Thank you for all your guidance, advice, love and understanding. I also do not want to miss expressing my gratitude for all the financial support they have provided me with over all those years, without which my studies could have never been successful. For all these reasons, I dedicate this dissertation to you.

Finally, I would like to thank my wife Sandra for always encouraging and supporting me, especially at times when completion of this dissertation seemed light years away. Her patience with occasional night shifts and her understanding for my unusual working hours have been outstanding. Thank you for always being there for me, even when we were living on two different continents.

Contents

List of Figures

List of Tables

Preface

Economic research for decades has relied on regarding decision makers as purely rational and selfish, sometimes having perfect knowledge of their environment. While proponents of present-day behavioral economic research like to point out that already Adam Smith's less well known book "The Theory of Moral Sentiments" constitutes a treatment of behavioral economics (Ashraf et al., 2005), it was only in the 1950s that mainstream economics started to systematically question some of the assumptions made. Examples include Allais (1953) who famously demonstrated the inability of many decision makers to behave consistently with the independence axiom, one of the fundamentals of Expected Utility Theory. Similarly, Strotz (1955) formulated a critique of exponential discounting, arguing that myopia could lead to time-inconsistent behavior.

Following the work by psychologists who were crossing lines into economics, primarily the group around Daniel Kahneman, behavioral economics slowly rose from being a niche to becoming a respected field within mainstream economics. By the 1990s, behavioral economics received widespread attention following influential theoretical contributions in major general interest journals that offered explanations for empirical observations standard economic theory could not account for. With the 2002 Nobel Prize in Economics being awarded to Daniel Kahneman and Vernon Smith, and the 2012 Nobel Prize to Al Roth, behavioral economics and also its sister field experimental economics finally reached center stage of economics research. Thanks to its prominence, many related disciplines and subfields of economics have picked modeling approaches that take non-standard behavior into account. Fields such as behavioral finance or behavioral public policy emerged, thanks to the possibility of considering systematic decision biases and their interactions with market environments or political interventions.

Behavioral finance can be considered a subset of traditional finance, in which assumptions about perfect markets and rational agents are relaxed, allowing for cognitive biases such as anchoring, mental accounting, overconfidence or probability weighting to play a role. Contributions in this field for example have made us understand, that the tendency to chase trends inherent to decision patterns in many people can repeatedly cause bubble-and-crash phenomena (Smith et al., 1988), and that this can be explained by variations in cognitive skills (Bosch-Rosa et al., 2015). Other insights include the tendency of people to perceive losses as more painful than the utility of gains of the same size (Kahneman and Tversky, 1979), which in combination with myopia may deliver an explanation for the equity premium puzzle (Benartzi et al., 1995).

Behavioral public policy instead has mainly focused on how behavioral idiosyncrasies contribute to outcomes of relevance for public economists (Chetty, 2015). For example, Chetty et al. (2009) describe how inattention to sales taxes can distort people's shopping behavior, and Choi et al. (2011) show that many people are prone to choosing dominated choices in their retirement plans. Recent contributions also make a link to topics of political economy, for example Dahl and DellaVigna (2009) investigate the relationship of movie violence and violent crimes, surprisingly finding that the attendance of violent movies acts as a substitute for violent behavior, and Card and Dahl (2011) find that domestic violence increases after the home team loses a football match although being favored by the odds. The rise of behavioral public policy and its success at debunking commonly held economic wisdoms has led to some countries establishing behavioral task forces, such as the United Kingdom's *Behavioral Insights Team* or the United States' *Social and Behavioral Sciences Team*, with the aim of providing legislators help in creating policies that can accommodate behavioral biases among their constituents.

This dissertation adds to both of these fields, behavioral finance and behavioral public policy, by contributing five chapters containing empirical studies. The first three chapters use laboratory experiments to identify behavioral anomalies and offer explanations for previously observed non-standard economic behavior in the environment of financial markets and risk-tasking behavior. The last two chapters are concerned with the malleability of political preferences and the effectiveness of gun policy instruments. They contain empirical studies in public policy using secondary data sets, both from historic, as well as contemporary sources.

Chapter 1, which is joint work with Martin Kocher, Stefan Trautmann and Yilong Xu, discusses the importance of selection effects in laboratory studies on the relationship of time pressure and risk-taking behavior. While earlier studies on risky decision-making usually allow subjects to take as much time as they want, to facilitate qualitatively good choices, this assumption rarely holds for many real-life decisions involving risk. This is for example the case for stock market traders, who have to react to incoming information within split-seconds and adjust their portfolio quickly. Therefore, recent studies have started to analyze how time pressure affects choices when subjects decide over risky prospects. Kocher et al. (2013) for example find strong shifts in behavior when subjects are exposed to time pressure. In the case of mixed gambles they show that participants are more likely to be attracted by prominent gains and shy away from prominent losses.

We build on the findings by Kocher et al., as we use mixed gambles and put subjects under time pressure. In contrast to their paper however, we apply the time pressure variation also within subjects instead of purely between subjects, allowing us to observe behavior of every subject also in the absence of time pressure. Because time pressure usually leads some subjects to violate the time constraint, the observed sample is potentially selected. With the additional within subjects variation, we can particularly investigate if those who violate the constraint behave differently in the absence of time pressure. Additionally, we elicit two dimensions of cognitive skills, intellectual capacity (IQ) and intellectual efficiency (IE). Intellectual capacity measures the cognitive reasoning power, i.e. the extent to which complex information can be handled, while intellectual efficiency measures the speed at which information is processed. Together with several other background variables that we collect, such as the locus of control questionnaire (Rotter, 1966), the self-efficacy scale (Schwarzer and Jerusalem, 1995) or a short version of the Big Five personality questionnaire (Gosling et al., 2003), we then try to predict which traits are associated with the ability to effectively cope with time pressure.

We find that subjects violating the time constraint indeed show fundamentally different behavior in the absence of time pressure. Because they make their decisions more carefully and spend more time on each gamble, they realize significantly higher expected values. When facing time pressure however, this advantage vanishes. Because of taking too much time initially, violators only have little time left for later decisions. This can also be seen in the variance of time spent on any decision, which is much higher for violators, than for non-violators. Although the choices that violators make early on yield higher expected values, the choices that they make with only little time left are so much worse, that when comparing all choices made between violators and non-violators, there is no significant difference. The fact that they violate the time constraint however makes them eventually worse off, as we designed the violation to come with the lowest possible payoff from that decision.

With respect to predictability of violating behavior, our results are more mixed. Our analysis does not find strong predictive power of either age, gender, IQ, locus of control, or Big Five. Only IE and self-efficacy seem to somewhat differ between violators and non-violators, a finding consistent with our expectations and other results from the literature. When zooming in more closely, we see that while IE is predictive of being a violator, it is not of decision quality, which is largely influenced by IQ. We then estimate a simplified parametric decision model of Cumulative Prospect Theory for unconstrained choices and see how well this predicts constrained choices. This prediction we then relate to our elicited background characteristics. We find that higher measures of IE, self-efficacy, and less time-use in the absence of time pressure make the fitted model predict behavior under time pressure reasonably well, lending credibility to the idea that observables relate to decision styles that subjects implement. In multivariate regression analyses, we then seek to identify which characteristic predicts success in choices under time pressure best. While we again find some weak evidence of IE, self-efficacy, and time-use in the absence of time pressure to be predictive, the overall variation in our outcome variables that we can explain remains very low.

Our findings suggest that different decision styles are a ubiquitous feature of decision making under risk. Subjects who can cope with time pressure clearly apply different strategies than those who don't, an important factor that future studies should take into account. The fact that behavior is not well predicted by easily obtainable characteristics such as questionnaire answers also matters for practitioners. If applicants for certain types of jobs are to be filtered for their ability to cope with time pressure, measures as the ones employed in our study may lack the predictive power to successfully do so.

In Chapter 2, which is joint work with Martin Kocher and Konstantin Lucks, we turn from individual decisions to markets. The tendency of human subjects to create bubble-and-burst phenomena in stylized experimental asset markets experiments has been a long known fact (Smith et al., 1988). Guidebooks on the psychology of investing and successful investors such as Warren Buffet have suggested that a lack of self-control abilities can be damaging when making investment decisions. We test this claim by implementing an experiment, in which we exogenously reduce some participants' ability to exert self-control, based on the paradigm of ego depletion from psychology (Baumeister et al., 1998). After this intervention, subjects both with and without a reduction in self-control capabilities engage in trading of a virtual asset in a continuous double auction with open order books, and each market consists of only participants with either high or low self-control resources. In an additional experiment where we seek to explore the mechanism through which lower self-control might affect trading behavior, we reduce self-control capacities of traders differentially within markets and elicit details about their emotional state.

We find that markets with traders low in self-control exhibit significantly more overpricing. Markets with non-depleted traders show signs of excessively high prices too, but to a lesser extent. Our treatment does not affect cognitive skills or risk aversion, thus excluding them as a driver of the result. When populating markets with an equal population of depleted and non-depleted traders, we observe the same extent of overpricing as if all traders were depleted, suggesting that already a moderate share of traders low in self-control can substantially inflate prices. We discover no difference in behavior among depleted and non-depleted traders after period one, but traders low in self-control tend to bid lower and ask higher than non-depleted traders during the first trading period. This suggests that traders low in self-control are primarily interested in realizing profits from trading, thereby fueling the bubble. The fact that differences disappear after period one could therefore be due to non-depleted participants following the market behavior of depleted participants more strongly. Additionally we find that profits do not differ due to the rapid convergence of behavior of our two groups of traders. While being low in self-control therefore seems to be detrimental to markets, it does not necessarily drive traders low in self-control out of the market, therefore generating a persistent negative effect on prices.

Using additional questions regarding participants' emotional state we can show that depleted traders feel significantly more excitement, fear, and joy at the end of the trading period. We also present suggestive evidence that their trading behavior might be driven by emotions more strongly than they are aware of. While previous literature has found that cognitive skills are strongly predictive of earnings in experimental asset markets (Bosch-Rosa et al., 2015), we also find that our treatment sidelines this channel, as cognitive skills have no beneficial impact on profits for depleted traders. We see this as suggestive evidence that traders low in self-control rely more strongly on the impulsive system 1, rather than reflective system 2 (Kahneman, 2011).

Implications from our findings are twofold. First, we present evidence that reduced self-control abilities can contribute to overpricing, delivering an additional possible explanation for why overpricing is sometimes observed in real markets. This seems particularly important in countries where large parts of the population engage in asset market speculation, such as in China, where other forms of gambling are illegal. Second, our results offer advice to traders to be aware of potential self-control problems. Cooling-off periods, or avoiding longer stretches of hunger or sleep deprivation might be beneficial to pricing patterns in asset markets.

Chapter 3 analyzes institutional features common in today's asset markets. With the advent of high-speed internet connections and large cost decreases in computing power, more and more trading has been shifted from human traders to computers. Approximately three quarters of trading volume in the United States stem from computerized trading and high frequency trading (HFT) generates trading profits of \$ 3 billion annually. These traders often trade using limit order based strategies. At the same time, market structures have largely converged to a pricing scheme called maker-taker pricing, in which those that accept outstanding offers (market orders) are levied a fee and those that create new offers to market participants (limit orders) are given a rebate.

Given that the pricing structure and the presence of HFTs have changed and now often coexist, I answer the question of how this affects human traders. In an experiment, I let human traders trade assets of a fictitious firm and vary whether they encounter an HFT in addition to other human traders, whether they are subject to maker-taker pricing, or both. The setup allows me to answer how traders adjust their behavior to these additional institutional features, and how this in turn affects market outcomes. Additionally, I elicit cognitive abilities, risk preferences and other demographics I can use as controls.

I find that maker-taker pricing does not impact individual behavior much. The evidence of how this pricing structure affects prices at which subjects trade is weak and inconclusive. If anything, maker-taker pricing leads traders to spend more per asset, but not significantly so in most specifications. Also, cognitive skills do not seem to influence the choice of using limit versus market orders when this particular fee/rebate structure is present. The results are stronger for the presence of HFTs. When HFTs are present, traders make much more use of market orders relative to limit orders. They also trade much more shares per period, but only if in addition to the HFT they also face the maker-taker fee structure.

When analyzing profits it becomes apparent that traders learned to game the HFT. In all treatments in which HFTs were present, average profits strongly increase over the trading periods. There is substantial heterogeneity in profits, suggesting that some subjects understand how to beat the algorithm, while others do not. Overall prices, however, are not different across treatments suggesting that individual changes in behavior did not necessarily translate to aggregate market efficiency.

My findings are important in two dimensions. First, they show that market prices do not change much in reaction to individual adjustments to either HFTs or maker-taker pricing, which is reassuring for market designers. Second, the fact that some traders learn to game the algorithm suggests a possible explanation for market crashes following a withdrawal of liquidity: HFTs usually involve algorithms that make them withdraw from trading if they accumulate losses. This in turn will remove a substantial amount of liquidity from the market after which trading could break down. The Wall Street Journal (2010) argues that this contributed to and exacerbated the 2010 Flash Crash.

In Chapter 4, which is joint work with Mark Westcott, we provide evidence on how racial attitudes can persistently be altered. Following the work by Bisin and Verdier (2001) that established a theoretical foundation for why certain types of preferences may be transmitted across generations, several empirical papers have reported persistence, for example of mistrust (Nunn and Wantchekon, 2011), or antisemitism

(Voigtländer and Voth, 2012). Psychologists in turn have long argued that prejudice towards minorities can be effectively reduced by increasing interaction between majority and minority, a framework called *intergroup contact theory* (Allport, 1954).

We provide a test of the persistence of this intergroup contact on racial attitudes using archival data. During World War II, more than one million American soldiers were transferred to the United Kingdom to help in the preparation of defeating Germany and its allies. Approximately 10% of these soldiers were African Americans, serving almost exclusively in segregated support units. Most British had not seen black people before, as the United Kingdom had only a very small black pre-war population of 8,000 that lived mostly in London and the port cities. Since deployment of troops was done according to military needs, we assume interaction with African American soldiers to be exogenous to pre-existing racial attitudes. We exploit the variation in the deployment patterns to construct a measure that reflects the extent of contact between the British population and African American soldiers. Using a non-representative internet survey and a complete membership list of the *British National Party* (BNP), an extremist party widely considered to be racist, we investigate the effects of historical contact with African American soldiers on contemporary racial attitudes.

We find that areas in which the likelihood of contact with black G.I.s during World War II was higher, see a significantly lower rate of BNP membership in 2007. Adding fixed effects and several control variables does not change this effect. Varying the clustering level, using different treatment indicator definitions, splitting the sample and using different specifications also provide stable and significant estimates. We then construct a radius around each geographic unit in which we observe racial attitudes and count the black troops located within this radius as an alternative treatment measure. Expanding the radius makes the effect fall almost linearly, as the likelihood of interaction between population and troops is decreasing in distance. Data from an internet survey confirms our initial findings, as respondents from areas with more exposure to African American soldiers state to have "warmer feelings towards black people".

To our knowledge, the findings in this chapter are the first that confirm a persistent positive effect of intergroup contact on stated and revealed racial attitudes. Especially in the light of the current Syrian refugee crisis in Europe, and the revitalized nationalistic movements in many developed countries, these insights carry important policy relevance. If prejudice between any minority and majority groups exists, bringing these groups together can offer an effective remedy. Integration policy should therefore also focus on contact-enhancing actions.

Chapter 5 is joint work with Christoph Koenig and investigates the relationship of gun ownership and crime rates founded on behavioral arguments. The debate about tougher gun legislation is fiercely fought in the political discourse of the United States, as access to firearms is a constitutional right for every citizen. Proponents of the right to bear arms usually argue that armed citizens can provide a credible deterrent to criminals, such that crime rates actually decrease. Gun control advocates however reason that the high rates of gun violence that the United States experience is largely due to the easy access to firearms. While some prior studies have tackled the relationship of gun prevalence and crime in predominantly correlational studies, the results are mixed.

We provide a novel approach in that we look at the gun demand shock that followed the shooting at Sandy Hook Elementary School in Newtown, Connecticut in December 2012. Fearing tougher gun legislation and perceiving a higher need for self-defense capabilities, millions of citizens across all US states were interested in acquiring firearms. Some states however had implemented waiting periods or installed other bureaucratic hurdles that prevented purchasers of firearms to receive their guns instantly, which we hypothesize eventually discouraged some citizens to buy any firearms at all. We therefore investigate the effect of gun purchase delay legislation on gun purchases and then use the results to estimate an impact on crime rates.

Gun purchases increase significantly stronger after the shooting at Sandy Hook in states in which no purchase delay legislation was in place. The effect is not driven by violations of the parallel trends assumption, as we include state-specific time trends and conduct a synthetic control exercise (Abadie et al., 2012). Most importantly, using Google search data we show that the ex ante interest in acquiring a firearm was not significantly different across the states, and that only the eventual buying decision differed. This is consistent with present-biased consumers postponing their purchase, but not with arguments from standard economic theory such as transaction costs alone. We can also rule out that demand moved from primary markets (gun stores) to secondary markets (gun shows), as both supply and demand for gun shows were unaffected.

We then investigate the effect on crime rates. All crime rates seem unaffected by the differential firearm purchasing rates, except for murder. Murder rates increase significantly in states that don't have purchasing delays in place. Our estimates suggest that approximately 98 lives could have been saved in each month of 2013 alone. The effects are again significant to numerous robustness checks. Including the 2012 Presidential election in the analysis shows that the effect is similar, suggesting that there was no direct effect of the mass shooting on crime rates. Using additional data, we are then able to show that all of those additional murders were committed with firearms and that a disproportionate share of victims were women, that were predominantly killed by men.

While our results should not be understood as an argument towards abolishing gun ownership rights, they should rather be seen as insights into how impulsive acts of violence can be prevented. Being able to save almost 1200 lives per year, of which a substantial fraction are women, seems substantial compared to a total of about 30,000 lives that are being lost in the United States each year in firearm-related deaths. Additionally, the robust relationship of gun ownership and murder rates suggests that there is scope for improving filter mechanisms to determine who should own a gun and who should not.

All five chapters of this dissertation are self-contained, they have their own introductions and can be read independently. Each chapter has its own appendix, and all appendices are added after Chapter 5. The bibliography containing all references can be found at the end of this document.

Chapter 1

Risk, Time Pressure & Selection Effects[*]

1.1 Introduction

Managerial decision making is often made under adverse conditions, including exposure to stress and time pressure (e.g. Claessens et al., 2007; Maruping et al., 2015, and references therein). In contrast, experimental studies of managerial decision problems (e.g., risk taking, or bargaining) typically provide decision makers with ample time to make their choices. They possibly allow the decision makers to revise and correct their choices, and sometimes provide learning opportunities in the form of repeated trials. A decision maker should find herself in an optimal setting to make good decisions. Because of the often unfavorable conditions in the field, researchers interested in the descriptive aspects of managerial decision-making have started to study decisions in controlled experimental settings that try to mimic aspects of these unfavorable decision environments. An important aspect of decisions in the field that has been transferred to controlled laboratory settings is the presence of time pressure in decision making (Spiliopoulos and Ortmann, 2015). Allocating people randomly into time-constrained and time-unconstrained decision environments, researchers have identified the causal effects of time pressure on various types of decisions (e.g., Sutter et al. (2003), on bargaining; Kocher and Sutter (2006), on beauty contests; Baillon et al. (2013), on decisions under ambiguity; Kirchler et al. (2014), on risky decisions; Buckert et al. (2015), on imitation in strategic games; El Haji et al. (2016), on bidding in auctions).

Observing decisions in adverse, but controlled environments is important, because it provides insights into decision processes that help to develop externally valid, descrip-

[*]This chapter is based on joint work with Martin Kocher, Stefan Trautmann and Yilong Xu.

tive models of decision making. However, the existing approach, with an exogenous variation of the aspect that affects the decision environment, has two potentially serious problems relating to issues of self-selection. First, and specifically for the case of time-constraints, if time pressure is supposed to be substantial and relevant, some people may violate the time constraint. The sample of decisions observed in the data set is self-selected.[1] Failure to take these selection effects into account may therefore result in a false interpretation of the observed behavior in terms of population averages. For example, Tinghög et al. (2013) argue that failures to replicate time pressure effects on cooperation in public goods games may be due to the original studies excluding about half of the participants because of failure to meet the time constraint (Casari et al. (2007), make a similar observation in the context of auction bidding). The second problem applies more broadly to any aspect of adverse environments implemented in randomized experiments. Outside the laboratory, people self-select into occupations and thus into job-related decision making environments.[2] In contrast, participants in experiments are exogenously assigned to a treatment condition that may not fit well with their tastes and skills (Al-Ubaydli and List, 2015). External validity of the observed experimental behavior for similar decision environments outside the laboratory thus cannot be taken for granted. Despite similarity of the experimental and the natural decision environments, the decision makers may systematically differ across the two settings in a self-selected way. Importantly, while external validity is an issue in any empirical study, it is a more central aspect in experiments that explicitly aim to mimic natural decision environments.

Observing that selection issues are at the heart of behavioral experiments with time pressure and other adverse conditions, we are the first to study (1) the empirical relevance of selection effects and (2) whether there are individual-level correlates based on observable background variables that can be used as predictors of the ability of a decision maker to cope with time pressure. We use the term time-pressure resistance for such ability. It relates to differences in the decision process, including the decision maker's time management. Our study aims to provide insights into these processes, and how they differ between decision makers. To this end, we collect data on risky decisions under time pressure, augmenting a design used in Kocher et al. (2013) to allow for both between-subject and within-subject analyses of behavior across time-constraint conditions. Thus, we observe each decision maker's risky choice behavior both in the presence and in the absence of time pressure for a similar set of risky alternatives.

[1]A similar problem obtains in studies on the effect of stress on decision making, where analyses of data are typically restricted to those participants who show a cortisol reaction under stress (Trautmann, 2014).

[2]External validity may be less of a problem in the context of time-limited offers for consumers, where self-selection seems less likely (e.g. Sugden et al., 2015).

We assess participants' scores on a measure of cognitive ability, on a score of cognitive efficiency, and a set of personality traits. We test whether these individual differences predict decision quality (measured in terms of expected payoffs in risky decision making, discussed in detail below) under time pressure and in the absence of time pressure. Importantly, while performance under time pressure can be measured in many ways, a risky decision task requires complex reasoning and has no obvious solution from the perspective of the decision maker (because optimal choices depend on preferences). Consequently, the decision maker has to choose a decision strategy, and this strategy may be affected by time pressure. Our performance measure aims to detect such shifts in strategy. Alternatively, we interpret a fitted decision model as a proxy for the person's decision strategy, and observe how the model performs across different time-pressure conditions. The details of the experimental design, including our measures of cognitive ability and efficiency are described in much detail in the next section.

Employing this design, we provide the following results. First, we find clear differences in decision styles across people that can be observed in the absence of time pressure and which affect the success in mastering the time constraint when it is present. That is, selection is very relevant. Second, while various observable characteristics correlate with success in decision making and with the decision maker's ability to maintain her decision style in the presence of time pressure, there is still an important role for unobserved factors. In other words, it difficult to predict who is time-pressure resistant and who is not, leading to modest predictive power of our models.

1.2 Studying self-selection in an adverse environment: Experimental design

We implement an experimental structure that allows us to observe both between- and within-subject differences in risky decision-making behavior in an adverse and a regular decision environment. This is how we can causally identify the effects of time pressure on risky decision making at the individual level. As decision makers are likely to show different reactions to adverse decision environments, we also measure potentially selection-relevant individual characteristics that may explain the different reactions. More specifically, for each participant we observe (i) risky choices in the absence of time pressure; (ii) risky choices in the presence of time pressure; (iii) a measure of cognitive ability ("IQ"); and (iv) a measure of intellectual efficiency ("IE"). We discuss the different tasks in detail below. The general structure of the experiment carefully counterbalances the order of the different parts as shown in Table 1.1, in order to avoid

Table 1.1: Treatment Design

Treatment (#obs)	Part 1: Indiv. Differences	Part 2: Risky Choices Set 1	Part 3: Risky Choices Set 2[a]	Part 4: Indiv. Differences
1 (93)	IQ	Time pressure	No time pressure	IE
2 (94)	IQ	No time pressure	Time pressure	IE
3 (96)	IE	Time pressure	No time pressure	IQ
4 (96)	IE	No time pressure	Time pressure	IQ

Notes: IQ: measurement of cognitive ability; IE: measurement of intellectual efficiency; a: a set of pure gain choices was added after Set 2 to give subjects the possibility to earn back potential losses in sets 1 and 2 (see section 1.2.2). Note that the IQ and IE tasks allowed subjects to move back and forth across items while this was not possible in the risky choice tasks.

unobservable order effects in our design. Our setup allows us to predict behavior under the adverse influence of time pressure by behavior in the absence of time pressure, controlling other relevant observable characteristics.

Each set of risky choices (Set 1 and Set 2) consists of 24 binary choices (see Table A.1 in the appendix for a full description). Time pressure was imposed by setting time limits for each of two 12-item subsets of choices within each of these two sets of risky choices. In particular, each set consists of (i) one subset of 12 binary choices that compare pure loss lotteries with mixed lotteries of lower expected value ("prominent gain"); and (ii) one subset of 12 binary choices that compare pure gain lotteries with mixed lotteries of higher expected value ("prominent loss"). A detailed description and motivation of these choice tasks is given in Section 1.2.2.[3] An important feature of the time-pressure implementation is that the time limit is imposed on the subset level, not on each choice item. That is, participants could go through the items in each subset at their own pace and therefore had to organize the allocation of time to the different choices efficiently. However, subjects were not allowed to go back and reconsider earlier choices. This setup requires subjects to manage their time use efficiently over a set of decisions, creating a trade-off between making more careful/thoughtful decisions on some items, and a higher chance to be able to complete all problems.

To make time pressure and no time pressure conditions as similar as possible in the presentation of the instructions and the task design, the unconstrained task also involved a time limit. However, this limit was selected such that it would not provide an actually binding constraint for subjects, namely at 420 seconds in all subsets. The extent of the time constraint in the time pressure conditions was calibrated in pre-test sessions such that there would be significant time pressure, while not making it impossible for the subjects to perform the decision task. In particular, under time pressure, the time limits were set at 120 seconds for the set with the prominent gains and at 80 seconds for the set with the prominent losses. In both cases, this value

[3]Because losses were possible in the lottery choices, we included a set of lottery choices after the main task (Set 2 / Part 3) that gave subjects the possibility to earn back any losses from sets 1 or 2. See section 1.2.2 for details.

implied a 20% reduction of the median decision times in the absence of constraints that we observed in six pilot sessions (details about the pre-test sessions are given in the online supplement).

1.2.1 Cognitive ability, intellectual efficiency, and personality measures

We employed the Raven's Advanced Progressive Matrices (APM) test (Raven and Court, 1998) to measure cognitive ability ("IQ") and intellectual efficiency ("IE"). Cognitive ability assesses a subject's cognitive reasoning power, i.e. the extent to which complex information can be processed. Intellectual efficiency measures cognitive reasoning speed, i.e. how fast incoming information can be processed. It gives us a measure for the ability of a decision maker to cope with time pressure.

Raven's progressive matrices are often used as a nonverbal assessment of cognitive reasoning power or general intelligence; its advanced version is aimed at subjects in the high cognitive ability ranges such as university students. In each item, subjects are presented with a 3-by-3 matrix of abstract symbols, with the symbol in the lower right corner missing. They are asked to choose, among eight possible alternatives, the one that completes the pattern in the matrix. We communicated to subjects that the items in the task were arranged in ascending order of difficulty and that they could go back and forth within the time limit to revise their answers. An example can be seen in Figure 1.1, where the correct answer is option 3.

Instead of running the full 48-item test, a short-form[4] containing 12 selected items from the APM test was administered to obtain a measure of IQ, as it has been argued that conducting the full APM does not add much predictive power (the correlation between the two formats is $\rho = 0.88$, see Bors and Stokes (1998)). As we are interested in the cognitive capacities of subjects, we allowed subjects to answer all twelve items at their own pace. To keep instructions as close as possible to our measure of intellectual efficiency (details below), we implemented a non-binding time constraint of 25 minutes, which was again calibrated in pre-tests.

We use the remaining 34 items[5] from the APM to construct a measure of cognitive efficiency, i.e. the speed of cognitive reasoning (Raven and Court, 1998). By imposing

[4] The short form of the APM test we used here was introduced by Bors and Stokes (1998), consisting of items 3, 10, 12, 15, 16, 18, 21, 22, 28, 30, 31, and 34 of the APM (Set II). It is more difficult, and therefore suits university students better, than the other short version proposed earlier by Arthur and Day (1994).

[5] We use the first two items in Set I of the APM test as instructional items, leaving 34 items for our intellectual efficiency measure.

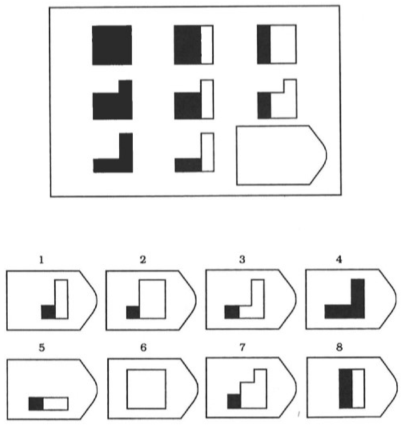


Figure 1.1: A Sample Screen from Raven's APM

a severe time limit on subjects (13 minutes to solve all 34 items), we measure how fast and how efficiently they can process information. They could again reconsider earlier choices at any time. The timing constraint proved to be binding in pre-tests: no participant was able to finish all questions within the time limit.

Based on the tasks, we define our measures IQ and IE as the number of correct items in the cognitive ability and the intellectual efficiency tasks, respectively. There is no additional reduction of the score for wrong answers or missing items. Note that, as shown in Table 1.1, the IQ and IE tasks were counterbalanced separately for each ordering of the time pressure tasks. The IQ and IE tasks were incentivized such that (1) a higher score yields a higher chance to win a monetary prize, and (2) subjects could never identify their number of correct answers exactly. We provide more details on and the procedure in Appendix A.3.

At the end of the experiment, we elicited several personality measures that have often been linked to managerial decision making (e.g. Nadkarni and Herrmann, 2010), and are potentially relevant to the ability of coping with time pressure. The Generalized Self-Efficacy Scale (Schwarzer and Jerusalem, 1995) is a ten-item questionnaire predicting subjects' perceived ability to cope with daily hassles and adaptation after stressful events. Time pressure is thus a natural environment in which self-efficacy may have an effect on decision-making quality. In particular, in line with previous research showing a positive correlation between measures of self-efficacy and good financial planning behavior (Kuhnen and Melzer, 2014), we hypothesized that higher level of self-efficacy may be associated with better decision quality under time pressure. Rotter's Locus of Control questionnaire (Rotter, 1966) is a 28-item survey that assesses the extent to which individuals believe that they have control over events that affect them in their lives. These beliefs may become especially important when control is exogenously manipulated, as was the case under time pressure in our experiment.

A Big Five Ten Item (TIPI) questionnaire (Gosling et al., 2003), which measures the Big Five personality characteristics using only ten questions was also administered. Here we predicted that traits such as conscientiousness could become a burden under time pressure, as subjects would not be able to implement the degree of diligence they would typically prefer. Finally, we elicited some general demographics and background data.

1.2.2 Risk preference measures

Our main task involves binary risky choices. We build on the design in Kocher et al. (2013), who analyzed risky decisions under time pressure. That study found strong time pressure effects for lottery choices involving mixed gambles, i.e., including both gains and losses. In particular, under time pressure, decision makers seem to be prone to prefer mixed gambles over pure loss gambles with higher expected value (thus being drawn by the prominent gain in the mixed gambles); similarly, decision makers seem to prefer pure gain gambles over mixed gambles with a higher expected value (thus being repelled by the prominent loss in the mixed gamble). Conte et al. (2016) find similar prominence effects under time pressure. Because we want to study the role of selection effects under time pressure, we deliberately employ this particular structure of lottery choices, expecting to induce robust time pressure effects with the design. As described before, we present subjects with two sets of choices, one set being time-constrained, and the other de facto unconstrained. Each set consisted of a subset of 12 choices of the prominent-gain format, and 12 choices of the prominent-loss format. The two subsets were separately time-constrained as described above. Within each subset, subjects had to proceed through the choice problems in a given order (fixed over all subjects and conditions) and could not go back to revise previous choices. A list of all choices is given in Appendix A.1. A screenshot of the presentation of the choices is given in Appendix A.2.

Subjects made as many choices as possible within the time constraint. At the end of the experiment, one choice was selected randomly from all the potential choice problems in Set 1 and Set 2, and payoffs depended on the decision made, i.e. the lottery chosen, in this choice problem (this procedure prevents wealth or house money effects, which we considered relevant in the context of risky choice). If a subject violated the time constraint and thus failed to answer some of the questions, she would receive the lowest possible outcome (i.e., the highest possible loss) if one of the unanswered decision problems was selected for payment. For example, if the selected choice problem involved a choice between the lottery (15%: €-15, 85%: €-11) and (15%: €12, 85%:

€-17) (see S1 / G11 in Table A.1), the earnings for a person who did not submit a decision was €-17.

Because the risky choices of sets 1 and 2 involve potential losses, we needed to endow the participants with sufficient funds to cover any losses they might incur in parts 2 and 3 of the experiment. Therefore, an additional task was added after Set 2 that involved six risky choices between lotteries in which the lowest possible gain amounted to €20. By adding the endowment task after all Part 2 and Part 3 choices had been made, and by endowing with the help of risky choices, we hoped to prevent subjects from integrating the endowment easily with the loss outcomes in the choices in earlier parts. This method was adapted from Kocher et al. (2013). We did not impose a time limit on the endowment task; at the end of the experiment, one of the six choices was randomly selected for payment, and earnings were added to earnings from the lottery selected from sets 1 and 2. When working on Set 1 and Set 2, subjects were not aware how the subsequent task would look like; they only knew that other parts were to follow and that they would not incur overall net losses from the experiment. In the analyses below, we always report performance based on behavior in sets 1 and 2, thus not incorporating the endowment task.

1.2.3 Laboratory details

The experiment was programmed using z-Tree (Fischbacher, 2007) and recruitment was done with the help of ORSEE (Greiner, 2015). We conducted 16 experimental sessions at the MELESSA laboratory at the University of Munich in July and September 2014. In total, 379 subjects took part in the experiment, up to 24 in each session. They were mostly undergraduate and graduate students from a diverse set of programs that the university offers. Payoffs were determined by randomly selecting Part 1, Part 4 or Part 2/3 for payment, with payment details then depending on the procedures described in the previous subsections. A typical session lasted for about 75 minutes, and subjects earned on average about €16.63 (approximately $21.32 at that time). In addition, we ran several pilot studies in Munich and Tilburg to calibrate the appropriate timing constraints. Information on the pilots is given in the supplementary material.[6] Experimental instructions can also be found in the supplement.

[6]See https://dl.dropboxusercontent.com/u/11242744/KSTX_webappendix_sep2016.pdf.

1.3 Time pressure and risky decisions: Manipulation check

We first consider whether the time pressure manipulation for the risky decisions was effective in terms of time-use, in terms of the number of participants violating the time constraint, and in terms of the number of unanswered decision items. Table 1.2 shows the results. Clearly, subjects made substantially faster decisions under time pressure (Mann-Whitney test, $p < 0.01$), were more likely to violate the time constraint (Mann-Whitney test, $p < 0.01$), and had more missing items (Mann-Whitney test, $p < 0.01$). The manipulation of time pressure was successful in providing a highly adverse decision environment.

Table 1.2: Time Pressure Manipulation for Risky Decisions

Treatment	Actual time used (average, in sec.)	# of subjects violating the time constraint	# of missing items per person
Time pressure (N=189)	164	58	0.87
No time pressure (N=190)	279	2	0.01

Notes: Decision times reported show the sum of time used for the two subsets. Total time constraint was 200 seconds under time pressure, and 840 seconds in the absence of time pressure.

Table 1.3: Time Pressure Effects on Risky Choices

Treatment	Expected value (choices made; €)^{a,b}	Expected value (all decision problems; €)^{a,c}	Percent of choices avoiding prominent loss	Percent of choices seeking prominent gain
Time pressure (N=189)	-1.31	-1.61	59.85%	62.50%
No time pressure (N=190)	-1.14***	-1.15***	50.83%***	55.66%**

Notes: a: averages reported; b: numbers reflect the expected value implied by choices actually made; c: numbers reflect the expected value implied by all choice problems, including missing items; *,**,*** indicates significance of difference from time pressure condition at the 10%, 5%, and 1% level, two-sided Mann-Whitney tests.

We next consider the effect of time pressure on risky decisions. Table 1.3 shows, for time-constrained and for unconstrained choices, the average expected payoff that is implied by the choices the subject actually made, the average expected payoff that is implied by all choice problems including missing items (which count as the highest loss)[7], the percentage of choices that avoid a prominent loss, and the percentage of choices that seek a prominent gain. The latter two percentages are conditional on the items that a person has answered. We observe that time pressure significantly reduces decision quality. The expected value (EV) implied by choices actually made is lower under time pressure. Clearly, missing items lead to additional losses and further reduce payoffs under time pressure. Under time pressure, participants make significantly more

[7] As a benchmark for the subsequent analyses we observe that the highest realizable expected value was €-0.39, and the lowest was €-2.28 if all choices were actually answered. Not answering any item would yield an expected payoff of €-11.23.

choices that avoid a prominent loss or seek a prominent gain, at a loss of expected value, than in the absence of time pressure. That is, choices are more heuristic under time pressure, being affected by salient attractive aspects of the lottery and sacrificing expected payoff. We observe that, despite differences in the implementation of the time constraint and in the design of the lottery choices, these comparative findings exactly replicate those reported in Kocher et al. (2013). Note however, that the share of choices for the lower-EV option (loss avoiding or gain seeking, respectively), is higher in the current experiment than in Kocher et al. (2013).

We observe that participants realize a lower expected value under time pressure. In the subsequent analyses, we consider expected value as a core measure of decision quality. This interpretation is supported by the direct (inverse) link of expected value to heuristic choices (loss avoiding and gain seeking). Expected payoff is also a criterion that is applied in many professional settings outside the laboratory to assess decision success. However, participants may not necessarily aim to maximize expected value in the experiment. In the supplementary material, we thus present the main results also under the alternative assumption that participants' decisions may reflect cumulative prospect theory preferences.

1.4 Results: Identifying selection

We first approach the question whether selection is relevant under time pressure. To this end, we compare those decision makers who violate the time constraint under time pressure ($N = 90$) to those who do not ($N = 289$). A violator is defined as a decisions maker who ran out of time before making all 12 choices, in at least one of the two subsets of risky decisions in her time-constrained part. Clearly, these two groups will thus differ under time pressure. However, the within-person design also allows us to study whether these groups differ when they are not time-constrained.

Table 1.4 shows results of the comparison between the two groups for various behavioral and performance measures. The left panel shows behavior in the absence of time pressure. Subjects who violate the time constraint differ substantially from those who do not violate the constraint in the way they approach the risky decision task. In terms of decision processes, violators use more time and distribute their time less evenly across choices. Moreover, violators are less affected by salient loss or gain features of the lotteries. They consequently perform significantly better on average in terms of the implied expected value of their choices than non-violators, when not exposed to time pressure. In the right panel of Table 1.4 we compare the two groups in the presence of time pressure. Also under a time constraint, violators use more time and have a higher variance of time used across choice problems. They perform significantly worse on the full set of choices. This effect is driven by the relatively strong punishment for not answering a choice problem, which they seem not to take sufficiently into consideration in their strategy. Moreover, under time pressure, violators do not anymore perform better than non-violators on the choices they actually made. However, violators do not perform significantly worse than non-violators on these items either.

Table 1.4: Differences between Time-constraint Violators and Non-violators

	No time pressure		Time pressure	
Performance measure	Violators (N=90)	Non-violators (N=289)	Violators (N=90)	Non-violators (N=289)
Actual time used (in sec.)	323.79	221.13***	188.87	148.72***
Variance of time used per item	77.75	37.29***	30.82	10.30***
Expected value (choices made; €)	-1.06	-1.21***	-1.29	-1.22
Expected value (all decision problems; €)	-1.08	-1.21***	-2.29	-1.22***
Expected value (items w/o violations; €)^a	-3.69	-3.90***	-3.72	-3.90***
Percent of choices avoiding prominent loss	49.63%	54.79%*	52.72%	55.45
Percent of choices seeking prominent gain	46.96%	60.67%***	47.95%	60.84%***

Notes: Violator status for each subject is assigned if at least one item in at least one time-constrained subset was not answered; *,**,*** at the entries for non-violators indicate that these values differ from those for violators, at the 10%, 5% , and 1% significance level, two-sided Mann-Whitney tests. a: expected value calculated on the basis of those items in the prominent gain and prominent loss subsets that all subjects were able to answer under time pressure.

Table 1.4 also shows the expected value over the set of choices for which no subject violated the time constraint (row five).[8] This includes the first seven choices in the prominent gain sets and the first three choices in the prominent loss sets. Apparently, on this subset of early choices, the violators perform much better than the non-violators do, and this holds true in both the time pressure and the no-time pressure condition. Moreover, comparing this performance measure across time-pressure conditions, we observe that the expected payoffs do not differ for either group of decision makers (for both groups, $p > 0.7$, Wilcoxon signed-rank test). Thus, under time pressure, initially the violators can fully implement the same decision strategy as in the absence of time pressure. However, as they move on and run out of time, they lose out, harming their overall performance for the choices they actually make (shown in row three in Table

[8]For this row, we removed one subject, a violator, from the analysis in this table, as she is the only participant violating the time constraint already at the fourth item for the subset of prominent gain lotteries, while others violate only after the seventh item. The results remain the same if we include all subjects but we lose a substantial amount of information for the subset of prominent gain lotteries.

1.4) and even more significantly for the full set of choices (shown in row four in Table 1.4).

Table 1.5: Differences between Time-constraint Violators and Non-violators across Choice Items under Time Pressure (Expected Value of Choices Made)

Choice (pr gain)	Violators	Non-Violators	Choice (pr loss)	Violators	Non-Violators
1	-9.89 (N=90)	-10.08 (N=289)	13	8.56 (N=90)	8.55 (N=289)
2	-5.09 (N=90)	-5.14	14	8.93 (N=90)	8.59*
3	-12.37 (N=90)	-12.56*	15	4.61 (N=90)	4.49
4	-5.57 (N=89)	-5.75**	16	8.33 (N=89)	8.30
5	-14.17 (N=89)	-14.56***	17	6.12 (N=89)	6.15
6	-8.94 (N=89)	-9.11**	18	9.42 (N=88)	9.39
7	-3.22 (N=89)	-3.44***	19	11.74 (N=87)	11.57
8	-11.71 (N=87)	-11.85**	20	5.61 (N=84)	5.68
9	-9.42 (N=83)	-9.70**	21	7.50 (N=77)	7.54
10	-13.11 (N=77)	-13.24	22	4.31 (N=66)	4.22
11	-12.28 N=(63)	-12.34	23	4.25 (N=47)	4.41
12	-8.37 (N=45)	-8.38	24	8.02 (N=19)	8.03

Notes: Entries are expected values (€) of choices made; averages over participants in the subgroup; *,**,*** at the entries for Non-violators indicate that these values differ from those for Violators, at the 10%, 5% , and 1% significance level, Mann-Whitney test. Number of observations in parentheses (constant for Non-violators).

This dynamic pattern of performance is shown in more detail in Table 1.5. The table shows for each item, in the order of appearance, the implied expected value of the choices made by violators and non-violators under time pressure. In the set of prominent gains (items 1-12), violators perform better early on. In the set of prominent losses (item 13-24), the effect is less pronounced, but points in the same direction. As they move on with the task and time runs out, violators do not make better decisions than the non-violators anymore. This is especially true in the set with prominent gains. Additionally, violators at some point violate the time constraint (shown by the decrease in sample sizes indicated for each choice item), leading to significant losses in expected value over all decision problems. We also observe that violation of the time constraint for prominent-loss choices leads to an additional loss of expected payoffs for the set of choices made. This is caused by the fact that lotteries in this subset had positive expected payoffs, and thus a participant's average expected payoff over choices made is harmed by simply reducing the number of prominent-loss choices that are completed. This effect leads to the negative effect on expected value for choices actually made under time pressure, shown in row three of Table 1.4.

Figure 1.2 provides further illustration of the time-use strategies of violators and non-violators. The figure plots for each item the average decision time for violators and for non-violators. In the absence of time pressure (upper panel), violators spend more time on each item than non-violators, and use a substantial amount of time for certain, apparently difficult items, leading to higher variance in terms of the average time-used per item. Under time pressure (lower panel in Figure 1.2), violators initially use too much time and then almost monotonically reduce their time used per item as they progress through the task.[9] Compared to the non-violators who remain relatively stable in their time used per item as they proceed through the task, violators thus initially use more time and later have even less time than the (on average) much faster non-violators take for the last few items. That is, given the significant punishment for violation of the time constraint, violators exhibit poor time management.

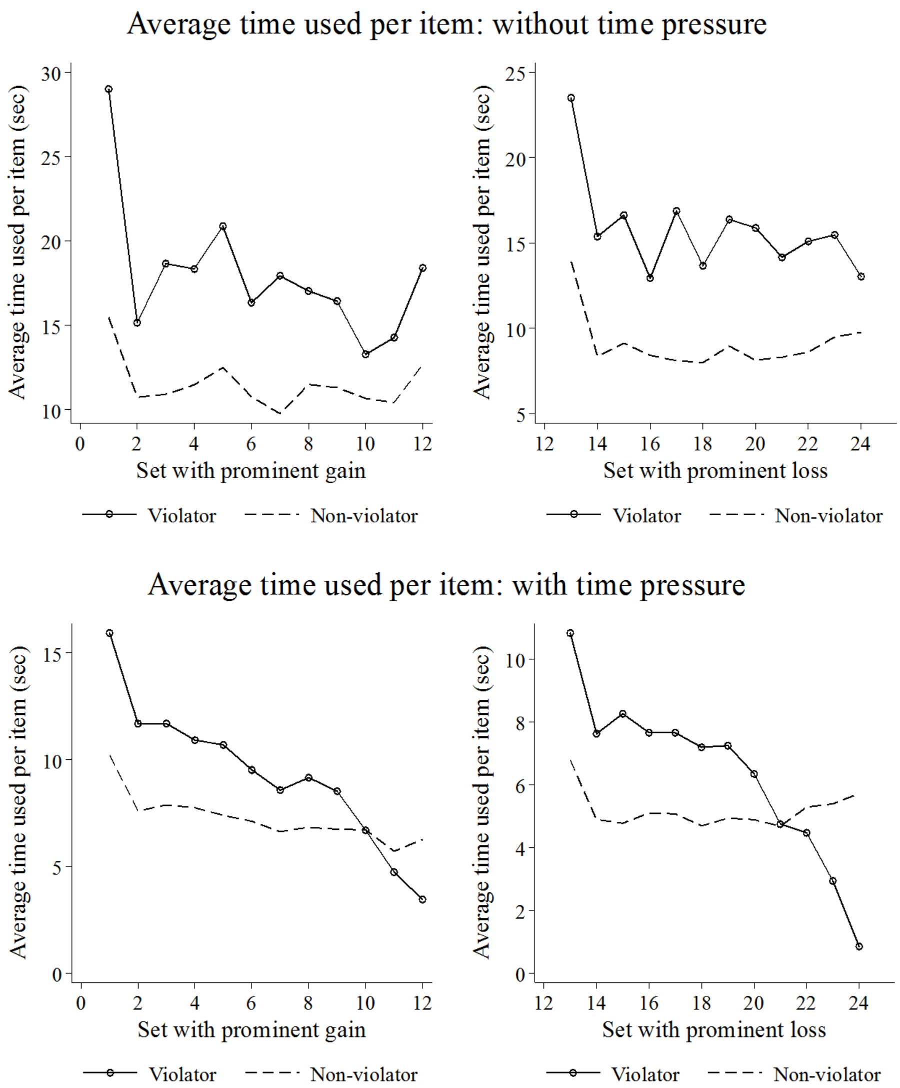


Figure 1.2: Time Use of Violators and Non-violators

An important question regarding the external validity of experimental observations of choice behavior concerns the within-person correlation of behavior in time-

[9]All items that a person cannot answer because time ran out are counted as zero time used. This is consistent with the person having indeed used zero seconds to make the decision. Note that the non-zero time use for later items in each subset by each group of violators is caused by the definition of violator based on a violation in at least one of the two subsets. Thus, not all violators run out of time in the prominent gain subset (prominent loss subset).

constrained and unconstrained decision environments at the individual level. Tables 1.4 and 1.5 showed that violators cannot implement their decisions as successfully under time pressure as in its absence. Table 1.6 displays the correlation of various behavioral measures across time pressure conditions; correlations are given separately for violators and non-violators. We observe that behavior and outcomes are positively correlated across environments, though correlations are larger for non-violators than for violators. While non-violators seem to be able to implement similar decision strategies both in the absence and in the presence of time pressure, violators are less able to sustain the same strategy in the different decision environments, especially for the last few items when they run out of time.

Table 1.6: Correlations of Behavior across Time Pressure Conditions

Performance measure	Spearman rank Correlation (only violators, N=90)	Spearman rank Correlation (only non-violators, N=289)
Actual time used (in sec.)	0.42***	0.63***
Variance of time used per item	0.47***	0.51***
Skewness of time used per item	0.03	0.09
Expected value (choices made; €)	0.37***	0.86***
Percent of choices avoiding prominent loss	0.69***	0.81***
Percent of choices seeking prominent gain	0.68***	0.82***

Notes: *,**,*** indicate that the correlation of behavior with time pressure and without time pressure is different from zero at the 10%, 5% , and 1% significance level, Spearman's rank correlation coefficients.

To sum up, we document significant selection effects under time pressure: those participants who cannot cope with the time constraint have very different time-use and decision strategies in the absence of time pressure. They make better decisions when sufficient time is available (i.e., under no-time pressure conditions) than those subjects who are not violating time-constraints under time pressure. However, under time pressure they are not able to sustain these strategies anymore and therefore lose out against the non-violators in terms of performance.

1.5 Results: Predicting who can better cope with time pressure

1.5.1 The effect of observable traits: Non-parametric analyses

Having observed that there are systematic differences in behavior between those who can and those who cannot cope with time pressure well, we now investigate whether there are observable traits or characteristics that allow predicting time-pressure resistance to in decision making. Table 1.7 shows that there are some differences in intellectual efficiency and self-efficacy between violators and non-violators. While not significant on conventional levels, differences in gender look suggestive. Moreover, we have already shown the pronounced differences in time use strategies between violators and non-violators, which are also observable in simple measurement tasks.

Table 1.7: Characteristics between Time-constraint Violators and Non-violators (Individual Differences)

Individual background variable	Violators (N=58)	Non-violators (N=131)
IQ	9.38	9.47
IE	22.47	23.27**
Big-Five dimensions		
Openness	5.64	5.53
Conscientiousness	5.56	5.68
Extraversion	4.62	4.78
Agreeableness	5.23	5.19
Neuroticism	4.82	4.95
Locus of control	12.64	12.27
Self-efficacy	28.36	29.80**
Gender (% male)	29.31 %	40.46 %
Age	23.93	23.75

Notes: *,**,*** at the entries for Non-violators indicate that these values differ from those for Violators, at the 10%, 5% , and 1% significance level, two-sided Mann-Whitney tests.

In the following we therefore study the differences between low and high IQ, low and high IE, low and high self-efficacy, small and large amounts of time used / variance of time used (in the absence of time pressure), and between males and females. Importantly, these differences are observed for all subjects, irrespective of whether they violate the deadline under time pressure or not. That is, we can also make use of variation in performance within the groups of violators and non-violators. Below we will use these measures also jointly in a multivariate analysis as predictors of decision performance under time pressure.

Effects of cognitive ability and intellectual efficiency. We consider the effects of IQ and IE on risky behavior, in time-constrained and unconstrained settings. Despite being positively correlated ($\rho = 0.5760$, $p < 0.01$, Spearman rank correlation), the correlation between IQ and IE is far from perfect, suggesting that the two measures capture separate traits. Since we are interested in outcomes and in selection effects, we report effects on the implied expected value from the choices made and from all choice problems, the percentage of time-constraint violators, the number of missing items per subject, and the incidence of avoiding prominent losses and seeking prominent gains. To allow for direct group comparisons, we split the sample at the median values of IQ and IE.[10] Table 1.8 reports the results for IQ, and Table 1.9 reports the results for IE.

As shown in the tables, in the absence of time pressure, low IQ and low IE subjects perform worse in terms of expected payoffs, and they are affected more strongly by salient features of the lottery than high IQ and IE groups. These results are consistent with previous findings in the literature (e.g., Dohmen et al. (2010)). The table also suggests that the amount of time used by high IQ/IE subjects and relatively low IQ/IE subjects are similar, in both conditions. In fact, there is no significant correlation between actual time used and IQ or IE.[11] Interestingly, the right panels of tables 1.8 and 1.9 show that IQ is correlated with decision quality under time pressure, while IE is not. This is counter to our prediction that IE measures the exact skill to make difficult decisions under a severe deadline. The fact that IQ effects remain significant under time pressure suggests that the absence of an effect for IE is not simply due to a larger noise under time pressure. Note that we consider univariate correlations, and IE and IQ may be differently affected by other variables that are related with behavior under time pressure (e.g., gender), which will be controlled for in the multivariate analyses below.

Effects of self-efficacy. Table 1.10 shows results for self-efficacy. Although self-efficacy directly measures individuals' ability to cope with hassles environment, similar to IE, there is no raw correlation with behavior and performance under time pressure and without time pressure.

Effects of gender. Table 1.11 shows pronounced gender differences both in the absence and in the presence of time pressure. Males take more time for decisions, are less

[10] The median IQ is 10. We split the sample such that IQ < 10 defines the "low" group. The median for IE is 23. We split the sample such that IE < 23 defines "low" group.

[11] A potential explanation for such weak correlation might be that high cognitive ability participants more carefully consider the decision problem, while low cognitive ability participants need more time to understand the problem.

Table 1.8: Effects of IQ

Performance measure	No time pressure (840 sec = 420+420)		Time pressure (200 sec = 120 +80)	
	IQHigh	IQLow	IQHigh	IQLow
Actual time used (in sec.)	257.72	228.62	159.29	156.82
Expected value (choices made; €)	-1.10	-1.27***	-1.14	-1.37***
Expected value (all decision problems; €)	-1.11	-1.27***	-1.41	-1.56**
Percent time-constraint violators	1.36%	0	24.09%	23.27%
Number of missing items per person	0.01	0	0.70	0.68
Percent of choices avoiding prominent loss	49.17%	59.64%***	50.33%	61.00%***
Percent of choices seeking prominent gain	52.17%	64.68%***	54.56%	62.24%***

Notes: *,**,*** at the entries for IQLow indicate that these values differ from those for IQHigh, at the 10%, 5% , and 1% significance level, two-sided Mann-Whitney tests. The median IQ score is 10. Total time was 840 in no-time pressure and 200 in time pressure conditions.

Table 1.9: Effects of IE

Performance measure	No time pressure (840 sec = 420+420)		Time pressure (200 sec = 120 +80)	
	IEHigh	IELow	IEHigh	IELow
Actual time used (in sec.)	250.96	237.38	157.44	159.48
Expected value (choices made; €)	-1.13	-1.24***	-1.19	-1.30
Expected value (all decision problems; €)	-1.13	-1.24***	-1.42	-1.54
Percent time-constraint violators	1.32%	0	22.47%	25.66%
Number of missing items per person	0.01	0	0.63	0.78
Percent of choices avoiding prominent loss	49.71%	59.32%***	51.92%	59.11%**
Percent of choices seeking prominent gain	54.56%	61.68%**	57.57%	58.09%

Notes: *,**,*** at the entries for IELow indicate that these values differ from those for IEHigh, at the 10%, 5% , and 1% significance level, two-sided Mann-Whitney tests. The median Intellectual Efficiency score is 23. Total time was 840 in no-time pressure and 200 in time pressure conditions.

affected by salient losses, and realize higher expected payoffs. Males seem to aim at maximizing expected payoffs, explicitly taking the time to calculate expected values. This is in line with self-reports of risk aversion, which are lower for males than for females (average degree of risk aversion equals 3.33 for males and 3.78 on a Likert scale ranging from minimum 1 to maximum 5, $p < 0.01$, two-sided Mann-Whitney test). Under time pressure, males are no more likely than females to violate the time constraint, despite their apparent maximization strategy. Males make worse decision under time pressure than in its absence (Wilcoxon signed-rank test, $p = 0.088$for choices made; $p < 0.01$ for all choices). However, the data suggests that they are on average still better able to cope with time pressure than females, despite their more time-consuming decision strategy.

Effects of time-use strategies. As seen before, time-use strategies differ strongly across subjects and correlate with violator status. Measuring the total time used and the variance across choice items in the conditions with no time pressure for all subjects,

Table 1.10: Effects of Self-Efficacy (SE)

Performance measure	No time pressure (840 sec = 420+420)		Time pressure (200 sec = 120 +80)	
	SEHigh	SELow	SEHigh	SELow
Actual time used (in sec.)	357.02	230.75	158.38	158.10
Expected value (choices made; €)	-1.17	-1.18	-1.28	-1.18
Expected value (all decision problems; €)	-1.17	-1.18	-1.47	-1.48
Percent time-constraint violators	0.47%	1.20%	20.66%	27.71%
Number of missing items per person	0.005	0.012	0.62	0.78
Percent of choices avoiding prominent loss	52.82%	54.52%	54.58%	55.08%
Percent of choices seeking prominent gain	57.52%	57.28%	57.99%	57.50%

Notes: *,**,*** at the entries for SELow indicate that these values differ from those for SEHigh, at the 10%, 5% , and 1% significance level, two-sided Mann-Whitney tests. The median Self-Efficacy score is 29. We split the sample such that SE < 29 defines the low group. Total time was 840 in no-time pressure and 200 in time pressure conditions.

Table 1.11: Effects of Gender

Performance measure	No time pressure (840 sec = 420+420)		Time pressure (200 sec = 120 +80)	
	Female	Male	Female	Male
Actual time used (in sec.)	224.57	277.13**	156.68	160.63
Expected value (choices made; €)	-1.23	-1.09***	-1.27	-1.18**
Expected value (all decision problems; €)	-1.23	-1.10***	-1.53	-1.38***
Percent time-constraint violators	0%	1.99%**	36.32%	19.87%
Number of missing items per person	0	0.02**	0.75	0.60
Percent of choices avoiding prominent loss	59.14%	45.14%***	60.33%	46.45%***
Percent of choices seeking prominent gain	59.28%	54.59%	58.60%	56.54%

Notes: *,**,*** at the entries for Male indicate that these values differ from those for Female, at the 10%, 5% , and 1% significance level, two-sided Mann-Whitney tests. Total time was 840 in no-time pressure and 200 in time pressure conditions.

we confirm the importance of the measures. Table 1.12 shows strong effects on violator status and missing items: those who make more careful decisions (higher time use and variance) are more likely to violate the time constraint and miss out an answering items. However, these careful decision makers do not perform worse on average than the less careful ones. If they manage to meet the time constraint, they realize a higher performance. Moreover, more careful decision makers are less prone to salience effects and obtain higher expected payoffs on the choices they make.

1.5.2 Parametric Decision Model

Tables 1.7 to 1.12 demonstrate that observable characteristics are strongly associated with risky behavior in the presence and absence of time pressure. Moreover, we found that these behaviors are correlated between the two environments, implying that subjects try to adhere to similar decision processes in both situations (Table 1.6). A natural question is whether we can link observable traits to a person's ability to maintain her

Table 1.12: Effects of Variance of Time Used Per Item and Total Time Used (in the Absence of Time Pressure) on Outcomes under Time Pressure

Performance measure	VARHigh (N=189)	VARLow (N=190)	TimeHigh (N=191)	TimeLow (N=188)
Actual time used (in sec.)	172.55	144.04***	176.46	139.77***
Expected value (choices made; €)	-1.15	-1.32***	-1.15	-1.32***
Expected value (all decision problems; €)	-1.53	-1.42	-1.57	-1.37
Percent time-constraint violators	34.39%	13.16%***	36.65%	10.64%***
Number of missing items per person	1.06	0.33***	1.19	0.19***
Percent of choices avoiding prominent loss	50.80%	58.79%**	49.57%	60.12%***
Percent of choices seeking prominent gain	49.73%	65.78%***	48.74%	66.97%***

Notes: *,**,*** at the entries for VARLow and TimeLow indicate that these values differ from those for VARHigh and TimeHigh, at the 10%, 5% , and 1% significance level, two-sided Mann-Whitney tests. The median variance of time used per item under no time constraint is 12.74. We split the sample such that VAR < 12.74 is the low variance group; and Time < 201 seconds is the low time (faster) group. Total time was 200 (in time pressure conditions).

decision processes under a tight time constraint. To answer this question, we estimate a simplified cumulative prospect theory model (CPT, (Tversky and Kahneman, 1992)) for each participant on the basis of her 24 time-unconstrained choices, and assess how well the fitted model predicts her behavior under time-pressure. Predictive success is then related to observable characteristics.

Our CPT model assumes linear utility and identical probability weighting parameter for gains and losses. Thus, we estimate two parameters, loss aversion and curvature of the weighting function (estimation code provided in the web supplement).[12] Because some individuals failed to submit all decisions under time pressure, we report both predictive success over choices made (missing items ignored), and predictive success over all decision problems (missing items counted as a failure to implement the decision model). The fitted CPT model has good predictive power. The average success rate of predicting choices under time pressure using time-unconstrained fitted parameters equals 67.71% for behavior over choices made and 65.71% over all decision problems, respectively. Both predictions are significantly better than the random prediction (both $p < 0.01$, Wilcoxon signed-rank tests).

The relation of observable characteristics with our measures of predictive success is shown in Table 1.13. We find strong links between observables and predictive success,

[12] We imposed constraints $\lambda > 0$ for loss aversion and $\gamma > 0$ for the weighting functions. We lose 37 observations because we could not recover reasonable parameter values for these participants from the estimation.

with virtually identical results for both prediction measures (i.e. for actual choices made and for all choices). For decision makers with higher measures of IE, higher measures of self-efficacy, and less time-use in the absence of time pressure, the fitted model predicts behavior under time pressure more successfully than for those in the relevant comparison categories. For IQ and variance in decision times, we find insignificant results. Moreover, males are better able to maintain their decision strategy, as are those who did not violate the time constraint. There is clear evidence that observables are related to a person's time pressure resistance. Importantly, in contrast to an evaluation based on expected payoffs, the current approach presumes no normative measure of success; each person is evaluated on the basis of her own choice behavior in the absence of time pressure, possibly showing loss aversion or probability weighting.

Table 1.13: Correct Predictions of the Individual Decision Model under Time Pressure

	Choice made		All choices	
	High	Low	High	Low
IQ	68.33%	66.27%	66.12%	64.75%
IE	70.17%	65.42%**	68.14%	63.43%**
Self-efficacy	69.50%	66.04%*	67.50%	64.03%*
Total time used under NTP	66.07%	72.98%***	63.50%	72.80%***
Variance of time used under NTP	68.19%	66.87%	65.37%	66.30%
Gender	Female	Male	Female	Male
	65.87%	70.60%***	63.81%	68.68%**
Violator	Yes	No	Yes	No
	63.70%	69.08%***	55.83%	69.08%***

Notes: Entries are percentages of correct predictions by the individual-specific CPT models (as estimated under no-time pressure) for choices under time pressure. *,**,*** at the entries for Low column indicate that these values differ from those for High column, at the 10%, 5% , and 1% significance level, two-sided Mann-Whitney tests.

1.5.3 Multivariate analyses

We finally provide multivariate analyses for our main dependent variables of interest. We study the partial effects of IE, IQ, self-efficacy, gender, as well as our two time-use measures on expected payoffs for the choices made (columns 1 and 2 in Table 1.14) and on expected value over all choices (columns 3 and 4 in Table 1.14), under time pressure. That is, we aim to identify whether using a set of observables allows us to predict outcomes under time pressure. In addition, we also conduct multivariate analyses for predictive success of the fitted CPT model under time pressure, both for choices made (column 5 in Table 1.14) and for all decision problems (column 6 in Table 1.14).

The results confirm our earlier findings, but show overall modest explanatory power of background variables for variations in expected payoff and predictive power of individual decision models under time pressure. IQ and IE, which are positively correlated, are significant predictors of expected value and predictive success. As expected, IE seems more relevant for all choices (including missed items), while IQ is more relevant for choices made. F-tests suggest that IE and IQ jointly determine the decision quality, with higher ability participants making better choices (EV) and more consistent choice across time pressure settings. Total time used under no time pressure predicts better decisions over choices made, but lower predictive success over all choices. We find no significant effect of gender. Self-Efficacy has a significant effect only for predictive success over all choices made.

Table 1.14: Multivariate Analysis of Performance Measures

Covariates	Expected value (choices made; €)	Expected value (choices made; €)	Expected value (all decision problems; €)	Expected value (all decision problems; €)	Predictive success (choices made; %)	Predictive success (all decision problems; %)
IE	0.0140	0.0178	0.0348	0.0333	0.0046	0.0062
	(0.0127)	(0.0128)	(0.0164)**	(0.0164)**	(0.0032)	(0.0033)*
IQ	0.0399	0.0321	0.0019	0.0048	0.0008	-0.0007
	(0.0169)**	(0.0175)*	(0.0228)	(0.0224)	(0.0050)	(0.0052)
F-test IE = IQ = 0	F=7.18***	F=5.73***	F=3.76**	F=3.60**	F=1.82	F=2.56*
Self-Efficacy	-0.0085	-0.0097	0.0036	0.0040	0.0027	0.0035
	(0.0074)	(0.0068)	(0.0070)	(0.0070)	(0.0017)	(0.0018)*
Female	-0.0414	0.0276	-0.0779	-0.1067	-0.0273	-0.0333
	(0.0696)	(0.0740)	(0.0863)	(0.0875)	(0.0199)	(0.0208)
Variance of time used per item when no time-constraint		-0.0001 (0.0006)		-0.0001 (0.0006)	0.0001 (0.0001)	0.0001 (0.0001)
Total time used when no time-constraint		0.0012 (0.0003)***		-0.0005 (0.0005)	-0.0001 (0.0001)	-0.0002 (0.0001)*
Controls	No	Yes	No	Yes	Yes	Yes
# obs	379	379	379	379	342	342
R^2	3.61%	11.05%	2.60%	3.10%	8.60%	10.25%

Notes: Results show coefficients from OLS regressions using robust standard errors, reported in parentheses. Controls include swiftness on the computer (Cappelen et al., 2015) and math score. All regressions control for treatment (Table 1.1); *,** and *** indicate significance at the 10%, 5% and 1% significance level, respectively.

Overall, we can explain only a small amount of the variance in expected value and in predictive success of the fitted CPT model. Although IE, IQ and time used strategy under unconstrained conditions are helpful in predicting decision, our results still emphasize the necessity of finding better instruments to predict the ability to perform under time pressure.

1.6 Discussion and conclusion

We set out to study the potential role of selection in adverse decision environments, and how it is linked to observable characteristics of the decision maker, including those characteristics that can be made observable using survey and experimental techniques. Clearly, different decision styles play an important role. Those who can and those who cannot easily cope with the time constraint in risky decisions differ along various dimensions. Those who violate time constraints, i.e. those who have lower time pressure resistance, make more careful (more variance, more time used), and consequently more successful decisions (higher EV, less affected by salient outcomes) when unconstrained. They also initially perform better under time pressure. However, as they run out of time, they cannot implement their strategy anymore, leading to considerable losses. Consequently, their performance and behavior are also much less correlated between the time pressure and the no-pressure conditions than it is the case for non-violators. A fitted decision model of their behavior in the absence of time pressure is less predictive of their decisions under time pressure. Violators try to make good decisions, sacrificing time, and violating the time constraint despite severe punishment (payment of maximum loss in the current design): they have a poor time management.

Our experiment aimed at making time-pressure relevant differences in decision style observable by considering measures of ability, personality and decision strategy. We find that various measures correlate with outcomes under time pressure and with the proneness to being attracted by salient features of prospects. Including these variables in a multivariate analysis, we identify IQ, IE and time-use in the absence of time pressure as moderate predictors of success under time pressure. Corroborating previous results by Kuhnen and Melzer (2014), self-efficacy is the only personality trait that has a weak systematic influence on decision-making. Overall, predictive power of observables is low, suggesting that we miss out important unobservable aspects of the decision strategies. This seems particularly important for practical applications, as identifying people based on tests for cognitive abilities and standard questionnaires seems to be of limited use when selecting agents that should perform well under time pressure.

We obtain results regarding the determinants of risky decision strategies. The finding that cognitive ability relates to higher realized expected payoffs is consistent with the extant literature. In a representative sample of the German population, Dohmen et al. (2010) find that subjects with higher cognitive skills are willing to take more risks. Similarly, Benjamin et al. (2013) report for Chilean high school students a significant correlation of risk aversion and cognitive capacities. With the average participant being risk averse, these directional effects are consistent with higher expected payoffs in our setting. However, other researchers have questioned the evidence on cognitive ability and risk taking. Andersson et al. (2014) provide evidence that these correlations may be spurious. They assert that, in fact, cognitive capacities are related to making errors and that the specific design of choice lists triggers the interpretation of differential risk attitudes. Our design does not involve choice lists, suggesting the cognitive ability effects are not merely driven by these design issues. However, our results suggest that the link between cognitive ability and risk behavior may be more moderate compared to the effects reported by Dohmen et al. (2010).

In conclusion, we find that selection is a very important factor in adverse decision environments. The relevance of selection effects has implications for the interpretation of the average laboratory behavior in terms of population parameters, and for the interpretation in terms of external validity. We try to predict who makes good decisions under time pressure. We find that cognitive ability measures, self-efficacy and observable time use measures correlate with performance under time pressure. However, more work is needed to make aspects of decision style and the use of heuristics predictable. If behavioral measures are shown to be of limited explanatory power, neurological markers may provide an interesting alternative (e.g., Buckert et al. (2014); Kandasamy et al. (2014)). Identifying people's ability to cope with time pressure is not a straightforward task. Tests for cognitive ability or efficiency and standard questionnaires need to be accompanied by additional measures.

Chapter 2

Unleashing Animal Spirits – Self-Control and Overpricing in Experimental Asset Markets*

2.1 Introduction

> "Even apart from the instability due to speculation, there is the instability due to the characteristic of human nature that a large proportion of our positive activities depend on spontaneous optimism rather than mathematical expectations, whether moral or hedonistic or economic. Most, probably, of our decisions to do something positive (...) can only be taken as the result of animal spirits – a spontaneous urge to action rather than inaction, and not as the outcome of a weighted average of quantitative benefits multiplied by quantitative probabilities."[1]
>
> *John Maynard Keynes*

Keynes famously saw "animal spirits" at the root of many (financial) decisions, potentially causing price exaggerations on the aggregate market level. As often in Keynes' work, the term "animal spirits" is not well-delineated. It alludes to optimism, instincts, urges, emotions, and similar concepts. In this chapter we assess the notion that a *lack of self-control abilities* may lead to price exaggerations on asset markets, and we analyze how the lack of self-control abilities is associated to emotions and trading behavior. In psychology, self-control abilities and willpower are defined as the capacities to override or inhibit undesired behavioral tendencies such as impulses and

*This chapter is based on joint work with Martin Kocher and Konstantin Lucks.

[1]Source: Keynes (1936), p. 136.

to refrain from acting on them (Tangney et al., 2004). Self-control is necessary to guard oneself against undue optimism, actions motivated by emotional responses, and impulsive decisions. Furthermore, self-control is required in order to stick to plans made in the past.

That self-control is considered relevant for investor success is also evident from statements of investors and from popular guidebooks on the psychology of investing. For instance, Warren Buffet emphasizes that "success in investing doesn't correlate with I.Q. once you're above the level of 25. Once you have ordinary intelligence, what you need is the temperament to control the urges that get other people into trouble in investing."[2] Similarly, anecdotal evidence from rogue traders show that they completely lost their self-control abilities at some stage. In a study by Lo et al. (2005) involving day traders from an online training program participants stated attributes related to self-control as the most important determinants of trading success.[3] In a similar spirit, Fenton-O'Creevy et al. (2011) report distinct differences in emotion regulation strategies among traders of different experience and performance levels from qualitative interviews with professional traders. Therefore, correlational evidence suggests that self-control matters for trading success on an individual level.

This contribution is the first to provide empirical evidence on the *causal* effect of a variation in self-control abilities on trading outcomes.[4] The major challenge to overcome is to exogenously vary self-control abilities in order to obtain causal inference on the impact of self-control abilities on behavior and market outcomes. A first step is to use the experimental laboratory and affect *state* self-control levels of traders. Most of the available techniques draw on the concept of self-control depletion or exhaustion. Our experimental identification rests on the assumption that self-control is a limited resource and that it is variable over time on the individual level. Evidence for these two characteristics is abundant (e.g. Baumeister et al., 1998; Gailliot et al., 2012), although it has also been questioned lately (Carter and McCullough (2013)). While validated survey measures for *trait* self-control exist, they can only provide correlational inference.

In the spirit of Keynes we concentrate on aggregate market outcomes in a first experiment and extend our analysis to individual behavior and performance in a second experiment. We use a well-established financial market setup in the experimental laboratory (Smith et al., 1988; Kirchler et al., 2012; Noussair and Tucker, 2013; Palan, 2013; Eckel and Füllbrunn, 2015) to investigate whether an exogenous variation in

[2] http://www.businessweek.com/1999/99_27/b3636006.htm

[3] They quote attributes such as persistence, tenacity, perseverance, patience, discipline, planning, controlling emotions, and (lack of) impulsivity as crucial (Lo et al., 2005, table 3).

[4] However, there is a quickly growing empirical literature on the effects of self-control abilities on decision making in other domains relevant to economists (see, for instance, Beshears et al. (2015).)

self-control abilities of traders leads to overpricing. This experimental asset market is known for its basic tendency to exhibit overpricing; it features a dividend-bearing asset with decreasing fundamental value.

In order to deplete self-control abilities before the start of the market, we employ the Stroop task (Stroop, 1935), which is one of the most commonly used tasks in psychology experiments for modulating self-control (Hagger et al., 2010). It is easy to administer, it can be implemented in an exhausting/depleting version and in an easy version (i.e. a placebo version), and it allows for additional controls. The majority of studies that use both survey measures and behavioral measures of self-control conclude that the effects of state self-control interventions are qualitatively similar to those of trait self-control levels (e.g. Schmeichel and Zell, 2007). Hence, even if our experiment is confined to the laboratory setting and to a variation in state self-control, it is likely that it extends to situations outside the laboratory in which also trait self-control matters.

A drop in self-control abilities can increase the extent of overpricing on a market through different channels. One psychological transmission mechanism runs through an increased influence of the impulsive decision making system. A consequence could be that traders' behaviors become more easily swayed by observing others' behaviors on the market (for instance, a more pronounced tendency to momentum trading). Another behavioral mechanism relates to an heightened influence of emotions (for instance, the excitement after seeing the prospect of making more money, or a stronger psychological reward of interim gains). Yet another option, potentially related to impulsivity, would be a stronger role of biases in decision making such as myopia, limiting the ability to correctly foresee the declining fundamental value and thus creating histories of overpricing on the market.

Our main finding is a significantly higher level of overpricing in markets where traders' self-control abilities have been depleted, compared to markets with traders whose self-control abilities have not been depleted. If markets are populated by both depleted and non-depleted traders the effect is similar in size and also highly significant. Obviously, having some self-control depleted traders on a market suffices to create the additional over-pricing effect.

Behavior in markets is path-dependent, choices are endogenous to other choices, and traders imitate each other. Nonetheless, we are able to provide robust evidence from control variables, from trading and from survey questions that can explain the additional overpricing with depleted self-control abilities. First, there is no direct effect of self-control depletion on risk attitudes or cognitive abilities of traders, which could explain our findings. Second, self-control depleted traders do not trade significantly less than non-depleted traders, ruling out a simple exhaustion effect. Third, several indicators show that self-control depleted traders trade more aggressively early in the market. In other words, they contribute more to the creation of overpricing histories, and non-depleted traders jump on this bandwaggon. Fourth, stronger emotional arousal of individuals on the market is related to being self-control depleted. In short, traders become more impulsive and potentially rely less on cognitive skills, when they cannot resort to their full self-control resources.

The remaining chapter is organized as follows: Section 2.2 gives an overview of the related literature, and in section 2.3, we explain and motivate our experimental design. Consequently, section 2.4 presents the results from our main experiment, and section 2.5 reports on an additional experiment that allows us both to test the robustness of our initial results and to better understand how self-control depletion translates into overpricing and how traders' behavior and decision processes might be affected by the treatment. We discuss potential channels explaining our findings in section 2.6. Section 2.7 concludes the chapter.

2.2 Related Literature

Our literature overview focuses on the two aspects in the economics and psychology literature that are most relevant for our study: self-control and experimental asset markets. As already said, self-control abilities and willpower are defined as the capacities to override or inhibit undesired behavioral tendencies such as impulses and to refrain from acting on them. There are different theoretical approaches in psychology and in economics that take self-control abilities and potential self-control problems into account.

First, self-control can straightforwardly be related to dual-systems perspectives of decision making. As outlined by Kahneman (2011), these perspectives share the general assumption that structurally different systems of information processing underlie the production of impulsive, largely automatic forms of behavior, on the one hand (system 1), and deliberate, largely controlled forms of behavior, on the other hand (system 2). System 2 is effortful and requires self-control resources.[5] Thus, if resources are low, reflective operations may be impaired, leading to a dominance of impulsive reactions that could be in conflict with objective reasoning. From this perspective, reducing self-control abilities can be interpreted as increasing the role of the (impulsive) system 1 in decision making (Hofmann et al., 2009).

[5]Note that the division of system 1 as automatic and system 2 as controlled describes a tendency; there are both automatic and conscious processes involved in exerting self-control and giving in to temptation, respectively (cf. Kotabe and Hofmann, 2015).

Second, and very much related to dual-system perspectives, economists have used dual-self models of impulse control (see, for instance, Thaler and Shefrin (1981) and Fudenberg and Levine (2006)) in order to describe self-control problems. These models study the interaction of two selves, a rational (long-term) and an impulsive (short-term) self. Such models can account for time inconsistent behavior (for instance, in connection with quasi-hyperbolic discounting) and for the fact that cognitive load makes temptations harder to resist. Third, willpower as a depletable resource has been modeled directly in economics. Ozdenoren et al. (2012) look at a consumption smoothing model that views willpower as a depletable resource, and Masatlioglu et al. (2011) consider lottery choices.

Is there empirical evidence for self-control abilities or willpower to be indeed limited or depletable resources? Many researchers in psychology have shown that exerting self-control consumes energy and consequently diminishes the available resources for other acts that require self-control.[6] Self-control can involve either cognitive control, or affective control, or both (Hagger et al., 2010). Self-control abilities regenerate through rest, can be trained, and differ between people (Baumeister et al., 1998; Muraven et al., 1999; Muraven and Baumeister, 2000; Tangney et al., 2004; Muraven, 2010).

Our experimental identification relies on self-control depletion. We reduce self-control abilities by exposing experimental participants to a self-control demanding task before the main task (known as the dual task paradigm). Such setups have been used in other domains in economics, mainly in the context of individual decision making. For example, the consequences of self-control variations in decision making under risk have been studied. Several papers report increased risk aversion following self-control depletion (Unger and Stahlberg, 2011; Kostek and Ashrafioun, 2014). However, a number of studies also reveal an increase in risk taking following similar manipulations (Bruyneel et al., 2009; Freeman and Muraven, 2010; Friehe and Schildberg-Hörisch, 2014). Both Stojić et al. (2013) and Gerhardt et al. (2015) find no significant effect of self-control manipulations on risk preferences elicited from choice lists. Bucciol et al. (2011, 2013) show in field experiments with children and adults that self-control depletion leads to reduced productivity in subsequent tasks. De Haan and Van Veldhuizen (2015) find no effect of a repeated Stroop task on the performance in an array of tasks in which framing effects – such as anchoring effects and the attraction effect – are typically observed.

Recently, experiments have looked at the effects of self-control variations on other-regarding preferences. Achtziger et al. (2016) report a strong but heterogeneous impact of reduced self-control on offers and accepting behavior in ultimatum games, presum-

[6] For recent overviews about the ongoing discussion in psychology and models of the underlying processes involved in self-control see Inzlicht and Schmeichel (2012) and Kotabe and Hofmann (2015).

ably depending on what an individual's more automatic reactions are. In a similar vein, Achtziger et al. (2015) provide evidence for reduced dictator giving after a reduction in self-control abilities.[7]

Existing studies also suggest a relationship between self-control abilities and financial decision making. However, we are not aware of experimental studies in this context. Using survey evidence, Ameriks et al. (2003, 2007) consider the connection between wealth accumulation and trait self-control in a sample of highly educated US households. Ameriks et al. (2003) attribute differences in savings among households to differing "propensities to plan" – i.e. different individual costs of exerting self-control. Ameriks et al. (2007) use the difference between planned behavior and expected behavior in a hypothetical scenario as a measure for self-control problems. They find a positive correlation between better self-control abilities and wealth accumulation, in particular for liquid assets. Gathergood (2012) conducts a similar study in the UK with a representative sample. He reports a positive association between lower levels of self-control and consumer over-indebtedness.

Our asset market is based on the seminal paper by Smith et al. (1988), who were the first to observe significant overpricing in an experimental double auction market. Many studies have followed up on these early findings.[8] Trader inexperience and confusion have been considered as one of the aggravating factors of overpricing (Dufwenberg et al., 2005; Kirchler et al., 2012), and Bosch-Rosa et al. (2015) for example show that grouping traders by cognitive skills leads to increased overpricing for groups with low cognitive sophistication. Nadler et al. (2015) provide evidence that giving testosterone to a group of male participants significantly increases prices, and Petersen et al. (2015) find that inducing stress decreases overpricing.

Since emotion regulation is correlated with self-control abilities (Tice and Bratslavsky, 2000), the influence of emotions on prices in asset markets is also relevant to our research question: Andrade et al. (2016) find that inducing excitement before trading triggers overpricing in asset markets stronger in magnitude and higher in amplitude than other emotions and a neutral condition. In a similar study, Lahav and Meer (2012) show that inducing positive mood leads to higher deviations from fundamental values and thus more overpricing. The role of emotions in experimental asset markets has also been evaluated using self-reported emotions on Likert scales (Hargreaves Heap and Zizzo, 2011) and face reading software (Breaban and Noussair, 2013), instead of inducing specific emotions exogenously. Results from these experiments indicate that

[7] Martinsson et al. (2014) analyze the relationship between self-control and pro-sociality in an indirect way, but their findings are also in line with the idea that pro-social behavior requires self-control. A similar result is provided by Kocher et al. (2016).

[8] Recent surveys can be found in Noussair and Tucker (2013) and Palan (2013).

Figure 2.1: Treatment Differences in the Stroop Task

excitement and a positive emotional state before market opening are correlated with increased prices relative to fundamental values. Moreover, fear at the opening of the market is correlated with lower price levels.

2.3 Experimental Design

This chapter reports the results from two experiments. The design of Experiment I is described in this section. Experiment II is a natural extension of Experiment I and described in greater detail in section 2.5. Experiment I consists of four independent parts: (i) instructions and dry runs of the asset market without monetary consequences and without the possibility to build reputation for the parts to come; (ii) the main treatment variation in self control, the Stroop task (Stroop, 1935) in two treatment versions; (iii) elicitation of risk attitudes and cognitive abilities, both incentivized; and (iv) a fully incentivized experimental asset market.

Our identification of the effects induced by a variation in self-control abilities in market prices relies on the comparison of behavior in markets following two different versions of the Stroop task. A tough version lowers self-control abilities, whereas a placebo version should leave self-control abilities largely unaffected. We implement a condition in which all market participants are subjected to the tough version of the Stroop task (henceforth *LOWSC* for low self-control) and a condition in which all participants were subjected to the placebo version (henceforth *HIGHSC* for high self-control). Except for this treatment variation in part (ii), the two experimental conditions are identical in all other parts.

The Stroop task follows a simple protocol: participants are instructed to solve correctly as many problems as possible within five minutes. An example of such a problem is displayed on the left-hand side of Figure 2.1. The task is to select the color of the font the word is printed in. A selection of six color buttons – always the same and in the same order – is given on the bottom right of the screen, and subjects are instructed to click on the correct one. As soon as they make a selection, the next word-color combination appears. Consecutive word-color combinations always differ from each other. The difficulty of this task is that the words always describe one of the six colors; the incongruence between the color of the word and the word itself causes a cognitive conflict, since reading the word is the dominant cue. Common explanations for the conflict are automaticity of reading the word or relatively faster processing of reading than color perception (MacLeod, 1991). The conflict has to be resolved, and resolution requires self-control effort. Applying this effort depletes self-control resources and leaves participants with lower levels of willpower and/or self-control resources after the five minutes.

The Stroop task is one of the most commonly applied methods to deplete self-control resources (Hagger et al., 2010). It can be easily implemented in a computer laboratory, is straightforward to explain, requires only basic literacy skills, and generates additional data on the number of correctly solved problems and the number of mistakes. The difference between the Stroop task in *LOWSC* and *HIGHSC* is the frequency with which a conflicting word-color combination occurrs.[9] All screens in *LOWSC* exhibit such a conflict, while in *HIGHSC* only every 70th screen does. Experimental participants do not receive any information on the frequency of such a conflict, and the instructions for the two versions of the task are identical. By having an occasional word-color incongruence in *HIGHSC* we are able to ensure that subjects take the task seriously. If anything, our setup reduces the potential treatment difference, because in *HIGHSC* some self-control depletion might still take place, making the potential result of a significant difference between the two conditions more difficult to obtain.

We decided to provide participants with a flat payment of € 3 for the Stroop task in order to signal that we were interested in their performance. We do not use a piece-rate or any other competitive payment scheme because it might create different wealth levels after the treatment variation, and wealth differences might be correlated with the treatment. Hence, treatment differences might potentially be confounded with wealth effects.[10] Upon completion of the five minutes, we ask experimental participants how hard they perceived the task on a six-point Likert scale.

[9]The right-hand side of Figure 2.1 shows an example of congruence between font color and word, as we use it in the placebo Stroop task in *HIGHSC*.

[10]Achtziger et al. (2015) find no differences in depletion effects between flat payments and incentivized versions of a related self-control manipulation. We are confident that subjects took the task

Self-control resource depletion can influence several relevant variables for the subsequent experimental asset market. We control for two mechanisms directly: cognitive ability and risk attitudes.[11] Eliciting control variables takes place after the self-control manipulation but before the experimental asset market for two reasons: Firstly, if these measures were to follow the asset market, there might be spillover effects due to experiences during the asset market and secondly the effect of our self-control manipulation might wear off since the asset market part of the experiment lasts a considerable amount of time during which self-control could start to regenerate (Muraven and Baumeister, 2000). In order to avoid that the self-control variation wears off before the asset market interaction starts, it is a requirement that measuring the control variables does not take much time. Two tasks that fit this requirement are the Cognitive Reflection Test (CRT) for measuring individual cognitive abilities (Frederick, 2005) and a simple multiple price list lottery design for eliciting individual risk attitudes (Dohmen et al., 2011).

First, our subjects answer the three questions of the standard CRT. It is well-known that CRT responses are correlated with more time-consuming measures of cognitive ability, risk and time preferences (Frederick, 2005), as well as with decisions in a wide array of experimental tasks such as entries in p-beauty-contest games (Brañas-Garza et al., 2012) and performance in heuristics-and-biases tasks (Toplak et al., 2011). Furthermore, Corgnet et al. (2014) and Noussair et al. (2014) find that the CRT is a good predictor of individual trader's profits in asset market experiments.[12] Subjects are paid € 0.5 for every correct answer but do not learn their CRT results and thus earnings until the end of the experiment.

Second, we elicit individual certainty equivalents (CE) for a lottery using a multiple price list as a measure for individual risk attitudes. Differences in risk attitudes can be a rational reason for trade (Smith et al., 1988) and might explain initial underpricing of assets on the market, thus sparking off later price increases and overpricing (Porter and Smith, 1995; Miller, 2002). Furthermore, Fellner and Maciejovsky (2007) find that more risk averse individuals trade less frequently. On a single computer screen, our experimental participants have to choose ten times between a lottery that pays either € .20 or € 4.20 with equal probability and increasing certain amounts of money that are equally spaced between the two outcomes of the lottery. Subjects are allowed to switch

seriously; only two participants in Experiment I tried less than 114 screens and one answered less than 110 items correctly. Most of our subjects answered many more – see appendix B.3 for details.

[11] For evidence of potential effects of self-control depletion on complex thinking see Schmeichel et al. (2003). As mentioned in the previous section, evidence on the relationship between self-control abilities and risk attitudes is rather inconclusive. Emotions as a potential transmission mechanism will be assessed in Experiment II.

[12] The CRT is regarded as a measure of cognitive ability and thinking disposition (Toplak et al., 2011). We will discuss the CRT results and their implications in more detail when we discuss our results in section 2.6.

at most once from the lottery to the certain amounts. At the end of the experiment, the computer randomly picks one of the ten decisions of each individual as payoff-relevant and implements the preferred option, potentially simulating the lottery outcome.

Immediately after risk elicitation the main part of the experiment, the asset market, opens. The asset market featurs a dividend-bearing asset with decreasing fundamental value over ten trading periods (lasting 120 seconds each) in a continuous double-auction market design with ten traders and with open order books, following Kirchler et al. (2012).[13] This is a simplified version of the markets in Smith et al. (1988). Before the first trading period, five subjects in a given market receive 1000 experimental points in cash and 60 assets, and the other five receive 3000 points in cash and 20 assets as their initial endowment. Assignment to the two initial asset allocations is random.

During each trading period, traders can post bids and asks as well as accept open bids and asks. Partially executed bids and asks continue to be listed with their residual quantities and inactive orders remain in the books until the end of the current period. At the end of every period, the asset pays a dividend of either ten or zero experimental points with equal probability. The dividend payment is added to each trader's cash holdings. Assets have no remaining value after the last dividend payment, i.e. they display a declining (expected) fundamental value. This design feature is explicitly stated and highlighted in the instructions. To make things clear, the instructions provide a detailed table with the sum of remaining expected dividend payments per unit of the asset at any point in time. Assets and cash are carried from period to period. Short selling and borrowing experimental points are not allowed. After every period, the average trading price as well as the realizations of the current and all past dividends are displayed on a separate feedback screen. At the end of the ten periods, experimental points are converted into euros, using an initially announced exchange rate of 500 points = € 1 .

At the end of the experiment, subjects learn about their payoffs from all parts of the experiment. We ask them to fill in a short questionnaire concerning demographics and background data. We also ask participants how tired they feel after the experiment and how hard they have perceived decisions over the course of the entire experiment on a 6-point Likert scale. Then, all earnings are paid out in private, and the subjects are dismissed from the laboratory.

Experiment I was conducted in October 2013. 160 participants took part in ten experimental sessions – four with one markets and six with two markets. Hence, we obtained 16 independent observations, eight for each of our treatment conditions. The

[13] Appendix B.7 provides the experimental instructions, including a screen shot and a description of the trading screen.

experiment was programmed using z-Tree (Fischbacher, 2007), and recruitment was done with the help of ORSEE (Greiner, 2015). Experimental sessions lasted for about 90 minutes, and participants earned € 18.18, on average. We only invited students who had never participated in an asset market experiment before. We also excluded students potentially familiar with the CRT or the Stroop task.[14] Prior to the start of the experiment, subjects received written instructions for all parts of the experiment. These were read aloud to ensure common knowledge. Remaining questions were answered in private.

2.4 Experimental Results

2.4.1 Manipulation Check

The data suggest that our treatment manipulation was successful: First of all, during the Stroop task participants attempted fewer problems, achieved fewer correctly solved problems and made more mistakes in the *LOWSC* condition than in the *HIGHSC* condition (all Mann-Whitney tests $p < 0.01$).[15] Participants perceived the Stroop task as significantly more demanding in the *LOWSC* condition than in the *HIGHSC* condition (Mann-Whitney test $p < 0.01$). Finally, we do not find any differences in background characteristics such as field ($p = 0.416$) and year of study ($p = 0.9162$), age ($p = 0.1709$) and gender ($p = 0.9558$) between our two treatments (Mann-Whitney tests and Pearson's χ^2 test for field of study), suggesting that random assignment to treatments was successful.

2.4.2 Definitions and Measures

In order to calculate mean prices one can use either an adjustment that takes trading volumes into account (henceforth: volume-adjusted prices) or an adjustment that takes the number of trades into account (henceforth: trade-adjusted prices). The former is an average price per asset, whereas the latter is an average price per trade. Our results remain unaffected by the choice of adjustment; in line with the literature, we mainly display results based on volume-adjusted prices in the following.

[14] Of our 160 subjects, one suffered from some form of dyschromatopsia, i.e. a color vision impairment. We asked for it in the post-experimental questionnaire in order to make sure that it is not a common phenomenon.

[15] Detailed distributions on these variables can be found in section B.3 of the appendix. All tests reported in this chapter are two-sided unless stated otherwise.

In order to quantify the tendency of markets to exhibit irrational exuberance we compare trading prices with the fundamental value of the asset. In the following we adopt the approach of Stöckl et al. (2010) and assess the market price developments using *Relative Absolute Deviation* (RAD) (in equation 2.1) and *Relative Deviation* (RD) (in equation 2.2) as measures for general mispricing and overpricing, respectively.

$$\text{RAD} = \frac{1}{T}\sum_{t=1}^{T}\frac{|P_t - FV_t|}{\bar{FV}} \tag{2.1}$$

$$\text{RD} = \frac{1}{T}\sum_{t=1}^{T}\frac{P_t - FV_t}{\bar{FV}} \tag{2.2}$$

P_t is the volume-adjusted mean price in period t, FV_t is the fundamental value of the asset in period t, and $\bar{FV}$ denotes the average fundamental value of the asset over all periods.

RAD is constructed as the ratio of the average absolute difference of mean market price and fundamental value, relative to the average fundamental value of the asset. RD is the ratio of the average difference between mean market price and fundamental value, relative to the average fundamental value. The difference between the two measures is how the difference between mean market price and fundamental value enters the calculation: For RAD the difference enters in absolute terms, thus all deviations from the fundamental value – overpricing and underpricing – increase RAD, making RAD a measure of average mispricing. For RD the wedge between market price and fundamental value retains its sign, thus periods with overpricing and underpricing can cancel each other out. Hence, RD provides the dominant direction of mispricing, making it, in effect, a measure of average overpricing.

Both measures are straightforward to interpret: A RAD of .1 means that prices are on average 10% *off* the fundamental value, while a RD of .1 indicates that prices are on average 10% *above* the fundamental value. Both measures are independent of the number of periods and the fundamental value.

2.4.3 Aggregate Price Development

Figure 2.2 shows how average market prices in *LOWSC* and *HIGHSC* evolve over the ten trading periods. In both conditions, average market prices start out at a similar level, displaying a moderate level of underpricing. However, from the third period onwards, average prices in both conditions exceed the fundamental value. Eventually,

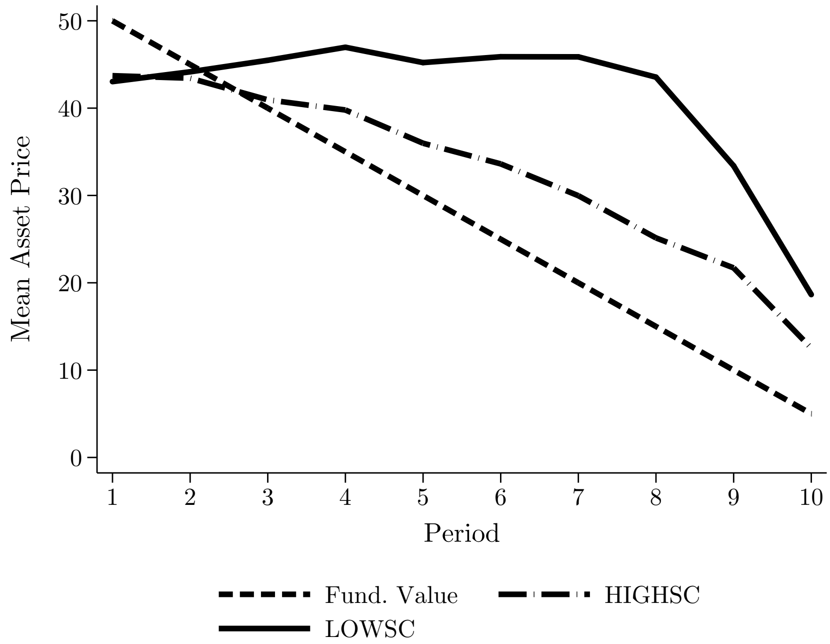


Figure 2.2: Mean (Volume-adjusted) Trading Prices in the Two Treatments

average market prices drop sharply, but do not drop below the fundamental value again.

The most conservative comparisons between the two treatments are based on market averages over all traders and over all ten periods. This is the approach we apply for all non-parametric tests regarding aggregate market outcomes. These averages are statistically independent in the strict sense, and test statistics are based on eight observations for each treatment. A Wilcoxon signed-ranks test confirms the impression from eyeballing, i.e. that market prices in both conditions are significantly different from the fundamental value (*HIGHSC*: $p = 0.0929$, *LOWSC*: $p = 0.0173$). Figure 2.2 suggests more pronounced overpricing in the *LOWSC* condition than in *HIGHSC*, which is confirmed by a Mann-Whitney test (*HIGHSC*: $\bar{RD} = 0.1885$, *LOWSC*: $\bar{RD} = 0.4990$; $p = 0.0742$)[16]. A comparison of RD tells us that while in *HIGHSC* overpricing is on average 19%, in *LOWSC* prices exceed the fundamental value by almost 50%. Thus, trade among individuals with low self-control leads to overpricing which is more than twice as high as in the baseline *HIGHSC*.

Furthermore markets in the *LOWSC* condition exhibit higher levels of mispricing (*HIGHSC*: $R\bar{A}D = 0.3253$, *LOWSC*: $R\bar{A}D = 0.5890$; Mann-Whitney test: $p = 0.0460$). According to RAD, prices in the *HIGHSC* condition deviate by about 33% from the fundamental value, whereas they deviate by about 59% from the fundamental value in the *LOWSC* condition.

[16]Both measures are significantly different from zero for both conditions.

Figure 2.3 displays the price evolution of single markets in the two conditions. There is a high degree of path-dependence and endogeneity in price evolution in the markets and a lot of heterogeneity among markets in the same condition. Therefore, finding a significant difference between the two conditions for the most conservative test in terms of statistical independence is the more striking. The left panel represents the markets from the *HIGHSC* condition, while the right panel shows the *LOWSC* markets. Price paths in *HIGHSC* markets often follow a rather flat or declining development, while in *LOWSC* a number of markets display a hump-shaped price evolution that initially increases and peaks in later trading periods. The emergence of overpricing oftentimes can be attributed to constant prices despite decreasing fundamental values (Huber and Kirchler, 2012; Kirchler et al., 2012) – a description that fits price paths in our *HIGHSC* markets better than those in *LOWSC* markets.[17]

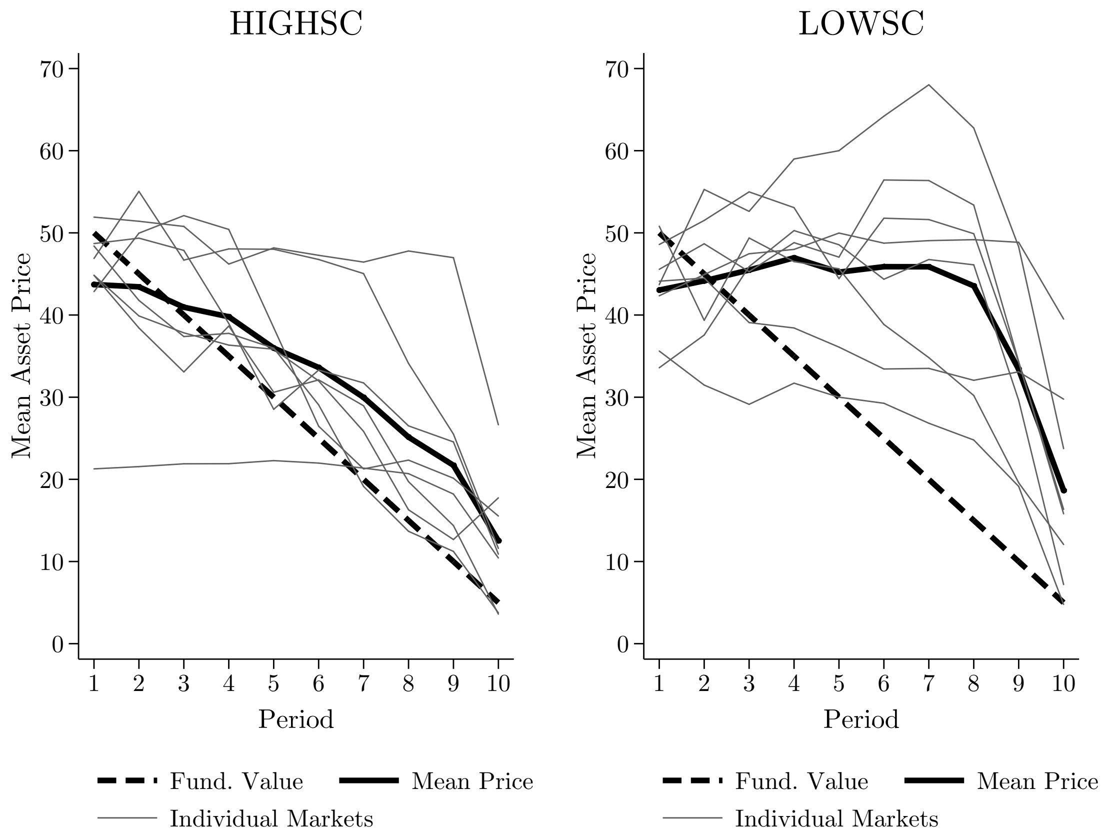


Figure 2.3: Evolution of Individual Market Prices in *HIGHSC* and *LOWSC*

[17]Section B.1 in the appendix shows a comparison of overpricing measures across treatments for each period separately. Overpricing in *LOWSC* significantly exceeds overpricing in *HIGHSC* in periods 6-9.

2.4.4 Potential Transmission Mechanisms of the Treatment Effect

Having established a significant treatment effect, the next step is to look at potential channels via which self-control variations could have had an effect on market outcomes. Detailed descriptive results on the variables considered in this section can be found in sections B.4ff. of the appendix.

Cognitive Abilities and Risk Attitude

Self-control depleted participants might not be willing to think as hard and thus provide the (wrong) intuitive answers in the CRT. The average number of correct answers in the CRT was 1.05 in *HIGHSC* and 1.14 in *LOWSC*. The difference in CRT score between the two conditions is not significant according to a Mann-Whitney test ($p = 0.7223$). We conclude that the Stroop task did not have an impact on our incentivized version of the CRT.[18] Risk attitudes might be affected by self-control depletion. The average certainty equivalent we elicited is close to the lottery's expected value: 2.2 in *HIGHSC* and 2.15 in *LOWSC*. Like the literature exploring the effect of reduced self-control on risk attitude that has come to inconclusive results (e.g. Bruyneel et al., 2009; Unger and Stahlberg, 2011; Gerhardt et al., 2015), we also find no significant effect (Mann-Whitney test, $p = 0.4083$) of our treatment variation on risk attitudes as measured by the multiple price list certainty equivalent elicitation.[19]

Trading Activity

An additional channel through which our results could be explained is changes in trading activity, i.e. the number of traded shares per trading period. People low in self-control have been reported to become more passive (Baumeister et al., 1998, Experiment 4). Increased passivity and thus a thinner market in *LOWSC*, where few trades could drive overpricing, could be responsible for our results. Thus we compare the number of shares traded in the two conditions. Figure 2.4 illustrates the evolution of average shares traded per period. Traders in *HIGHSC* traded slightly more overall: while the average trader traded 13.02 shares per period in *HIGHSC*, only 11.39 shares changed hands on average per trader in each period in *LOWSC*. However, according to a Mann-Whitney test, there is no significant difference between amounts traded between the two conditions ($p = 0.3446$).[20] When analyzing the results of Experiment

[18] If we include the observations from our second experiment, the CRT scores of the two groups become 1.0875 and 1.1375 respectively with $p = 0.7442$ from a Mann-Whitney test.

[19] Including observations from Experiment II does not provide significant differences between the two groups.

[20] An additional regression analysis in Table B.2 in appendix B.2 reinforces this conclusion.

II, we shall take a closer look at trading strategies of self-control depleted traders versus non-depleted traders.

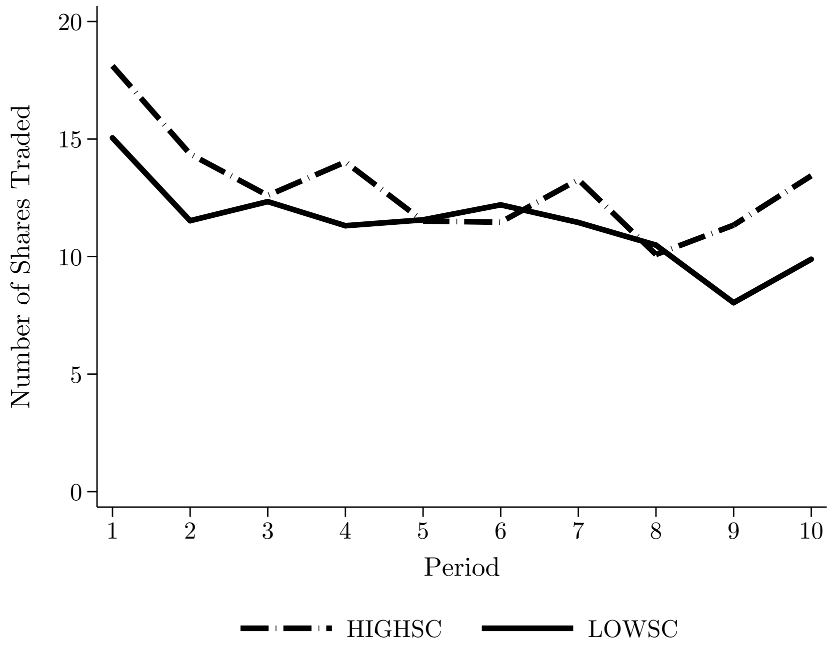


Figure 2.4: Evolution of Average Shares Traded per Trader by Condition

Regressions Controlling for Potential Channels

Although our control variables seem unaffected by our treatment, they could still possess explanatory power for the difference in overpricing that we observe. We therefore run regressions, including controls as indepedent variables. To avoid endogeneity problems across trading periods and between subjects, respectively, we aggregate overpricing measures over all ten periods on the individual level and use robust standard errors clustered at the market level. We do this separately for sales and purchases, since selling above fundamental value results in an expected profit, while buying above fundamental value results in an expected loss. We define measures for individual overpricing for purchases and sales, which we call $IndRD_{purchases}$ and $IndRD_{sales}$, respectively. Similar to the measure RD they are defined as the percentage of buying (selling) prices exceeding the asset's fundamental value pooled over all periods, but for each subject's buying (selling) activity separately instead of on the market level as before. We report results on $IndRD_{purchases}$ as the dependent variable in the regressions in Table 2.1 In appendix B.2, we provide robustness checks for our chosen approach for sales and both aggregated sales and purchases.

In all four models we are interested in the effect of the explanatory variables on $IndRD_{purchases}$, our measure of an individual's overpricing tendency. Throughout all

Table 2.1: Determinants of Individual RD Based on Purchases

	(1)	(2)	(3)	(4)
		$IndRD_{purchases}$		
LOWSC	0.400**	0.390**	0.816***	0.843***
	(0.140)	(0.134)	(0.131)	(0.125)
CRT		-0.0708*	-0.0952	-0.0912
		(0.0392)	(0.0558)	(0.0547)
CE		-0.0188	0.0684	0.0719
		(0.0459)	(0.0441)	(0.0455)
CRT × *LOWSC*			0.0612	0.0628
			(0.0821)	(0.0831)
CE × *LOWSC*			-0.224***	-0.237***
			(0.0712)	(0.0709)
Female				0.0666
				(0.0690)
Constant	0.0933	0.194	0.0255	-0.0353
	(0.0971)	(0.120)	(0.0597)	(0.0682)
Observations	110	110	110	110
R^2	0.275	0.307	0.364	0.370

OLS regression, dependent variable is Individual Relative Deviation (IndRD) for purchases, an individual equivalent to market level Relative Deviation (RD) restricted to purchases only. *LOWSC* is a dummy where 1 stands for *LOWSC* and 0 for *HIGHSC*. CE is an individual's certainty equivalent. CRT denotes the number of correct answers on the CRT. Subjects who indicated they knew one or more of the CRT questions before were excluded. Heteroskedasticity robust standard errors clustered at market level in parentheses, *** p<0.01, ** p<0.05, * p<0.1

specifications, we observe a significant treatment effect: Being in *LOWSC* increases an individual's propensity to buy at excessive prices. In specification (2), our measure of risk attitude is not significant, but if we also include interactions with our treatments in specifications (3) and (4), relative risk seeking is correlated with lower individual overpricing when self-control capabilities are reduced. Performance on the CRT has the expected effect of reducing the tendency of buying at prices above fundamental value in all specifications where it is included, and its effect does not significantly differ between participants in *LOWSC* and *HIGHSC* markets.[21] Hence, introducing measures for risk aversion and cognitive skills and their interactions with our treatments do not reduce the size or significance of the treatment coefficient. We conclude that neither changes in cognitive skills nor in risk preferences after self-control depletion can explain our main result of excess overpricing after self-control depletion.

2.5 Experiment II: Mixed Markets

2.5.1 Motivation and Design

The results reported in section 2.4 referred to markets, in which either all market participants underwent the tough Stroop task or none of them, i.e. either everyone's self-control resources had been reduced or no one's. In this section we report results from markets, in which only half of the participants' self-control resources were depleted. Each market consisted of five participants randomly assigned to the easy (placebo) Stroop version from the *HIGHSC* condition and five participants randomly assigned to the tough Stroop version from the *LOWSC* condition. We call this new condition *MIXED* and for simplicity refer to traders facing the tough version of the Stroop task as *MIXLO* traders and to those facing the easy version of the Stroop task as *MIXHI* traders. The motivation for this additional experiment is twofold. First, asset market experiments are zero sum games and behavior is highly path-dependent and endogenous to market prices, which makes it technically impossible to analyze differences in behavior resulting from reduced self-control in our homogeneous markets. Therefore, we wanted a condition in which traders under both conditions are active at the same time. It allows us to assess differences in trading behavior and performance between *MIXLO* traders and *MIXHI* traders. Second, since in real-world settings – either due to dispositional differences or due to differential previous demands on self-control resources – it is likely that individuals high and low in self-control interact, we want

[21] Note that we exclude subjects who were familiar with the CRT from these analyses, since such knowledge might have inflated correct CRT responses and thus obfuscate any effects of CRT scores. The regression results are qualitatively very similar when including these subjects.

so see whether the effect of reduced self-control observed in *LOWSC* markets can be replicated with a smaller share of depleted traders in *MIXED* markets.

We conducted eight additional sessions with 16 markets in April 2014 and November 2015. In the last four sessions we added several questions to the experimental questionnaires dealing with participants' emotions. We were interested whether our variation of self-control had taken effect via changes in emotional states. In order to reduce experimenter demand effects and as is common in experiments analyzing emotions, we confronted subjects with several emotions of which some were not relevant at all to our question of interest. Apart from the assignment to the respective version of the Stroop task within a market and the additional questions in the questionnaires of the last four sessions, the experimental protocol remained exactly the same as in Experiment I. Experimental participants were not aware of the different versions of the Stroop task, i.e. they were unaware of the fact that half of the traders performed the tough version and half of the traders the easy version.

2.5.2 Aggregate Price Evolution

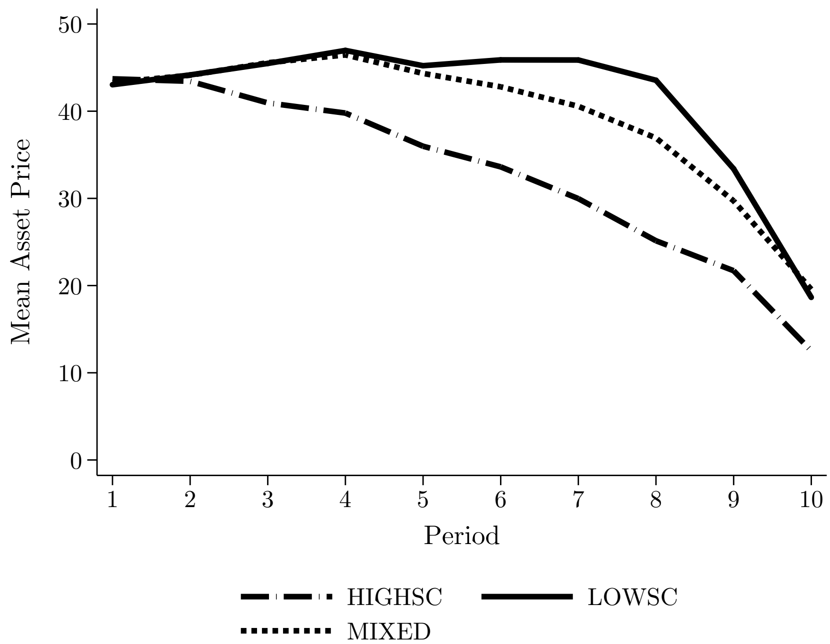


Figure 2.5: Trading Price Evolution Including *MIXED*

Figure 2.5 shows the evolution of average trading prices in all three treatments of Experiment I and II. Interestingly, the effect of reduced self-control on mispricing and overpricing does not seem to be changed if only part of the trader population is self-control depleted. Both *LOWSC* and *MIXED* on average display more overpricing than *HIGHSC*. For *MIXED* we observe an average RAD of 0.551 and an average RD of 0.430. A Mann-Whitney test confirms that the mispricing measure RAD in *MIXED* is significantly different from *HIGHSC* ($p = 0.0500$) but cannot be statistically distinguished from *LOWSC* ($p = 0.8065$). This result also holds for our overpricing measure: RD in *MIXED* differs significantly from *HIGHSC* ($p = 0.0864$), but not from *LOWSC* ($p = 0.5006$).[22]

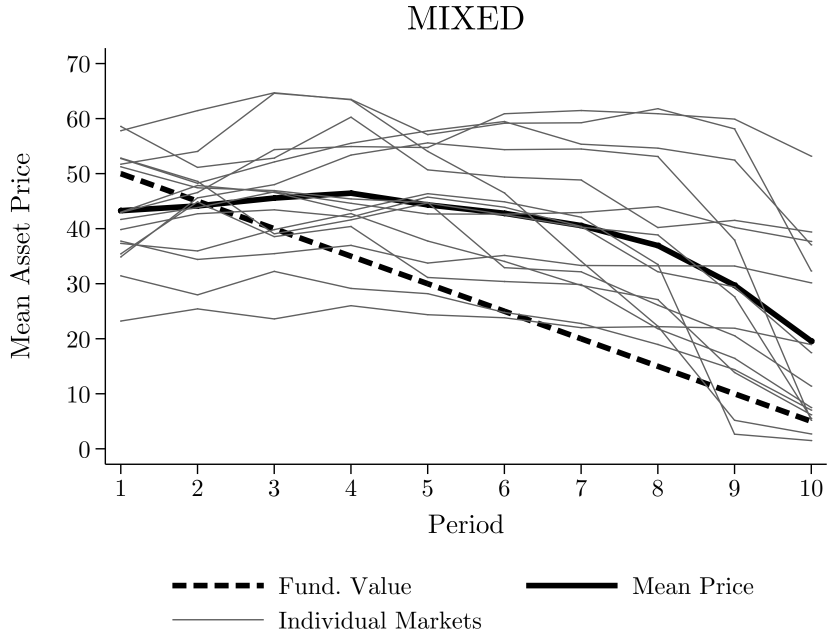


Figure 2.6: Price Evolution in Individual Markets in *MIXED*

Figure 2.6 illustrates the evolution of mean trading prices for the 16 individual markets in the *MIXED* condition. Qualitatively, we get similar results as in *LOWSC*. That is, in some of these markets prices exhibit a hump-shaped development, initially increasing and peaking in some intermediate period. Thus already the presence of a moderate share of traders with depleted self-control abilities is sufficient to reproduce the excess overpricing we observed when all traders' self-control levels were depleted.

2.5.3 Differences in Trading Behavior and Outcomes

Trading Behavior

Differences in market outcomes in the *MIXED* condition compared to *HIGHSC* markets must result from different actions of *MIXLO* traders. However, when analyzing trading

[22]The results of these comparisons also hold when looking at quantity- or trade-adjusted mean prices.

behavior, distinguishing cause and effect is particularly difficult, as already mentioned earlier. A particular deviation in behavior by some traders in the early phases of a market might shift behavior of other (non-depleted) traders. We therefore start by focusing on the very first trading period, where dependencies are less relevant than in later periods. Table 2.2 compares several variables concerning trading activity between *MIXLO* and *MIXHI* traders. Remember that we conduct all statistical tests based on the most conservative definition of independence (the market level), and hence significant effects are usually associated with large absolute differences.

Table 2.2: First Period Differences in Trading Behavior

	Group Mean		
	MIXHI	*MIXLO*	p-value
$\overline{p_{bid}}$	36.377	28.487	0.035**
$\overline{p_{ask}}$	49.931	54.478	0.196
$\overline{q_{bid}}$	16.109	17.788	0.660
$\overline{q_{ask}}$	14.389	15.202	0.796
$\overline{time_{bid}}$	60.425	47.000	0.017**
$\overline{time_{ask}}$	50.230	50.383	0.796
$\overline{firsttime_{bid}}$	51.517	39.846	0.048**
$\overline{firsttime_{ask}}$	34.635	34.435	0.959

Variables starting with a p denote prices, q quantities and time variables refer to the time passed in the current period, thus lower values indicate behavior earlier on. *bid* and *ask* refer to posted bids and asks, p-values from Wilcoxon signed-rank tests with data collapsed on market and treatment level, *** p<0.01, ** p<0.05, * p<0.1

According to Wilcoxon signed-rank tests *MIXLO* traders make significantly lower bids initially ($p = 0.035$) and post these bids earlier than their non-depleted peers ($p = 0.017$). They are also quicker in posting their first bid at the beginning of the period ($p = 0.048$). While not significant, there also seems to be the tendency that *MIXLO* traders (while bidding low) ask for a higher price than the *MIXHI* traders ($p = 0.196$). After period one, these differences vanish, suggesting that non-depleted traders start imitating the behavior of self-control depleted traders.[23] The averages in Table 2.2 suggest an initially more aggressive trading pattern of *MIXLO* traders, trying to buy lower and sell higher than *MIXHI* traders. From trading period two on, however, their behavior has incited non-depleted traders to behave similarly and hence set many markets on an entirely different trajectory.

[23]Results for period two are reported in table B.8 of the appendix indicating that these initial trading differences disappear, while *MIXLO* traders display significantly higher asking prices in period 2.

Table 2.3: Rank Correlations of First Period Behavior with Overpricing

	ρ	p-value
$\overline{p_{bid}}$	0.436	0.104
$\overline{p_{ask}}$	0.488	0.055*
$\overline{q_{bid}}$	0.486	0.066*
$\overline{q_{ask}}$	-0.229	0.393
$\overline{time_{bid}}$	-0.607	0.016**
$\overline{time_{ask}}$	0.262	0.327
$\overline{firsttime_{bid}}$	-0.421	0.118
$\overline{firsttime_{ask}}$	0.079	0.770

Rank correlations of average first-period behavior over all market participants with average relative deviation over periods 2-10 for *MIXED* markets, *** p<0.01, ** p<0.05, * p<0.1

Table 2.3 presents evidence that the observed differences in first period behavior between our treated and non-treated traders are also those behaviors that are correlated with later overpricing. While Table 2.2 has shown that low-self control traders bid earlier in period one and also post their first bid significantly earlier, Table 2.3 shows that markets in which bidding occurs early in period one, are those that exhibit more overpricing over the course of the experiment.

Profits

On average, *MIXLO* traders earned € 8.16, and *MIXHI* traders earned € 7.84 in the experimental asset market – a difference that is not significant (Wilcoxon signed-rank test, $p = 0.9794$). We consider this as evidence that inhibited self-control abilities affect overpricing, but that depleted traders are not necessarily driven out of the market. Instead, as shown previously, they might goad non-depleted traders into more aggressive trading behavior, making everyone end up with similar profits. While this suggests that a lack of self-control abilities is not necessarily detrimental to trading performance, it shows how negative the effect can be for markets on which traders potentially imitate each other's behavior.

2.5.4 Increased Emotional Reactivity

In the experimental sessions that we conducted in November 2015, we asked participants a number of questions relating to their emotional experience during the asset market. In particular, we asked participants to rate how strongly they felt a number of emotions at the beginning of the first period and at the end of the last period, respectively. We asked participants at the end of the experiment, requiring them to recollect their emotions.[24]

Table 2.4 reports the results for those emotions that have previously been connected to overpricing in experimental asset markets (Hargreaves Heap and Zizzo, 2011; Andrade et al., 2016; Lahav and Meer, 2012; Breaban and Noussair, 2013). Note that we collapsed all the emotional measures on the treatment group level within each market and test for differences with Wilcoxon signed-rank tests. Strikingly, the intensity of every single measure of experienced emotions is higher in the *MIXLO* than in the *MIXHI* group, with many measures being statistically significant. At the beginning of period 1, *MIXLO* participants report to feel borderline significantly more surprise ($p = 0.103$) and significantly more joy ($p = 0.058$). Remember that Lahav and Meer (2012) found that inducing positive mood before trading leads to higher deviations from fundamental values and thus larger levels of overpricing and that correlational studies also suggest such a relationship (Breaban and Noussair, 2013; Hargreaves Heap and Zizzo, 2011). Furthermore, at the end of the final trading period, *MIXLO* traders report significantly higher levels of excitement, fear and surprise than *MIXHI* participants (all $p < 0.05$).

We also asked participants in the post-experimental questionnaire explicitly about how strongly they felt their behavior was driven by emotions and how much they had tried to suppress the influence of emotions on their trading behavior (see final panel of 2.4). Even though the difference in averages goes in the expected direction, given the responses to the questions on experienced emotions, they fail to reach significance on conventional levels. The results indicate that the behavior of the traders with depleted self-control abilities might have been driven by emotional factors to a larger degree than they were aware of themselves.

2.5.5 Reduced Cognitive Control

Experiment I did not show a direct effect of the Stroop task on incentivized CRT performance. Condition *MIXED* gives us the possibility to look at the issue again, in particular at the association between CRT, the treatment (*MIXLO* and *MIXHI*), and performance in terms of profits.

[24]We also provided participants with a questionnaire regarding their trading behavior which we do not report here. The average responses to all the emotion-related questions and the test statistics can be found in Table B.4 of the appendix. Average values for changes in emotions over time can be found in Table B.5.

Table 2.4: Ex-post Reported Emotions of Traders in *MIXED*

	MIXHI	MIXLO	p-value
Beginning of the First Period			
Excitement	4.200	4.500	0.400
Fear	2.100	2.175	0.395
Surprise	3.600	4.050	0.103
Joy	3.625	4.375	0.058*
End of the Last Period			
Excitement	3.425	4.200	0.042**
Fear	1.900	2.575	0.014**
Surprise	2.450	3.400	0.030**
Joy	3.375	4.125	0.207
Self-Evaluation of Emotional Reactivity			
Emotion driven	2.475	2.725	0.362
Suppressed emotions	5.300	4.950	0.205

Data collapsed on the treatment level per market; responses were on 7 point Likert scales; test results from Wilcoxon Signed Rank tests; *** p<0.01, ** p<0.05, * p<0.1

Previous research has shown that CRT scores correlate positively with individual participants' profits in similar experiments (Corgnet et al., 2014; Noussair et al., 2014). Toplak et al. (2011) find that CRT scores are correlated with measures of cognitive ability, thinking disposition and executive functioning. Thus, we can interpret the CRT score as a measure of cognitive control. In order to check whether the effect of CRT performance on profits is similar here, we ran additional regressions which we report in table 2.5. Note that we excluded participants who had indicated that they knew at least one of the CRT questions at the end of the experiment. The knowledge of CRT questions before the experiment might have inflated correct CRT responses and thus obfuscate any interaction effects between treatment and CRT scores.[25]

In specification (1) we reproduce the finding that there is no statistically significant difference between the profits of traders in *MIXLO* and *MIXHI*. Specification (2) confirms findings from earlier studies that higher CRT scores are positively related to higher overall profits for both *MIXLO* and *MIXHI*. However, when we separate this effect by treatment by including an interaction of the *MIXLO* dummy with the CRT score, we obtain a larger effect of the CRT score on profits for *MIXHI* traders, while for *MIXLO* traders the effect of CRT scores on profits is significantly smaller

[25] 72 subjects in *MIXED* markets reported to know at least one of the CRT questions. Including these subjects makes the coefficient of the interaction term CRT $\times$ MIXLO insignificant with $p = 0.110$ and $p = 0.197$ respectively in the specifications parallel to (3) and (4) as these subjects water down the effect.

($p < 0.05$) and in fact cannot be distinguished from zero overall (post-estimation Wald test, $p = 0.43$).

Thus, *MIXLO* subjects' trading seems to be relying less on their underlying ability for cognitive control. Together with the results indicating higher emotional valence and reactivity, this suggests an interpretation of trading behavior of *MIXLO* participants as relatively more relying on impulsive system 1 processes than on reflective system 2 processes (Kahneman, 2011).[26]

Table 2.5: Determinants of Profits in *MIXED*

	(1)	(2)	(3)	(4)
	Profit			
MIXLO	1.036	1.040	4.342*	4.301*
	(0.770)	(0.795)	(2.222)	(2.215)
CRT		1.084**	1.882***	1.757**
		(0.497)	(0.621)	(0.691)
CE		0.473	0.867	0.685
		(0.550)	(0.768)	(0.753)
CRT × MIXLO			-1.660**	-1.547**
			(0.642)	(0.690)
CE × MIXLO			-1.031	-1.051
			(1.125)	(1.098)
Female				-1.381
				(0.888)
Constant	7.035***	5.302***	3.936***	5.326***
	(0.441)	(1.097)	(1.323)	(1.638)
Observations	88	88	88	88
R^2	0.016	0.079	0.120	0.145

Participants who indicated to know at least one of the CRT questions excluded; robust standard errors clustered on the market level in parentheses; *** p<0.01, ** p<0.05, * p<0.1

2.6 Discussion

We observe a large main effect of self-control depletion on overpricing in both experiments. The difference in overpricing cannot be explained by a change in risk attitudes

[26] Hefti et al. (2016) argue that good performance in an asset market requires two dimensions of cognitive capabilities, mentalizing and cognitive abilities. Self-control depletion could in principle affect both dimensions and lead subjects to become more impulsive. We deem this an interesting question for future research.

or a simple change in cognitive abilities. Experiment II gives us additional power to assess potential explanations for the excess overpricing after self-control depletion.

First, there are differences in trading behavior. Self-control depleted traders trade slightly less on average, and their initial trading behavior seems more aggressive. For instance, the fact that self-control depleted traders post bids significantly earlier supports the notion that their behavior is driven by a higher degree of impulsivity than the behavior of non-depleted traders. In an environment in which early aggressive activity is potentially imitated by others on the market, not much is needed to set a market on an overpricing trajectory. Notably, trading behavior is strongly path-dependent in experimental asset markets, and the evolution of prices follow different forms and different timings on different markets. We could have presented additional empirical evidence for effects of self-control depletion on trading behavior and trading strategies, but such evidence requires assumptions that are somewhat arbitrary. Hence, we decided to present fewer analyses and only those in whose robustness we are confident.

Second, there are differences in the reported intensity of emotions and relevance of emotions. Due to existing findings, initial differences before the opening of the asset market (and after the Stroop task) are one channel via which depleted self-control could have affected overpricing. Apart from the pre-market emotional state, differential emotional reactions during the market could be driving our results. Emotion regulation has been shown to draw on self-control resources (Baumeister et al., 1998; Hagger et al., 2010). We have evidence that participants displayed more intense emotional states, in particular at the end of the asset market. We interpret our treatment effect as the result of an increased sensitivity towards emotions triggered by self-control depletion. Our effect is in line with the literature on self-control depletion. For example, Bruyneel et al. (2006) have shown that people whose self-control has been reduced rely more on affective and less on cognitive features in product choice. Similarly, in our setting traders with low self-control levels could rely more heavily on affective features of the asset, e.g. the thrill from its recent price increase or from speculation, than on cognitive features, e.g. the knowledge that the fundamental value of the stock is decreasing. Thus emotional responses could be responsible for more myopic decision making, a higher level of overconfidence/overoptimism (Michailova and Schmidt, 2016), and more speculative trading.

Third, cognitive abilities could be different after the two versions of the Stroop task. However, the issue is not as straightforward as we expected. For our sample, we cannot provide evidence on a direct impact of the treatment on CRT performance. This might be because the monetary incentives to do well in the CRT are relatively high, and it is well-known that people can temporarily overcome self-control problems if the motivation is sufficient (Muraven and Slessareva, 2003; Vohs et al., 2012). However, there is evidence in our data for an indirect effect of self-control depletion on cognitive abilities. We find that the CRT carries predictive power for traders' profits, but only if their self-control has not been depleted previously.

There are additional explanations that we cannot pin down fully and have to leave for verification in future research. Self-control depleted traders, for instance, report significantly higher levels of surprise after the last period of the market. This could be an indication for a reinforcement of myopic behavior when being self-control depleted. Another candidate explanation for our treatment effect is a problem to stop, i.e. to sell early enough and not to stick too long to the expectation of a future price rise. Self-control depletion could lead to a reluctance to sell an asset whose price is rising. Similarly, it could lead to undue overoptimism.

2.7 Conclusion

In this chapter, we provide causal empirical evidence for the notion that a lack of self-control can fuel overpricing on asset markets. We consider experimental continuous double auction markets for which Smith et al. (1988) first reported a tendency for overpricing. We exogenously reduce market participants' ability to exert self-control using a tough version of the Stroop task, which has previously been shown to deplete people's ability to exert self-control in subsequent tasks (Baumeister et al., 1998). Comparing two market settings in which either everyone's or no one's self-control was reduced, we observe significantly more mispricing and overpricing as the result of a reduction in self-control abilities than without this reduction.

Self-control depletion affects trading behavior and the perception of the trades and market outcomes. We provide evidence that in markets populated by self-control depleted and non-depleted traders initial trading strategies of the former are more aggressive than of the latter. However, the evidence is not entirely conclusive. Trading is path-dependent on experimental asset markets, and it is difficult to pin down the exact reasons for overpricing to emerge without making arbitrary assumptions. We do not observe a performance difference between traders with depleted self-control and traders with full self-control abilities, suggesting that low self-control traders might not be driven out of the market, but rather incite other traders to engage in more aggressive trading. In addition, we have evidence for an emotional channel that explains our main result. Self-control depleted traders show stronger emotions, in general, but in particular stronger emotions that have been linked to overpricing in previous studies that induce emotions or that measure emotions while trading. Finally, we find that our measure for cognitive skills loses predictive power for the profits of low self-control traders. This might indicate that even though cognitive skills seem unaffected by self-control depletion (as are risk attitudes), different cognitive processes play a role in traders with low self-control levels. These results are in line with a dual systems perspective of self-control: self-control depleted participants seem to have acted more on the basis of emotions and less on the basis of cognition, thus driving up prices.

Our findings have relevant implications: First, with differences in self-control levels, we add a potentially important explanation to the existing explanations for overpricing on asset markets. We have shown that already a moderate number of participants with low self-control levels are sufficient to more than double the extent of overpricing in terms of relative deviation from fundamental value. Second, our results can be regarded as indicative of the role of self-control in markets outside the laboratory – there, both temporary reductions in self-control as well as the personality trait self-control might play an important role in determining trading behavior and perception. Self-control might also be an important attribute on which individuals self-select into trading. However, low self-control traders might not be as easily exploitable by high self-control traders as one would think. In our case, they would not have been driven out of the market quickly. Several practical implications of our results for investing and trading activities come to mind. Given our findings, investment decisions should not be taken under limited self-control or willpower conditions. For instance, cognitive load, food or sleep deprivation, and self-control effort in unrelated domains have been shown to be correlated with limited self-control abilities. If such conditions are unavoidable, decision aides to sustain self-control such as commitment devices should prove useful to circumvent the potentially negative consequences. This might be particularly relevant in fast-paced markets.

Our experiment opens up interesting paths for future research: It would be interesting to see to what extent our results are robust to changes in alternative market mechanisms such as call markets and to changes in the fundamental value process such as a constant fundamental value process, which has been shown to reduce overpricing (Kirchler et al., 2012). Finally, the role of self-control for traders in markets outside the laboratory remains largely unexplored. One can imagine field experiments or using quasi-experimental variations of self-control abilities to study decisions of traders on real markets.

Chapter 3

High-Frequency Trading, Maker-Taker Pricing and Behavioral Adjustments: An Experimental Study of Pricing Structures in Fast-Paced Markets

3.1 Introduction

Financial markets have transformed drastically in the past two decades. While the image of human traders making orders by phone or in face-to-face interactions on the trading floor has become the stereotype of stock market trading, more and more trading has been transferred into dedicated data centers where orders are being executed by computers, sometimes in milli-, micro- or nanoseconds. Hendershott et al. (2011) report that in 2009, already 73% of US trading volume was conducted by computers, a share that has presumably been rising since. The advent of low-latency connections between trading venues and affordable high-power computers enabled the rise of the so-called high frequency traders (HFTs). HFTs are programmed to realize gains across platforms by eliminating arbitrage opportunities, within platforms by using statistical relationships in co-movement of stocks, or by interpreting imbalances in order books as demand and supply shifts that translate into future price changes, usually using limit order based strategies (O'Hara, 2015; Xu, 2015). Brogaard (2010) reports that HFTs made $ 3 billion annual gross trading profits on US equity markets alone in the years 2008 and 2009 and most studies estimate their market share to be within 1/3 and 2/3 of overall trading volume (Biais and Foucault, 2014). These trading profits result primarily from large turnovers, Baron et al. (2012) estimate that the average profit per contract for HFTs is only around $.70.

Using limit order based strategies has traditionally been regarded as having positive impacts on markets, as they act as liquidity suppliers (Foucault et al., 2005; Handa and Schwartz, 1996). For this reason, many markets have adopted maker-taker pricing (MTP), a pricing structure in which limit orders receive rebates on the trading price, and market orders are levied a fee, the spread between the two generating revenue for the trading venue. Maker-taker pricing has become the dominant pricing scheme in the largest equities exchanges (Securities and Exchange Commission (SEC), 2015). There has been some concern that HFTs do not make good liquidity providers, as they may cease to submit limit orders anytime, thereby inducing and perpetuating market instability (Madhavan, 2012). With HFTs dominating limit order based trading, the corresponding favorable pricing scheme and their ability to close arbitrage opportunities faster than any human trader could, human investors need to presumably adjust their trading behavior and react to the shifts in trading strategies and speeds that HFTs induce. The field of behavioral finance has in the past however established that humans in general suffer from several cognitive limitations that might thwart such adjustment processes, such as limited attention (Hirshleifer and Teoh, 2003) or over- and underreaction to signals (Palfrey and Wang, 2012) that might prevent incoming information from being processed correctly.

This chapter investigates how human traders react when high frequency traders enter their market environment, and what role the pricing structure plays in this adjustment process, as it heavily favors HFTs. Since the presence of HFTs and the pricing structure cannot be easily manipulated in real markets, I rely on a laboratory experiment. The stylized environment in a laboratory experiment is highly adequate for this study, as the purpose is to observe general tendencies in human behavior that become apparent whenever the mechanism in question is present. Using an experiment, I can cleanly vary the pricing structure and enable or disable HFTs. In four treatments, I manipulate whether traders face an HFT in their market, whether limit (market) order traders are given a rebate (levied a fee), or both. I then compare these conditions by looking at prices and trading strategies to detect behavioral changes.

My results show that neither maker-taker pricing nor the presence of HFTs significantly influence price paths. Across all treatments trading prices react sluggishly to changes in asset valuation, but there is neither much under- nor overpricing. HFTs substantially shift behavior from limit to market orders, but maker-taker pricing has no such effect. Overall, maker-taker pricing does not impact trading at all, other than that it increases the average number of trades in markets with HFTs. When HFTs are present, I observe that human traders learn to game the algorithm and realize substantially higher profits.

Several papers relate to the present study. One strand of literature deals with the presence of computerized traders. Akiyama et al. (2013) use automated traders in call auctions to investigate the effects of strategic uncertainty. They find that replacing human subjects by computer traders makes price forecasts more accurate, but not perfectly so. Grossklags and Schmidt (2006) and Farjam and Kirchkamp (2016) investigate the role of beliefs about the presence of computerized traders. The former paper uses a market order based trading mechanism for the computer, while the algorithm used by the latter paper is unknown. Both papers report reduced overpricing when automated traders are present. Finally, De Luca and Cliff (2011) and Das et al. (2001) compare different trading strategies using limit orders in markets with automatic execution to see which strategy outperforms humans best. It is important to note, that none of these contributions was particularly interested in the changes in trading behavior due to computerized traders making fast limit order based trades without automatic execution. The only experimental contribution in the area of pricing structures that explicitly models maker-taker pricing is Bourke and Porter (2015), although some empirical studies with less experimental control but a more natural setting exist (Malinova and Park, 2015; Cardella et al., 2015). Bourke and Porter investigate the impact of the pricing on several variables, but only find book depth to increase.

My results add insights on the psychology of traders that are so desperately needed to explain the deviations from rational behavior that lead to stock market crashes and recessions. The finding that price paths are neither influenced by HFTs nor fee structure is reassuring for the design of actual equity markets. More concerning however is the fact that HFTs realize losses. This can be seen as a precursor of market crashes due to a lack of liquidity, as many HFTs would exit the market if they consistently realized losses, which is exactly what happened during the 2010 Flash Crash (Wall Street Journal, 2010). In markets with many uninformed traders that do not close the wedge between prices and values, this problem will likely be exacerbated.

This chapter proceeds as follows, section 3.2 explains the experimental design and sets up testable hypotheses, section 3.3 presents the findings and section 3.4 discusses the implications and concludes.

3.2 Experimental Design & Hypotheses

3.2.1 Setting & Treatments

In the experiment, subjects acted in an asset market where they could trade shares of a single fictitious company over the course of 10 periods. In 4 different between-subjects treatments, implemented using a 2×2 design, I varied whether computer traders were present (*HFT*), whether a pricing structure rewarding market making was implemented (*MTP*), or both (*HFT-MTP*). Table 3.1 lists all treatments.

Table 3.1: Overview of Treatments

		Computerized Traders	
		Yes	No
Maker-Taker	Yes	*HFT-MTP*	*MTP*
Pricing	No	*HFT*	*CONTROL*

Each treatment began with a 2 period dry-run of the asset market after subjects had correctly answered several control questions to test their proper understanding of the market structure. After completion of the dry-run, the actual asset market commenced for a total of 10 periods. Finally, two additional parts elicited subjects' risk preferences using a choice list and measured their cognitive skills. Before receiving their pay, subjects filled in a standard questionnaire on background demographics and answered several questions regarding their trading behavior.

3.2.2 Detailed Description: Asset Market

The experimental asset market used in my experiment closely resembles SSW-style asset markets that have been heavily used in the literature to study overpricing phenomena (Smith et al., 1988; Kirchler et al., 2012; Eckel and Füllbrunn, 2015; Kocher et al., 2016), but also borrows important features from markets concerned with optimal belief updating (Palfrey and Wang, 2012; Bonn et al., 2016; Lindner and Schindler, 2016). More specifically, subjects were initially endowed with assets and cash, and they could act as buyers and sellers for these assets in a ten period double auction market populated with 6 traders, where each period lasted exactly 120 seconds. To encourage trading, the ex ante expected value of each subject's portfolio was held constant, but every trader was endowed with a different percentage of wealth in assets, either 20 assets and 3,000 in cash, or 60 assets and 1,000 in cash. Because the initial asset value was fixed at 50, every trader started the experiment with a wealth of 4,000. In order to trade their assets, subjects could submit bids and asks to a public order book or accept any standing bid or ask for one asset at a time. Additionally, they could remove their own standing offers from the order book at any time during the trading period. Traders also saw a price chart of the price evolution in each period, the current market price and a list of their own past transactions to guide their decisions.

In contrast to many asset market experiments (for a survey see Palan (2013); Noussair and Tucker (2013)), I did not implement dividend payments with increasing cash-to-asset ratio and decreasing fundamental value. Instead, the asset was repurchased at the end of the market (similar to (Kirchler et al., 2015)), but to a randomly determined value. While the asset value before trading began was exactly 50, after each period, the asset value decreased or increased by exactly 20 points with equal probability, but never exceeding 100 and never going lower than 0. Subjects were fully aware of this adjustment process and its limits, and the actual asset value would be publicly announced at the end of each period. The choice for this deviation from the literature was driven by the desire to observe a market that does not create overpricing simply due to subject confusion or inattention. I rather wanted to observe subjects' behavioral adjustments to the presence of a pricing structure and a high speed computerized trader, if they fully understand the setting.

A computerized trader was present in two of the treatments (see above) and automatically traded at very high speeds, significantly faster than subjects. The computer was programmed to post 3 bids slightly below the current market price, and 3 asks slightly above the current market price at any time during a period.[1] At any action of any subject (such as posting a bid or ask, accepting any bid or ask, or deleting a current bid or ask) the computer would reevaluate his position, delete bids above the current market price from the order book, delete asks below the current market price from the order book and post new bids and asks (up to 3 each). The computer would never accept open bids or asks from the order book. Subjects were fully aware of the presence of the computer and the instructions explained the trading behavior of the computer in detail. Furthermore, the computer trader was able to access unlimited cash and assets for trading purposes, a feature that human traders could not make use of, as short-selling and borrowing for participants were ruled out by design. The programming of the computer in many ways resembles high frequency traders in that a) trading happens at faster speed than any other market participant is capable of trading at, b) trading strategies are limit order based, and c) computer traders actions are strategy-based (O'Hara, 2015).

Two of the treatments contained a fee structure (see above) that levied a fee on any trader that accepted an open bid or ask from the order book, and gave a rebate to any trader that had submitted the corresponding limit order in that trade.[2] The fee and rebate were set at 1 point per trade, such that they were revenue-neutral and would allow comparison of my treatments without the fee structure. This pricing structure is called maker-taker pricing, because the liquidity providing party (the market maker) is receiving a rebate, while the liquidity removing party (the market taker) is levied the

[1] How far away from current market price bids and asks were was determined by drawing of several random numbers, that were added or subtracted from the current market price.

[2] In this setting, a market order denotes any transaction where an open bid or ask is accepted from the order book. A limit order in turn means submitting a new bid or ask to the order book.

fee. The fee or rebate were not included in any trading prices and automatically applied upon execution of a trade. For both treatment variations, all subjects received instructions explaining computerized traders and the fee structure, and only later revealed if their session would contain one, any or both of those variations. This particular design was chosen, such that treatment differences due to differently complex instructions could not arise.

After each period, subjects had to answer four questions intended to elicit their beliefs. Specifically, I asked them to indicate 1) the probability of the asset's value increasing in the next period, 2) the average answer from all traders in their market to question 1), 3) the asset's value at the end of the experiment, and 4) the average estimated asset's value among all other traders in their market. While question 1) served as a test whether subjects could successfully remember or understand that the probability of a value increase was always 50%, question 2) elicited the extent to which subjects believed other traders would remember or understand this, serving as a measure for rational expectations. Question 3) tested for particularly negative or positive expectations of asset valuation developments for the trader, while question 4) assessed the trader's expectations about other traders' expectations. All questions were incentivized using a quadratic scoring rule, which exhibits some theoretically undesirable features Manski (2004), but doesn't perform worse in practice compared to other elicitation methods (Trautmann and Kuilen, 2015).

At the end of the asset market, the final asset value was announced, all assets were repurchased and the revenue added to cash holdings, which were then converted into Euros using an ex ante announced exchange rate.

3.2.3 Detailed Description: Additional Tasks

After completion of the asset market, I elicited subjects' risk attitude using a multiple choice list with 20 items comparing a lottery and an increasing fixed amount (Dohmen et al., 2011). Subjects were only allowed to switch once from the lottery to the fixed amount, and I treat the switching point as their individual certainty equivalent, allowing to categorize subjects into risk-averse, risk-neutral and risk loving. After they indicated their preferred choice for each of the 20 items, one item was randomly picked by the computer and the preferred choice was implemented (either the lottery or the fixed amount).

Subjects then took on the Raven's Advanced Progressive Matrices (RAPM) test. For the duration of 5 minutes, they were given a number of 3×3 matrices of symbols, where the symbol in the lower right hand corner was missing. They then had to select from 8 options the only symbol that would meaningfully complement the matrix. This test has been used to measure cognitive abilities, as completing the matrices requires logical thinking. Subjects received a positive payment for each correct answer, and a negative payment of the same amount for each incorrect answer. After answering one matrix, the next matrix immediately appeared, until the entire 5 minutes had passed, upon which the task ended immediately.

3.2.4 Procedures

In April 2016, I conducted the experiment with 4 treatments and 3 sessions per treatment to a total of 12 sessions. 264 subjects participated in sessions of 24 subjects each at the computerized laboratory MELESSA of the University of Munich. Sessions lasted approximately one hour and subjects received an average payoff of € 14.46. Before each part, subjects received written instructions that were read aloud to ensure common knowledge.[3] Remaining questions were answered in private and subjects received additional support if they encountered difficulties with the practice questions. All payments were privately made in cash before subjects were dismissed from the laboratory.

3.2.5 Hypotheses

If all traders were perfectly rational and had identical risk preferences, there should not be any trade, as all traders carry the same asset valuation, even if the latter changes after each trading period. Allowing for varying degrees of risk aversion among traders however will also generate differences in valuation, as future price paths of the asset become uncertain. Such a setting requires the most risk loving trader to hold all assets as his willingness to pay for any asset is higher than that of any other trader in the market. Importantly, the presence of computerized traders should have no impact of trading behavior in such a setting. The computer generates offers above and below market price, but in a setting with perfectly rational traders, the market price will exactly be the difference between the risk premia of the two most risk loving traders – a number potentially decreasing in the number of traders in a market. A fee structure in turn would have an impact on rational traders to the extent, that it constitutes a shift in prices. Traders executing market orders would trade at prices that reflect their valuation minus the fee, while limit order traders execute trades at prices reflecting their valuation plus the rebate.

[3]English translations of the instructions can be found in the appendix.

The experimental data allows me to test a number of hypotheses—about the impact of maker-taker fees, the presence of high-speed computerized traders, and the joint influence of the two. Since fees and rebates are revenue-neutral, they constitute simply a shift of the price from the market order trader to the limit order trader. Given the complexity of the trading environment (Huber and Kirchler, 2012; Kirchler et al., 2012), we might expect that not all traders are fully aware of the fees at all time due to limited attention and might trade at higher (or lower) prices than they would in the absence of fees. If limited attention is a feature of cognitive abilities as I measure them, I should also expect a differential effect for different levels of cognitive skills. This leads to the first hypothesis:

Hypothesis 1 *Market order traders will trade at higher prices in the presence of a fee structure, while limit order traders will trade at lower prices. The effect is more pronounced for subjects with lower cognitive abilities.*

Alternatively, traders with limited cognitive skills might out of fear of being involved in disadvantageous trades instead rely more strongly on limit orders instead of market orders, in order to avoid incorporating the fee structure in their trading decisions.

Hypothesis 2 *Traders with lower cognitive abilities use more limit orders in the presence of a fee structure.*

The presence of high-speed computerized traders in turn will automatically allow subjects to make more use of market orders, as the computer mechanically provides 6 limit orders at all times (3 bids and 3 asks). This can also facilitate trading speed, as open bids and asks can be accepted continuously, without waiting for other human traders to generate acceptable offers.

Hypothesis 3 *In the presence of high-speed computerized traders, human traders will rely more strongly on market orders than limit orders. They will also engage in more trades per period than in the absence of automated traders.*

Given that computerized traders strategically maximize trading profits by posting offers around the market price to eliminate arbitrage opportunities, they should have a stabilizing effect on prices and generate smaller profits for human traders. Such a stabilizing effect should also be expected to be present in the within-period variance of market prices.

Hypothesis 4 *Profits are smaller in the presence of computerized traders and the variance of market prices within a period is smaller.*

When both a fee structure and computerized traders are implemented, the predictions become less clear. While the fee structure could drive subjects towards using more limit orders, automated traders could induce the opposite effect. Due to the comparatively small fee, I hypothesize that the latter effect is stronger.

Hypothesis 5 *When both high-speed computerized trader and a fee structure are implemented, subjects rely on market orders more strongly, but to a smaller extent than they do in the absence of fees.*

3.3 Results

3.3.1 Prices

Before I look at the detailed behavioral adjustments that traders make before they are faced with the respective market structure change, I investigate the price paths observed in the different treatments. Figure 3.1 shows the evolution of prices in all four treatments, compared to the actual fundamental value (FV). Remember, while the initial value of the asset is 50 points, the fundamental value of the asset adjusts by plus or minus 20 points after each period and therefore varies over the course of 10 periods. The first striking result that can be seen is that prices evolve more or less constantly, reacting somewhat to the changes in fundamental value, but clearly to a lesser extent than the fluctuations in value. The price adjustment in my experiment therefore seems similar to the way beliefs are updated in Bonn et al. (2016); Palfrey and Wang (2012). To quantitatively test whether this observed difference is statistically significant, I calculate for each period the increase in fundamental value over the previous period, and the increase in prices over the previous period, and compare the two.[4] Wilcoxon signed-rank tests confirm that price adjustments are indeed smaller than the adjustment of fundamental value in all four treatments ($p = .0745$ in *CONTROL*, $p = .0033$ in *MTP*, $p = .0044$ in *HFT*, and $p = .0029$ in *HFT-MTP*).

Due to prices being almost constant while the fundamental value varies, the question arises if the resulting prices constitute over- or underpricing. Figure 3.1 at least suggests that prices stay above and below fundamental value for at least some time. I therefore test whether prices and fundamental coincide in each of the four treatments using Wilcoxon signed-rank tests. While prices are on average statistically indistinguishable from the fundamental value in treatments *CONTROL* ($p = .1688$), *MTP* ($p = .7221$), and *HFT* ($p = .1307$), they significantly lie below the fundamental value in *HFT-*

[4]All non-parametric tests in the results section are two-sided and conducted after collapsing the variables to the market level to obtain conservative estimates, unless otherwise noted.

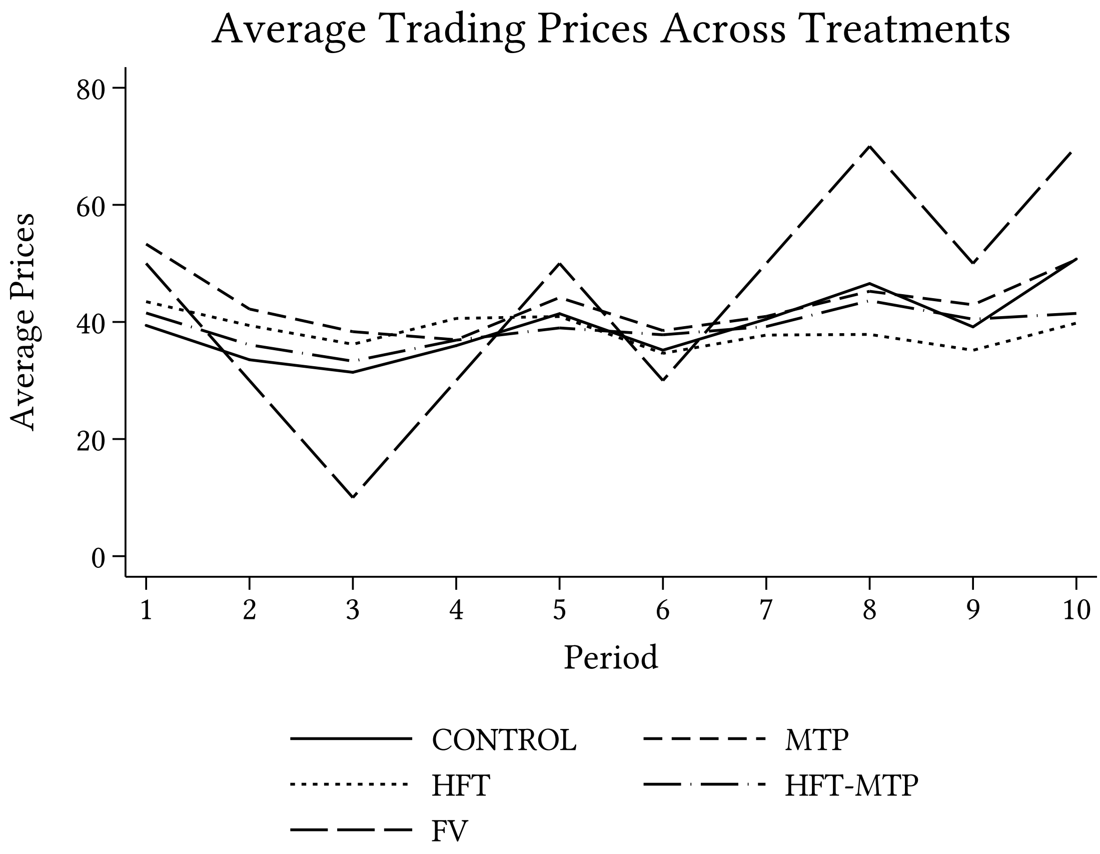


Figure 3.1: Price development across treatments

MTP ($p = .0712$). This significant difference is presumably mainly due to the strong increase of the fundamental value in the last period, while prices slightly drop. In all other periods, prices in the different treatments seem to be pretty close together. A Kruskal-Wallis test indeed confirms that no price path is significantly different from the others ($p = .6224$).

Overall, the price evolutions suggest that neither the presence of high-speed computerized traders, nor a fee structure has much of an impact of prices in this setting. Subjects trade at almost constant prices and react very little to changes in fundamental value. This means that behavioral changes on the individual level that I will look at more closely below, do not translate into aggregate price changes. This is reassuring, as it means that aggregate market outcomes may not necessarily depend on these features of the market structure.

3.3.2 The Effects of Fees

To study the effects of a fee structure with maker-taker pricing, I first investigate if traders who use more market orders exhibit substantially different prices. I therefore build the average price at which each trader traded and relate it to the average number of those transactions that are market orders. A Spearman correlation on the subject level then detects a directionally negative correlation, that is however not statistically significant ($\rho = -.1055$, $p = .4107$). It therefore does not seem that traders take the fee structure into account, although the very small fee of one point makes detecting treatment differences unlikely to begin with.

To further test Hypotheses 1 and 2, I turn to regression analysis. Table 3.2 reports results from an OLS regression at the individual level with heteroskedasticity-robust standard errors, only using data from my *CONTROL* and *MTP* treatments. Column 1 regresses the average trading price of a trader on a treatment dummy, where 1 denotes *MTP*. Maker-taker pricing seems to increase trading prices in this specification, a finding that did not show up in the non-parametric comparisons, presumably due to more conservative testing. Adding the share of market orders in column 2 does not alter the results by much, and confirming my non-parametric finding from above, there does not seem to be a relation between the share of market orders and prices.[5] Column 3 adds the interaction of the share of market orders and the *MTP* dummy. This coefficient particularly answers how prices were influenced by subjects that used more market orders in *MTP*. The coefficient is negative but not yet significant ($p = .107$), delivering at most weak evidence that traders using more market orders in *MTP* might actually trade at lower prices in line with hypothesis 1. Column 4 adds gender, which has been predictive of trading behavior (Eckel and Füllbrunn, 2015). While gender itself has no predictive power, it drives the coefficient of the interaction term further away from significance. Finally, column 5 adds cognitive skills as the overall performance in the Raven task. This reduces the p-value of the interaction term further, while itself cognitive skills do not seem to matter for prices.[6]

While there is no significant evidence that traders using more market orders also trade at higher prices, there is weak evidence that adding a fee structure has the hypothesized effect of driving traders in this direction. The results are however not significant at conventional levels, which could be due to the rather small fee levied on the transaction, and will therefore be treated as non-results. Cognitive skills do not seem to possess much explanatory power itself. I therefore turn to Hypothesis

[5]Note that the number of observations drops in column 2 as not all traders submitted market orders.

[6]Including a third level of interactions using the results from the Raven task does not deliver additional insights. Results available from the author upon request.

Table 3.2: Regressions of trading prices on fee structures

	(1)	(2)	(3)	(4)	(5)
VARIABLES			Trading Price		
Fees	3.913**	3.503*	125.663*	125.052	114.587
	(1.704)	(1.772)	(74.740)	(75.863)	(79.009)
Market Orders		-17.881	112.752	108.489	101.829
		(82.672)	(125.583)	(129.253)	(137.680)
Fees × Market Orders			-244.290	-243.186	-221.710
			(149.290)	(151.527)	(157.750)
Male				.571	.446
				(1.954)	(1.939)
Cog. Skills					-.900
					(1.104)
Constant	35.550***	44.956	-16.940	-15.150	-9.840
	(2.624)	(41.629)	(62.956)	(64.654)	(68.374)
Observations	126	121	121	121	121
R-squared	.040	.032	.041	.042	.048

Coefficients from a OLS regression of individual trading prices using the data from treatments *CONTROL* and *MTP*. Fees denotes treatment *MTP*, Market Orders is the share of trades conducted as a market order, Male is a gender dummy, and Cog. Skills is the performance in the Raven task. Robust standard errors reported in parentheses, and * $= p < .10$, ** $= p < .05$ and *** $= p < .01$.

2 by regressing the share of individual market orders on my measure of cognitive skills, as well as other covariates. Column 1 shows the baseline, in which only a dummy indicating whether treatment is *CONTROL* or *MTP*. By construction, subjects use a market order share of close to 50 % in both treatments.[7] Adding cognitive skills in column 2 and the interaction in column 3 does not add much power and hardly influences the results, indicating that cognitive skills play no role on the decision whether to rely on market or limit orders. Adding additional controls in columns 4 and 5 does not change this finding.

While my previous results at least allowed to speak of suggestive evidence because some tests only marginally missed conventional significance levels, the results in this table are more clear. It does not seem that traders low in cognitive skills evade the incorporation of a fee structure by switching from market to limit orders. Overall, and in line with the findings in Bourke and Porter (2015), a fee structure does not seem to change much of traders' behavior, apart from weak evidence that it entices them to trade at higher prices.

3.3.3 The Effects of High-Speed Traders

The presence of high-speed computerized traders opens the possibility for subjects to switch from market making to market taking, as the computerized trader will provide sufficient liquidity through its limit orders. Figure 3.2 compares the share of transactions being market orders across all four treatments. While our treatments

[7] In the absence of HFTs, each human market order must correspond to a human limit order.

Table 3.3: Regressions of market orders on cognitive skills

VARIABLES	(1)	(2)	(3)	(4)	(5)
	Share of Market Orders				
Fees	-.001	-.001	-.006	-.006*	-.006
	(.001)	(.001)	(.004)	(.004)	(.004)
Cog. Skills		.000	.001	-.001	-.001
		(.001)	(.002)	(.002)	(.002)
Fees × Cog. Skills			.002	.002	.002
			(.002)	(.002)	(.002)
Male				.002	.002
				(.001)	(.001)
Age					.000
					(.000)
Constant	.502***	.501***	.502***	.501***	.498***
	(.002)	(.002)	(.003)	(.003)	(.008)
Observations	121	121	121	121	121
R-squared	.007	.008	.020	.034	.039

Coefficients from a OLS regression of the share of market orders using the data from treatments *CONTROL* and *MTP*. Fees denotes treatment *MTP*, Age denotes the trader's age in years, Male is a gender dummy, and Cog. Skills is the performance in the Raven task. Robust standard errors reported in parentheses, and * = $p < .10$, ** = $p < .05$ and *** = $p < .01$.

where computerized traders are absent show a share of market orders of 50%, this share strongly increases to about 60% in all treatments where an automated trader is present. When comparing the share of market orders between treatments *CONTROL* and *HFT*, this difference is statistically highly significant (Mann-Whitney test, $p = .0001$), and so is the difference between *CONTROL* and *HFT-MTP* (Mann-Whitney test, $p = .0001$).

These results are clearly support for the hypothesized shift of trading behavior from limit orders to market orders when the computer is supplying sufficient liquidity through active orders in Hypothesis 3. Interestingly, there is still a substantial extent (approximately 40%) of trading being conducted using limit orders. To test the second part of this hypothesis, I will compare the number of trades per period across all treatments. Figure 3.3 shows exactly this comparison for all four treatments. There is a clearly visible difference in the Figure, but the claim that this difference is driven by adding computerized traders cannot be confirmed (comparing the average number of trades in *CONTROL* and *HFT* does not deliver significant estimates, Mann-Whitney test, $p = .3600$, nor does the comparison between *MTP* and *HFT*, $p = .2244$). Strikingly, the entire increase in trades seems to stem from the combination of maker-taker fees and having computerized traders, as Mann-Whitney tests show highly significant differences: $p = .0138$ when testing against *HFT*, $p = .0006$ when testing against *MTP*, and $p = .0010$ when testing against *CONTROL*. This suggests that both features are necessary to induce subjects to trade more volume, but interestingly, the increased trading activity does not lead to significantly higher prices.

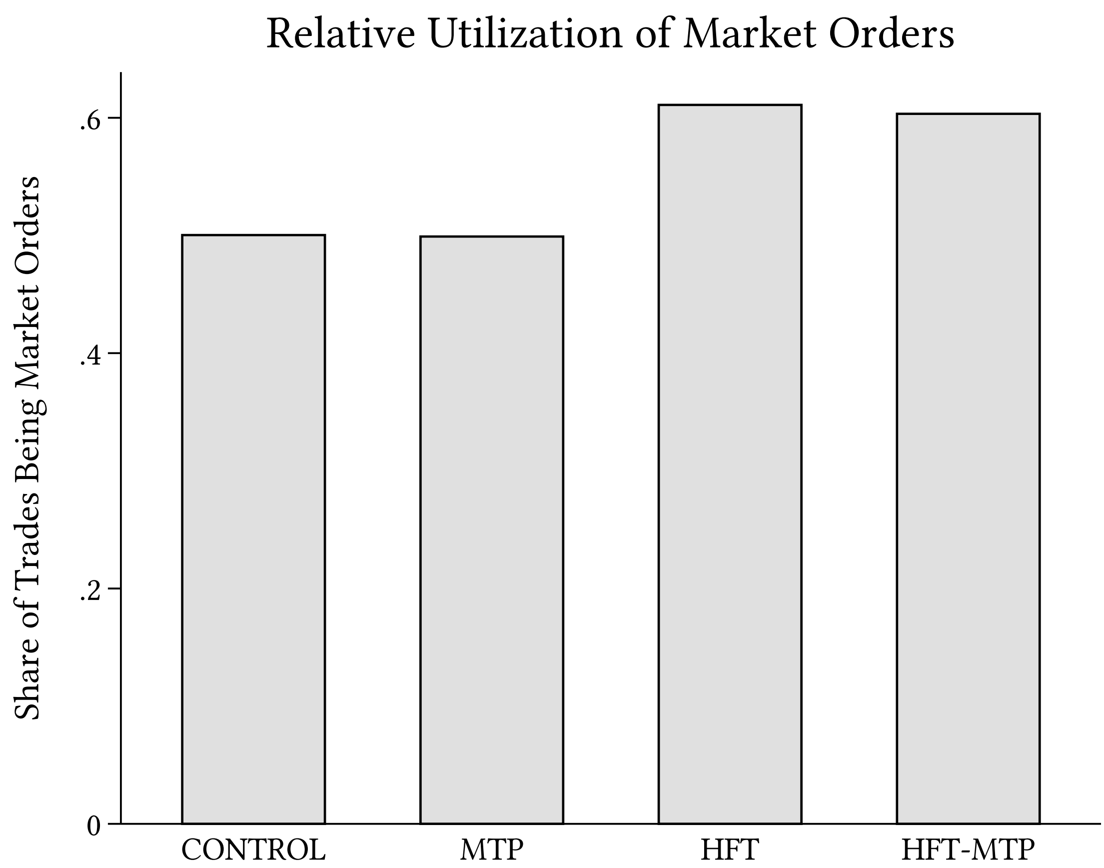


Figure 3.2: Share of market orders across treatments

The findings with respect to Hypothesis 3 are therefore straightforward. High-speed computerized traders indeed push human traders to make increasing use of market orders. They are however not sufficient to increase trading volume, which also needs maker-taker pricing to lead to significant differences. I now test whether these differences in trading behavior can also influence traders' profits and the variance of market prices.

Figure 3.4 shows the evolution of profits in all four treatments. Because of the asset market being a zero-sum game in the absence of computers, average profits always equal €4 in treatments *CONTROL* and *MTP*. In both treatments where computer traders are present, however, subjects make substantially higher profits that accumulate over time. Apparently, some subjects strategically used a weakness of the computer algorithm. Because all offers were relative to the current market price, subjects executed market orders at low prices initially and purchased large quantities of assets from the computer. In later periods (and even later within periods), they traded at higher prices to make the computer create offers at high prices, in which they sold all their previously cheaply bought assets back to the computer. The fact that the computer had no

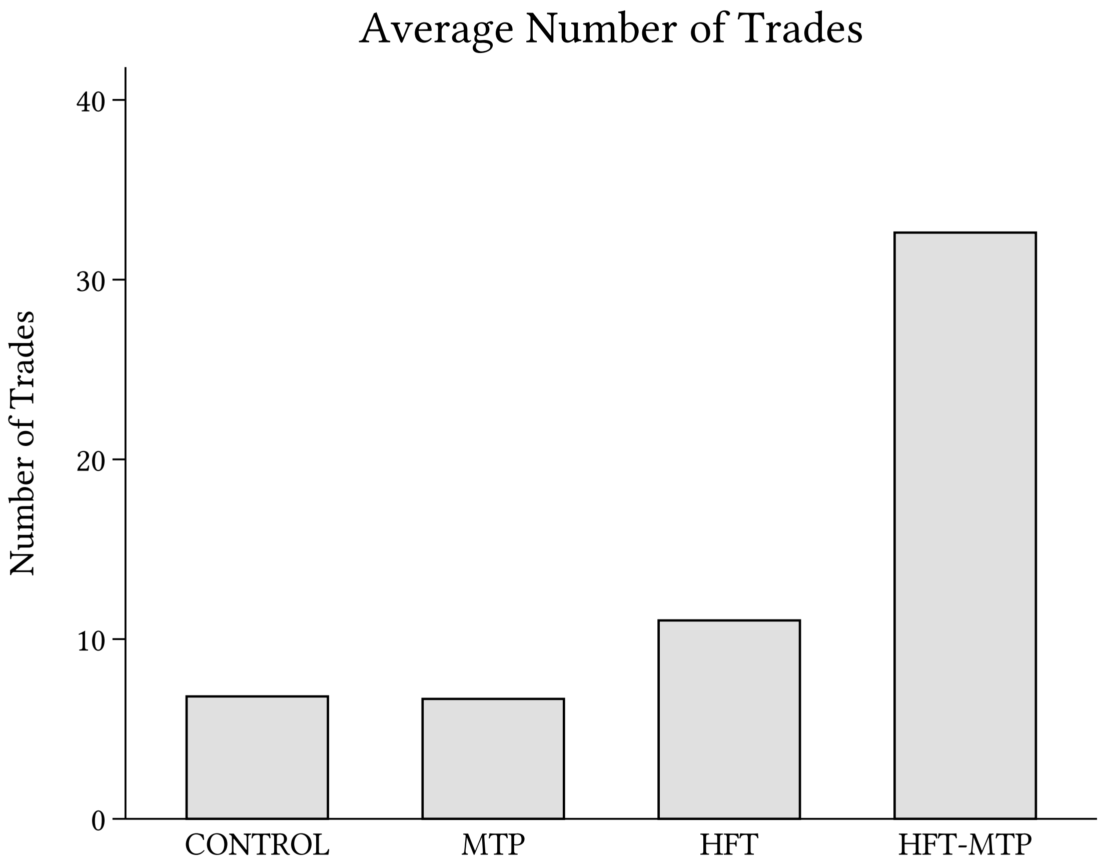


Figure 3.3: Number of trades across treatments

participation constraint, but supplied assets and cash constantly, could thus be used to realize substantial profits. This offers insights into how bubble-and-crash phenomena in such a setting could occur: An actual high frequency trader would react similarly, but would usually possess a routine that makes him exit the market, whenever he consistently realizes losses. This could lead to a sudden drop in liquidity provision that makes the market collapse.

Since developing a strategy to game the algorithm and then colluding to adjust market prices requires substantial cognitive effort, I test the relationship of cognitive skills and profits. A Spearman rank correlation shows that higher scores on the Raven task tend to be correlated with higher profits in the asset market but not significantly so: $\rho = .3468$, $p = .1050$. I find a similar picture when comparing the share of market orders with profits from the Raven task. While the correlation is positive and barely misses conventional significance levels in treatments in which computerized traders are present: $\rho = .3473$ and $p = .1045$, there is no such relationship in treatments where computer traders are absent ($\rho = -.2471$ and $p = .2802$). While some traders therefore strategically exploited the computer, others did presumably not realize the arbitrage

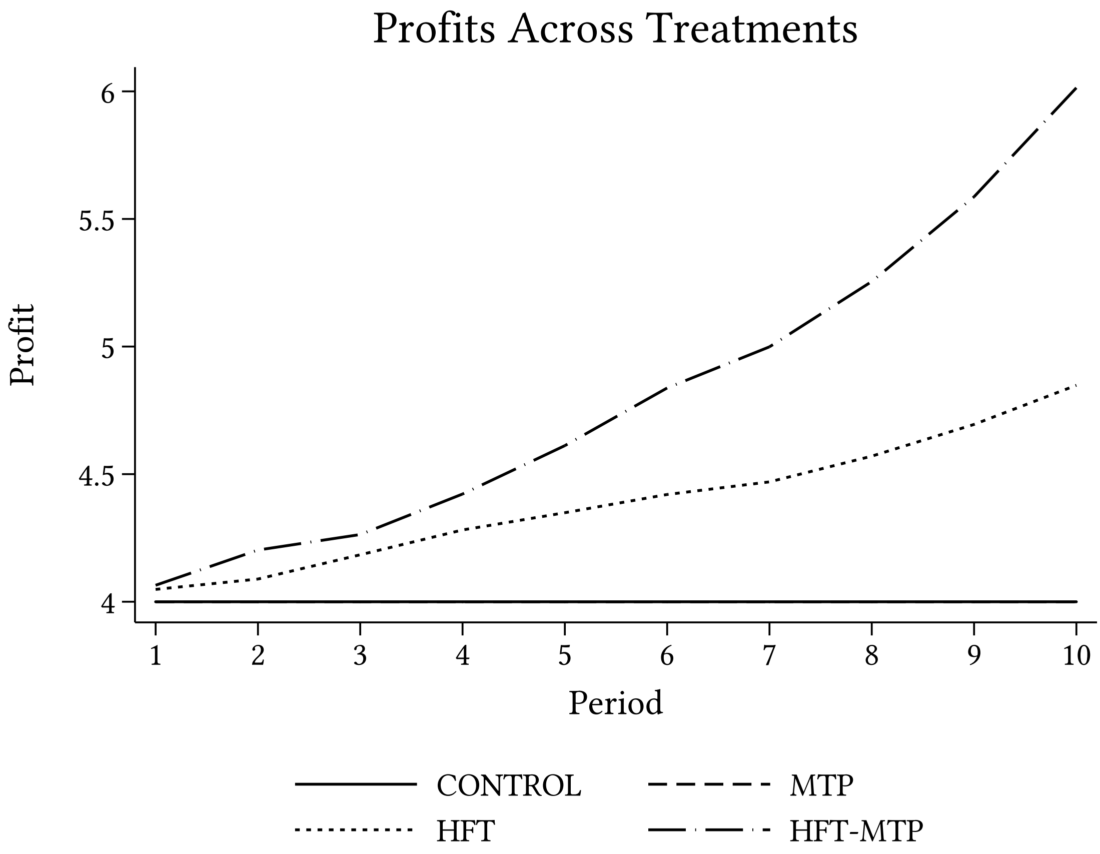


Figure 3.4: Profits across treatments

opportunities created by the wedge between prices and value. This phenomenon would be a particular problem in markets in which many traders are uninformed about underlying values of assets, such as in China, where large parts of the population engage in asset trading as a substitute for gambling, which is illegal.

It is also not the case that high-speed computer traders have a mitigating impact on price fluctuations. I compute the within-period variance of market prices for each treatment as a measure of price volatility. A Kruskal-Wallis test however does not find one condition to be significantly different ($p = .5553$). Given the increasing profits for traders in the settings with computerized traders, there could be counteracting effects. While computerized traders stabilize the price around the current market price by offering bids and asks close to the current prize, they also destabilize the price when human traders try to exploit them.

Summarizing, while I find support for behavioral shifts towards more market orders, computer traders alone neither contribute to increased trading activity, nor do they stabilize the prices. Owing to some subjects colluding and exploiting the computerized trader, the algorithm actually suffers from severe losses.

3.3.4 Effects in HFT-MTP

Since I have previously established that the share of market orders does not change following the introduction of maker-taker pricing, and that market orders suddenly increase when computerized traders are present, there should not be much of a mitigation of that effect in a setting, where both treatment variations are present. Figure 3.4 has already shown that there is a difference between *HFT-MTP* and *CONTROL*, as well as *HFT* and *CONTROL*. As expected from visual inspection, there is indeed no significant difference between the share of market orders in *HFT-MTP* and *HFT* (Mann-Whitney test, $p = .9020$), but the tendency is consistent with hypothesis 5, as the share of market orders is slightly lower in *HFT-MTP*.

Apart from the not significant difference in market order utilization, treatment *HFT-MTP* only differs from the others by a much steeper slope of profits over time (see Figure 3.4) and by significantly more trades (see Figure 3.3).

3.4 Conclusion

This chapter investigated the effects of high-frequency traders (HFT), maker-taker pricing and the joint effect of the two on behavioral adjustments by human traders. The topic is of particular importance, given that asset markets have undergone major change in the past decades. While more computerized trading is often argued to have made markets more stable and removed arbitrage opportunities, due to the boundedness of rationality in human traders, there could be unintended consequences.

My experiment therefore consists of four treatments, where I vary whether HFTs are present, whether maker-taker pricing is implemented or if both are present at the same time. Laboratory subjects engage in trading of one fictitious asset in each of the treatment conditions. HFTs are programmed to constantly create limit orders for purchases and sales slightly above and below the market price, thereby smoothing price paths. Maker-taker pricing takes the form of a fee to be levied on market orders, while the corresponding limit orders receive a rebate of the same amount.

I find that prices do not differ across the four treatments, suggesting that all treatments contribute to market efficiency equally. Maker-taker pricing does not appear to have much of an influence on any variables, as I do not observe changes in trading behavior. HFTs instead encourage human traders to engage much more strongly in market orders. Because subjects learn to game the algorithm, they end up with substantially higher profits. This is made possible, because other market participants fail to exploit the arbitrage possibilities due to the wedge between value and prices.

My findings provide important insights. First, it seems that human traders are not more prone to engage in speculative behavior depending on the presence of HFTs and the fee structure. This should be reassuring to market designers. Second, the fact that collusion can lead to strong losses for the algorithm, reemphasizes the need to ensure liquidity provision to markets. When algorithms face consistent losses, they usually withdraw from the market, as can be seen with the role of spoofing in the 2010 Flash Crash.

This study also has some limitations, that should be addressed. First, the fact that the pricing structure has no visible impact on trading behavior might be due to the fee being relatively small compared to market prices (and the variance thereof). This might have led subjects to just ignore the fee, and focus on realizing trading profits from other traders that do not value the asset correctly. Future experiments should therefore analyze this relationship in a setting where the trading fee is much more substantial. Second, the fact that HFTs can be exploited without exiting the markets upon realization of substantial losses suggests that crashes due to lack of liquidity might have been prevent. To also judge the influence HFTs have on human decision making when markets break down, future studies should design an HFT such that it exits the market when needed to cut losses. I leave these questions for future research.

Chapter 4

Shocking Racial Attitudes: The Cultural Legacy of Black G.I.s in Europe*

4.1 Introduction

The death of Martin Brown in Ferguson, Missouri and anti-migration marches in Germany have reawakened debates about racial prejudice in both the US and Europe. Whilst racism persists, minorities' welfare suffers through labour market discrimination, reduced access to public goods and hate crime. How can such prejudices be reduced? Allport (1954) suggested that contact with minorities could be enough to soften attitudes towards them, and this has been an important subject of study for social psychologists ever since (see Pettigrew 1998 for a review).

This chapter exploits a natural experiment to test Allport's theory of intergroup contact. During the course of World War II, around 150,000 African American GIs (troops) served in Europe. Like white soldiers, most arrived via ports in the west of England and spent time at one or more locations in England and Wales before the invasion of German-occupied France in June 1944. Whilst stationed in the UK, both black and white GIs came into frequent contact with the local population: "Got any gum chum?" became a popular refrain amongst British children and troops were frequently to be found in local pubs and dance halls. In areas where black troops were posted, many locals saw black people for the first time.

We create a measure of potential contact with black GIs during World War II using detailed information on military units and their locations at various points in time.

*This chapter is based on joint work with Mark Westcott.

We combine this with data on membership of the British National Party (the BNP, a far-right party) from Biggs and Knauss (2011) and with data collected by Project Implicit, a large online experiment (Xu et al., 2014), to demonstrate the effect of contact on contemporary racial attitudes. Individuals living in areas where significant numbers of black troops were located are less likely to join the BNP and report warmer feelings towards black people. Using data on those who have migrated to 'treated' areas, we show that persistence is primarily driven by 'horizontal' socialisation rather than intergenerational transmission. However, we find no persistent effect on contemporary implicit attitudes.

Our study adds to the literature on intergroup contact in two main ways. Firstly, our identification strategy provides quasi-experimental variation in potential contact throughout an extensive geographical area. In many papers, data on the ethnic composition of neighbourhoods has simply been combined with individual-level attitudinal data to show that those living in more diverse areas report less prejudiced attitudes (see Pettigrew and Tropp 2006). However, since migrants might avoid moving to areas with hostile populations, these studies suffer from potential selection bias (Dustmann and Preston, 2001). In our setting, decisions on the assignment of troops were made according to military needs without consideration of local racial attitudes. Black soldiers served almost exclusively in segregated black-only units, providing support services such as logistics, transport and food preparation. By necessity, these units (around 10% of the total) were stationed alongside white-only combat units. This led to a similar pattern of deployment between white and black troops but variation in the number of black units at any base, which our identification strategy can exploit. In addition, we add to the contact hypothesis literature by showing that a short period of intergroup contact can have persistent effects on attitudes decades later.

We also contribute to the literature on the persistence of cultural norms, within which individual preferences are seen as endogenous to social and family environments. A general model of preference formation is provided by Bisin and Verdier (2001). In this model, parents take costly investments in their children's preferences, resulting in 'vertical' transmission of values, but children are also socialised by the wider society in which they grow up, resulting in 'horizontal' transmission. Bisin and Verdier show the conditions under which heterogeneous cultural values can endure in the presence of these potentially opposing forces, and several recent empirical papers have now documented very long-run persistence in cultural values. Voigtländer and Voth (2012) show that anti-semitic attitudes in German towns and cities persisted over a time span of almost 600 years; individuals in locations which saw persecution of Jews during the middle ages were more likely to engage in anti-semitic behaviour immediately prior to and during World War II. They attribute the persistence to vertical inter-generational transmission. Using data from the 'Afrobarometer' project, Nunn and Wantchekon (2011) also show persistence of cultural traits across several centuries. They demonstrate systematically lower levels of trust in ethnic groups which were more heavily affected by the trans-Atlantic slave trade. They find evidence for both horizontal and vertical transmission of trust, but attribute a substantially larger effect to the vertical transmission mechanism. In contrast to both these papers, our results indicate that persistence in racial attitudes is primarily driven by horizontal transmission through the neighbourhood.

To the best of our knowledge, we are the first to look at the persistence of cultural norms not only at the explicit, but also at the implicit level. The distinction between explicit and implicit attitudes goes back to Greenwald and Banaji (1995), who define implicit attitudes as "traces of past experience [that] affect some performance, even though the influential earlier experience is not remembered in the usual sense—that is, it is unavailable to self-report or introspection" (Greenwald and Banaji, 1995, p. 4f.). Implicit attitudes, as measured by a computerised test ('IAT test'), have been shown to be predictive for outcomes in a variety of domains (for an overview see Uhlmann et al., 2009). Agerström and Rooth (2011) find that hiring managers' discrimination against obese job applicants is predicted by a 'weight' IAT that measures implicit attitudes towards the overweight. In an incentivised experiment, Stanley et al. (2011) show that white subjects' implicit attitudes predict their judgement about the trustworthiness of randomly matched black partners, even conditional on their reported racial attitudes. Our analysis shows that contact between black GIs and local populations has had no persistent effect on contemporary 'implicit' racial attitudes. So whilst our analysis finds evidence for the 'contact hypothesis' in two independent measures of explicit attitudes, implicit attitudes were either unaffected or have not been subject to the same transmission mechanisms.

The chapter proceeds as follows: Section 4.2 introduces the historical background and data, Section 4.3 describes some necessary preliminaries for the data analysis. Section 4.4 documents the effect of historical intergroup contact on contemporary support for the British National Party. Section 4.5 uses data from Project Implicit to explore channels of persistence and to test whether the historical episode affected implicit attitudes. Section 4.6 concludes.

4.2 Historical Background and Data

4.2.1 Historical Background

The United States entered World War II on December in 1941 following a declaration of war from Germany and its immediate reciprocation by the US Congress. A headquarters for the US military's European operation were established in London the next month. The first combat troops arrived via ports in Northern Ireland and VIII Bomber Command, later redesignated as the Eighth Airforce, was incorporated and dispatched to Britain to join the Royal Air Force in the aerial bombardment of Germany. It flew its first missions in June from bases in the east of the country, designated the Eastern Base Section. This was one of four base sections (areas) that had been created (East, West, South and Northern Ireland) in order to decentralize operations (See Waddell, 2010, p. 142 and Figure 4.1 for a map).

The UK also functioned as a staging post for the ground troops who would later liberate France and eventually Germany. These troops began arriving in the UK in May 1942 in preparation for a mid-1943 land offensive, but this operation was postponed and troop numbers declined from a peak of 230,000 to around 100,000 as troops were reallocated to North Africa or the Pacific (for figures on monthly troop strength see Ruppenthal 1978 and Figure 4.2). The 'friendly invasion' (Buckton, 2006) of Britain began in earnest in May 1943 once plans for a 1944 offensive were settled. By November, around 160,000 troops were arriving per month. Troop numbers reached their peak in June 1944 with one and a half million GIs stationed in the UK, with newly arrived troops concentrated particularly in the Southern and Western Base Sections.

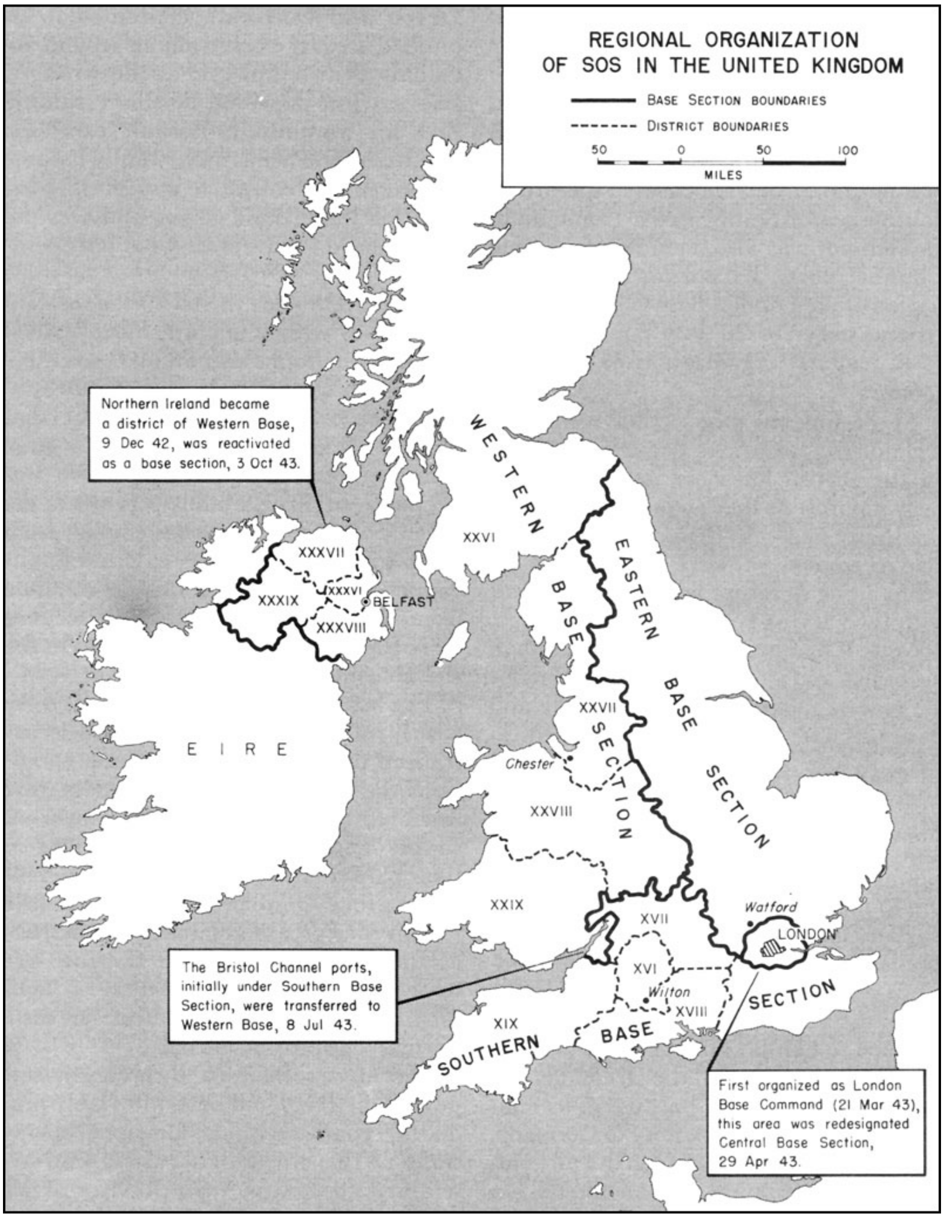


Figure 4.1: Base Sections in the UK, figure taken from (Ruppenthal, 1978, p. 85).

A map of the UK base sections showing the boundaries for the western, southern and eastern sections.

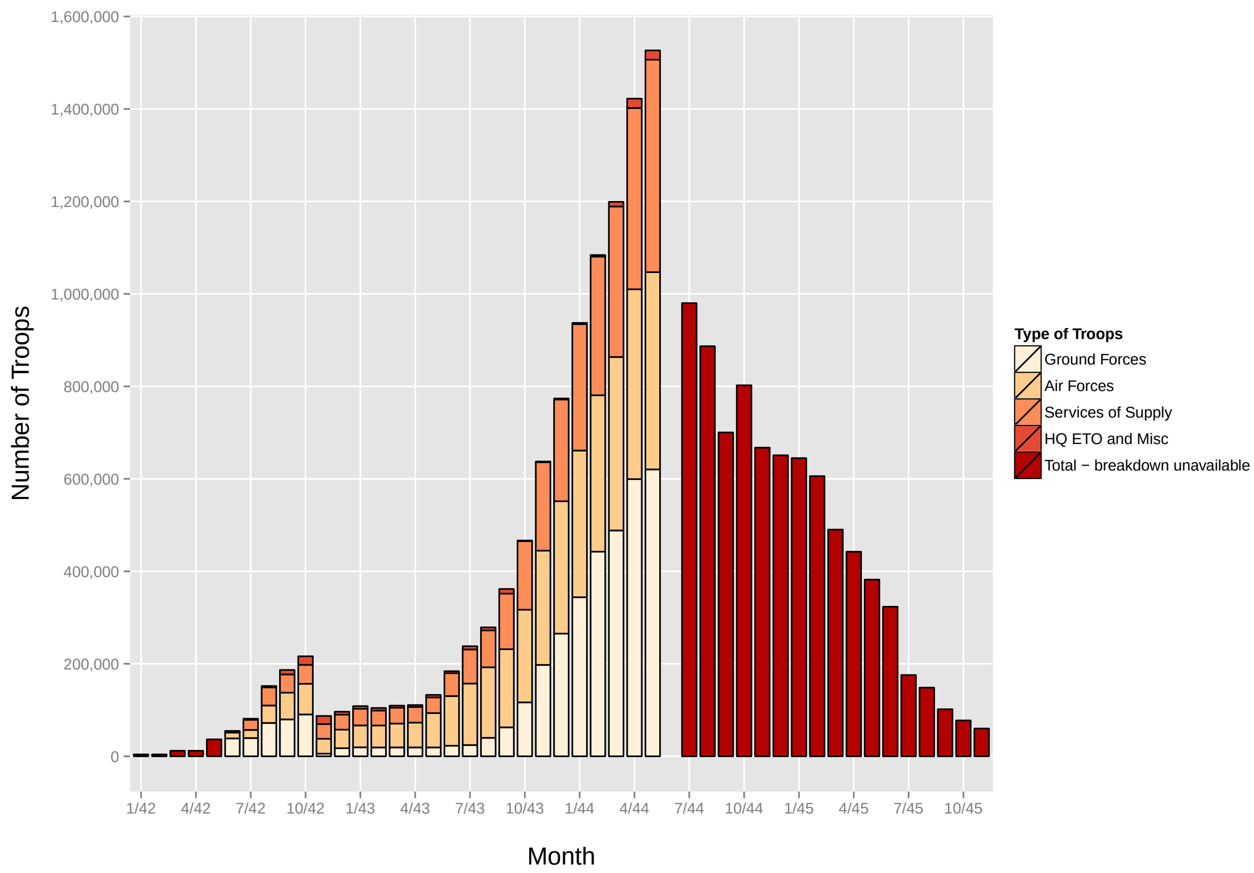


Figure 4.2: Build up of US Army strength in the UK

Figure shows US troop numbers in the UK from 01/1942 to 11/1945 and, where available a breakdown according to type. Data for 06/1944 is unavailable. Sources: Reynolds (1995, p. 410), Ruppenthal (1978, p. 232) and Pogue (1954, p. 541)

Around ten percent of GIs who served in the UK were black. African American soldiers had fought for the US Army and its predecessors since the American War of Independence. Over nine hundred thousand would serve during World War II, more than half of them overseas (Moore, 2013). As in previous wars, black soldiers served in racially segregated units, normally under command of white officers. Results from a pseudo-scientific aptitude test[1] were used to justify limiting these black units to non-combat 'labour' roles, most often supply, transport, food preparation and sanitation.

The British government had at first tried to discourage the US military from sending black troops to Britain. Ostensibly this was to avoid conflict between white GIs and British citizens, who might show "more effusiveness to the coloured people than the Americans would readily understand"[2], but concerns about sexual activity no doubt also played a role (Reynolds, 2006). The matter was taken up by Anthony Eden, the Foreign Secretary, and likely also by Churchill (Reynolds, 2006). Nonetheless, the government's request was refused, primarily because the need for labour units in Britain meant the use of black troops was a military necessity, but also out of domestic political concern. Suggestions were then made to limit black troops to port areas, where the UK's existent black population of around 8,000 people was concentrated, but these were also rejected. The policy instead was to "place them [black troops] where needed" (Rash, 1942).[3]

In practice, this meant black troops would serve throughout the country, in rural and urban areas alike. The influx of troops into a small country put huge strain on available accommodation; troops were stationed wherever space could be found[4]. Because of the support services they provided, black units had to be stationed in close proximity to combat (i.e. white) units to be useful. Nonetheless, the colour bar was maintained. Interaction between black and white soldiers was minimised: accommodation, dining and training facilities were all segregated. A 'pass system' was introduced in order to keep black and white troops apart during their leisure time, with black and white

[1]The Army General Classification Test, which tested for acquired skills and knowledge, and in which black soldiers scored on average substantially worse than their white peers, was often seen as proof of the innate incapacity of black soldiers to perform combat roles (Reynolds, 1995).

[2]E. Bridges to J. Martin, 21 July 1942, in Franklin D. Roosevelt Library, Hyde Park, New York, USA, Harry Hopkins papers, box 136.

[3]Units arriving in Britain were first assigned to one of several armies by the Headquarters of the European Theater. They would then be assigned to a base by their army's headquarters. Troops arriving in early shipment phases were mainly accommodated in newly constructed camps or ex-RAF or British Army quarters. Soldiers were also accommodated in private homes once the required accommodation started to outstrip supply. At first, this happened on a voluntary basis, but starting in late 1943 rooms in private homes were requisitioned. No more than ten percent of troops were billeted in private homes, and to the best of our knowledge black troops were never accommodated this way.

[4]Although Scotland was deemed unsuitable for military facilities and only small numbers of troops were ever stationed there (Lee, 1966, p. 623).

units allowed off base on different days of the week or assigned different venues to visit. It was during this leisure time that most contact between soldiers and local populations took place: GIs were frequently to be found in local bars, restaurants and dancehalls. GIs were often invited into locals' homes, and there are also frequent reports of younger women visiting troops on base, sometimes resulting in prosecutions for trespass. Officially black and white units had the same quota for leisure time, although some black soldiers complained about white southern officers restricting their passes (Smith, 1987, p. 134). The UK government was at pains to take no overt actions to enforce segregation, refusing for example to instruct police officers to recommend segregation to local bar and restaurant owners. There are however isolated examples of local authorities attempting to limit contact between black GIs and British women (Reynolds, 2006, p. 123).

In most areas where black GIs were stationed locals saw and interacted with black people for the first time.[5] Although some racial prejudice no doubt existed, the British on the whole warmly welcomed black GIs, whilst attitudes towards white soldiers were more mixed. Writing in Tribune, George Orwell remarked that "the general consensus of opinion seems to be that the only American soldiers with decent manners are the Negroes" and one countryman famously quipped, "I don't mind the Yanks, but I don't care much for the white fellows they've brought with them" (Olson, 2010, p. 287). A quarter of those providing information to 'Mass Observation', a social research organisation, said that they had become 'more pro-colour' as a result of their wartime experience (Smith, 1987, p. 123). According to a survey carried in November 1943, 80% of black troops had a 'favourable' opinion of the English, compared to 68% of whites, many of whom were unhappy that black GIs had been so warmly received by the local population (Smith, 1987, p. 134).

Most American ground troops left England in the course of Operation Overlord, the invasion of occupied Europe beginning on June 6th 1944. On the first day of the operation 150,000 troops landed in Northern France. By the time the operation ended in August 1944, 700,000 GIs were left in Britain, down from the June peak of one and a half million. Units continued to cross to Europe, but troop numbers in the UK did not decrease much further until the end of the war: the UK continued to serve as the headquarters of operations in Europe, as a base for Army Air Force units, as the point of entry for American troops bound for continental Europe and as the main location of

[5]General Eisenhower wrote to General Lee: "There is practically no coloured population in the British Isles. Undoubtedly a considerable association of colored troops with British white population, both men and women, will take place on a basis mutually acceptable to the individuals concerned." (September 5 1942, ETOUSA AG 291.2-B, available at `https://archive.org/stream/IndoctrinationOfPersonnelArrivingInTheUK/IndoctrinationOfPersonnelArrivingInTheUK_djvu.txt`)

military hospitals in Europe. However, by November 1945, almost all American units had left the United Kingdom.

4.2.2 Troop data

Data on the location of US Army units in England and Wales is sourced from monthly station lists produced by the US Army's Adjutant General's office, initially the "MRU United Kingdom Station Lists" (1942-1944) and subsequently the "MRU United Kingdom and Continental Station Lists" (1944 onwards). Each list provides an overview of all US Army units in the United Kingdom and (later) the wider European Theater at a given point in time.

The lists were created using data collected by the Army's Machine Records Units (MRUs), designated mobile units for collecting and processing personnel and troop information. Every day, large numbers of these mobile data-entry units punched 'morning reports' from each unit onto IBM punch cards, which were then transported to a central unit at the Adjutant General's office and formed the basis for the station lists. Around one hundred and fifty copies of the lists were distributed by the Adjutant General to various military commanders each month, listing all units stationed in the United Kingdom and their location at a snapshot in time, typically a few days before the production of the station list. Each entry in a station lists represents a unit. Units are identified with an abbreviation of their name (e.g. 1944 QM TRK CO for the 1944th Quartermaster Truck Company), and listed along with their current map coordinates[6], the nearest town (e.g. Watford), and a symbol to indicate if the unit contained black soldiers.

Most station lists for the period 1942-1953 survive and are housed at the US National Archives[7]. Sixteen station lists covering the United Kingdom have been digitised and shared with us by Captain Philip Grinton. The station lists cover sixteen months from June 1943 to December 1945[8] with a concentration in 1944, where troop numbers were highest.

[6]Map coordinates are provided in a coordinate system defined on the Cassini projection, for example WL5715 for Watford. The first two digits of the coordinate indicate an 100 by 100km square, the subsequent digits provide the northing and easting from the bottom left of that square, to an accuracy of 1 kilometer. We reproject coordinates to the British National Grid using a Cassini projection with false easting 500000, false northing 100000, central meridian -1.19276, scale factor 1.0 and latitude of origin 50.617708.

[7]Record Group 407, HMS Entry Number NM 3 377 A. Catalog entry at `http://research.archives.gov/description/6883370`.

[8]June 1943, August 1943, September 1943, November 1943, December 1943, February 1944, March 1944, April 1944, May 1944, June 1944, August 1944, October 1944, November 1944, December 1944, May 1945, December 1945.

Using these digitised station lists, we create a dataset of all 1,576 military bases/-camps (unique according to their coordinates) where we know personnel were stationed. Figure 4.3 shows the distribution of these locations across England and Wales. The size of the rectangle represents the frequency at which a location is mentioned in the station lists, which we later call the number of 'unit-months' and use as our measure of overall military presence in a location. The shade of the rectangles indicates the share of units stationed at a location which 'were black'; a completely black rectangle indicates that all mentions of a location in the station lists are for black units, a completely white square that black units were never posted in the location. It is clear that there is considerable geographic variation across locations in the share of black units stationed; it is this variation that our identification strategy exploits.

There are two obvious geographic patterns in Figure 4.3. Firstly, the absence of units in the South East of the England. This was a consequence of the plan for the invasion of occupied France: US vessels left from the west of the England, UK troops from the east, and landed in France on French beaches in the same formation. The second obvious feature is the large concentration of white units in East Anglia. This was the area used to launch bombing raids by the Army Air Forces, which had a lower proportion of black soldiers than the army overall (Army Air Force, 1945, p. 22). Figure 4.3 also shows a grid that we will later use to control for geographical fixed-effects to account for such geographical idiosyncrasies.

4.3 General Estimation Framework

We use two independent data sets with high geographical resolution to test the hypothesis that racial attitudes were affected by contact with black GIs. The first of these provides data on BNP party membership across all of England and Wales' 180,000 census output areas ('neighbourhoods'), the lowest level geography on which the national statistics agency collects data. The second is an extract from the data generated by 'Project Implicit', a website on which individuals can carry out a test for their implicit racial attitudes. As well as carrying out the test, individuals are prompted to provide self-reported racial attitudes and demographic information including their postcode. We describe the datasets in more detail in sections 4.4 and 4.5, but first lay down the shared estimation strategy that we apply to both datasets.

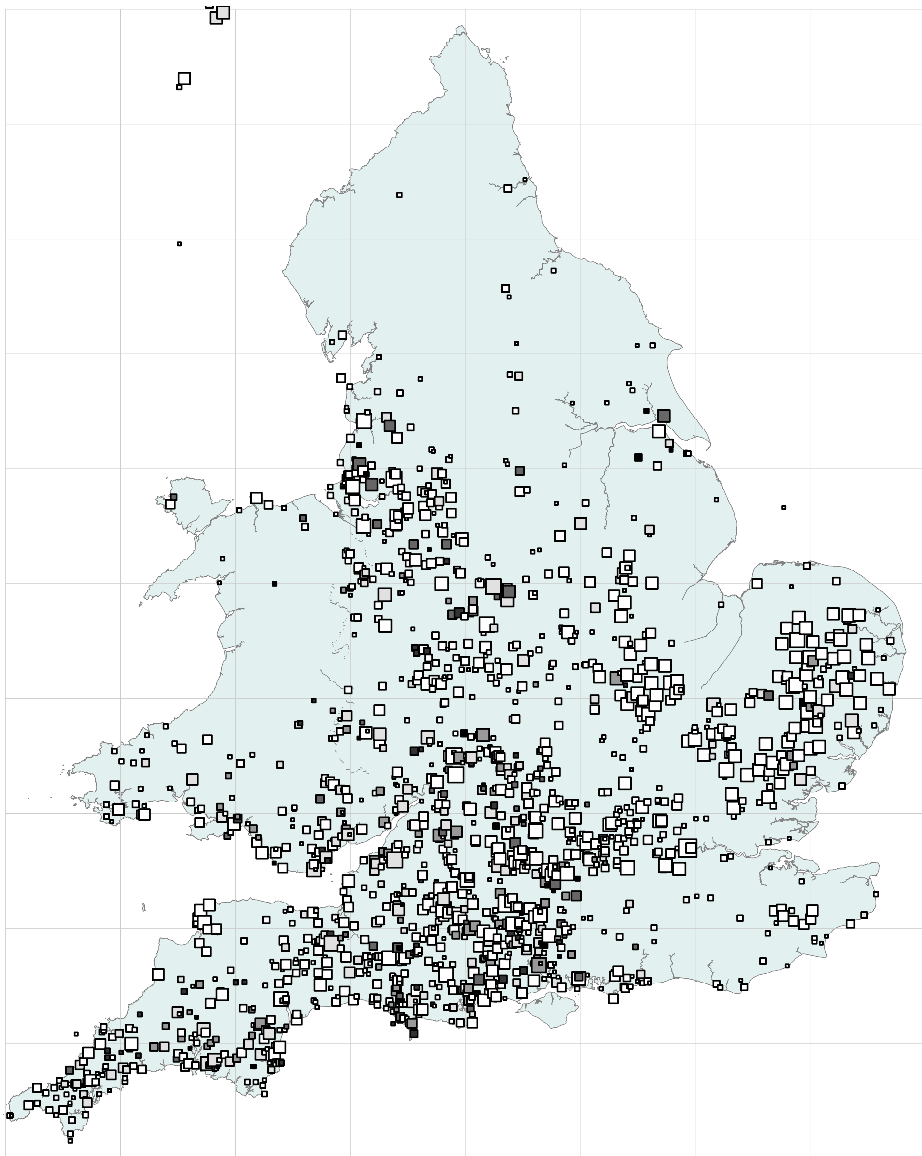


Figure 4.3: Troop Locations across England and Wales

This map shows locations of US Army troops during World War II. The size of the rectangle repesents the frequency at which a location is mentioned in the station lists. The shade of the rectangle indicates the share of black units at this location.

4.3.1 Population of interest

Our population of interest consists of contemporary populations living in locations which were close enough to US military bases that contact with troops was relatively likely. We know the precise locations of US military bases and all our potential outcome variables are observed at ultra fine geographical resolutions (census output areas or postcode). We must therefore define what it is for an area to be 'close' to a base.

Rather than simple geodesic distance, our preferred specification uses a definition of closeness which takes into account variations in population density. We argue that the probability of contact with troops stationed at a distance of k kilometers was lower in more densely populated areas: consider units posted in a rural area with a pub serving a wide geographic area compared to a city with pubs serving a hyper-local clientele. In order to capture this logic, we define a neighbourhood and base as being close if they lie within a common postcode district (N = 2,261). Postcode districts are more-or-less arbitrary districts defined by Royal Mail for mail sorting and delivery purposes; they are ideal for our purposes as they cover very small areas in cities but somewhat larger areas in urban areas (65% of the variance in log-area is explained by log-density), with a median size of 27 square kilometers. Using this definition of close, our population of interest consists of individuals living in postcode districts where one or more bases were located. Our treatment measure will also be defined on the postcode district level. We explore alternative methods of matching military bases to observations of outcomes in the appendix.

4.3.2 Treatment Definition

With random interactions, the probability of individuals coming into contact with troops increases in the number of troops stationed nearby and the length of time they were stationed for, something that our treatment measure should reflect. To achieve this we sum up the number of units stationed in each postcode district, both across locations and through time.[9]. When we sum only black units we generate our main treatment variable, 'black unit-months', and when we sum over all units we generate a control variable 'unit-months', which reflects the overall intensity of presence of US military forces. Formally,

$$\text{Unit-months in postcode district}_j = \sum_t \sum_b \text{Units}_{b,t}$$

[9] According to personal conversations with staff at the National Archives in College Park, Maryland, information on the number of soldiers in each unit was destroyed at the end of the war. We are therefore unable to weight units by the number of personnel they contained.

and

$$\text{Black unit-months in postcode district}_j = \sum_t \sum_b \text{BlackUnits}_{b,t}$$

where j is a postcode district, t ranges over time periods, b ranges all over bases in postcode district j and $\text{Units}_{b,t}$ ($\text{BlackUnits}_{b,t}$) is the number of units (black units) stationed at base b at time t.

'Unit-months' therefore measures measures the overall strength military presence in a postcode district, which we shall always control for, and 'black unit-months' the probability of contact with black GIs.

4.3.3 Identification Strategy

Our identification strategy requires that the number of 'black unit-months' be exogenous to anything else correlated with contemporary racial attitudes, controlling for the overall number of 'unit-months'. In other words, given a certain size of base, it was as good as random how many units stationed there were black and how many were white.

Our estimation results would be biased if black units were strategically stationed in areas with particular racial attitudes. There is no evidence for this being the case. Black soldiers served almost exclusively in segregated units with support (non-combat) roles: quartermaster units, laundry units, transport units etc[10]. For these units to carry out their roles effectively, they needed to be posted close to combat (i.e. white) units. Strategic behaviour would have required both precise knowledge of the variation in racist attitudes across geography and the willingness to accomodate these preferences when solving the optimal allocation of troops problem. Such an exercise would have been costly, without any obvious payoff, and would surely have left documentary evidence if carried out. As discussed in the introduction, there were some early suggestions to limit black troops to port areas (where Britain's tiny existing black population was concentrated), but these were rapidly dismissed, as indicated by a 1942 internal memo from the Services of Supply Headquarters, "The policy has been defined to place them [black units] where needed."

[10]The prevailing attitude among senior military figures at the time is summed up in a letter General Patton wrote to his wife: "A colored soldier cannot think fast enough to fight in armor." (Sasser, 2014, p. 104).

4.4 BNP Membership

We begin by proxying racial attitudes using data from Biggs and Knauss (2011), who geolocate members of the far-right British National Party (BNP) using a membership list published online in 2008[11]. The list was confirmed to be genuine and is understood to provide a complete listing of members of the party in November/December 2007, although the BNP has claimed that the list contains a number of ex- and prospective members too.

The British National Party was founded in 1982 as a splinter group from the National Front, an openly racist organization with links to European neo-Nazis. The party's founder was jailed for conspiracy to incite to racial hatred in 1986, and a senior official described the party as "100 per cent racist" in 1995 (BBC, 2001). Police officers and prison officials are banned by their employers from joining the party, and in 2009 a government minister attempted to introduce a similar ban for teachers, citing a desire to "keep racism . . . out of our schools" (The Guardian, 2009).

Attempting to increase its electoral relevance, the party began to outwardly reject claims of racism in 1999, but its ideology is still widely considered to be just that (The Spectator, 2009). Most tellingly, non-white members were banned from the party until this was deemed illegal by a court in 2010. Even whilst espousing a 'modernization' agenda, the party's leader from 1999 (who had been convicted for distributing material likely to incite racial hatred) continued to call for the "repatriation" of non-white Britons, who a party manual referred to as "racial foreigners".

We argue that party membership provides an excellent indication of racist attitudes. Members provide financial support to the party through yearly fees, they receive a membership card, regular party communiques and are invited to attend party meetings and events. So although there is no survey evidence on individual motivations for joining the party, membership likely indicates a high degree of ideological compatibility with the party. Membership is not contingent on the presence of a local branch; at the time the membership list was publicised, individuals could join the party by completing a paper or online form. Doing so, we argue, provides evidence of willingness to pay to support a party with extreme views on race.

The membership list comprises 13,009 individuals; 12,563 (97%) of the entries contain a home address with a valid UK postcode. There are over one million unique postcodes throughout the UK, so this allows BNP members to be geolocated extremely accurately. Using an official lookup table, Biggs and Knauss aggregate the data to

[11] Due to legal constraints in Germany, we are not able to process this information directly from the original, leaked membership list.

the 2001 'Output Area' level, the lowest level geography on which the UK statistical agency aggregates demographic and social data, and report the number of members within each area. The authors kindly shared their dataset. With the help of another lookup table, we match the data to the most recent output area definitions, giving data on the universe of the 181,408 '2011 Output Areas' across England and Wales, which we refer to as neighbourhoods. This provides a significantly finer level of geographic observation than would be the case using voting data, which is available only for ten regions (European Elections) or 573 constituencies (general elections). Of the 184,109 neighbourhoods in England and Wales (median population: 303), 12,513 (6.7%) include at least one BNP member. The maximum number of members per neighbourhood is eleven, an output area in Barnsley.

4.4.1 Results

In this section, we show the results of regressions with BNP members per 10,000 white residents as the dependent variable. The denominator arises since only whites were permitted to join the party at the time the membership list was published.

Our outcome is measured at neighbourhood level (N = 181,408), so we run our main regression at the neighbourhood level too, without further aggregation. Using neighbourhoods as the unit of observation allows us to accurately match each unit to larger historical geographies in order to generate control variables: neighbourhoods almost always fit completely inside historic boundaries, something which would not be the case for larger geographic units. In order to account for correlation in the error term between observations, we cluster standard errors at the modern local authority level, which divides England and Wales into 348 regions. Aggregating neighbourhoods to higher geographies or changing the clustering level does little to change the pattern of results (tables not reported).

As described in section 4.3.2, we consider a neighbourhood as potentially 'treated' if it lies within a postcode district where troops were posted. Table 4.1 shows some summary statistics. There were bases within the same postcode district as around a third of England and Wales' 181,408 neighbourhoods: these are the 60,028 neighbourhoods which will enter our sample. Around a third again (25,321) had some potential for contact with black troops. We observe troop locations in sixteen months between 1943 and 1945; on average troops were close to the effected neighbourhoods for seven of these. There are a mean of 2.58 BNP members per 10,000 white residents across all neighbourhoods in England Wales, and a mean of 2.31 close to US bases. It is apparent from the table that there are fewer BNP members in locations where black troops were once present, but also that black troops were more likely to have been present in areas

Table 4.1: Summary Statistics, BNP Membership

	Mean (absolute)	Mean if base in postcode district	Mean if black troops ever in postcode district
Population	309	305	302
Area (hectares)	83	126	153
In neighbourhood's postcode district:			
US military base	33%		
Black troops ever present	14%	41%	
Months with units stationed (max 16)	2.6	7.7	10.7
Months with black units stationed (max 16)	0.45	1.35	3.3
'Unit-months'	31	93	160
'Black unit-months'	1.26	3.8	9.15
BNP Members	0.067	0.062	0.053
BNP Members per 10,000 white inhabitants	2.58	2.31	1.9
N	181,408	60,028	25,321

Summary statistics for several measures, unconditional, conditional on ever having troops present and conditional on ever having black troops present. 'Unit-months' and 'Black unit-months' are the possible extent of contact with any troops and black troops, respectively.

with a larger military presence. This is entirely consistent with it being as good as random if any single unit posted was black or white. In order to control for this we turn to a regression framework, where we are also able to control for other variables that might plausibly have affected the allocation of black troops.

Our main results are shown in Table 4.2. Each column represents a simple OLS regression of the form:

$$\text{BNP members per 10,000 whites}_{ij} = \beta_0 + \beta_1 \text{UnitMonths}_j + \beta_2 \text{BlackUnitMonths}_j + X_i + u_i$$

where i represents a neighbourhood, j that neighbourhood's postcode district, UnitMonths_j and BlackUnitMonths_j are constructed as per section 4.3.2, X_i is a vector of neighbourhood level-controls and u_i is the error term. Regressions employ standard errors clustered on modern local authority level in order to account for the spatial correlation of the error term.

Column (1) includes no additional controls. We introduce further controls in a piecewise fashion, starting with column (2). First we impose an arbitrary eight by ten grid on the map of England and Wales and generate a dummy variable for each of these grid cells. The inclusion of these geographic fixed effects means only variation within each grid cell is used to estimate our effect, comparing neighbourhoods which

Table 4.2: Baseline Regressions

	(1)	(2)	(3)	(4)	(5)
Within postcode district:					
Black unit-months	-0.0199***	-0.0178***	-0.0170***	-0.0183***	-0.0182***
	(-4.49)	(-3.99)	(-3.69)	(-3.80)	(-3.88)
Unit-months (100s)	-0.108*	-0.0597	-0.0553	-0.0468	-0.0454
	(-1.82)	(-0.94)	(-0.87)	(-0.79)	(-0.76)
F-stat joint significance tests on:					
Grid cells		2624.8	26.99	14.11	13.91
Industrial Sectors (1931)			0.128	0.618	0.577
Population Density (1931)			0.121	0.305	0.139
Urban Rural status (1931)			0.263	0.0378	0.0324
Distances			0.748	0.816	0.754
Qualifications (2011)				2.304	2.390
Unemployment (2011)				0.0250	0.0597
Housing Tenure (2011)				1.996	1.899
Ethnic Backgrounds (2011)					0.180
Clusters	278	278	278	278	278
Observations	60028	60028	59994	59994	59994

Coefficients from OLS regressions. The unit of observation is the neighbourhood (2011 census output area). Outcome is BNP members per 10,000 white inhabitants. Independent variables are our measure for contact with black troops 'Black unit-months' and any troops 'unit-months' (divided by one hundred) respectively. Standard errors are clustered at the local authority district level. T statistics in brackets. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

are relatively close to one another but vary in terms of the number of black units posted nearby. This is a more flexible way to capture variables with regional variation than latitude/longitude controls.

In column (3), we match neighbourhoods to two historic geographies: 1931 local government districts (N = 1,800) and 1931 parishes (N = 14,267), using boundary data provided by the Vision of Britain Project. In most cases, neighbourhoods are completely contained within these geographies; if not, they are matched to the geography which contains the neighbourhood's population weighted centroid. Having matched neighbourhoods to these historic geographies, we generate control variables using data from the UK's 1931 census; population density (parish level), urban/rural status (local government district level) and working age population by occupational sector (agriculture, light industry, staple industry, professional, all at local government district level). We calculate the geodesic distance from each neighbourhood's population weighted centroid to the nearest historic local government district with urban status, to the nearest town/city with a 1939 population over 100,000 (N = 55), to the nearest city with a 1939 population over 300,000 (N = 8) and to the coast. None of these controls enter significantly into the regression.

In column (4), we include an extensive range of neighbourhood level socio-economic status variables from the 2011 UK census. Although these are 'bad controls' (Angrist and Pischke, 2008, p. 64), in the sense that they are measured after treatment and are thus potential outcomes, they should serve to pick up local characteristics which have persisted through time. We include variables which have been shown by Biggs and Knauss (2011) to correlate with BNP membership, calculating the proportion of individuals in each of the following categories: aged 30-65, without qualifications, with a university degree, in each of seven social classes, living in a home they own, living in a house rented from the council, living in an overcrowded house, unemployed. We calculate the non-white proportion on both the neighbourhood level and the higher local authority level. In column (5), we also add variables designed to capture contemporary ethnic compositions to the regression. Again following Biggs and Knauss (2011), we calculate the 'minority' population share (defined as non-white). We do the same at the higher local authority level (N = 348), where we also calculate a measure of segregation:

$$\frac{1}{2}\sum_{i=1}^{n}\left|\frac{w_i}{\sum w_i}-\frac{m_i}{\sum m_i}\right|$$

where i ranges over the n neighbourhoods in a local authority, w_i and m_i refer respectively to the white and non-white proportion of neighbourhood i. The measure, bounded between 0 and 1, can be interpreted as the portion of minorities that would need to move neighbourhoods in order for the non-white/white distribution to be geographically even through the local authority. The point estimate remains stable across all specifications and of similar magnitude.

In Table 4.3, we examine robustness to alternative treatment indicators. We progressively add controls in the same fashion as in table 4.2. In the top panel, we use a binary variable as our main regressor of interest; this is one if black units were ever posted in a neighbourhood's postcode district and zero otherwise. In the middle panel, we count the number of months (maximum sixteen) which black units were present for. The third panel reports the results of a 'horse race', including all indicators. Our preferred measure, 'black unit-months' stays highly statistically significant when using historical controls. It stays significant in all specifications, unlike our alternative treatment indicators. While the number of months in which black units were stationed is never significant, the binary indicator of black troops ever being present loses significance as we add further controls. Taken together, the table suggests that both the number of black troops and their duration of stay matter for contemporary outcomes, and that our outcome measure does a good job of picking up both of these factors.

To ensure that our results are not driven by regional idiosyncrasies, table 4.4 reports results on selected samples. In all cases, we use the same set of controls as in column (3) of Table 4.2: including geographic and pre-treatment variables, but no contemporary controls. First, we split the sample according to the East-West coordinate of the neighbourhood. Encouragingly, our coefficient of interest, that on 'Black unit-months', is very similar in the west (column 1) and the east of Britain (column 2). In column (3) we drop from the sample any neighbourhoods within twenty kilometers of the coast. This excludes all locations in which the small pre-war black population of England and Wales was concentrated. Results remain highly significant. Finally, we exclude London from the sample, column (4). Across all our specifications, 'Black unit-months' remains highly significant. We interpret this as evidence for the effect not being driven by regional outliers, but that the effect is consistently present throughout England and Wales.

All regressions up to this point have used a linear probability model. In Table 4.5, we show that our results are independent of alternative model specifications. In column (1), we transform our key variables by taking the natural logarithm (plus one, to deal with zeros). The presence of troops is skewed across locations, and a logarithmic specification will take this into consideration. In column (2), we dichotomise BNP membership and run a probit regression on that dependent variable. Spillovers (i.e. peer effects) within areas might cause BNP membership to be inflated. Consider neighbours where both are racist and one is a member of the BNP. Then, the non-member might feel that it is socially more acceptable to become a BNP member as

Table 4.3: Alternative Treatment Measures

	(1)	(2)	(3)	(4)	(5)
Within postcode district:					
Black-units ever stationed	-0.593***	-0.334***	-0.322***	-0.330***	-0.321***
	(-4.24)	(-2.68)	(-2.65)	(-2.79)	(-2.73)
Unit-months (100s)	-0.0994*	-0.0792	-0.0739	-0.0683	-0.0675
	(-1.83)	(-1.33)	(-1.25)	(-1.23)	(-1.21)
Within postcode district:					
Months with black units	-0.108***	-0.0736***	-0.0694***	-0.0804***	-0.0801***
	(-4.58)	(-3.31)	(-3.04)	(-3.65)	(-3.69)
Unit-months (100s)	-0.0895	-0.0617	-0.0571	-0.0447	-0.0433
	(-1.49)	(-0.95)	(-0.87)	(-0.73)	(-0.71)
Within postcode district:					
Black units ever stationed	-0.529***	-0.312**	-0.316**	-0.265*	-0.256
	(-3.14)	(-2.00)	(-2.05)	(-1.72)	(-1.64)
Black unit-months	-0.0124*	-0.0204***	-0.0206***	-0.0174**	-0.0174**
	(-1.80)	(-2.72)	(-2.83)	(-2.08)	(-2.07)
Months with black units	0.00716	0.0397	0.0450	0.0162	0.0157
	(0.18)	(0.95)	(1.10)	(0.37)	(0.35)
Unit-months (100s)	-0.0762	-0.0485	-0.0456	-0.0351	-0.0343
	(-1.34)	(-0.78)	(-0.73)	(-0.60)	(-0.59)
Controls:					
Grid cells	N	Y	Y	Y	Y
Industrial Sectors (1931)	N	N	Y	Y	Y
Population Density (1931)	N	N	Y	Y	Y
Urban Rural status (1931)	N	N	Y	Y	Y
Distances	N	N	Y	Y	Y
Qualifications (2011)	N	N	N	Y	Y
Unemployment (2011)	N	N	N	Y	Y
Housing Tenure (2011)	N	N	N	Y	Y
Ethnic Backgrounds (2011)	N	N	N	N	Y
Clusters	278	278	278	278	278
Observations	60028	60028	59994	59994	59994

Coefficients from OLS regressions. The unit of observation is the neighbourhood (2011 census output area). Outcome is BNP members per 10,000 white inhabitants. Independent variables are a dummy for whether black units ever being present (panel I), the number of months black units were present (panel II) and our measure for contact with black troops ‘Black unit-months’ (panel III). The measure for contact with any troops ‘unit-months’ is used throughout (military units are measured in hundreds). Standard errors are clustered at the local authority district level. T statistics in brackets. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

Table 4.4: Subsample Analyses

	(1) East	(2) West	(3) Inland	(4) Not London
Within postcode district:				
Black unit-months	-0.0146***	-0.0158**	-0.0244***	-0.0156***
	(-2.63)	(-2.41)	(-4.63)	(-3.50)
Unit Months (100s)	0.0369	-0.138**	0.00917	-0.0730
	(0.51)	(-2.16)	(0.12)	(-1.14)
F-stat joint significance tests on:				
Grid cells	61.68	3.252	108.5	28.73
Industrial Sectors (1931)	0.424	0.463	0.544	0.136
Population Density (1931)	0.817	1.363	2.519	0.216
Urban Rural status (1931)	0.000149	0.0415	0.0428	0.0314
Distances	3.266	1.356	0.646	1.204
Clusters	104	186	208	264
Observations	27491	32503	34141	58551

Coefficients from OLS regressions. Column (1) and (2) only use the west and east of England, column (3) restricts the sample to inland locations and column (4) excludes London. The unit of observation is the neighbourhood (2011 census output area). Outcome is BNP members per 10,000 white inhabitants. Independent variables are our measure for contact with black troops 'Black unit-months' and any troops 'unit-months' (divided by one hundred) respectively. Standard errors are clustered at the local authority district level. T statistics in brackets. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

Table 4.5: Alternate Models

	(1) Log specification	(2) Probit	(3) Negative Binomial
Within postcode district:			
Black unit-months	-0.141**	-0.00462***	-0.0106***
	(-2.36)	(-3.82)	(-3.80)
Unit Months (100s)	-0.0811	-0.00729	-0.0266
	(-1.34)	(-0.53)	(-0.79)
F-stat joint significance tests on:			
Grid cells	28.71	1384.4	2358.3
Industrial Sectors (1931)	0.130	0.951	0.753
Population Density (1931)	0.0287	4.615	0.134
Urban Rural status (1931)	0.129	0.389	0.527
Distances	0.752	3.910	2.388
Clusters	278	278	278
Observations	59994	59992	59994

Coefficients from several regressions. The unit of observation is the neighbourhood (2011 census output area). Outcome is BNP members per 10,000 white inhabitants, except for column (2) where it is a dummy whether the neighbourhood conatins any BNP members. Independent variables are our measure for contact with black troops 'Black unit-months' and any troops 'unit-months' (divided by one hundred) respectively, apart from column one, where the logarithm (plus one) of the measures has been applied. Standard errors are clustered at the local authority district level. T statistics in brackets. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

well than in an environment where no neighbours are BNP members. The probit approach takes care of these spillovers by reducing the variance in the outcome measure. In column (3), we run a negative binomial regression on the absolute number of BNP members per neighbourhood, constraining the log number of white residents to one in the regression. This model is appropriate for a count outcome containing many zeros, such as membership data for a small party. The coefficients are not comparable across specifications, but all coefficients on 'Black unit-months' are in the expected direction and statistically significant.

4.5 Implicit and Explicit Attitudes

In this section, we repeat the analysis on two different potential outcome measures from an entirely different sample, warmness of feelings towards black people ('thermology') and implicit racial biases. We make use of demographic data provided by respondents to examine mechanisms of persistence.

The data comes from Project Implicit, an American non-profit organization which hosts several country-specific websites allowing users to test their implicit attitudes using various Implicit Association Tests ('IATs'). IATs are provided on a range of subjects (e.g. gender, sexuality, weight), but by far the most popular test is that for race. After completing the test, subjects are asked to answer several questions regarding their attitudes towards religion, minorities, politics and supply their postcode and general demographics. About half of UK residents taking the IAT test do so on the US website. Following IRB approval, we received a dataset containing all data collected by the UK and US IAT websites between 2004 and 2013[12].

The dataset from the UK website contains around 240,000 started sessions, 90,000 of which were seen through to completion. Of these, valid postcode data is provided in 25,826 sessions. We combine these with 19,582 observations from the US website to complete our sample of 45,408 sessions. We make no claims about the representativeness of the sample: the median age is 29 years, compared to 40 for the population, and roughly two thirds of the participants are male. Subjects are also better educated than the population as a whole. Nonetheless, the sample contains individuals from throughout the UK (all 348 of England and Wales' local government districts are represented) and of a wide range of ages (13 to 89).

Curiosity seems to be the main driver for participation in the test. The most common stated reasons for coming to the website are the recommendation of a friend

[12] A public version of the US dataset, without zip codes, is available via the Open Science Framework at `https://osf.io/y9hiq/`

and mention in news articles. Around 20% of tests in the dataset were conducted in the days after the BBC publicised the test on its news website in April 2005 under the headline "Are you racist? The test that claims to know" (BBC, 2005).

4.5.1 Results: Thermology

We follow the same basic empirical strategy as in section 4.4, but now run individual-level regressions on responses to the question "Please rate how warm or cold you feel toward the following groups (0 = coldest feelings, 5 = neutral, 10 = warmest feelings)", the target group being black people (or, if the test is being taken on the US website, 'African Americans'). As per section 4.4, treatment intensity is measured by the number of 'black unit-months' in the respondent's postcode district. Throughout this section, we use the same set of controls as in our preferred section 4.4 specification, including a full set of pre-treatment controls and geographic fixed effects but no post-treatment controls. We control for basic demographic information provided by subjects (age, age squared and a gender dummy) in all regressions.

In line with the findings on BNP membership in the previous section, we find that individuals living in areas with more black soldiers report warmer feelings (a higher number) towards black people, even if the magnitude of the effect is somewhat small. Interestingly, the treatment effect is no stronger on whites than on the entire population, column (2). As per the regressions on BNP membership, controlling for contemporary ethnic compositions does not do much to change the point estimate (column 3); it is therefore not the case that persistence is driven by selective non-white migration to treated areas. In column (4), we exclude individuals born prior to 1947 from the sample and again the estimate is hardly changed. Individuals born after this date could not have come into direct contact with black troops, so any treatment effect must be the result of the transmission of values. In column (5), we use responses to the question 'what is the postcode in which you have spent most of your life' to limit the subsample to individuals who are now living in treated areas, despite having spent most of their life in areas where no troops were posted. The coefficient remains statistically significant, and is (insignificantly) larger in magnitude. Although not a perfect test, this provides evidence that the transmission mechanism driving persistence is not vertical (from parents to children), but horizontal (from the neighbourhood).

4.5.2 Results: Implicit Attitudes

Greenwald et al. (1998) developed the Implicit Association Test (IAT) as a way of measuring implicit attitudes. The test, which is computer administered, can be ap-

Table 4.6: Thermology Results

	(1) All	(2) Whites	(3) All	(4) DoB > 1947	(5) Movers
Within postcode district:					
Black Unit Months	0.00353**	0.00322*	0.00341**	0.00304**	0.00441**
	(2.44)	(1.91)	(2.24)	(2.12)	(2.40)
Unit Months (100s)	-0.0178***	-0.0152***	-0.0165***	-0.0181***	-0.0224***
	(-4.34)	(-3.22)	(-3.89)	(-4.33)	(-3.34)
F-stat joint significance tests on:					
Demographics	80.50	73.49	79.34	77.25	36.90
Grid cells	23.57	72.00	20.05	24.46	34.11
Industrial Sectors (1931)	0.124	0.415	0.131	0.172	1.110
Population Density (1931)	1.795	0.0828	0.650	2.487	3.185
Urban Rural status (1931)	2.740	0.845	1.797	2.433	0.762
Distances	1.493	1.648	1.644	1.749	1.163
Ethnic Backgrounds (2011)			1.832		
Clusters	273	273	273	273	268
Observations	12866	11611	12866	12486	8617

Coefficients from OLS regressions. The unit of observation is the individual. Dependent variable is the self-reported measure of attitudes towards black people. Independent variables are our measure for contact with black troops 'Black unit-months' and any troops 'unit-months' (divided by one hundred) respectively. All regressions control for age, age squared and gender of subjects. Column (2) restricts the sample to white people only, column (4) restricts the sample to those born after 1947 and (5) restricts the sample to people who have spent most of their life time at a different postcode. T statistics in brackets. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

plied to measure implicit associations on any topic, but the most common application has been to race. A consistent finding (at least in the US) is of a strong automatic preference towards white people, although there is considerable variation in race IAT scores between individuals.

The race IAT consists of five steps. In each step subjects have to assign a ‘stimulus’ (a word or a picture of a face) to one of two groups by pressing keys on their keyboard. In step one, subjects sort pictures of black and white faces into the categories ‘black’ and ‘white’. In step two, subject sort words into two categories, ‘positive’ or ‘negative’. The words are all easy to categorise (e.g. “terrible”, “joy”, “hurt” and “peace”). In step three, the tasks are combined – for example asking participants to assign black faces and positive words to one category, white faces and negative words to the other. In step four participants again assign faces to categories. Step five is akin to step three, in that participants have to sort both faces and words, but this time the groupings are reversed. So the participant might now have to assign black faces and *negative* words to one category, white faces and positive words to the other. If individuals require more cognitive effort to pair a) black faces and positive words and b) white faces and negative words than a) black faces and negative words and b) white faces and positive words, their response times will vary between blocks three and five. This is measured by the “D” score, calculated by comparing response times between step three and step five, normalised by the standard deviation of response times in all steps. Higher scores indicate a stronger automatic association of positive words with white faces, a score over zero indicates implicit racial bias toward whites (Greenwald et al., 1998). In our dataset, 98% of observations lie in the range -0.73 to +0.977. The mean score for those self-reporting as white is 0.37, and for those self-reporting as black as 0.15.

In table 4.7 we test whether contact with black GIs has had persistent impact on IAT scores. Controls are introduced piecewise, as in Table 4.2. We find significant effects in none of the specifications. The point estimates are close to zero, with relatively small confidence intervals. We can therefore reject the hypothesis that contact had meaningfuly sized persistent effects on implicit attitudes.

4.6 Conclusion

In this chapter we used a unique episode in British history to carry out a rigorous test of the intergroup contact hypothesis, a psychological theory that dates back to the 1950s and predicts that prejudices about other groups can be reduced by contact among groups. While many research articles in the psychology literature have carried out experiments and survey studies to validate the theory, we are the first to tackle this

Table 4.7: IAT Results

	(1)	(2)	(3)	(4)	(5)
Within postcode district:					
Black Unit Months	0.000236	0.0000515	0.0000273	0.0000862	0.0000389
	(0.67)	(0.14)	(0.07)	(0.22)	(0.10)
Unit Months (100s)	-0.0000637	0.000119	0.000567	0.000374	0.000128
	(-0.06)	(0.10)	(0.45)	(0.28)	(0.09)
F-stat joint significance tests on:					
Grid cells		6846.0	166.3	93.87	86.52
Industrial Sectors (1931)			1.505	1.504	1.421
Population Density (1931)			0.847	1.635	0.183
Urban Rural status (1931)			3.896	1.526	1.541
Distances			2.122	1.387	1.262
Qualifications (2011)				0.403	0.814
Unemployment (2011)				0.366	0.779
Housing Tenure (2011)				2.414	1.387
Ethnic Backgrounds (2011)					5.861
Clusters	273	273	273	273	273
Observations	13295	13295	13284	13284	13284

Coefficients from OLS regressions. The unit of observation is the individual. Dependent variable is the measure of implicit racist attitudes from the IAT. Independent variables are our measure for contact with black troops 'Black unit-months' and any troops 'unit-months' (divided by one hundred) respectively. All regressions control for age, age squared and gender of subjects. T statistics in brackets. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

test from an economist's perspective. Our identification is unique in providing quasi-experimental evidence from a large-scale historical episode. Furthermore, we do not solely rely on survey measures but analyze changes in preferences also by changes in observed choices. In addition, we show that the effect of intergroup contact is persistent even after several decades.

When the US Army arrived in England in the early 1940s to prepare for the invasion of occupied Europe, troops contained around 150,000 African American soldiers. Those black soldiers served almost exclusively in segregated black-only units with mainly supportive functions. Deployment of troops was made according according to military needs and hence orthogonal to existing racial attitudes.

Based on the geographical variation in the probability of encountering black soldiers, we can estimate the effect of intergroup contact on contemporary racial attitudes. We use membership data from the British National Party, a far-right party, widely considered to be extremist, as our main outcome variable. We find a persistent effect, in that BNP membership is lower in areas where black soldiers were stationed in the 1940s. Using several controls and robustness checks we show that the effect is not driven by outliers, geographical factors and does not hinge on the model specification.

We also contribute to the literature on the formation and persistence on preferences in economics. Using data from a large online survey (and hence a different sample) we show that self-reported attitudes also vary in the same fashion. These attitudes, as our results cautiously suggest, are not passed down vertically across generations, but are rather transmitted horizontally via the living environment.

The same dataset allows us to draw conclusions about implicit attitudes. We do not find an effect of intergroup contact on contemporary implicit attitudes. Our interpretation for this finding is that implicit attitudes presumably do not follow the same pattern of transmission as explicit attitudes.

This is especially striking as there exist geographical patterns in implicit attitudes that seem to be persistent across time. Further research is needed in determining the transmission mechanism of implicit attitudes.

Chapter 5

Dynamics in Gun Ownership and Crime – Evidence from the Aftermath of Sandy Hook*

5.1 Introduction

Gun control has been a polarizing topic in United States politics over the past decades. The debate between gun rights and gun control advocates is fiercely fought, often relying on only a small set of arguments. Supporters of the right to possess firearms, as well as some conservative politicians often argue that arming citizens and abolishing gun-free zones will lead to decreases in violent crime. In 2012, a few days after the shooting at Sandy Hook Elementary School in Newtown, Connecticut, NRA CEO Wayne LaPierre used the phrase "The only thing that stops a bad guy with a gun, is a good guy with a gun." (New York Times, 2012), to emphasize the importance of arming civilians to deter crime. To back up their argument, gun rights proponents usually argue that while gun ownership has risen over the last decades, violent crimes have dipped. On the other hand, gun control activists, including many liberal politicians, point to the high numbers of violent crimes in the United States, in particular homicides committed with guns, and argue that the significantly lower levels observed in virtually all similarly developed countries that have stricter gun laws must reflect a causal relationship of gun prevalence on crime rates (Brady Campaign, 2016). Additionally, they often assert that widespread availability of guns creates substantial risks to society if terrorists, convicted felons or domestic abusers obtain firearms easily and subsequently use them for criminal activities.

*This chapter is based on joint work with Christoph Koenig.

Because the political debate tends to be based on isolated observations and potentially spurious correlations, providing objective, scientific evidence is imperative. More than 30,000 US residents every year die of suicide, homicide or an accident involving a firearm. If a reduction or increase in gun ownership would effectively reduce these numbers, substantial welfare gains could be realized. This study therefore provides quasi-experimental evidence on the relationship between firearm purchases and crime rates. In contrast to the existing literature of similar scope from the fields of economics, public health and criminology that largely relies on correlational evidence, we use arguably exogenous variation in the demand for firearms from a subset of the population that can reasonably be expected to buy guns for lawful purposes, such as self-defense.

In particular, we use a country-wide exogenous shift in firearm demand in the United States following the shooting at Sandy Hook Elementary School. Fear of tougher gun legislation and an increase in perceived need of self-defense capabilities drove up gun sales across the entire United States (Vox, 2016; CNBC, 2012). In most US states, citizens have instant access to firearms, i.e. they can take their gun home immediately after purchase. Some states, however, have implemented legislation intended to delay purchases, either by imposing mandatory waiting periods between purchase and receipt of a gun or by introducing other time-consuming bureaucratic hurdles such as making government issued gun purchasing permits mandatory. These delays led to differential firearm purchases following the shooting at Sandy Hook, a feature that we exploit to estimate changes in crime rates in a standard differences-in-differences setup.

In a first step, we show that handgun purchases in states with instantaneous access to guns increased more strongly following the tragic events at Sandy Hook. We compare the change in the number of handgun purchase background checks per 100,000 population induced by the shooting between states that impose delays and those that do not. In the two year window around the shooting, states without delays saw significantly higher post-shooting handgun purchases. This finding is robust to increases of the time window and the addition of controls that are allowed to have a varying effect over time. While graphical inspection of the time series data does not show any differential pre-trends, adding state-specific time trends makes the results even stronger, and performing a synthetic control exercise (Abadie et al., 2012) also confirms the robustness of our findings.

Using Google search data, we then analyze if the expression of interest in buying a gun differed depending on whether the state imposed purchasing delays or not. We compare the difference in shooting-induced changes of daily searches for the term ‘gun store’ between both groups of states, a search expression that has been shown to be the best predictor of gun purchasing intentions (Scott and Varian, 2014). Using the same set of covariates as before, and stepwise varying the time window around the shooting at Sandy Hook from seven days up to 730 days does not reveal any differences in search patterns. In other words, consumers in all states did not show strongly different intentions to purchasing a firearm after the shock, but when it came to actually purchasing the gun, legislative delays led some consumers not to buy. This finding can be rationalized by the presence of procrastinating naïve present-biased consumers, but not by arguments from standard economic theory such as transaction costs alone. A present-biased consumer would show interest in purchasing a firearm, but then due to the delayed reward of receiving the gun postpone the purchase, while still intending to buy the gun in the future. Publicly known transaction costs however would cause consumers to not wanting to buy guns in the first place. Our findings suggest that insights from behavioral economics are important factors to consider in these kinds of public policy questions.

We then rule out that differences in primary market firearms purchases were counteracted by shifts to secondary markets (i.e. gun shows instead of licensed gun dealers), as the demand for gun shows did not tilt towards states with purchase delays. We again employ Google search data, using searches for the expression 'gun show'. When varying the time window around the shooting, we initially detect a higher demand in states without purchase delays. With a long enough time horizon, this effect vanishes such that both groups of states become statistically indistinguishable. This shows that states implementing delays do not experience stronger demand shifts towards secondary markets. If anything, we underestimate the true total effect of the shooting on gun purchases.

The second part of our empirical analysis looks at the impact of the differential firearm purchases on different types of crimes. Using monthly FBI data on several categories of crime, we estimate changes in the crime rate between states with and without delays after the shooting at Sandy Hook. To ensure an adequate treatment of possibly diverging pre-trends, we allow for control variables to have a differential impact at different dates and include county-specific time trends. We find that states granting instant access to firearms see significantly more murders after Sandy Hook. This effect is strongest and significant when including at least five and up to 14 months before and after the shooting, and is fading for longer time horizons, suggesting that the effect wears off quickly. After twelve months, allowing instant access to guns is associated with an estimated 14.47% increase in murder rates, which implies that approximately 98 lives could have been saved from murder in each month of 2013 if mechanisms to delay purchases had been in place in all US states. While murder rates increase, we also find that all other crime categories remain unaffected, and the share of assaults or robberies committed with guns does not change significantly. This provides tentative evidence against the claim of guns leading to a credible deterrence effect, in which criminals avoid certain types of crimes, fearing backlash from armed victims.

Additional analyses investigate the robustness of the effect on murders. To rule out that the result is purely seasonal, we conduct a placebo regression around the same date, but in 2010 instead of 2012, and find no effect. We also provide evidence on how the effect changes when varying the time horizon of our sample. As mentioned earlier, the effect only exists for medium time horizons (five to 14 months), as including only a few months of data leads to very imprecise estimates, and including more than 14 months before and after the shooting shows that the effect eventually washes out as we would have expected from a temporary shock. Furthermore, excluding states one-at-a-time yields similar coefficient estimates, and all estimates remain highly significant, no matter which state is removed, effectively showing that the effect is not caused by a single outlier state.

There might also be concern that the relationship between increases in gun ownership and increases in murder rates is not causal. This would be a problem if the shooting at Sandy Hook changed attitudes towards murder differentially in states with and without delays, while at the same time changing gun ownership. After all, a mass shooting itself constitutes an act of violent crime that through extensive media coverage might influence preferences towards violence for some individuals. To address this concern, we complement the analysis by looking at an additional event unrelated to violent crime but responsible for positive gun demand shifts: The 2012 Presidential election. While we cannot separate the election and the shooting at Sandy Hook, because the former happened in November 2012, and the latter in December 2012, we can however show that the coefficient for murders is much stronger if we pre-date the treatment event by one month. This is consistent with President Obama's re-election contributing to the effect in the same direction as the shooting. It is also reassuring that the effect becomes immediately insignificant if pre-dated or post-dated by more than one month.

We then repeat our estimations using an Instrumental Variables (IV) approach, in which we estimate the relationship of gun purchases and murder rates by instrumenting gun purchases with our differences-in-differences treatment indicator from our earlier analyses. Our results confirm the previous findings: Increases in gun ownership are significantly positively related with murder rates, but no such effect exists for other types of crimes. The IV estimates predict that every 2049 additional guns cause one additional murder, or that 97 lives could have been saved from murder in each month of 2013. This finding is remarkably close to the effect size of 98 lives that our earlier estimates predicted. The results from the IV approach therefore suggest that our set of control variables performs well in predicting the reaction to the demand shock in terms of compliance, such that intent-to-treat and treatment effect on the treated almost coincide.

Finally, we use supplementary homicide data to identify the circumstances of the additional murders. When analyzing murders separately by whether a firearm was used as the murder weapon or not, we are able to show that the entire effect can be attributed to murders by firearm and that murders, where the weapon of choice was not a gun, remain unaffected. We then investigate which gender was more likely to become a murder victim. Our estimates suggest that murders of both men and women increase, but that the relative increase is more than twice as large for women. This is primarily driven by men that murder women.

This study is related to several streams of research. Scholars from the disciplines of economics, criminology and public health have previously tried to find empirical support for the relationship of firearm ownership and violent crime rates, with mixed results. Several studies find that more guns lead to more crime (e.g. Cook, 1978; Cook and Ludwig, 2006; Duggan, 2001; Hemenway and Miller, 2000; Kaplan and Geling, 1998; Miller et al., 2002, 2007; Siegel et al., 2013; Sorenson and Berk, 2001), using either time series data on an aggregated level, cross-sectional or panel data.[1] Other studies tackle the issue more indirectly by estimating the effect of gun legislation (Fleegler et al., 2013) or gun shows (Duggan et al., 2011) on crime rates, based on the idea that those in turn might influence gun prevalence. The most prominent study to find a negative effect is the book by Lott (2013). He argues that the enactment of concealed carry laws has created a credible deterrent, such that criminals abstained from committing crimes, and that availability of firearms through this channel decreases violent crimes. His findings are supported by the results in Lott and Mustard (1997) and Moody (2001). Additional research suggests that there is no statistical relationship between gun prevalence and crime (Kates and Polsby, 2000; Kleck and Patterson, 1993; Lang, 2016; Moody and Marvell, 2005). All of these studies however rely on correlations for inference and thus give rise for omitted variables bias. No clear effect is reported in Kovandzic et al. (2013), in which the authors use an instrumental variables approach, but their suggested instruments for gun ownership seem unlikely to satisfy the exclusion restriction (voter share for the Republican party, share of veterans and subscriptions to gun-related outdoor magazines). We contribute to this literature by providing well-identified evidence from an exogenously timed gun demand shock.

This chapter is also closely related to literature exploring the role of legislative restrictions for gun sales such as waiting periods. Fleegler et al. (2013), for example, looks at the link between the strength of firearm legislation and firearm-related fatalities. It

[1] An excellent survey discussing in particular the early contributions is provided by Hepburn and Hemenway (2004), newer contributions are discussed by Kleck (2015).

finds that stronger firearm legislation is associated with fewer gun-related deaths, but is not able to determine if this relationship is causal. Kleck and Patterson (1993) studies a similar question but does not find any effect of gun control laws on violence, but rather that violent crimes lead to higher gun prevalence. In contrast to these studies, this chapter is not interested in a general measure of gun control, but more specifically in the induced delays by waiting periods and time-consuming permit processes that can eventually discourage some consumers. Closely related is the contribution by Ludwig and Cook (2000) that is interested in the effects of introducing waiting periods through the Brady Act. The paper does not present clear-cut evidence that waiting periods contribute to changes in violent crimes, but the approach is notably different to ours. We are primarily interested in differential firearm take-up and therefore study an exogenous event that influences behavior given an existing set of constraints. Ludwig and Cook (2000) instead analyzes the changes resulting from legislative advances assuming constant firearm demand. This might be a problematic approach if consumers can anticipate legislative changes and adjust the timing of firearm purchases. In the case of the Brady Act, almost three years passed between the introduction of the bill and when it went into effect. Our approach is also not just restricted to waiting periods, but includes mandatory purchasing permits that also delay the receipt of a firearm.[2]

Given the small number of studies on the effects of gun legislation, the Center for Disease Control in 2003 concluded that more evidence is needed to ultimately judge the effect of waiting periods on violent crimes (Hahn et al., 2003). Recent contributions to the literature concerned with the effects of legislative restrictions put stronger emphasis on providing credible identification to allow for causal inference, such as Dube et al. (2013) and (Chicoine, 2016) that look at the expiration of the Federal Assault Weapons Ban and subsequent violence in Mexican municipalities. They find that the availability of assault rifles following the expiration significantly increased violent crimes. While assault rifles seem of relative importance to organized crime such as cartels and gang violence, our study in contrast is concerned with a demand shift in legally acquired handguns. This distinction is important, because judging the relationship of gun prevalence and crime rates may yield very different results if the type of gun is primarily used by professional criminals instead of citizens for lawful purposes.

By offering an explanation for why gun purchasing delays may lead to fewer relative purchases that is based on findings from behavioral economics, we additionally contribute to the recent literature in public economics that analyzes the effect of be-

[2] Rudolph et al. (2015) analyzes the effect of the introduction of Connecticut's mandatory pistol purchasing permit in 1995 and finds a strong decrease in homicides. Because the passage of this legislation did not come unannounced, the same caveats apply as in the case of Ludwig and Cook (2000).

havioral biases for public policies (e.g. Chetty et al., 2009; Choi et al., 2011).[3] We also relate to studies in economics that link behavioral shortcomings with criminal activity, violent behavior and policing. Dahl and DellaVigna (2009) investigates the effect of movie violence on violent crimes and finds that attendance of movies serves as a substitute for violent behavior. Card and Dahl (2011) finds that unexpected losses of the home football team increase instances of domestic violence, and Mas (2006) observes a decline in policing quality after a lost salary arbitration by the respective police union in New Jersey.

Our results have important implications for other researchers and policy makers. First, we complement the literature on the relationship of gun ownership and crime by providing well-identified estimates of the effect of an exogenously timed change in firearm ownership rates. Since, according to anecdotal evidence, our sample of interest purchased firearms for lawful purposes and not for criminal activity, the significant increase in murder rates seems even more striking, especially since Fabio et al. (2016) report that only a small fraction of crimes are committed using legally acquired firearms. Second, our findings do not detect a deterrence effect, suggesting that criminals either do not anticipate changes in firearm ownership correctly, or that they have a negligible impact on their choices. Therefore, arming citizens to prevent crimes does not seem to be a very promising approach. Third, legislators deciding about gun control measures should take our findings into account, as designing smart regulations can have non-negligible welfare effects, and save a substantial number of lives. Waiting periods and bureaucratic hurdles appear very promising to at least somewhat reduce impulsive acts of violence, especially against women. Fourth, our contribution suggests that cognitive biases and limitations established by the growing field of behavioral economics can meaningfully be applied in a context of political economy and research on crime, and should thus be taken into account when modeling individual behavior.

This chapter is organized as follows: Section 5.2 provides details about gun laws in the United States, describes the tragic events at Sandy Hook Elementary School and the subsequent firearm demand shock. It furthermore delivers theoretical explanations for the differential firearm take-up of prospective gun owners. Section 5.3 in turn describes our data in detail and explains our estimation strategy. Results can be found in Section 5.4 and Section 5.5 concludes.

[3] Excellent summaries on the application of behavioral biases in public policy are provided by DellaVigna (2009) and Chetty (2015).

5.2 Background

5.2.1 Gun Laws in the United States

The Second Amendment to the United States Constitution together with nine other amendments, collectively called the Bill of Rights, was passed by the 1st United States Congress on September 25, 1789 and then adopted on December 15, 1791 through ratification by 75% of the states. It protects the right of citizens to keep and bear arms at the federal, and through the Fourteenth Amendment also at the state and local government level. While the official text of the Second Amendment says that "A well regulated militia being necessary to the security of a free state, the right of the people to keep and bear arms shall not be infringed", the United States Supreme Court clarified in its decision of *District of Columbia et al. v. Heller* (554 U.S. 570) that this right also refers to individuals (outside of militias) for traditionally lawful purposes such as self-defense.

The federal government, as well as state and local governments have in the past however enacted laws that make it harder or require more effort from citizens to acquire firearms. On the federal level, the most important pieces of legislation for this study are the Gun Control Act of 1968 and the Brady Handgun Violence Prevention Act. The Gun Control Act requires that all professional gun dealers must have a Federal Firearms License (FFL). Only they can engage in inter-state trade of handguns, are granted access to firearm wholesalers and can receive firearms by mail. The Brady Act was enacted on November 30, 1993, and mandated background checks for all gun purchases through FFL dealers. Initially, the bill also imposed a five-day waiting period on handgun purchases, which upon successful lobbying of the NRA, was set to expire when the National Instant Criminal Background Check System (NICS) took effect in 1998, a system operated by the FBI that handles all background checks related to the sales of firearms.

While there is little regulation regarding firearm ownership on the federal level compared to other similarly developed countries, there is substantial heterogeneity in restrictions imposed by the states. For example, many states invoke restrictions on the prerequisites and responsibilities of gun dealers, such as whether they require an additional state license to operate their business or whether they are supposed to keep centrally stored electronic records of transactions. Other legal restrictions concern buyers, as states can for example decide if they want buyers to be able to purchase guns in bulk, if buyers need a permit prior to purchase, if they have to undergo background checks (for transactions exempted from federal background check requirements), or if buyers are required to wait a certain amount of time between purchasing and receiving their gun. Finally, there exists legislation concerned with restrictions on carrying firearms in public places, including schools and the workplace.[4]

Most of the constraints on private firearm ownership at the state level attempt to either prohibit convicted felons or otherwise potentially dangerous people from acquiring guns for non-lawful purposes, or restrict the usefulness of firearms for non-lawful purposes independent of the buyer. One restriction that is of substantial interest to our study is the imposition of mandatory waiting periods. While the establishment of waiting periods with the Brady Act aimed to give law enforcement agencies sufficient time to conduct background checks, they also provide a "cooling-off" period and can therefore help to prevent impulsive acts of violence (Cook, 1978; Andrés and Hempstead, 2011). These impulsive acts of violence might be directed towards other persons (such as homicides and assaults), but they could also be directed towards the gun holder itself (e.g. suicides).

As of 2016, nine states and the District of Columbia have imposed mandatory waiting periods. California and D.C. require ten days, Hawaii 14 days, Rhode Island seven days and Illinois between 24 hours (long guns) to 72 hours (handguns) on all firearm purchases. Minnesota is the only state to require seven days wait between purchase and pickup of handguns and assault rifles only. Maryland and New Jersey impose seven days for handguns, while Florida and Iowa impose a three day waiting period for handguns. Wisconsin has repealed its 48 hour waiting time on handguns in 2015.

Additionally, some states require a license to possess or buy a firearm prior to the actual purchase, which due to bureaucratic hurdles can also impose a waiting time. In Connecticut, a handgun eligibility certificate may take up to 90 days before being issued. Before buying a gun in Hawaii, prospective gun owners have to obtain a permit to purchase which can take up to 20 days to be issued. Buyers in Illinois have to obtain a Firearm Owner's Identification card (FOID) before being allowed to purchase an unlimited number of firearms in the following ten years. Obtaining an FOID can take up to 30 days. The state of Maryland requires buyers to hold a Handgun Qualification License which will be issued or denied within 30 days of application. In Massachusetts, authorities may take up to 30 days to process a request for a license to carry or a Firearm Identification Card (FID), where the former allows unlimited purchases of any firearms without additional paperwork and the latter is restricted to rifles and shotguns. Citizens in New Jersey in turn must obtain a permit to purchase a handgun for each purchase separately, while they can purchase unlimited shotguns and rifles with a Firearms Purchaser Identification Card (FPIC). Authorities may take

[4]An excellent overview of all restrictions in the respective states can be found in The Brady Campaign (2013), NRA (2016) and Law Center to Prevent Gun Violence (2016).

up to 30 days to issue such a permit. In New York, a license to possess or carry a handgun is necessary for each gun and obtaining one can take up to six months. In North Carolina, a license to purchase a handgun can take up to 14 days to be issued, and it is valid for one gun only. Citizens of Rhode Island need to wait up to 14 days to receive their pistol safety certificate (blue card). Table 5.1 summarizes the waiting periods and license requirements for handguns across states.[5]

Table 5.1: Handgun waiting periods and handgun purchasing license delay by state

State	AL	AK	AZ	AR	CA	CO	CT	DE	FL
Mandatory Waiting Period	0	0	0	0	10	0	0	0	3
Maximum Purchasing Permit Delay	0	0	0	0	0	0	90	0	0
State	GA	HI	ID	IL	IN	IA	KS	KY	LA
Mandatory Waiting Period	0	14	0	3	0	3	0	0	0
Maximum Purchasing Permit Delay	0	20	0	30	0	0	0	0	0
State	ME	MD	MA	MI	MN	MS	MO	MT	NE
Mandatory Waiting Period	0	7	0	0	7	0	0	0	0
Maximum Purchasing Permit Delay	0	30	30	0	0	0	0	0	0
State	NV	NH	NJ	NM	NY	NC	ND	OH	OK
Mandatory Waiting Period	0	0	7	0	0	0	0	0	0
Maximum Purchasing Permit Delay	0	0	30	0	180	14	0	0	0
State	OR	PA	RI	SC	SD	TN	TX	UT	VT
Mandatory Waiting Period	0	0	7	0	0	0	0	0	0
Maximum Purchasing Permit Delay	0	0	14	0	0	0	0	0	0
State	VA	WA	WV	WI	WY	DC			
Mandatory Waiting Period	0	0	0	2**	0	10			
Maximum Purchasing Permit Delay	0	0	0	0	0	0			

Mandatory Waiting Period refers to the amount of time in days to pass between the purchase and the receipt of a firearm. If a state has different waiting periods for different types of firearms, the number refers to the purchase of handguns. Maximum Purchasing Permit Delay refers to the maximum time in days that can pass before a permit that will allow the holder to purchase one or more handguns will be issued or denied. 0 means that no permit is needed or will be issued instantaneously. * Repealed in 2015. Source: `http://smartgunlaws.org`

5.2.2 The Shooting at Sandy Hook Elementary School

On the morning of December 14, 2012, then 20-year-old Adam Lanza, a resident of Newtown, Connecticut, first shot and killed his mother at their home before driving to Sandy Hook Elementary School, where he shot and killed six adult school employees and 20 students, who were between six and seven years old. Although the carnage only lasted about five to ten minutes, Lanza was able to discharge his firearms (a semi-automatic AR-15 type assault rifle and a pistol) 156 times, averaging approximately one shot fired every two to four seconds. He committed suicide shortly after the first law enforcement officers arrived at the scene. Eyewitness and police reports describe

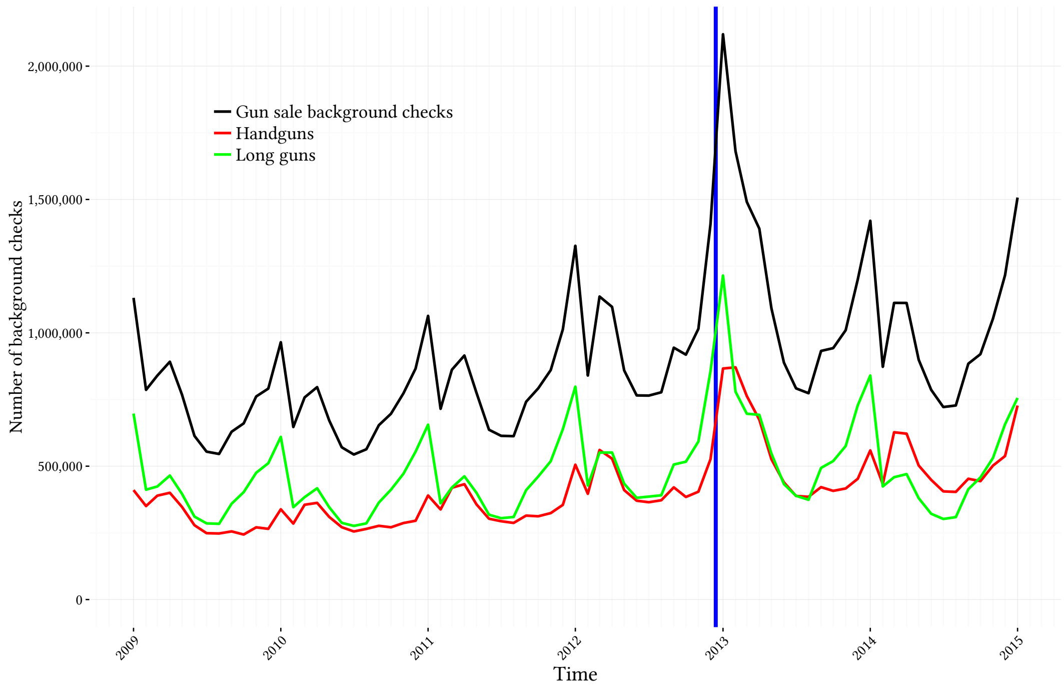
[5]Michigan and Nebraska also require a permit, but only if the gun is purchased from an unlicensed seller. Since we are interested in gun sales from licensed sellers only, we treat these states as not imposing delays. Iowa's permit is issued instantly and can be used after the aforementioned three day waiting period.

that Lanza was acting calmly throughout and killed his victims with targeted shots to the head. Even after several years, his motives are still not fully understood. Lanza did not leave any documents that could explain his thoughts, but it has been suggested that he had a history of mental illness. His father reported to have observed strange and erratic behavior in Lanza that he might have falsely attributed to his son's Asperger syndrome, rather than a developing schizophrenia (New Yorker, 2014).

The shooting being the deadliest shooting at a US high or grade school and the third deadliest mass shooting in US history at the time, combined with the fact that most of the victims were defenseless children, sparked a renewed debate about gun control in the United States. A few days after the shooting, President Barack Obama announced that he would make gun control a central issue of his second term. A gun violence task force under the leadership of Vice President Joe Biden was quickly assembled with the purpose of collecting ideas how to curb gun violence and prevent mass shootings. The task force presented their ideas to President Obama in January 2013, who announced to proceed with 23 executive actions, aimed at improving background checks, addressing mental health issues and insurance coverage of treatment thereof, as well as enhancing safety measures for schools and law enforcement officers responding to active shooter situations. Additionally, the task force proposed twelve congressional actions, including renewing the Federal Assault Weapons Ban, expanding criminal background checks to all transactions, banning high capacity magazines, and increase funding to law enforcement agencies. The proposals were met by fierce opposition from the NRA and some Republican legislators. At the end of January 2013, Senator Dianne Feinstein introduced a bill aimed at reinstating the Federal Assault Weapons Ban. While the bill passed the Senate Judiciary Committee in March 2013, it eventually was struck down on the Senate floor 40-60 with all but one Republicans and some Democrats opposing the bill. A bipartisan bill to be voted on at that same day, introduced by Senators Joe Manchin and Pat Toomey, aimed at introducing universal background checks, also failed to find a majority with 54-46, leaving federal legislation eventually unaffected.[6]

While no new federal regulations eventually followed the events at Sandy Hook Elementary School, gun sales soared in the months after the shooting. Fear of tougher gun legislation and a higher perceived need of self-protection drove up sales for both, handguns and rifles (Vox, 2016). While gun sales had surged after every prior mass shooting during the Obama administration, the increase in sales was unprecedented after the

[6] Three states tightened their gun laws in the wake of the Sandy Hook Shooting: Connecticut, Maryland and New York. Connecticut and New York tightened their existing assault weapons ban, required universal background checks and outlawed high capacity magazines. Maryland outlawed assault rifles and high capacity magazines, and now requires handgun licensing and fingerprinting for new gun owners.



Monthly federal NICS gun sale background checks plotted over time between 2009 and 2015 in absolute numbers. The blue vertical line marks the date of the shooting at Sandy Hook Elementary School. The red line shows background check for handguns, the green line for long guns, and the black line displays the sum of the two.

Figure 5.1: NICS background checks before and after Sandy Hook

shooting at Sandy Hook. The extreme demand shift even created supply problems for some dealers, who were hoping to see sales increases of a magnitude of up to 400% (CNBC, 2012; Huffington Post, 2013). Several executives in the gun industry have openly admitted that they view mass shootings as a boon to their business, attracting especially first-time gun owners. Tommy Millner, CEO of Cabela's in response to the Sandy Hook shooting said "the business went vertical ... I meant it just went crazy [... We] got a lot of new customers." and James Debney of Smith & Wesson explained that "the tragedy in Newtown and the legislative landscape [...] drove many new people to buy firearms for the first time." (The Intercept, 2015). But the increased interest in firearms was not just restricted to brick-and-mortar stores. In a recent paper, Popov (2016) shows that prices for gun parts in an online market sharply increased by 20% after Obama's announcement for tougher gun legislation. Figure 5.1 shows the spike in gun sales – it displays the evolution of firearm background checks over time, before and after the Sandy Hook shooting. While gun sales generally increase at the end of the year, the spike following the Sandy Hook shooting is much more pronounced than in the years immediately before and after. It is these new customers and their differential propensity to acquire firearms in some states that we seek to exploit for our data analysis.

5.2.3 A Behavioral Motivation for Firearm Purchase Delays

There exist several theoretical approaches that could explain why a purchasing delay would result in some individuals not buying guns at all. In the following, we therefore provide insights from standard economic theory, as well as behavioral economics that predict differential purchasing decisions depending on whether delays exist or not. We also establish more precise conditions under which these theories hold, such that our data analysis can deliver supporting or opposing evidence for the theories in question. Note that the purpose of this exercise is not to support our empirical findings, since our results hold independent of any theory. We rather aim to provide a theoretical foundation to shed light on possible reasons behind the observed facts, and as a by-product are able to test which theory best explains the patterns evident in the data.

In standard economic theory, the primary reason for differential purchasing reactions given identical preferences root in differing transaction costs. Without delays, purchasing a gun requires the prospective buyer to travel to the gun store, file the necessary paperwork, and take home their gun which creates costs. Delays however require substantially more effort. In the case of waiting periods, the prospective buyer not only has to to travel to the gun store and file the necessary paperwork, they would also have to come back after a few days to pick up their gun. The additional costs associated with a second visit to their gun dealer can outweigh the net benefit of purchasing, and therefore prevent some marginal customers to buy a firearm. If for example each trip to a gun dealer generates utility losses of c, and a gun provides utility gains of v, the assumption $2c > v > c$ is sufficient to observe a differential reaction depending on whether the state implemented waiting periods or not. Implementing mandatory handgun permits that are only issued after a delay creates a similar effect. Before prospective buyers can undertake their trip to the gun store, file the paperwork and take home their gun, they have to travel to their closest public authority commissioned with issuing these types of permits, file the necessary paperwork there and wait for the permit to be issued. If the costs associated with getting a permit are k, $k + c > v > c$ will lead to states with delayed permits experiencing lower rates of gun purchases. Importantly, transaction costs in a full information setup with perfectly rational agents require that there already exists a differential interest in buying a firearm between the onset of the demand shock and the act of buying the gun.[7] This is due to the fact that buyers incorporate transaction costs in their decisions and then either buy a gun or not.

[7] We will discuss the appropriateness of the assumption of full information about delay legislation when we present our results regarding the intention to buy a firearm.

Another reason why we would observe a differential reaction depending on the implementation of delays is due to naïve present-biased consumers (Laibson, 1997). Present-biased decision makers prefer consuming today over tomorrow, but not necessarily tomorrow over in two days, i.e. they change their assessment of preferred options at different points in time, depending on when they are being asked about it. For example, when being asked whether to have fruit salad or cake tomorrow, a time-inconsistent decision maker deciding for fruit salad today, could revise the decision to cake when being asked again tomorrow. Naïvete refers to the decision makers' inability to know about their tendency to behave time-inconsistently.

Present-bias affects several dimensions of inter-temporal choice and has been frequently observed in decision-making contexts: Meier and Sprenger (2010) report that people who hold credit card debt are more likely to be present-biased, Augenblick et al. (2015) find evidence of present-bias in effort allocation decisions among workers, Choi et al. (2011) observe dominated investment behavior in retirement decisions due to time-inconsistency, and Duflo et al. (2011) build a model of present-bias in fertilizer use that predicts actual fertilizer purchases in rural Kenia quite well.[8]

Following the approach by O'Donoghue and Rabin (1999), where future utility streams are uniformly discounted by $\beta \leq 1$, we argue that delaying the reward from owning a gun can lead some consumers to not buy a firearm. Assume that the costs of purchasing a firearm are c and the utility of having a firearm is v. In the absence of delays, prospective buyers will decide to buy if $v > c$. When delays are present however, this condition becomes stricter $\beta v > c$, because the utility from owning the gun has been shifted to the future and is therefore discounted by the prospective buyer. Depending on the ratio of v and c, and their relationship with β, some present-biased consumers will therefore not buy a gun when facing delays.

The difference between this theory and simply assuming transaction costs is however that with naïve present-bias, the ex ante interest of buying a gun should be identical between states that impose delays and states that do not (given identical preferences). This can easily be seen in a two-period model, in which a gun can be purchased in either period 1 or period 2. While a naïve present-biased prospective buyer will not buy in period 1 because $\beta v - c < \beta(v - c)$ when $\beta < 1$, they still believe in period 1 that they will buy eventually in period 2, because $\beta v - \beta c > 0$. This behavior arises because the prospective buyer underestimates how heavy the costs of buying the gun will weigh in the future. He therefore might make plans to purchase a gun in the next period, but never follows through.

[8]More examples are discussed in Sprenger (2015) and Frederick et al. (2002).

We can test whether transaction costs or present-biased consumers are primarily responsible for causing a divergence in purchasing behavior. If we find evidence that prospective buyers exhibit strong differences when it comes to their intention to buy a gun after the shock, this can be reconciled with transaction costs but not with present-bias. In contrast, if we observe differential purchasing behavior while the interest in buying a gun is independent of whether the state implemented a delay, this is evidence for present-biased consumers.

Also note that several other theories cannot account for differential purchasing behavior. Projection bias (Loewenstein et al., 2003) predicts that consumers' current taste is projected to the future, where it may not be accurate anymore. As a policy implication, offering return periods allows these consumers to make better decisions, as the purchase can be reversed when the realization of tastes is substantially different to the anticipated taste. Waiting periods for firearm purchases however will not interact with projection bias. Consumers interested in buying a gun (based on their projected utility) will do so, no matter whether they have to wait for their gun or not. Since the purchasing decision is made when conducting the background check and filing the paperwork, and not when picking up the firearm, there is no scope for projection bias influencing the decision. Similarly, transient emotional reactions can hardly be expected to drive the difference, as the purchasing decision is made when filing the paperwork. This should lead to differential pick-up rates of purchased guns, but not necessarily to differences in purchases per se.[9]

5.3 Data & Estimation Strategy

5.3.1 Estimation Strategy & Identification

Following the shooting at Sandy Hook, firearm demand in the United States increased strongly, both for fear of tougher legislation, as well as a higher perceived need of self-protection. As some states allow their citizens to instantly purchase the guns of their choosing, the higher demand in those states could immediately translate into increased sales. States that were imposing mandatory waiting periods or that had a time-consuming application process for purchasing permits, however, were able to delay transactions, possibly discouraging buyers from eventually buying any guns.

We therefore define all states that had a positive waiting period for handguns or that require a time-consuming permit to be issued prior to purchase as "delayed states", as

[9] Section E.3.1 in the appendix discusses the assumptions under which competing theories could offer an explanation and argues why we believe these assumptions to be unrealistic.

listed in Table 5.1: California, Florida, Hawaii, Illinois, Iowa, Maryland, Massachusetts, Minnesota, New Jersey, New York, North Carolina, Rhode Island, Wisconsin and the District of Columbia. All other states we subsume under "instant states", and we remove Connecticut from all samples, because the state might have been affected differently by the shooting at Sandy Hook, as Newtown lies in Connecticut.[10] We first show that delayed states have a smaller increase in gun sales than instant states, and that this increase is not driven by pre-existing trends. Then, we continue with investigating the effect of differential firearm purchases on crime rates. There exist several potential outcomes for such an analysis. First, crime rates could increase as new gun owners might turn criminal. For example, a domestic dispute otherwise gone unnoticed to law enforcement might suddenly turn violent with one spouse shooting and killing the other. It is also conceivable that if new gun owners are marginally law-abiding in the sense that their low income is weakly preferred to being criminals, any income shock might turn them criminal, a profession potentially more lucrative for someone armed. Second, crime rates could decrease, because armed citizens serve as a credible deterrent to criminals. Robbing someone on the street or burglarizing someone's apartment becomes more dangerous if the likelihood of the victim being armed is higher, therefore decreasing the relative profitability of being a criminal. Additionally, engaging in bar fights or similar altercations suddenly becomes more risky if the opponent has a gun at their disposal. Third, crime rates could overall stay unchanged, but there could be a shift from less severe to more severe crimes. Speaking in the examples above, the domestic dispute or the bar fight due to the availability of lethal force could turn from assault to murder, effectively reducing crimes in one category and raising it in the other.

Our data analysis uses a classical differences-in-differences (DiD) approach, in that we compare crime changes due to the Sandy Hook shooting between delayed and instant states. Our estimation equation reads

$$y_{it} = \alpha + \gamma_i + \lambda_t + \beta(POST_t \times INSTANT_i) + \delta \mathbf{X_{it}} + t_i + \epsilon_{it} \tag{5.1}$$

where y_{it} is our measure of gun purchases in the first step of the analysis, and our measure of crime rates in the second step, γ_i denotes location and λ_t time fixed effects. Our coefficient of interest is β, capturing the joint influence of dummy $POST_t$ (time periods after the shooting at Sandy Hook) and dummy $INSTANT_i$ (locations that belong to instant states). $\mathbf{X_{it}}$ denotes a vector of control variables with time-varying

[10]None of our results depend on this decision. Appendix Section E.2 contains the results when including Connecticut for the two most important tables from our results section. All other tables are available from the authors.

influence and t_i is a time trend, allowed to vary within each location unit. ϵ_{it} is the error term.

To assure proper identification and validity of our DiD estimator, we need two assumptions to be fulfilled: First, we must not have differential pre-trends in our outcome measures to ensure that divergence between instant and delayed states is indeed driven by the treatment and not just the continuation of a pre-existing divergence. We will address this concern by graphical visualization, synthetic control exercises and allowing for location-specific time trends. Second, there must not have been other events responsible for the divergence that occur at approximately the same time as the treatment. Placebo regressions and shifting the onset of the treatment will deliver evidence that the effect is particular to a very small time window that coincides with the shooting. We furthermore argue that the timing of the shooting at Sandy Hook is entirely exogenous to any relevant outcome variables: The perpetrator Adam Lanza apparently had prepared the crime for years, investigators believe he downloaded videos and other materials related to the shootings at West Nickel Mines School (2006) and Columbine High School (1999) on his computer. The shooting came as a surprise to law enforcement and Lanza's social environment and apparently were not triggered by any other public events. Therefore, we assume strict exogeneity of the event to our outcome variables.

Additionally, in order to establish a causal link between the demand shock and crime rates, proper identification requires that the shock only affected firearm demand and not crime rates directly other than through the changes in gun prevalence. Given that a mass shooting itself is a crime and might therefore differentially influence attitudes towards violence, we will address this potential issue by extending our analysis to include the date of the 2012 Presidential election, to see if the re-election of President Obama contributes to our estimated coefficient from the shooting at Sandy Hook.

5.3.2 Datasets

Gun Ownership

Unfortunately, no reliable information about gun ownership and inflow of new guns exists at a sufficiently fine geographic (e.g. county) level. Researchers therefore have relied on proxies for gun ownership levels, including subscription to a gun magazine (Duggan, 2001), fraction of suicides committed with a firearm (Cook and Ludwig, 2006; Azrael et al., 2004; Kleck, 2004), and gun ownership questions from the General Social Survey (GSS) (Glaeser and Glendon, 1998).

Since we are interested in the increase in gun ownership rather than the stock of guns, we use applications from the National Instant Criminal Background Check System (NICS), which is available from the FBI for each state and month since 1998 and has been used for this purpose before (Lang, 2016).[11] Each purchase of a new firearm at a federally licensed firearm dealer triggers an application for a background check.[12] In total, we obtain monthly data for background checks of firearm purchases in all 50 states (and the District of Columbia) between November 1998 and December 2014.

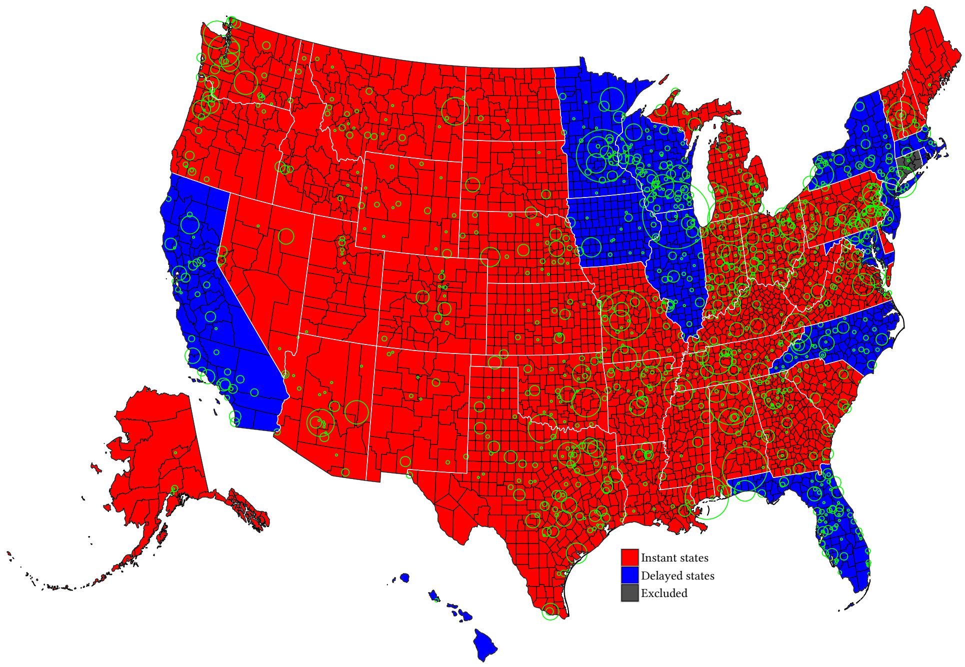
While the NICS data gives us a good idea of actual firearm purchases, we would also like to measure the ex ante interest in buying a firearm. If the assumption of present-biased consumers abstaining from buying firearms in delayed states were true, they should still be affected by the demand shock in the sense that, ceteris paribus, their desire to own a firearm increased by just as much as for consumers in instant states. We therefore extract daily search data from Google Trends (`http://google.com/trends`) for the expression ‘gun store’ for each state between 2009 and 2014. Google Trends is a data service that reports the relative frequency of specific Google search expressions across time and geography. The search term ‘gun store’ has been shown to be highly predictive of the willingness to purchase a firearm (Scott and Varian, 2014). Unfortunately, Google Trends data is always rescaled and sometimes censored, such that some manual conversions are needed to make meaningful comparisons. First, for each query, Google rescales data to be between 0 and 100, where 100 is assigned to the largest value in the entire time window of the query. Thus, we only obtain data on the relative occurrence of the expression in one particular state on each date within a 3 month time window. To correctly adjust the results in the time dimension, we designed the queries such that months were overlapping, for example the first query downloaded January to March data, the second March to May, the third May to July and so on. We then rescaled the data based on the overlaps using January of 2009 as a baseline. In order to also make the cross-section comparable, we designed a query for January 2009 for each state to obtain initial relative weights for each state. Second, Google Trends data is censored to zero if the number of searches falls below a certain, not publicly known threshold. Since we have no reason to believe that censoring affects states across our treatment groups differently, we do not employ corrective action. Due to censoring, we only obtain daily data for in total 34 states and the District of Columbia.[13]

[11] The data can be downloaded at `https://www.fbi.gov/file-repository/nics_firearm_checks_-_month_year_by_state_type.pdf`.

[12] State permit holders to purchase a gun are exempt from the background check, if the process of obtaining their permit involved passing a background check (since November 30, 1998 only NICS background checks qualify).

[13] We are missing data for the states Alaska, Arkansas, Delaware, Hawaii, Idaho, Maine, Montana, North Dakota, Nebraska, New Hampshire, Rhode Island, South Dakota, Vermont, West Virginia and Wyoming, representing both instant and delayed states. Connecticut is also removed from the sample.

Note that our NICS measure of changes in gun ownership potentially misses some trades on secondary markets (e.g. through gun shows). 2016 Democratic nominee for President Hillary Clinton's campaign has suggested multiple times that up to 40% of guns are purchased at gun shows, a number that has however been criticized to be misleading (Washington Post, 2015). Duggan et al. (2011) instead reports that in 1993/1994, transactions at gun shows only accounted for about 4% of all firearms transactions. A study by the Bureau of Alcohol, Tobacco and Firearms (1999) furthermore estimates that 50-75% of dealers at gun shows are federally licensed, therefore being required to perform background checks on all transactions, even at gun shows. Additionally, eighteen states and DC have passed laws that require federal background checks in most or all private transactions. Therefore, the majority of transactions at gun shows should be reflected in the NICS background checks. Since most of the states that employ some form of purchase delay are also those states that require background checks for private transactions, we would at most underestimate the difference in firearm acquisitions between instant and delayed states.



Map of the United States showing the distribution of gun shows in 2012 and 2013. Red states denote instant states. Blue states denote delayed states. Connecticut is shown in gray as we exclude the state from the sample. Each location with a gun show is represented by a green circle, the size of the green circle indicates the number of gun shows held at this location.

Figure 5.2: Locations of gun shows

To alleviate remaining concerns, we collect daily data from Google Trends for the search expression 'gun show' for each state between 2009 and 2014. We employ the same scaling procedure as explained above and thus obtain a measure that allows us to investigate the temporal development of demand for gun shows across states with or without waiting periods or delays. In total we obtain data from 40 states and the District of Columbia.[14] Addtionally, we collect data on the location and timing of gun shows. The website `http://www.gunshowmonster.com/` provides a database of future and past gun shows across the United States. The website allows users to make submissions, which after editorial approval will be published and therefore provides decent coverage: our sample contains 8764 gun shows between July 2009 and December 2014 in almost all US states. We aggregate gun shows on the county level for each month. Note that the sample is surely incomplete and possibly even skewed towards certain states with easier access to guns. We therefore only use this data in some supplementary estimations to show that the effects regarding the supply and demand for gun shows go in similar directions. Figure 5.2 shows the locations of gun shows in 2012 and 2013 using green circles, where the circle size increases in the number of gun shows at a certain location.

Crime

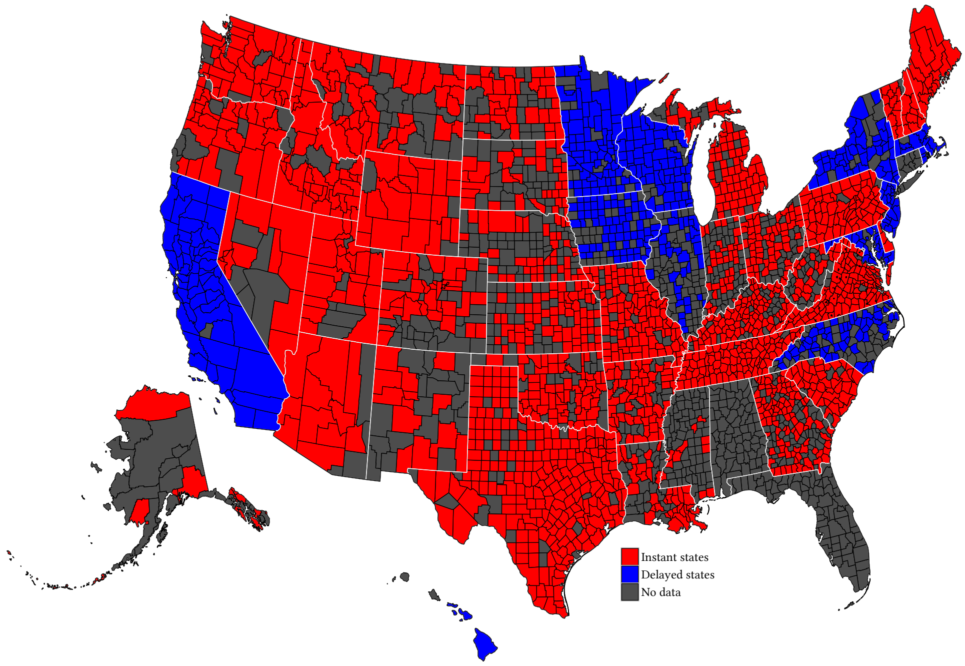
As our outcome measure in the second step where we investigate the effects of gun ownership on crime rates, we use the FBI's Uniform Crime Reports (UCR): Offenses Known and Clearances by Arrest (USDOJ: FBI, 2014a; USDOJ: FBI, 2015a).[15] Approximately 18,000 federal, state, tribal, county and local law enforcement agencies voluntarily submit detailed monthly crime data, either through their state's UCR program or directly to the FBI, about offenses known to these agencies. Variables include the monthly count of different types of crime for each law enforcement agency, such as murder, manslaughter, rape, assault, robbery, burglary, larceny and vehicle theft. The data also comprises variables that distinguish the type of weapon used (e.g. firearm, knife, strong arm) for example in robberies and assaults, and they allow to distinguish between severity for some crimes, such as simple assault versus aggravated assault, or forcible and non-forcible rape.

Unfortunately, some agencies in the data set are not reporting consistently. Common reporting mistakes include large negative absolute values for crimes, or continuously reporting zero crimes. We address this issue by following the guidelines for properly cleaning UCR data from Targonski (2011): First, we determine truly missing data points. An entry of zero could either mean that no crimes occurred, or that the agency was not reporting any crimes. An additional reporting variable however indi-

[14]Censoring removes Delaware, Hawaii, Maine, Nevada, New Hampshire, New Jersey, North Carolina, Rhode Island, Wyoming, again comprising instant and delayed states. Connecticut is also taken out of the sample.

[15]For a placebo regression using different years later in the main text we use (USDOJ: FBI, 2012; USDOJ: FBI, 2013).

rectly indicates, whether data was submitted. If no data was submitted, this reporting variable will have missing values for that specific date. We thus exclude all observations showing zero crimes, where the additional reporting variable contains missing values. Second, there are some obvious cases of data bunching, as there exist agencies that report their data only quarterly or (semi)annually, but no data in the months between. We identify those observations using an algorithm designed by Targonski and we also exclude them from the analysis.[16] Third, some smaller agencies choose to not report crimes themselves, but through another agency. In that case, they show up as reporting zeroes, although their counts are reflected in the data of the reporting agency. We drop those observations. Fourth, we apply the rule of 20 to identify wrongly reported zero crimes. Whenever an agency reports on average 20 or more crimes per month, it seems unlikely they experienced zero crime in any given month. Such data points are also excluded from our analysis. Fifth, we delete all observations with outlier values 999, 9999 and 99999 from the sample. Sixth, we remove all data containing negative values smaller than -3.[17]



Map of the United States showing our UCR sample. Red counties are located in instant states. Blue counties are located in delayed states. Grey counties are not present in the sample.

Figure 5.3: Counties represented in the UCR sample

[16]The algorithm is not part of Targonski (2011) but we received instructions and rules for the algorithm from Joe Targonski in a personal email exchange. The algorithm basically identifies any county (with absolute annual crime reports above 10) that report crimes only in March, June, September and December (or a subset of those for (semi-)annually reporters), and zero crimes in all other months. The email exchange is available from the authors upon request.

[17]In line with Targonski (2011) we ignore small negative values of at least -3. Those are usually corrections for misreporting in previous months.

In addition to the approach by Targonski (2011), we drop data from all counties that do not report for the full time period of our study, and all counties that always report zero crimes. If all reporting agencies in any county cover less than 50% of the county's population, we also remove these data points to make sure that control variables reflect the characteristics of the sample.[18] We then aggregate the crime data for all cases of murder, manslaughter, total rapes, forcible rapes, attempted rapes, total robberies (independent of type of weapon used), robberies with firearms, total assaults (independent of type of weapon used), aggravated assaults, simple assaults, assaults with firearms, total burglaries (including forced and non-forced entry), and total larceny cases, for each US county that we have data for, for each month of the years 2012 and 2013. This generates a data set of 13 types of crimes in 2,084 counties and approximately 50,000 observations. Figure 5.3 shows all counties that remain in our sample.

Crime: Supplementary Data Set

The UCR data only contains murder counts, but does not distinguish between the type of weapon used and does not provide additional information with respect to the circumstances of the crime. In order to determine who is most likely to be murdered as a result of the demand shock, we utilize the UCR Supplementary Homicide reports for 2012 and 2013 (USDOJ: FBI, 2014b; USDOJ: FBI, 2015b). These reports are compiled from information that law enforcement agencies voluntarily submit to the FBI and they contain detailed information about each homicide, such as gender of victim and offender and weapon used.

In addition to offering more detailed information on each homicide, they might also lead to a different count of murders than the UCR data due to the reporting issues mentioned above. Replicating the effect found in the UCR data with this additional data set would therefore be reassuring in the sense that our results are not driven by an inappropriate data cleaning procedure.

We aggregate the data for murder from the report to monthly counts on the county level. In addition, we generate counts for each of these crimes committed with any type of firearm, versus those committed without the use of a gun. We also build a count of murders for each gender separately and we count the number of murders in which the offender is male and the victim is female.

[18]The decision to set the cutoff at 50% was made arbitrarily, but even keeping all counties does not qualitatively change our results. Details will be given in the results section.

Controls

To control for potential confounds and account for differences in counties, we obtain several covariates. In selecting variables, we follow the choice of controls from the many correlational studies that investigate the relationship of firearm prevalence and crime (e.g. Cook and Ludwig, 2006; Kovandzic et al., 2012, 2013). From the 2010 US Decennial Census, we use (log) population, area, % rural, % African Americans, % Hispanics, and % below the poverty line, all on the county (and when necessary aggregated to the state) level.

5.4 Results

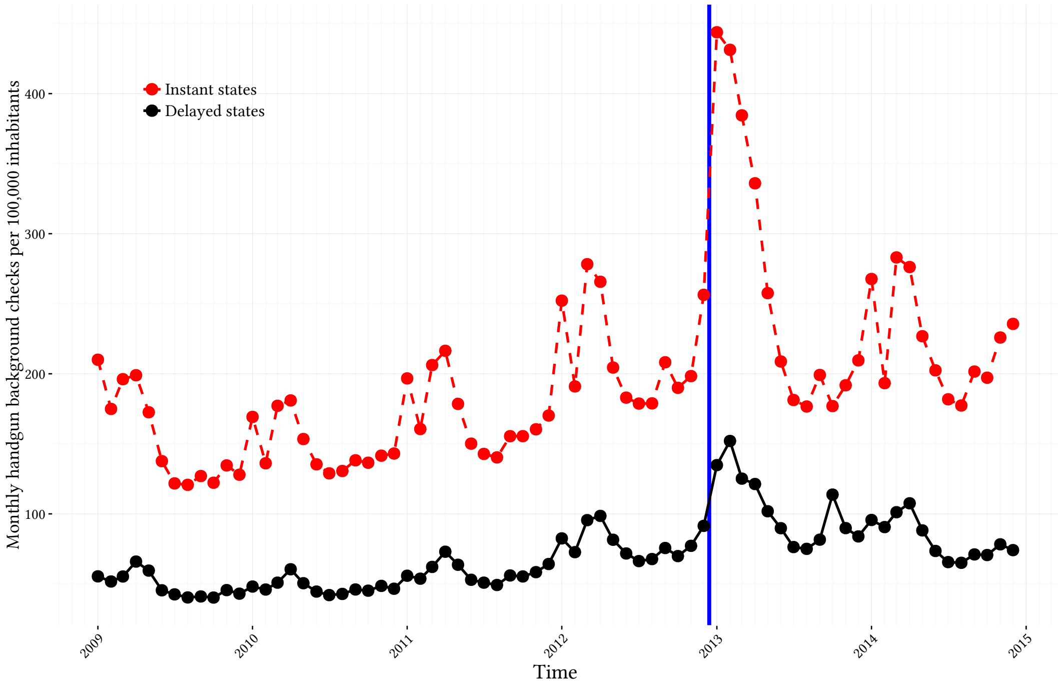
5.4.1 The Firearm Demand Shock After Sandy Hook

Since we predict some people to abstain from firearm purchases in delayed states, as a first step to our analysis, we investigate the effect of the shooting at Sandy Hook on NICS background checks for handguns. Higher demand after the shooting should translate into more purchases in both groups of states, but the increase should be more pronounced in instant states. Figure 5.4 shows a massive spike in background checks for instant states right after the shooting at Sandy Hook Elementary School.[19] There also seems to be a stronger than usual spike for the delayed states, but it is clearly smaller.

Employing our DiD estimation strategy, we estimate the differential effect of the shooting in instant and delayed states using regression analysis. Table 5.2 reports the results of a linear regression of the number of monthly background checks per 100,000 population on being in instant states after the shooting at Sandy Hook, as explained in section 5.3.1 above. Columns 1 through 3 estimate the effect for handguns, while columns 4 through 6 show the results for long guns. In all specifications, we include the years 2012 and 2013, i.e. approximately one year post and prior to the shooting at Sandy Hook. We chose this time frame to reduce the risk of picking up trend breaks that the linear trends cannot account for (for example due to other events).[20]

[19]Section E.1 in the appendix shows the respective figures for long gun purchases and total gun purchases.

[20]The results for handguns (our primary weapon of interest) do not depend on the choice of the time frame. Table E.3 in the appendix reports results for all available years of the Obama administration. Clearly, the effect decreases in the inclusion of additional time periods, but stays significant throughout. The results for long guns are less clear, but remember that our classification of instant and delayed states depends on delayed purchases of handguns which does not necessarily coincide with delays for long guns.



Monthly NICS background checks per 100,000 inhabitants for handguns in delayed states (black) and instant states (red) between 2009 and 2014. The blue vertical line marks the date of the shooting at Sandy Hook Elementary School.

Figure 5.4: Background checks for handguns in delayed vs instant states

Columns 1 and 4 contain a baseline specification without any controls or time trends, and just include month and state fixed effects. We cluster standard errors at the state level to account for serial correlation in outcomes, and regressions are weighted by the state population to not give less densely populated states more explanatory power than high density states. The effect for handguns and long guns is positive and significant, showing that purchases increased as a result of the shooting at Sandy Hook by more in instant states than in delayed states. In columns 2 and 5, we add our set of controls to take potential confounds into account. Included variables are the logarithm of population, area, % rural, % African American, % Hispanics, % below the poverty line (all interacted with month fixed effects to allow for a potentially different impact each month). In doing so, we account for the fact that gun ownership dynamics might be different in more rural versus more urban states, and might differ along dimensions such as income and education levels, for reasons not present in our estimation equation. The choice of these controls is not arbitrary, but follows the approach of (Cook and Ludwig, 2006; Kovandzic et al., 2012, 2013) addressing the critique of (Kleck, 2015) that most prior studies have either used irrelevant covariates or too few if any. Adding controls does not affect our results for handguns by much, but the coefficient for long guns drops while the standard error increases slightly, leading to an insignificant estimate. Columns 3 and 6 then also add a linear time trend for each state to account for potentially different pre-trends in the states. This strongly raises

Table 5.2: Handgun background checks

	Monthly handgun sale background checks per 100,000 inhabitants					
	Handgun			Long gun		
	(1)	(2)	(3)	(4)	(5)	(6)
Instant × Post	26.281***	25.525***	102.850***	20.997**	14.727	45.215**
	(8.116)	(9.358)	(28.486)	(8.539)	(10.127)	(21.180)
State FE	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	Y	N	Y	Y
State FE×t	N	N	Y	N	N	Y
States	50	50	50	50	50	50
Observations	1,200	1,200	1,200	1,200	1,200	1,200
Mean DV	162.61	162.61	162.61	189.48	189.48	189.48
R^2	0.882	0.905	0.919	0.890	0.924	0.937

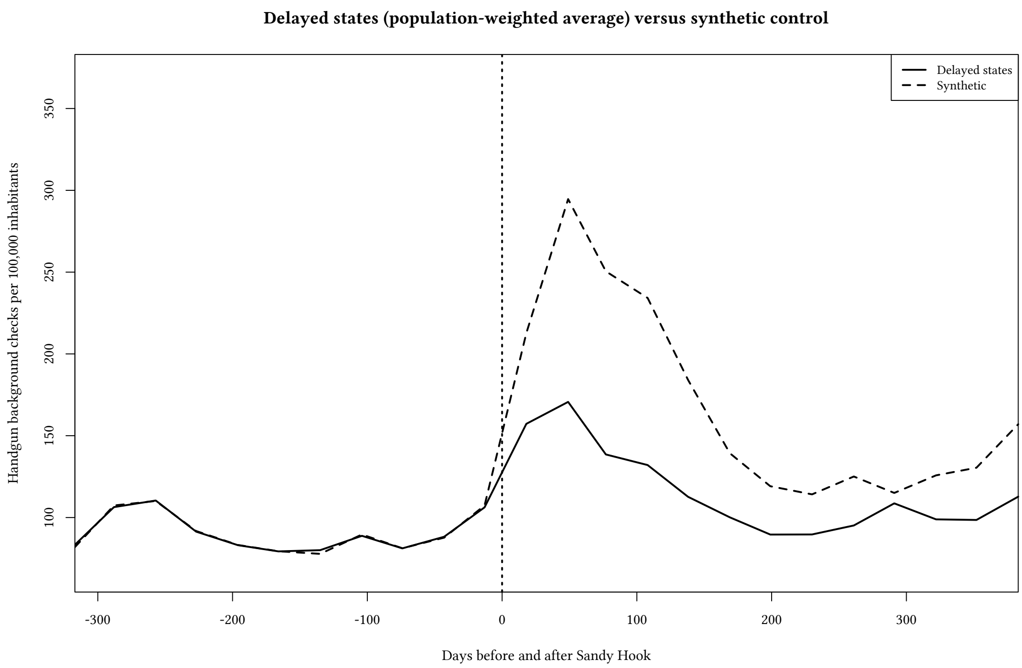
Notes: Coefficients from linear panel regressions of monthly handgun sale background checks per 100,000 inhabitants on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes months after the shooting at Sandy Hook. The level of observation is the state level. We include data from the years 2012 and 2013. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by state population.

coefficients for both types of guns, and both are highly significant, effectively showing that diverging pre-trends are not a concern in our setting. Note that each column indicates the weighted mean of the dependent variable to allow for a judgment of the treatment magnitude. In the case of our specification in column 3, our estimated effect implies a 63% stronger increase in handgun sales in instant over delayed states.

While all three specifications show a significant effect that led to more handgun purchases in instant over delayed states following the shooting at Sandy Hook, the specifications in columns 3 and 6 are our preferred estimates. Including location-specific time trends ensures that differential pre-trends do not bias the results, and allowing controls to vary over time provides a very flexible approach to dealing with regional idiosyncrasies that may amplify seasonal patterns. Therefore, we continue using this specification throughout the rest of the chapter, and provide robustness checks relative to this estimation strategy.

Although including state-specific trends should effectively deal with potential non-parallel trends, we provide an additional robustness check: Figure 5.5 provides a comparison of delayed states with a synthetic control of instant states. The method of synthetic control studies has first been used by Abadie and Gardeazabal (2003) to model the impacts of terrorism in the Basque country, and subsequently been formalized by Abadie et al. (2012). Since its inception, it has become a popular method for comparative case studies in many fields, where only some geographical region or object has received treatment, including but not limited to public finance (Kleven et al., 2013), financial economics (Acemoglu et al., 2016), and macroeconomics (Cavallo et al., 2013).

The method is a powerful tool to control for pre-treatment violations of the common trends assumption, as it constructs a synthetic control from the set of all control regions matching the pre-treatment outcomes of the treated regions. Because we have more instant than delayed states, and in contrast to our earlier regression analysis, we redefine delayed states to be the treated group and create a synthetic control from the set of instant states. Since our data comprises more than one treated state, we first average handgun sales over all delayed states, weighted with their respective population. Then, we construct a synthetic control from all instant states to match the handgun sales from the single delayed state. While we obtain a good match of pre-treatment outcomes using this method, Figure 5.5 shows that handgun sales in the synthetic control of instant states rise strongly and remain above delayed states for the full course of a year after the shooting at Sandy Hook.[21] We consider this more reassuring evidence that pre-trends are not biasing our results.



Handgun background checks per 100,000 inhabitants in delayed states and a synthetic control of instant states in the months before and after the shooting at Sandy Hook Elementary School. The vertical line indicates the date of the shooting.

Figure 5.5: Handgun sales in delayed states and synthetic control states

To test whether present-bias or transaction costs are more likely to explain our findings, we analyze if the ex ante willingness to purchase a gun was affected similarly by the demand shock. We test this assumption using daily Google Trends data on searches for the word ‘gun store’, which has been shown to be a good predictor of firearm purchasing intentions (Scott and Varian, 2014). Table 5.3 uses our preferred

[21] Figure E.3 in the appendix shows that this does not necessarily hold for long guns.

regression specification from Table 5.2, but now utilizes a varying time window around the shooting at Sandy Hook for our Google search expression of interest. Column 1 inspects the variation in Google searches for the first seven days before and after the shooting, which we expand to 30 days in column 2, 90 days in column 3, 365 days in column 4 and finally 730 days (= 2 years) in column 5. None of the specifications show a difference in firearm purchasing interest that would reject our notion that instant and delayed states were equally affected from the demand shock.[22] In fact, the coefficients across all specifications are very similar and small compared to the mean of the dependent variable. We therefore feel confident arguing that although gun purchases differed substantially, the intent to buy a gun was largely similar across delayed and instant states.

As we explained earlier, this similar intent to buy a firearm cannot be reconciled by publicly known transaction costs alone, but rather needs naïve present-biased gun buyers who credibly believe to purchase a gun in future periods. The important assumption under which this finding holds however is that prospective buyers are well-informed about their state's gun laws, especially about waiting periods and delayed purchasing permits. We argue that this assumption is reasonable for several reasons. First, searches for 'gun store' do not capture interest in learning about gun laws, but rather intend to locate the closest gun store. Since most people presumably know that there exist differing gun laws in their respective states, they will in most cases research the process of obtaining a gun before finding a local dealer. Second, since a transaction cost argument requires prospective buyers to be marginal after the shock, they will potentially at other points in time have thought about owning a gun already and should therefore be more likely to be familiar with gun laws. This is especially true if the shock did not extremely shift preferences for firearms. Third, since the estimated effect is not significant in any of our specifications, and also very small compared to baseline levels, the share of users not knowing about gun laws would need to be substantial. We conducted a short survey on surveymonkey.com in which we asked 119 participants about the gun laws in their state. Of the 113 respondents that completed all questions, 88 (= 78%) correctly identified whether their state implemented waiting periods or required purchasing permits. Excluding all participants that are sure that they will not buy a gun in the next two years increases the share of respondents to correctly identify their state's gun laws only slightly to 82% (46 of 56). We therefore deem the assumption of knowledge about gun laws at the time of search for gun store locations as viable and argue that transaction costs alone are therefore not responsible for the divergence, but rather time-inconsistent plans by prospective buyers due to procrastination.

[22]Figure E.4 in the appendix shows the development of Google searches between November 2011 and January 2013 graphically.

Table 5.3: Google searches for 'gun store'

	Daily normalized Google searches for 'gun store'				
	7 days	30 days	90 days	365 days	730 days
	(1)	(2)	(3)	(4)	(5)
Instant × Post	0.600	1.715	1.073	−0.629	0.068
	(2.011)	(2.069)	(0.886)	(0.545)	(0.474)
State FE	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y
State FE×t	Y	Y	Y	Y	Y
States	35	35	35	35	35
Observations	525	2,135	6,335	25,553	51,103
Mean DV	9.36	10.19	9.47	10.17	10.07
R^2	0.865	0.780	0.754	0.709	0.690

Notes: Coefficients from linear panel regressions of daily Google searches for the expression 'gun store' on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes days after the shooting at Sandy Hook. The level of observation is the state level. Each column indicates how many days prior and post to the shooting at Sandy Hook we use. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by state population.

We have already argued that transactions at gun shows should be largely reflected in our background check data or should bias the effect downwards. To ensure that the demand for gun shows had not changed more strongly in delayed states, we use the regression specification from Table 5.3 using Google Trends data on the search term 'gun show'. Table 5.4 reports the findings of stepwise expanding the time window around the shooting. In the first 90 days (columns 1 through 3) we pick up a positive effect that is significant mostly at the 10% level, suggesting that if anything, shifts to the secondary market were stronger in instant than in delayed states.[23] Expanding the time frame further makes the effect insignificant, suggesting that demand did not shift to secondary markets more strongly in delayed states. Section E.3.2 in the appendix additionally provides some tentative evidence that the supply of gun shows did not tilt towards delayed states either and the results qualitatively match the findings for gun show demand.

Overall, while citizens in instant and delayed states were similarly interested in acquiring firearms, we find that purchases were affected more strongly in instant states, as some buyers in delayed states postponed their purchases, leading to differential ownership dynamics. These results are not driven by differences in pre-existing trends or by diversion to secondary markets. In the following, we use this differential firearm take-up to investigate changes in crime rates.

[23]Figure E.5 in the appendix depicts the evolution of Google searches graphically. It also shows that there are some gaps in searches between the states already before and also after the shooting event that appear to be non-systematic.

Table 5.4: Google searches for 'gun show'

	Daily normalized Google searches for 'gun show'				
	7 days	30 days	90 days	365 days	730 days
	(1)	(2)	(3)	(4)	(5)
Instant × Post	4.417*	4.091**	2.912*	0.984	0.687
	(2.605)	(1.965)	(1.642)	(1.077)	(0.897)
State FE	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y
State FE×t	Y	Y	Y	Y	Y
States	41	41	41	41	41
Observations	615	2,501	7,421	29,971	59,901
Mean DV	8.82	12.05	11.73	12.12	12.53
R^2	0.844	0.775	0.748	0.715	0.704

Notes: Coefficients from linear panel regressions of daily Google searches for the expression 'gun show' on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes days after the shooting at Sandy Hook. The level of observation is the state level. Each column indicates how many days prior and post to the shooting at Sandy Hook we use. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by state population.

5.4.2 The Effect of Firearm Availability on Crime

Since following the shooting at Sandy Hook Elementary, there were more new guns owned in instant than in delayed states, we would like to know how this subsequently affected crime rates. Table 5.5 reports results from linear regressions, using our preferred specification from our first step analysis regarding changes in gun ownership. The regressions include the years 2012 and 2013, i.e. the years prior and after the shooting at Sandy Hook.[24] In column 1 we regress the sum of all crimes per 100,000 inhabitants in each county on our treatment indicator (being in an instant state after the shooting at Sandy Hook). We include county and month fixed effects, and report robust standard errors clustered at the state level and to not give less densely populated counties more explanatory power, we weigh the regressions by population. The effect on overall crime is positive, but far from statistical significance. In column 2, we add our set of control variables to account for potential confounds. The coefficient now turns negative, but is still not significant. We include a linear time trend for each county in column 3. This approach insures us against simply picking up diverging pre-trends, and the effect stays insignificant. Column 4 and 5 of Table 5.5 split up the results from column 3 in the most common violent and non-violent crimes according to the UCR definition: A violent crime is defined as either being murder, non-negligent

[24] This choice of time frame is motivated by two arguments: First, the previous section established that the effect of the shooting on differences in gun purchases is strongest for this shorter time frame, suggesting that the effects eventually wash out; and second, a shorter time frame reduces the risk of picking up trend breaks that the linear trends cannot account for. Robustness checks below will vary this time frame.

manslaughter, forcible rape, robbery, or aggravated assault. Non-violent crimes are all remaining crimes. Even now, our regressions do not pick up significant effects. Importantly, the estimated coefficients in all specifications seem extremely small compared to the baseline value of the dependent variable.

Table 5.5: Baseline: All crimes

	Monthly incidents per 100,000 inhabitants				
	Total			Violent	Nonviolent
	(1)	(2)	(3)	(4)	(5)
Instant × Post	1.012	−7.376	3.937	1.057	2.880
	(4.534)	(13.826)	(7.242)	(0.811)	(6.577)
County FE	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y
Controls	N	N	Y	Y	Y
County FE×t	N	Y	Y	Y	Y
Counties	2084	2084	2084	2084	2084
Observations	50,016	50,016	50,016	50,016	50,016
Mean DV	331.03	331.03	331.03	30.16	300.87
R^2	0.941	0.951	0.956	0.917	0.952

Notes: Coefficients from linear panel regressions of crime incidents per 100,000 inhabitants on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes months after the shooting at Sandy Hook. The level of observation is the county level. We include data from the years 2012 and 2013. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by county population.

To investigate the issue further, we break up violent crimes into their components. Table 5.6 shows the regressions in our preferred specification for each type of violent crime.[25] The first column is just a repetition of column 4 in the previous table. Column 2 reports the results on murder and shows that instant states experience significantly higher murder rates after the shock in gun demand following the shooting at Sandy Hook. The estimated effect size of 0.055 additional murders per 100,000 population constitutes a 14.47% increase in murders over delayed states. Since the population of instant states comprises approximately 178 million people, this implies that gun purchasing delays prevented approximately 98 murders in each month of 2013, and up to 1180 lives could have been saved from murder in instant states in only one year had such legislation been in place. Our other specifications in columns 3 through 6 show no significant effect on manslaughter, rape, robberies, or assaults and the estimated coefficients are much smaller compared to their baseline value than for murder. We thus find no evidence of a deterrence effect: None of the violent crimes experiences a significant downshift to counteract the uptick in murder rates.

[25] Table E.4 in the appendix shows for murder, manslaughter and aggravated assault that the results remain qualitatively unchanged if we use state time trends instead of county time trends in our regressions. In fact, the coefficients and standard errors remain almost unchanged. Using time trends for large geographical regions however strongly decreases the coefficients for all crimes and makes the

Table 5.6: Violent crimes

	Monthly incidents per 100,000 inhabitants					
	All	Murder	Mansl'ter	Rape	Robbery	Agg. Assault
	(1)	(2)	(3)	(4)	(5)	(6)
Instant × Post	1.057	0.055***	0.003	0.106	0.458	0.438
	(0.811)	(0.021)	(0.005)	(0.139)	(0.534)	(0.414)
County FE	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
County FE×t	Y	Y	Y	Y	Y	Y
Counties	2084	2084	2084	2084	2084	2084
Observations	50,016	50,016	50,016	50,016	50,016	50,016
Mean DV	30.16	0.38	0.01	2.34	9.05	18.39
R^2	0.917	0.363	0.095	0.478	0.945	0.859

Notes: Coefficients from linear panel regressions of crime incidents per 100,000 inhabitants on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes months after the shooting at Sandy Hook. The level of observation is the county level. We include data from the years 2012 and 2013. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by county population.

Table 5.7 reports our findings for the most common non-violent crimes. Column 1 repeats the last column from Table 5.5. Column 2 reports results on burglary, column 3 on larceny and column 4 on vehicle theft. None of these crimes see a significant in- or decrease in instant states following the shooting at Sandy Hook and the estimated effects are small. We find the non-results on larceny and vehicle theft particularly reassuring, since changes in these crimes could easily be spurious correlations—finding an ex ante reason for why they should be affected by changes in gun ownership seems challenging (Lott, 2013, p. 29).

A non-effect in assaults or robberies could however still mean that a higher share of these crimes were now committed using a firearm. In Table 5.8, we therefore split up assaults and robberies in those committed with a gun and those without. Column 1 reports the overall effect on robberies and column 2 reports the effect on robberies at gunpoint. Column 3 and 4 report the same specification for assaults. All four models lack statistical significance and exhibit coefficients close to zero, suggesting that the increased gun ownership did not lead to a significantly higher or lower share of assaults or robberies being committed with a firearm. It therefore seems that the only criminal activity these newly acquired guns were used for were homicides, a finding consistent with impulsive acts of violence, which gun purchasing delays aim to prevent.

Our results show that murder rates strongly increased in instant states as compared to delayed states after the gun demand shock following the shooting at Sandy Hook

results for murder insignificant. Combining these findings suggests that relevant convergence processes happen at the state level.

Table 5.7: Non-violent crimes

	Monthly incidents per 100,000 inhabitants			
	All	Burglary	Larceny	Veh.Theft
	(1)	(2)	(3)	(4)
Instant × Post	2.880	0.951	5.145	−0.429
	(6.577)	(2.170)	(4.227)	(0.588)
County FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y
County FE×t	Y	Y	Y	Y
Counties	2084	2084	2084	2084
Observations	50,016	50,016	50,016	50,016
Mean DV	300.87	53.07	155.92	19.34
R^2	0.952	0.870	0.924	0.925

Notes: Coefficients from linear panel regressions of crime incidents per 100,000 inhabitants on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes months after the shooting at Sandy Hook. The level of observation is the county level. We include data from the years 2012 and 2013. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by county population.

Elementary School.[26] The estimated effect size suggests that a substantial number of lives could have been saved had access to guns been delayed in all states. To ensure that our effect on murder rates is not driven by confounding factors, the following section provides detailed robustness checks to our regressions from this section.

5.4.3 Robustness Checks

To address concerns that our regressions are merely picking up seasonal effects that are different in delayed and instant states respectively, we repeat our regression from Table 5.6, but now instead of December 14, 2012 use December 14, 2010 as the event, and include only the years 2010 and 2011 in our estimation to mimic the data range used previously.[27] This placebo regression should only be picking up an effect if there is something particular about December and adjacent months that causes our estimates to be invalid. Table 5.9 reports the estimates for changes in violent crimes after December 14, 2010. None of the crimes seems to be significantly affected, and importantly, so is murder. The coefficient in fact is much closer to zero, suggesting that the previously uncovered effect can be attributed to the treatment rather than seasonal variation across groups of states.

[26] Table E.5 in the appendix presents the results for murder, manslaughter and aggravated assault if all reporting agencies (even if they cover less than 50% of the population of their respective county) are included. The results are qualitatively similar.

[27] We chose 2010, because 2010 and 2011 were the latest years that were free of large mass shootings with more than 20 victims.

Table 5.8: Crimes by type of weapon

	Monthly incidents per 100,000 inhabitants			
	Robbery		Assault	
	All	Gun	All	Gun
	(1)	(2)	(3)	(4)
Instant × Post	0.458	−0.120	0.438	0.012
	(0.534)	(0.170)	(0.414)	(0.171)
County FE	Y	Y	Y	Y
Month FE	Y	Y	Y	Y
Controls	Y	Y	Y	Y
County FE×t	Y	Y	Y	Y
Counties	2084	2084	2084	2084
Observations	50,016	50,016	50,016	50,016
Mean DV	9.05	3.39	18.39	3.92
R^2	0.945	0.914	0.859	0.799

Notes: Coefficients from linear panel regressions of crime incidents per 100,000 inhabitants on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes months after the shooting at Sandy Hook. The level of observation is the county level. We include data from the years 2012 and 2013. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by county population.

Even if the effect is not purely seasonal, it could also be caused by unrelated events taking place in our period of observation. Additionally one could be concerned that the effect is a construct of the choice of the time window around the event. Therefore, Figure 5.6 shows how the coefficient estimate on murder changes if we include data further away from the shooting. Initially, with only up to four months post and prior to the shooting, the effect is insignificant, but adding more data moves the coefficient in the right direction and drastically decreases standard errors. Increasing the time window around the shooting far enough leads to a positive estimate, starting at five months post and prior the shooting. The effect stays largely the same and significant throughout the first nine months, but then eventually diminishes until it is not significant anymore if more than 14 months post and prior the shooting are being used. We are therefore confident that our result is not simply an outlier driven by a favorable choice of the time window. Additionally, it seems unlikely that events further than six months before or after the shooting initiated the observed effect, effectively excluding the 2013 US Federal Government shutdown, the 2013 DC Navy Yard shooting, but not ruling out the November 2012 re-election of President Obama and the July 2012 shooting in a movie theater near Denver, Colorado, which constitute the most notable events in the United States in 2012 and 2013.

Both of the these events however would presumably contribute directionally similarly to the effect. As pointed out earlier, mass shootings during the Obama presidency have throughout increased gun sales. This would mean that our findings constitute the lower bounds of the effects, as the 2012 Denver shooting would have differentially

Table 5.9: Placebo regressions of violent crimes

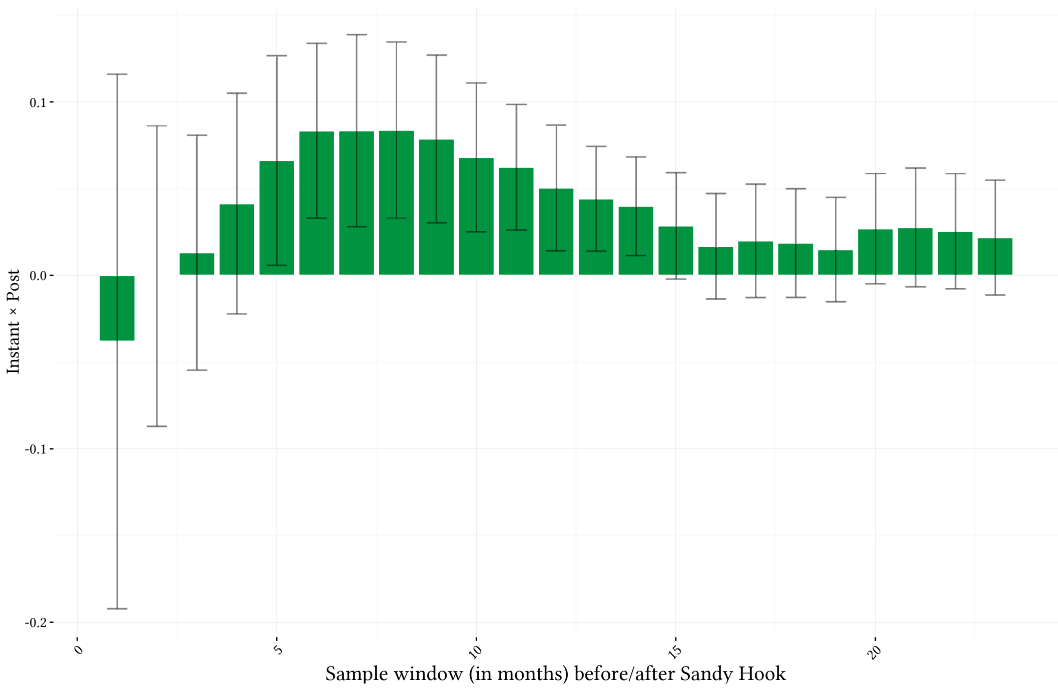
	Placebo: Monthly incidents per 100,000 inhabitants					
	All	Murder	Mansl'ter	Rape	Robbery	Assault
	(1)	(2)	(3)	(4)	(5)	(6)
Instant × Post	0.995	0.002	0.001	0.074	0.670	0.249
	(1.130)	(0.045)	(0.006)	(0.072)	(0.608)	(0.640)
County FE	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
County FE×t	Y	Y	Y	Y	Y	Y
Counties	2086	2086	2086	2086	2086	2086
Observations	50,064	50,064	50,064	50,064	50,064	50,064
Mean DV	31.32	0.38	0.01	2.2	9.48	19.27
R^2	0.923	0.376	0.094	0.445	0.945	0.869

Notes: Coefficients from linear panel regressions of crime incidents per 100,000 inhabitants on our treatment indicator and controls. Since this is a placebo regression, Instant denotes a dummy for instant states, Post denotes months after the placebo date of December 14, 2010. The level of observation is the county level. We include data from the years 2010 and 2011. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by county population.

affected gun sales already, making the effects due to the Sandy Hook shooting look smaller than they actually are. Similarly, the November 2012 election increased gun sales as citizens feared tougher gun legislation from a Democratic president (CNN Money, 2012).

There could also be the concern that the shooting at Sandy Hook not only influenced firearm demand, but also had a direct impact on crime rates. When comparing the differences in crime rates across delayed and instant states, we implicitly assume that they were driven by the different gun ownership dynamics and not directly changed by the event. It could however be conceivable that a violent act such as a mass shooting influenced attitudes towards violence differently in instant and delayed states, as the states differ across many dimensions. One way to address this concern is to find a non-violence related shock to firearm demand. One example is the 2012 Presidential election. Former Governor Mitt Romney took a more a liberal position on gun rights than President Obama, which earned him the endorsement of the NRA. After the election, gun sales increased strongly (CNN Money, 2012).

Unfortunately, the 2012 Presidential election happened shortly before the shooting at Sandy Hook Elementary School in November 2012. Separating the two events therefore seems impossible, such that we can only determine if the election contributed to the overall observed effect. Figure 5.7 reports the coefficients and corresponding 90% confidence intervals of the murder regression from Table 5.6 when the treatment date is shifted to an earlier or later month. Clearly, the effect is strongly positive and significant at date 0, i.e. right after the shooting at Sandy Hook took place. Shifting



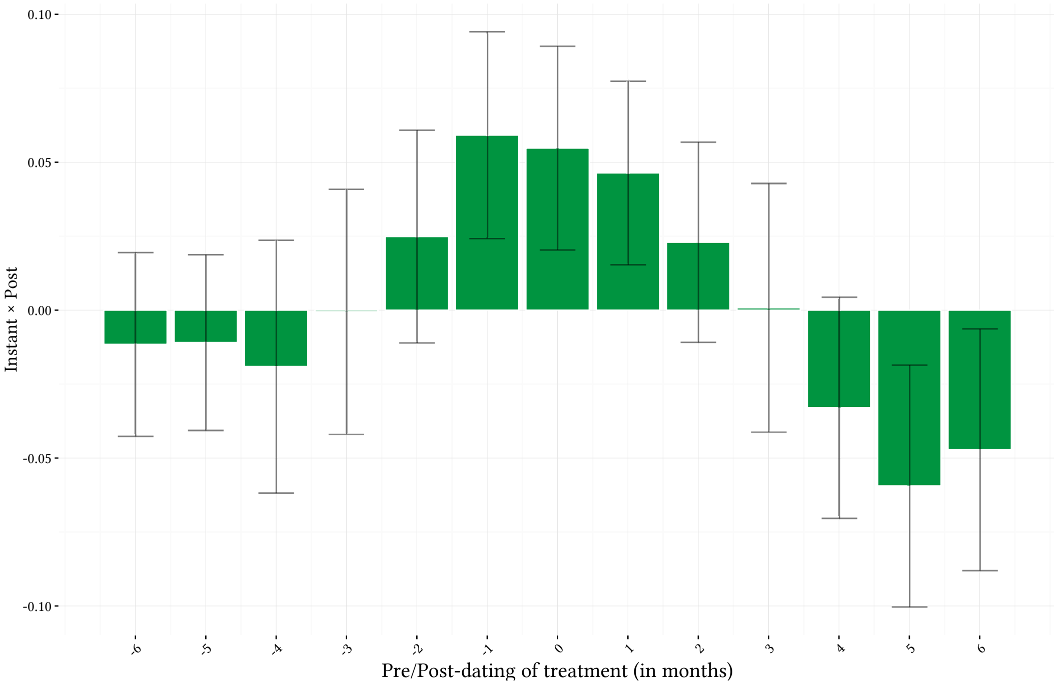
Coefficients on murder including an increasing number of months before and after the shooting at Sandy Hook in the regression and corresponding 90% confidence intervals. 1 means the month post and prior to the shooting are included (=2 months in total), 2 means that 2 months prior and post the shooting are included (=4 months in total), etc.

Figure 5.6: Time window on murder coefficient

the date to the future makes the effect less strong and it becomes insignificant if post-dated by more than one month. Reassuringly, the coefficient also falls monotonically for the first four months. Post-dating the event even further then makes the coefficient jumpy, as we probably pick up other events. Predating the treatment event by one period (to include the election in November 2012) however strongly increases the effect and it becomes more significant. This suggests that the 2012 election indeed positively contributed to the effect. Pre-dating the treatment even further than one month also delivers insignificant effects which is reassuring as the effect does not result from other events than the shooting and the election. We are therefore confident that our estimated effect on murder rates resulting from the shooting at Sandy Hook is indeed due to differences in firearm ownership rates across instant and delayed states.[28]

Finally, we would like to know how robust our estimation is to the exclusion of states. If the coefficients become very unstable upon exclusion of single states, we should be very cautious with interpreting the effect of delays to be a general finding across the United States. Figure 5.8 reports the results from our murder regression with 90% confidence intervals, removing one state at a time for all states in our sample. The first bar shows the estimated effect for the full sample, the second bar does not

[28]Section E.3.3 in the appendix provides directionally similar, yet insignificant results for the 2008 election.



Coefficients on murder when temporally shifting the onset of the treatment and corresponding 90% confidence intervals. 0 denotes the shooting at Sandy Hook. Positive numbers denote months after the shooting, while negative numbers denote months prior to the shooting.

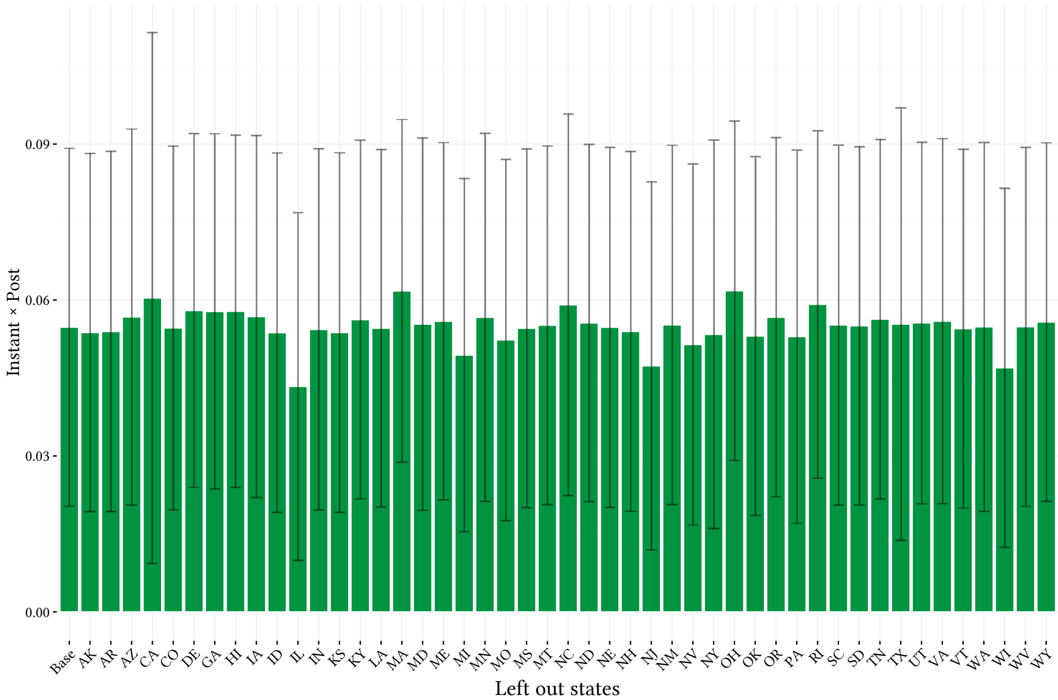
Figure 5.7: Timing of treatment onset

include Alaska, the third bar shows the full sample minus Arkansas, and so forth. Across all specifications, the coefficient is of similar magnitude. In fact, it never falls below .04 and rarely exceeds .06 and overall seems to exhibit relatively little variance. The confidence bands show that the effect is significant across all specifications.

5.4.4 Instrumental Variables Approach

One valid criticism towards our estimation strategy so far could be that our effect is more of an intention-to-treat rather than a true treatment effect on the treated. We have shown that "offering treatment" (being in an instant state after Sandy Hook) leads to "treatment" (experiencing higher gun sales) and that "offering treatment" leads to the effect (experiencing higher murder rates). This however disregards the variation in gun sales ("(non)-compliance") that can possible dilute our treatment effect.

For this reason, we provide results from a two stage least squares (2SLS) estimation. We first divide handgun sales by 1,000 to ease the interpretation of our estimated coefficients. Our first stage regresses these gun sales on the dummy of being in an instant state after Sandy Hook and our usual set of covariates, fixed effects and time trends. The predicted values we then use to estimate our second stage: How gun ownership affected crime rates. We do this separately for murder, manslaughter and



Coefficients on murder removing a single state (denoted on the x-axis) from the sample and corresponding 90% confidence intervals. Base denotes the inclusion of all states.

Figure 5.8: Coefficients on murder leaving out states

aggravated assault using two stage least squares (2SLS) and as a comparison estimate the structural equation with OLS to visualize the size of the problem that endogeneity would have caused when estimating the relationship of gun purchases and crime rates directly.

Table 5.10 presents the results. Column 1 shows the first stage and not surprisingly delivers a coefficient highly significant and of similar magnitude as earlier.[29] The F-statistic of 21.01 reported at the bottom of the column suggests that we do not face a weak instrument. Column 2 presents the OLS results of the structural equation for murder, while column 3 takes the first stage into account by using the 2SLS estimator. The coefficient implies that every 2049 additional guns lead to one additional murder.[30] Note that we would not have found an effect for murders had we not dealt with the endogeneity of gun sales (even after including control variables). Columns 4 and 5 repeat the exercise for manslaughter. A simple OLS estimation would have shown an effect, but the 2SLS estimator clarifies that this effect only arises due to endogeneity.

[29] Note that the sample is slightly different than in Table 5.2, because we have no data for Alabama and Florida in the second stage.

[30] Note that in the second stage both variables, gun sales and crimes, are expressed per 100,000 population, and gun sales are additionally expressed in thousands. This changes the interpretation of the coefficient compared to our earlier regressions. The coefficient now tells you by how much murders (in absolute numbers) increased for every 1,000 additional handguns sold.

Columns 6 and 7 provide results for aggravated assaults and neither column reports a significant effect.

Table 5.10: Instrumental variables regression

	Monthly incidents per 100,000 inhabitants						
	First Stage	Murder		Mansl'ter		Agg. Assault	
		OLS	2SLS	OLS	2SLS	OLS	2SLS
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Instant × Post	0.112***						
	(0.024)						
Handgun sales per 1,000		0.058	0.488***	0.037**	0.030	−2.829	3.901
		(0.067)	(0.186)	(0.018)	(0.045)	(2.195)	(3.684)
County FE	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y
County FE×t	Y	Y	Y	Y	Y	Y	Y
Counties	2084	2084	2084	2084	2084	2084	2084
Observations	50,016	50,016	50,016	50,016	50,016	50,016	50,016
R^2	0.118	0.363	0.363	0.095	0.095	0.859	0.859
IV F-stat.	21.01						

Notes: Coefficients from Instrumental Variables and OLS regressions of crime incidents per 100,000 inhabitants on gun sales per 100,000 inhabitants in thousands, where gun sales is instrumented by our treatment indicator. The first stage is reported in column 1. Instant denotes a dummy for instant states, Post denotes months after the shooting at Sandy Hook. OLS regressions estimate the structural equation only. The level of observation is the county level in the second stage and the state level in the first stage. We include data from the years 2012 and 2013. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by county/state population.

The results from our Instrumental Variables approach clearly show that stronger increases in gun ownership lead to more murders, supporting our earlier findings. Furthermore, the Instrumental Variables approach shows a similar magnitude of the effect. While the first stage implies that instant states saw gun sales that were elevated by approximately 200,000[31] in the months after the shooting at Sandy Hook, the effect from the second stage predicts that roughly 97 murders could have been prevented in each month of 2013, which is almost the 98 murders that our reduced form regressions predicted. This suggests that non-compliance does not play a large role in our setting, presumably due to a sensible choice of controls, and that therefore intent-to-treat and treatment effect on the treated almost coincide. Our results also emphasize the importance of credible identification in this setup, as the OLS results of the structural equation vary significantly from the 2SLS estimates and can therefore lead to wrong conclusions about the actual causal relationship between gun ownership and crime rates.

[31] The calculation involves a population estimate of 178 million for instant states.

5.4.5 Identifying Murder Circumstances

Our previous analyses have only predicted that instant states see additional murders, but we haven't learned much about the characteristics of those murders. The crime data employed so far was not detailed enough to distinguish between murders with guns and murders with other types of weapons. This section therefore provides results from regressions using the UCR Supplementary Homicide reports.

Columns 1 and 2 of Table 5.11 compare the results using our original UCR crime data set (UCR) with the UCR Supplementary Homicide Reports (SHR). Clearly, the coefficients are significant, positive and of similar magnitude in both cases. Note however that the R^2 is somewhat different and coefficients and standard errors are not identical, suggesting that there are some differences between the SHR and the UCR data. Although we performed extensive data cleaning on the UCR data, these differences could arise if the cleaning procedure did not eliminate all irregularities. We find it reassuring that the results are so similar, suggesting that there are only few misreports left in the remaining UCR sample.

Table 5.11: Results from Supplementary Homicide reports

	Monthly murders per 100,000 inhabitants							
	UCR: All	SHR: All	with Gun	w/o Gun	Victim Female	Victim Male	Men killing Women	All Other
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Instant × Post	0.055***	0.059***	0.070***	−0.010	0.018***	0.052***	0.015***	0.055***
	(0.021)	(0.017)	(0.018)	(0.015)	(0.007)	(0.018)	(0.005)	(0.018)
County FE	Y	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y
County FE×t	Y	Y	Y	Y	Y	Y	Y	Y
Counties	2084	2084	2084	2084	2084	2084	2084	2084
Observations	50,016	50,016	50,016	50,016	50,016	50,016	50,016	50,016
Mean DV	0.38	0.42	0.3	0.12	0.04	0.25	0.03	0.26
R^2	0.363	0.631	0.644	0.188	0.146	0.687	0.130	0.679

Notes: Coefficients from linear panel regressions of murders per 100,000 inhabitants on our treatment indicator and controls. Column 1 uses UCR data, all other columns are based on SHR data. Instant denotes a dummy for instant states, Post denotes months after the shooting at Sandy Hook. The level of observation is the county level. We include data from the years 2012 and 2013. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by county population.

We have yet to provide evidence that the observed increase in murder rates stems from the use of guns. Columns 3 and 4 of Table 5.11 split the SHR results by murder weapon, column 3 performs the regression for all murders in which any type of firearm was involved, while column 4 shows the results for all murders with all other weapons. The results propose that the entire effect can be attributed to murders with guns, as the coefficient in column 3 is positive and highly significant, while the coefficient in column 4 is comparably small and far from significance.

Columns 5 and 6 break up murder victims (where the weapon of choice was a firearm) by gender. The additional victims seem to be drawn from both genders, but when comparing the effect size to the mean of the dependent variable, it becomes apparent that women are relatively more affected than men: Murder rates for men increase by only 20.8%, while the increase for women is more than twice as high with 45%. This suggests that women face a (relatively) higher risk to be murdered from the additional guns that purchasing delays keep out of the hands of citizens. Columns 7 and 8 complement these findings by providing evidence from a somewhat different angle. Column 7 reports results for all gun murders in which the offender was male and the victim was female, while column 8 depicts the results for all other murders (female offenders and male offenders killing men). The results are very similar: The rate of men murdering women increases by 50%, while the rate of all other murders rises by only 21%, less than half as much.

While not providing definitive evidence, the results are suggestive of an increase in deadly domestic violence. While the typical murder victim and offender are both male, the fact that women are disproportionately affected indicates that relationship through blood or marriage between victim and offender could play a role. This would also align with the general observation that self-control capabilities are correlated across domains (Tangney et al., 2004). If this applied to both the time-inconsistency that keeps some citizens from buying guns and the lack of self-control that leads to the commission of impulsive acts of violence, it would deliver an explanation for our observations. This is of course highly speculative, but we think that the importance of this issue warrants further investigation of this mechanism which is beyond the scope of this chapter.

5.5 Conclusion

With the political debate on gun control gaining traction amid several recent mass shootings, understanding the consequences of legislation that limits access to firearms is imperative. The existing literature has delivered contradicting results and there exists reasonable doubt about identification in many previous contributions. This chapter contributes to this important topic by presenting well-identified effects on crime rates of an exogenous country-wide gun demand shock that led to differential gun purchases depending on whether states had implemented mechanisms to delay purchases or not.

Our results show that in states without any firearms purchasing delays, gun sales rose stronger following a country-wide gun demand shock as a result from the shooting at Sandy Hook Elementary School in Newtown, Connecticut. The strong effect is robust to a variety of controls and alternative specifications and does not seem to be caused by pre-existing time trends. Additional evidence suggests that the effect might be driven by time-inconsistent consumers postponing their gun purchases rather than simply transaction costs, as results show that the intention to acquire firearms right after the shooting was not significantly different across groups of states, but large differences appear in actual gun sales. Our analysis also addresses transactions on secondary markets and provides evidence that sales at gun shows are unlikely to confound our results.

We then use these findings to explore how crime rates changed. While all but one crime categories remain unaffected, murder rates increase significantly in states where gun purchases reacted more strongly. Several additional regressions and analyses address possible robustness and endogeneity issues. Our findings suggest that in each month of 2013 alone, approximately 98 murders could have been prevented if all states had implemented delays for gun purchases. The null finding regarding other crime categories suggests that armed citizens might not provide a strong deterrent effect. Including the 2012 United States Presidential election as an additional event that drove up gun sales makes the effect stronger, suggesting that our results are not due to a direct effect of mass shootings on crime rates, but rather through increased firearm ownership rates. We also look more closely at the additional murder victims and find that women are relatively more likely to being murdered. This is primarily driven by male offenders.

At this point we would like to emphasize that the aim of this chapter is not to speak in favor of the abolition of Second Amendment rights. The topic of allowing citizens to own guns is obviously very complex and has many dimensions that need to be taken into consideration, such as whether armed citizens can provide protection against exploitation by the government, or whether stripping legal gun owners from their rights will in the short run lead to more crime as potential victims become defenseless. Additionally, our results do not address the issue of widespread firearm access of criminals or structures of organized crime that are so pervasive in some larger US cities and that have their own roots and causes. Our study merely provides evidence on two aspects of the gun control debate, i.e. that more widespread gun ownership positively contributes to murder rates and that delaying gun purchases can help to prevent murders, especially of women. These findings should therefore be taken into account when deciding about the extent and the means of gun control legislation.

Finally, we see our study as a good starting point to investigate additional, related issues. First, additional direct evidence on the circumstances under which gun ownership leads to increased violent crime is needed. While we conclude that time-inconsistency is consistent with our findings, it would be interesting to see if differences in present-bias can actually be seen in individual gun buyers. Second, it would be interesting to determine the exact relationship of victims and offenders. While we can only speculate why women are more at risk of being murdered after such a shock, we would expect to see immediate family members and spouses to be more heavily affected if the murders result from impulsive acts of violence. Third, the effect of the Sandy Hook shooting on suicides remains completely unknown. Given the large evidence on the relationship of gun prevalence and suicides, knowing whether guns not particularly bought for the purpose of killing oneself actually impact suicide rates. We leave these issues for future research as they constitute different research questions or cannot yet be answered given the available data.

Appendix A

Risk, Time Pressure & Selection Effects

A.1 List of Binary Risky Choices

Table A.1: Lotteries Used in Risky Choice Tasks

#Set/# Lottery	Lottery A					Lottery B				
	Payoff 1	Prob. 1	Payoff 2	Prob. 2	EV	Payoff 1	Prob. 1	Payoff 2	Prob. 2	EV
S1 / G1	- 19.00	27.14 %	- 5.00	72.86 %	- 8.80	4.50	21.50 %	- 15.50	78.50 %	- 11.20
S1 / G2	- 8.00	15.00 %	- 4.00	85.00 %	- 4.60	3.00	15.00 %	- 7.00	85.00 %	- 5.50
S1 / G3	- 16.00	23.34 %	- 10.00	76.66 %	- 11.40	8.50	13.00 %	- 16.50	87.00 %	- 13.25
S1 / G4	- 9.00	40.00 %	- 2.00	60.00 %	- 4.80	7.00	40.00 %	- 15.00	60.00 %	- 6.20
S1 / G5	- 15.50	30.00 %	- 12.50	70.00 %	- 13.40	5.50	12.08 %	- 18.50	87.92 %	- 15.60
S1 / G6	- 15.00	12.00 %	- 7.50	88.00 %	- 8.40	5.00	27.50 %	- 15.00	72.50 %	- 9.50
S1 / G7	- 7.50	14.55 %	- 2.00	85.45 %	- 2.80	3.00	33.33 %	- 7.50	66.67 %	- 4.00
S1 / G8	- 16.00	40.00 %	- 8.00	60.00 %	- 11.20	4.00	10.00 %	- 14.00	90.00 %	- 12.20
S1 / G9	- 15.00	28.89 %	- 6.00	71.11 %	- 8.60	3.50	24.86 %	- 15.00	75.14 %	- 10.40
S1 / G10	- 16.00	10.00 %	- 12.00	90.00 %	- 12.40	6.00	10.00 %	- 16.00	90.00 %	- 13.80
S1 / G11	- 15.00	15.00 %	- 11.00	85.00 %	- 11.60	12.00	15.00 %	- 17.00	85.00 %	- 12.65
S1 / G12	- 11.00	30.00 %	- 6.00	70.00 %	- 7.50	5.00	30.00 %	- 15.00	70.00 %	- 9.00
S1 / L1	6.50	80.62 %	14.50	19.38 %	8.05	13.50	79.00 %	- 6.50	21.00 %	9.30
S1 / L2	5.00	80.00 %	15.00	20.00 %	7.00	15.00	75.00 %	- 5.00	25.00 %	10.00
S1 / L3	3.00	76.66 %	6.00	23.34 %	3.70	10.00	72.06 %	- 7.00	27.94 %	5.25
S1 / L4	6.00	75.00 %	14.00	25.00 %	8.00	14.00	80.00 %	- 12.00	20.00 %	8.80
S1 / L5	4.50	60.00 %	7.50	40.00 %	5.70	12.50	71.25 %	- 7.50	28.75 %	6.75
S1 / L6	7.50	70.00 %	12.50	30.00 %	9.00	14.50	80.44 %	- 8.50	19.56 %	10.00
S1 / L7	9.50	81.66 %	15.50	18.34 %	10.60	14.50	87.65 %	- 2.50	12.35 %	12.40
S1 / L8	4.00	75.00 %	9.00	25.00 %	5.25	14.00	70.00 %	- 11.00	30.00 %	6.50
S1 / L9	6.00	70.00 %	10.00	30.00 %	7.20	12.50	84.00 %	- 15.00	16.00 %	8.10
S1 / L10	2.00	65.00 %	7.00	35.00 %	3.75	7.00	85.00 %	- 7.00	15.00 %	4.90
S1 / L11	2.00	65.00 %	5.00	35.00 %	3.05	10.00	70.00 %	- 5.00	30.00 %	5.50
S1 / L12	5.00	75.00 %	15.00	25.00 %	7.50	10.00	90.00 %	- 4.00	10.00 %	8.60
S2 / G1	- 20.00	30.00 %	- 4.00	70.00 %	- 8.80	4.00	20.00 %	- 15.00	80.00 %	- 11.20
S2 / G2	- 8.00	32.00 %	- 3.00	68.00 %	- 4.60	4.00	13.64 %	- 7.00	86.36 %	- 5.50
S2 / G3	- 15.00	10.00 %	- 11.00	90.00 %	- 11.40	8.00	15.00 %	- 17.00	85.00 %	- 13.25
S2 / G4	- 10.00	30.67 %	- 2.50	69.33 %	- 4.80	7.00	41.33 %	- 15.50	58.67 %	- 6.20
S2 / G5	- 15.00	20.00 %	- 13.00	80.00 %	- 13.40	6.00	10.00 %	- 18.00	90.00 %	- 15.60
S2 / G6	- 14.00	20.00 %	- 7.00	80.00 %	- 8.40	4.00	25.00 %	- 14.00	75.00 %	- 9.50
S2 / G7	- 6.00	20.00 %	- 2.00	80.00 %	- 2.80	3.00	30.00 %	- 7.00	70.00 %	- 4.00
S2 / G8	- 17.00	31.76 %	- 8.50	68.24 %	- 11.20	4.50	12.11 %	- 14.50	87.89 %	- 12.20
S2 / G9	- 14.00	40.00 %	- 5.00	60.00 %	- 8.60	4.00	20.00 %	- 14.00	80.00 %	- 10.40
S2 / G10	- 16.00	20.00 %	- 11.50	80.00 %	- 12.40	6.50	9.78 %	- 16.00	90.22 %	- 13.80
S2 / G11	- 15.50	29.10 %	- 10.00	70.90 %	- 11.60	11.00	14.00 %	- 16.50	86.00 %	- 12.65
S2 / G12	- 12.00	30.77 %	- 5.50	69.23 %	- 7.50	5.00	26.31 %	- 14.00	73.69 %	- 9.00
S2 / L1	7.00	85.00 %	14.00	15.00 %	8.05	12.00	85.00 %	- 6.00	15.00 %	9.30
S2 / L2	4.50	72.22 %	13.50	27.78 %	7.00	13.50	80.56 %	- 4.50	19.44 %	10.00
S2 / L3	3.00	65.00 %	5.00	35.00 %	3.70	9.00	75.00 %	- 6.00	25.00 %	5.25
S2 / L4	6.50	82.35 %	15.00	17.65 %	8.00	15.00	77.86 %	- 13.00	22.14 %	8.80
S2 / L5	5.00	65.00 %	7.00	35.00 %	5.70	12.00	75.00 %	- 9.00	25.00 %	6.75
S2 / L6	8.00	80.00 %	13.00	20.00 %	9.00	15.00	80.00 %	- 10.00	20.00 %	10.00
S2 / L7	10.00	85.00 %	14.00	15.00 %	10.60	14.00	90.00 %	- 2.00	10.00 %	12.40
S2 / L8	3.50	70.83 %	9.50	29.17 %	5.25	14.00	70.59 %	- 11.50	29.41 %	6.50
S2 / L9	6.00	60.00 %	9.00	40.00 %	7.20	12.00	85.00 %	- 14.00	15.00 %	8.10
S2 / L10	3.00	85.00 %	8.00	15.00 %	3.75	8.00	80.63 %	- 8.00	19.37 %	4.90
S2 / L11	2.50	78.00 %	5.00	22.00 %	3.05	11.50	64.71 %	- 5.50	35.29 %	5.50
S2 / L12	4.50	66.66 %	13.50	33.34 %	7.50	11.00	82.86 %	- 3.00	17.14 %	8.60

Notes: Payoffs and expected values in €; S=set; G=prominent gain; L=prominent loss

A.2 Graphical Presentation of Risky Choices

A.3 Incentivization of Cognitive Ability Tasks

The IQ and IE tasks were incentivized in the following way. In each task we count the number of correct items. This number then determines the probability to win a fixed prize of €10 in the IQ and of €20 in the IE task. The probability is calculated as $P(\text{win } €10) = (\text{Number of correct items} + 1)/(12 + 1)$ in the IQ task; and as $P(\text{win } €20) = (\text{Number of correct items} + 1)/(34 + 1)$ in the IE task. That is, subjects have a clear incentive to solve as many items as possible, since their expected payoff is monotonically increasing in the number of correct items. At the same time, they always have a positive probability of winning the prize. The procedure was chosen to make sure that participants could never draw clear conclusions regarding their score or regarding the correct answer of single items. We wanted to avoid such inference because we did not want subjects to draw strong inference regarding their cognitive ability from our experiment. Note that subjects learned about the outcome of the random payment draw only at the end of the experiment if either Part 1 or Part 4 were randomly selected for real payment.

Appendix B

Unleashing Animal Spirits – Self-Control and Overpricing in Experimental Asset Markets

B.1 Period-specific Price Comparisons

Looking at single periods, it is possible to get a more precise picture of when the price differences between conditions arise. Table B.1 reports the per-period differences of volume-adjusted mean prices, trade-adjusted mean prices, RAD and RD between *LOWSC* and *HIGHSC*. The z-values from Mann-Whitney tests testing the equality of the respective measures across the two conditions are displayed in parentheses with significance levels indicated by asterisks. While in the first periods we see almost no price differences, starting from period five, markets in *LOWSC* exhibit significantly higher mean prices, mispricing, and overpricing, with the peak in period 8. There are no significant differences between the two conditions in the ultimate period. By definition, this implies a more pronounced bubble and burst pattern in *LOWSC* markets than in *HIGHSC* markets.

Table B.1: Period-specific Effects

Period	Δvolume-adjusted mean price	Δtrade-adjusted mean price	ΔRAD	ΔRD
1	-0.67	-0.85	0.0143	-0.0245
	(0.84)	(0.735)	(-0.63)	(0.84)
2	0.73	2.87	-0.0749	0.0266
	(0.105)	(-0.21)	(0.21)	(0.105)
3	4.53	3.38	0.0006	0.1646
	(-0.84)	(-0.525)	(-0.105)	(-0.84)
4	7.18	7.64 *	0.1720	0.2612
	(-1.47)	(-1.89)	(-1.26)	(-1.47)
5	9.24 *	9.03 *	0.2523	0.3359 *
	(-1.785)	(-1.785)	(-1.47)	(-1.785)
6	12.27 **	12.01 **	0.4186 **	0.4461 **
	(-2.205)	(-2.31)	(-2.205)	(-2.205)
7	15.90 **	15.84 **	0.5703 **	0.5781 **
	(-2.521)	(-2.415)	(-2.521)	(-2.521)
8	18.40 **	19.00 **	0.6573 **	0.6693 **
	(-2.521)	(-2.521)	(-2.521)	(-2.521)
9	11.69 **	11.78 **	0.4249 **	0.4249 **
	(-2.1)	(-1.995)	(-2.1)	(-2.1)
10	6.13	6.48	0.2007	0.2228
	(-1.26)	(-1.26)	(-1.05)	(-1.26)

Differences between *LOWSC* and *HIGHSC* and z-values (in parentheses) for Mann-Whitney tests. Volume-adjusted mean prices denote the average price per asset, while trade-adjusted mean prices denote average price per trade.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

B.2 Additional Regression Results

Table B.2: Determinants of Trading Activity

	(1)	(2)	(3)	(4)
	Average quantity traded			
LOWSC	-0.120	-0.0503	-2.915	-3.038
	(2.079)	(2.032)	(4.235)	(4.562)
CRT		-0.0287	-1.310*	-1.328*
		(0.842)	(0.700)	(0.716)
CE		0.685	0.558	0.542
		(0.802)	(0.820)	(0.868)
CRT × *LOWSC*			2.881*	2.874*
			(1.525)	(1.537)
CE × *LOWSC*			0.278	0.335
			(1.700)	(1.922)
Female				-0.295
				(1.942)
Constant	12.52***	11.06***	12.42***	12.69***
	(1.025)	(2.173)	(2.280)	(3.313)
Observations	110	110	110	110
R^2	0.000	0.006	0.035	0.036

OLS regression, dependent variable is individual average number of trades. LOWSC is a dummy where 1 stands for *LOWSC* and 0 for *HIGHSC*. CE is an individual's certainty equivalent. CRT denotes the number of correct answers on the CRT. Subjects who indicated they knew one or more of the CRT questions before were excluded. Heteroskedasticity robust standard errors clustered at market level in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table B.3: Determinants of Trading Activity (MIXED)

	(1)	(2)	(3)	(4)
	Average quantity traded			
MIXLO	-0.424	-0.828	-0.573	-0.586
	(2.252)	(2.157)	(5.110)	(5.143)
CRT		-1.309	0.579	0.541
		(0.807)	(1.429)	(1.462)
CE		1.653	1.217	1.161
		(0.973)	(1.191)	(1.289)
CRT × MIXLO			-3.984*	-3.950*
			(1.903)	(1.885)
CE × MIXLO			1.211	1.205
			(2.265)	(2.258)
Female				-0.422
				(1.602)
Constant	12.09***	9.805***	9.266***	9.691***
	(1.155)	(1.517)	(2.189)	(3.124)
Observations	88	88	88	88
R^2	0.001	0.040	0.084	0.085

OLS regression, dependent variable is individual average number of trades. MIXLO is a dummy where 1 stands for *MIXLO* and 0 for *MIXHI*. CE is an individual's certainty equivalent. CRT denotes the number of correct answers on the CRT. Subjects who indicated they knew one or more of the CRT questions before were excluded. Heteroskedasticity robust standard errors clustered at market level in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table B.4: Ratings of Emotions in MIXED Markets

	MIXHI	MIXLO	p-value
Excitement1	4.200	4.500	0.400
Fear1	2.100	2.175	0.395
Surprise1	3.600	4.050	0.103
Anger1	1.800	2.025	0.440
Relief1	2.825	3.250	0.161
Sadness1	1.525	1.725	0.324
Joy1	3.625	4.375	0.058*
Excitement2	3.425	4.200	0.042**
Fear2	1.900	2.575	0.014**
Surprise2	2.450	3.400	0.030**
Anger2	2.025	2.000	0.723
Relief2	3.275	4.150	0.233
Sadness2	1.950	1.725	0.622
Joy2	3.375	4.125	0.207
Emotion intensity	2.720	3.163	0.025**
Emotion valence	1.464	1.969	0.208
Emotion intensity1	2.811	3.157	0.123
Emotion valence1	1.754	2.069	0.123
Emotion intensity2	2.629	3.168	0.025**
Emotion valence2	1.604	1.944	0.400

Note: p-values from Wilcoxon-Signed Rank tests collapsing data on the market level by *MIXLO* and *MIXHI* respectively; emotion intensity is the average score over all emotion questions, emotion valence is the average score over all positive emotions minus the score over all negative emotions; variables ending in 1 or 2 relate to questions at the beginning (1) or the end (2) of the asset market, respectively; *** p<0.01, ** p<0.05, * p<0.1

Table B.5: Changes of Ex-post Emotion Ratings in MIXED Markets

	MIXHI	MIXLO	p-value
Diff excitement	-0.775	-0.300	0.232
Diff fear	-0.200	0.400	0.029**
Diff surprise	-1.150	-0.650	0.288
Diff anger	0.225	-0.025	0.575
Diff relief	0.450	0.900	0.441
Diff sadness	0.425	0.000	0.290
Diff joy	-0.250	-0.250	1.000

Note: p-values from Wilcoxon-Signed Rank tests collapsing data on the market level by *MIXLO* and *MIXHI* respectively; *** p<0.01, ** p<0.05, * p<0.1

Table B.6: Determinants of Individual RD Based on Sales

	(1)	(2)	(3)	(4)
		$IndRD_{sales}$		
LOWSC	0.355**	0.350**	0.605**	0.648**
	(0.146)	(0.143)	(0.243)	(0.221)
CRT		-0.0488	-0.0774	-0.0712
		(0.0395)	(0.0613)	(0.0608)
CE		0.00173	0.0584	0.0639
		(0.0551)	(0.0617)	(0.0634)
CRT × *LOWSC*			0.0684	0.0709
			(0.0800)	(0.0782)
CE × *LOWSC*			-0.146	-0.167
			(0.109)	(0.106)
Female				0.103
				(0.0655)
Constant	0.172	0.210	0.111	0.0164
	(0.106)	(0.147)	(0.104)	(0.110)
Observations	110	110	110	110
R^2	0.227	0.241	0.269	0.283

OLS regression, dependent variable is Individual Relative Deviation (IndRD) for sales, an individual equivalent to market level Relative Deviation (RD) restricted to sales only. *LOWSC* is a dummy where 1 stands for *LOWSC* and 0 for *HIGHSC*. CE is an individual's certainty equivalent. CRT denotes the number of correct answers on the CRT. Subjects who indicated they knew one or more of the CRT questions before were excluded. Heteroskedasticity robust standard errors clustered at market level in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table B.7: Determinants of Individual Miscpricing

	(1)	(2)	(3)	(4)
		$IndRD$		
LOWSC	0.383**	0.375**	0.723***	0.760***
	(0.138)	(0.134)	(0.160)	(0.147)
CRT		-0.0593	-0.0775	-0.0722
		(0.0349)	(0.0574)	(0.0568)
CE		-0.0155	0.0553	0.0600
		(0.0457)	(0.0461)	(0.0475)
CRT × *LOWSC*			0.0461	0.0483
			(0.0722)	(0.0728)
CE × *LOWSC*			-0.182**	-0.200**
			(0.0784)	(0.0774)
Female				0.0884
				(0.0584)
Constant	0.119	0.203	0.0648	-0.0159
	(0.0979)	(0.125)	(0.0713)	(0.0774)
Observations	110	110	110	110
R^2	0.299	0.326	0.370	0.382

OLS regression, dependent variable is Individual Relative Deviation (IndRD), an individual equivalent to market level Relative Deviation (RD). *LOWSC* is a dummy where 1 stands for *LOWSC* and 0 for *HIGHSC*. CE is an individual's certainty equivalent. CRT denotes the number of correct answers on the CRT. Subjects who indicated they knew one or more of the CRT questions before were excluded. Heteroskedasticity robust standard errors clustered at market level in parentheses, *** p<0.01, ** p<0.05, * p<0.1

Table B.8: Second-Period Differences in Trading Behavior

	Group Mean		
	MIXHI	*MIXLO*	p-value
$\overline{p_{bid}}$	29.400	31.803	0.510
$\overline{p_{ask}}$	50.383	55.751	0.039**
$\overline{q_{bid}}$	15.441	15.407	0.659
$\overline{q_{ask}}$	13.291	11.687	0.796
$\overline{time_{bid}}$	54.392	49.446	0.470
$\overline{time_{ask}}$	48.067	45.576	0.587
$\overline{firsttime_{bid}}$	42.484	40.912	0.683
$\overline{firsttime_{ask}}$	28.344	28.591	0.717

Variables starting with a p denote prices, q quantities and time variables refer to the time passed in the current period, thus lower values indicate behavior earlier on. *bid* and *ask* refer to posted bids and asks, p-values from Wilcoxon signed-rank tests with data collapsed on market and treatment level, *** p<0.01, ** p<0.05, * p<0.1

Table B.9: Determinants of Trading Profits

	(1)	(2)	(3)	(4)
	Trading Profits			
MIXLO	1.036	1.040	4.342*	4.301*
	(0.770)	(0.795)	(2.222)	(2.215)
CRT		1.084**	1.882***	1.757**
		(0.497)	(0.621)	(0.691)
CE		0.473	0.867	0.685
		(0.550)	(0.768)	(0.753)
CRT × MIXLO			-1.660**	-1.547**
			(0.642)	(0.690)
CE × MIXLO			-1.031	-1.051
			(1.125)	(1.098)
Female				-1.381
				(0.888)
Constant	7.035***	5.302***	3.936***	5.326***
	(0.441)	(1.097)	(1.323)	(1.638)
Observations	88	88	88	88
R^2	0.016	0.079	0.120	0.145

OLS regression, dependent variable is average trading profit from asset market in €˙ CRT denotes the number of correct answers in the CRT. Subjects who indicated they knew one or more of the CRT questions before were excluded. CE is the individual certainty equivalent, robust standard errors clustered on market level in parentheses. *** p<0.01, ** p<0.05, * p<0.1

B.3 Distribution of Answers in the Stroop Task

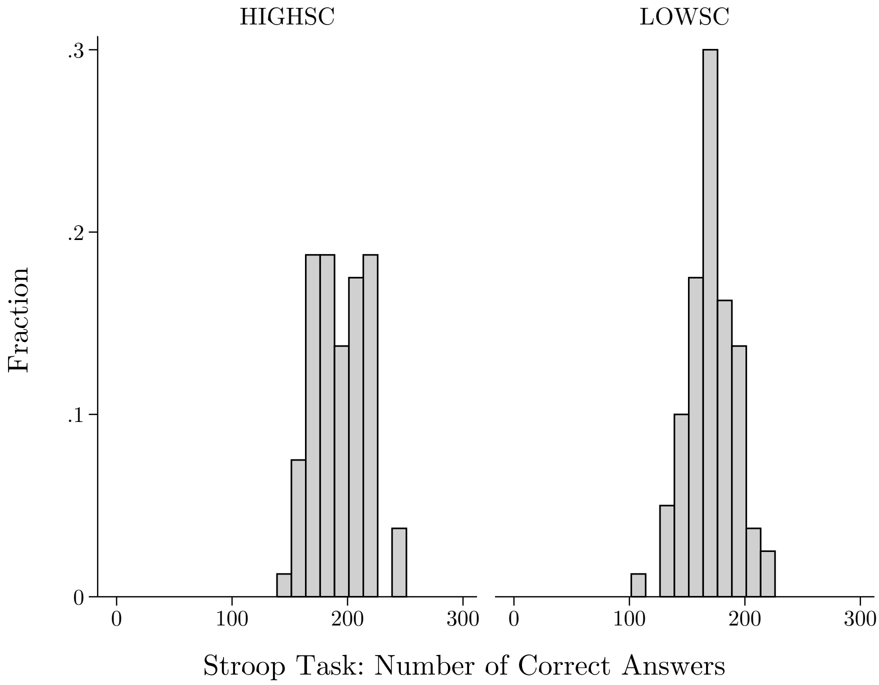


Figure B.1: Correct Stroop responses in *HIGHSC* vs. *LOWSC*

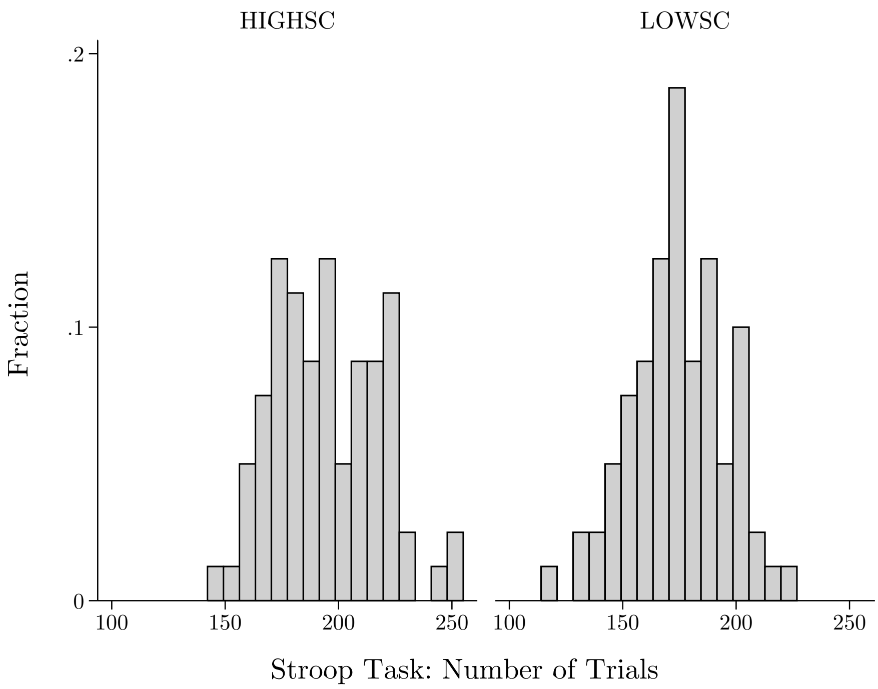


Figure B.2: Stroop Trials in *HIGHSC* vs. *LOWSC*

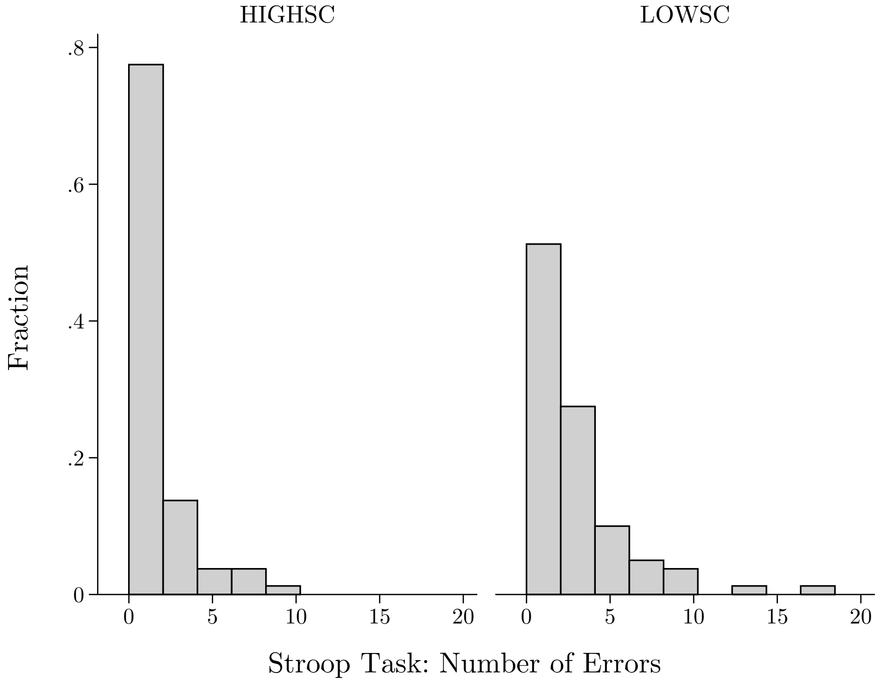


Figure B.3: Errors in the Stroop Task in *HIGHSC* vs. *LOWSC*

Table B.10: Distribution of Answers in the Stroop Task

Distribution of Answers in the Stroop Task

HIGHSC	Mean	Standard deviation
Correct Answers	192.65	22.6146
Trials	194.55	23.55973
Errors	1.9	1.879941
LOWSC	Mean	Standard deviation
Correct Answers	171.3125	20.68363
Trials	174.45	20.96948
Errors	3.14	2.971356

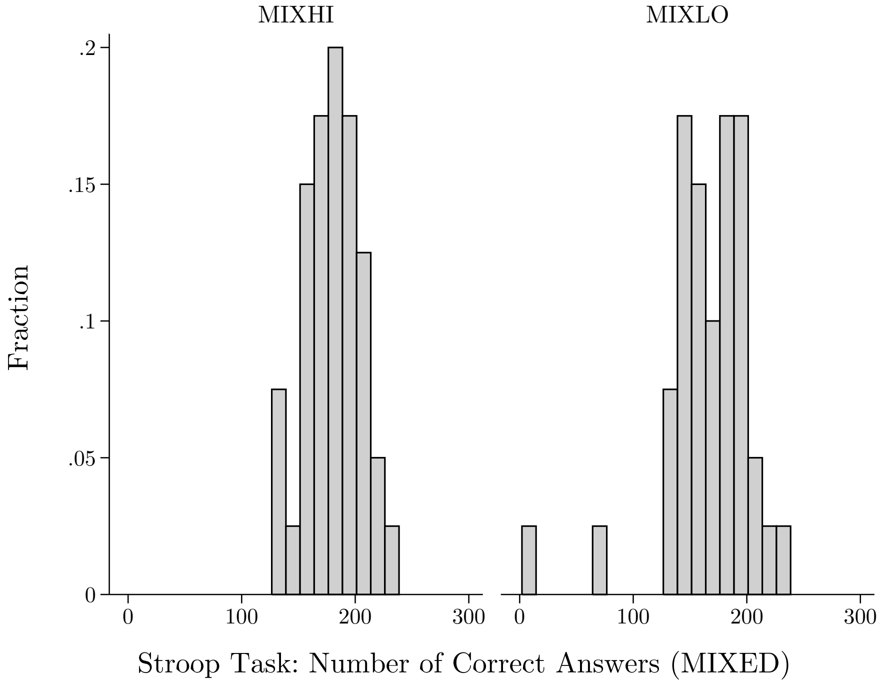


Figure B.4: Correct Stroop Responses in Treatment *MIXED* by Condition

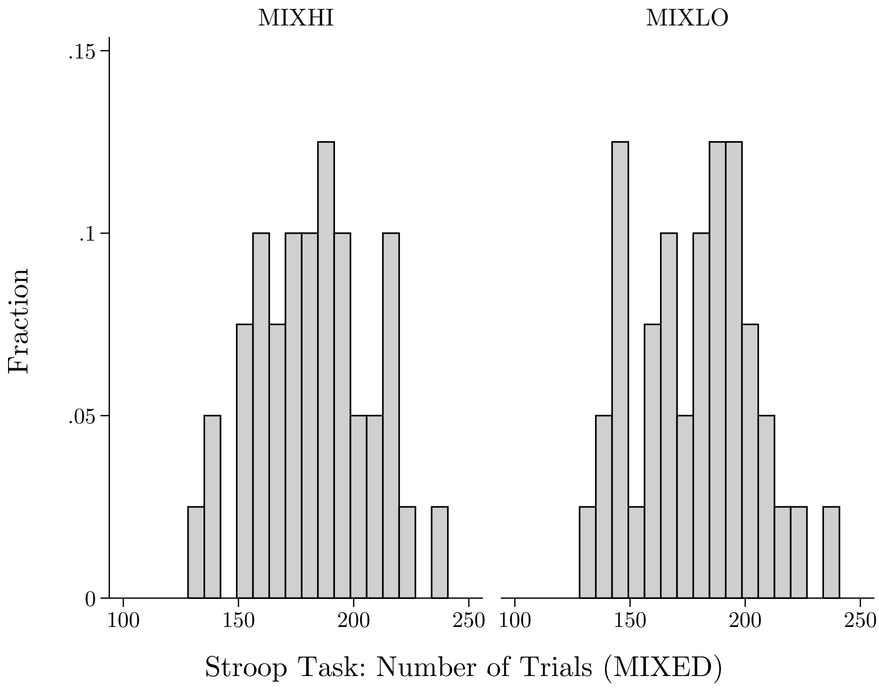


Figure B.5: Stroop Trials in Treatment *MIXED* by Condition

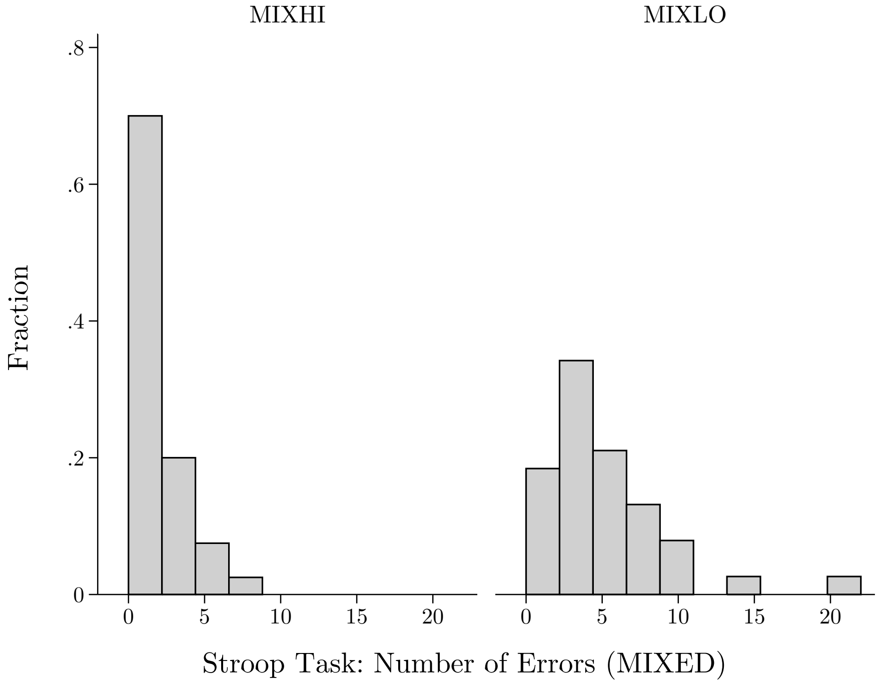


Figure B.6: Errors in the Stroop Task in Treatment *MIXED* by condition[1]

Table B.11: Distribution of Answers in the Stroop Task (MIXED)

Distribution of Answers in the Stroop Task (MIXED)		
HIGHSC	Mean	Standard deviation
Correct Answers	179.225	24.1135
Trials	182.65	24.59784
Errors	2.425	1.448031
LOWSC	Mean	Standard deviation
Correct Answers	164.05	39.93838
Trials	178.3	25.47518
Errors	13.25	36.44367

[1] Two outliers were dropped from this display in the *MIXLO* group, both of whom apparently did not fully understand the task. One had 123 errors and the other had 205 errors.

B.4 Distribution of Subjective Measures

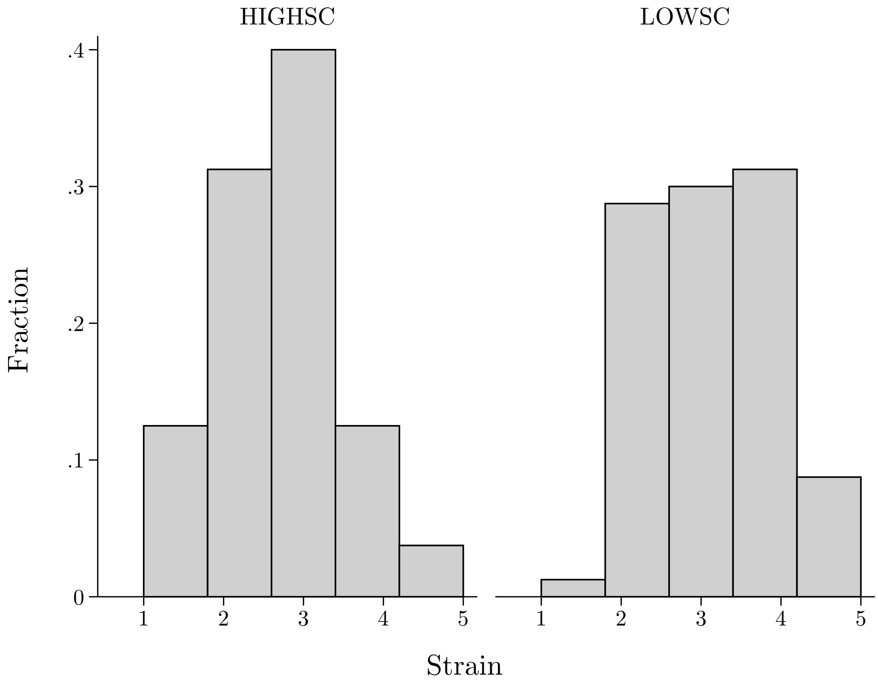


Figure B.7: Strain in *HIGHSC* vs. *LOWSC*

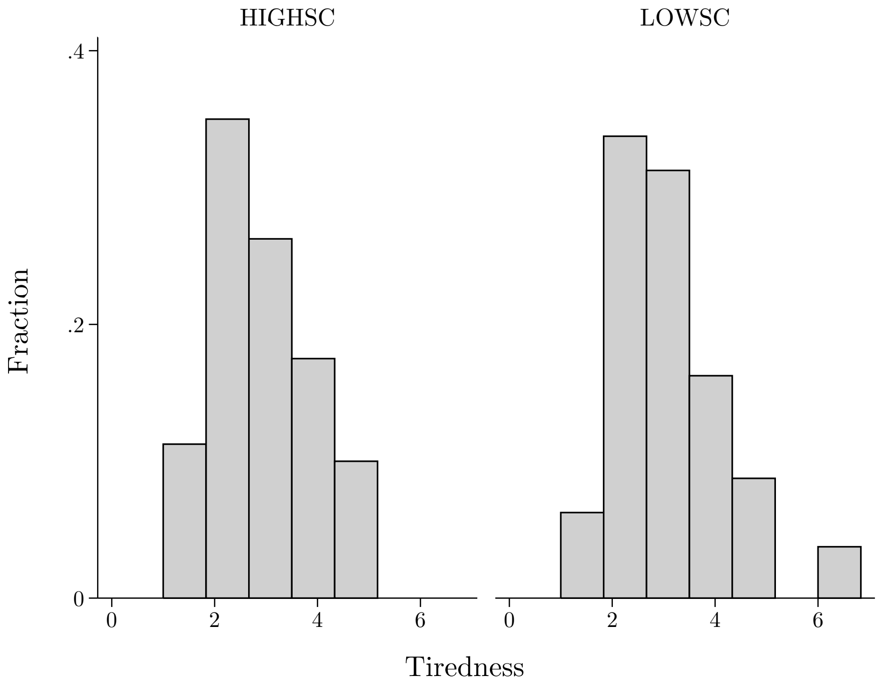


Figure B.8: Tiredness in *HIGHSC* vs. *LOWSC*

Table B.12: Distribution of Subjective Measures

Distribution of Subjective Measures

HIGHSC	Mean	Standard deviation
Strain	2.6375	0.9839696
Tiredness	2.8	1.162712
LOWSC	**Mean**	**Standard deviation**
Strain	3.175	0.9907803
Tiredness	2.9875	1.206457

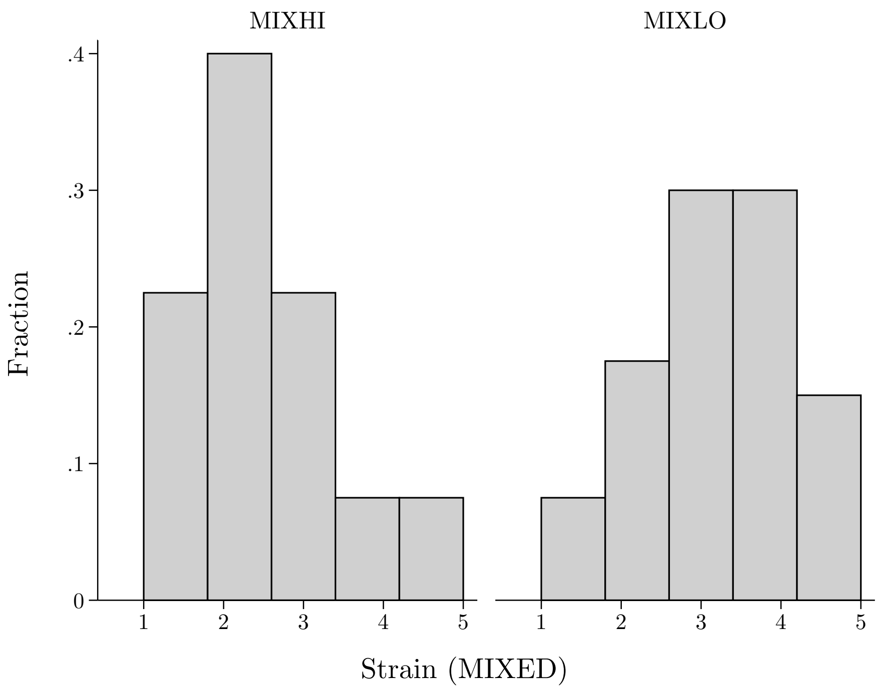


Figure B.9: Strain in Treatment *MIXED* by Condition

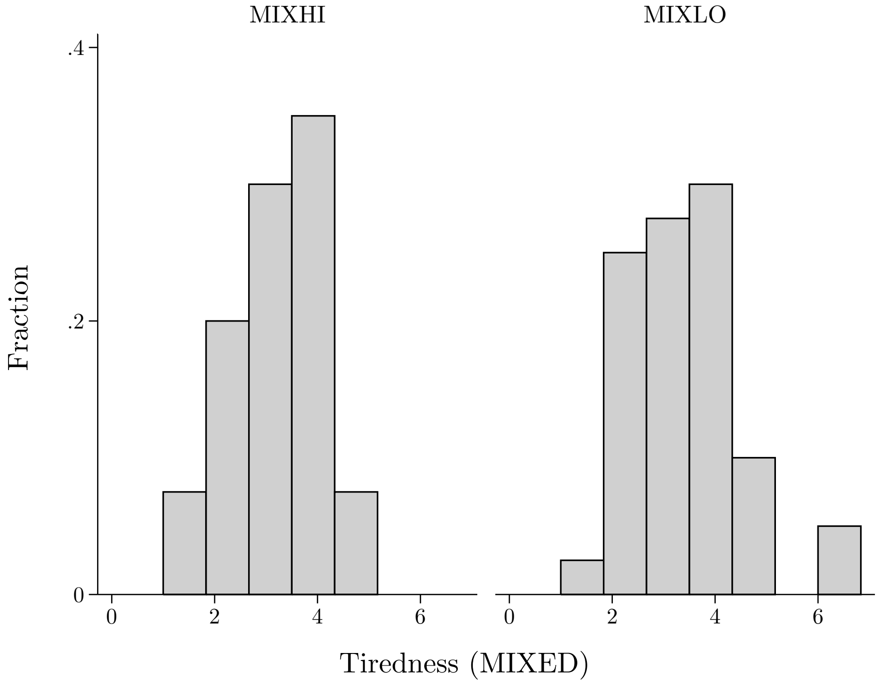


Figure B.10: Tiredness in Treatment *MIXED* by Condition

Table B.13: Distribution of Subjective Measures (MIXED)

Distribution of Subjective Measures (MIXED)

MIXHI	Mean	Standard deviation
Strain	2.375	1.14774
Tiredness	3.15	1.075365
MIXLO	Mean	Standard deviation
Strain	3.275	1.154423
Tiredness	3.35	1.188621

B.5 Distribution of Answers in the Cognitive Reflection Test

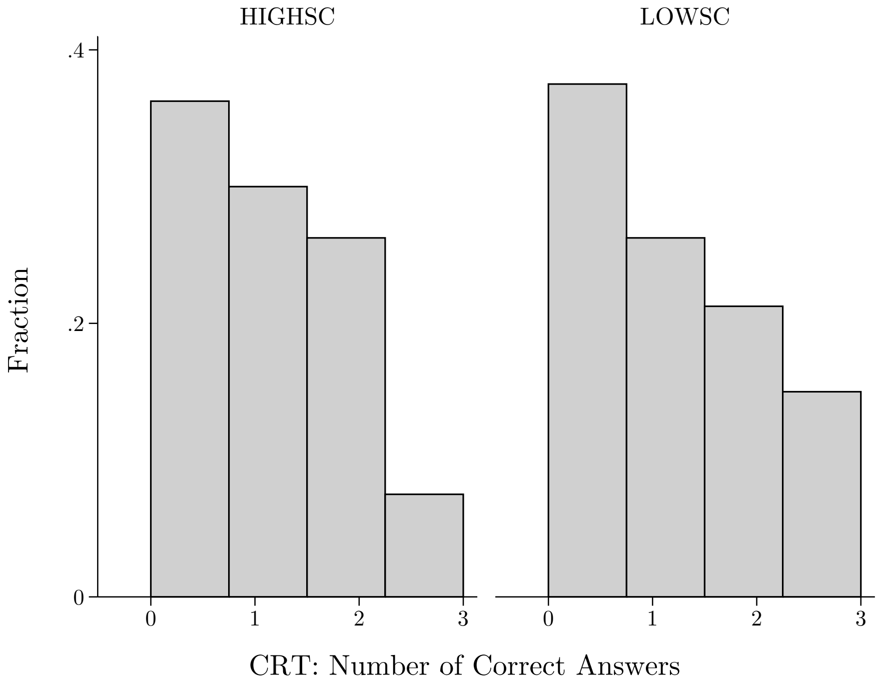


Figure B.11: Correct CRT Answers in *HIGHSC* vs *LOWSC*

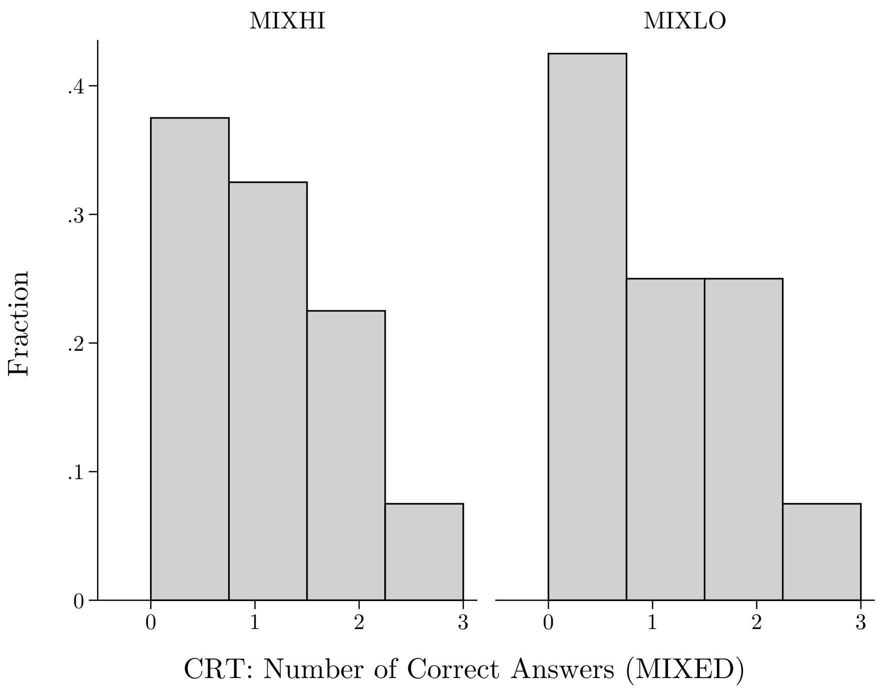


Figure B.12: Correct CRT Answers in *MIXED* by condition

Table B.14: Distribution of Answers in the Cognitive Reflection Test

Distribution of Answers in the Cognitive Reflection Test

	Mean	Standard deviation
HIGHSC	1.05	.9665284
LOWSC	1.1375	1.087836
MIXED	Mean	Standard deviation
MIXHI	1	.9607689
MIXLO	.975	.9996794

B.6 Distribution of Certainty Equivalents

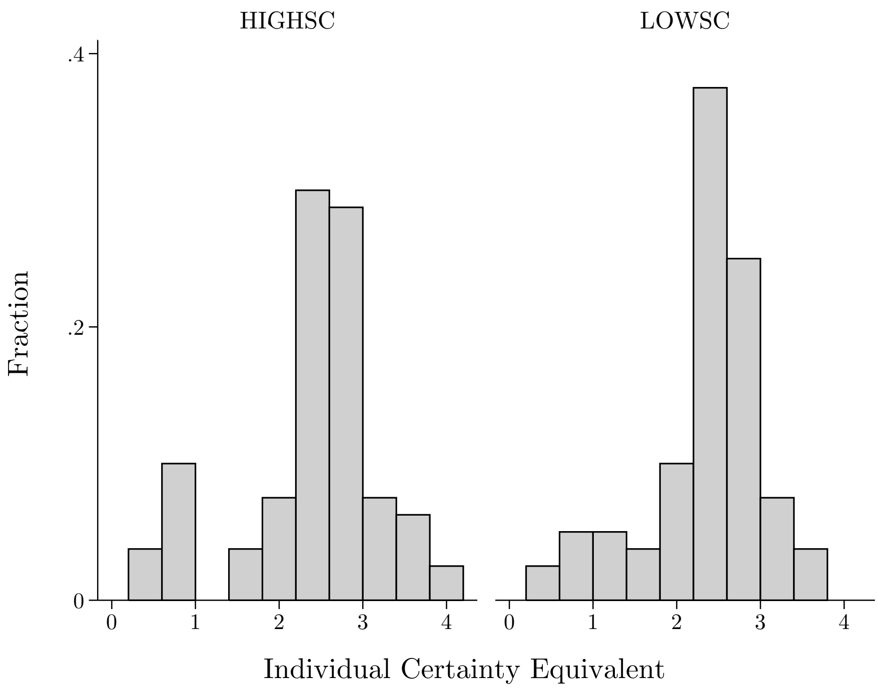


Figure B.13: Individual Certainty Equivalents in *HIGHSC* vs *LOWSC*

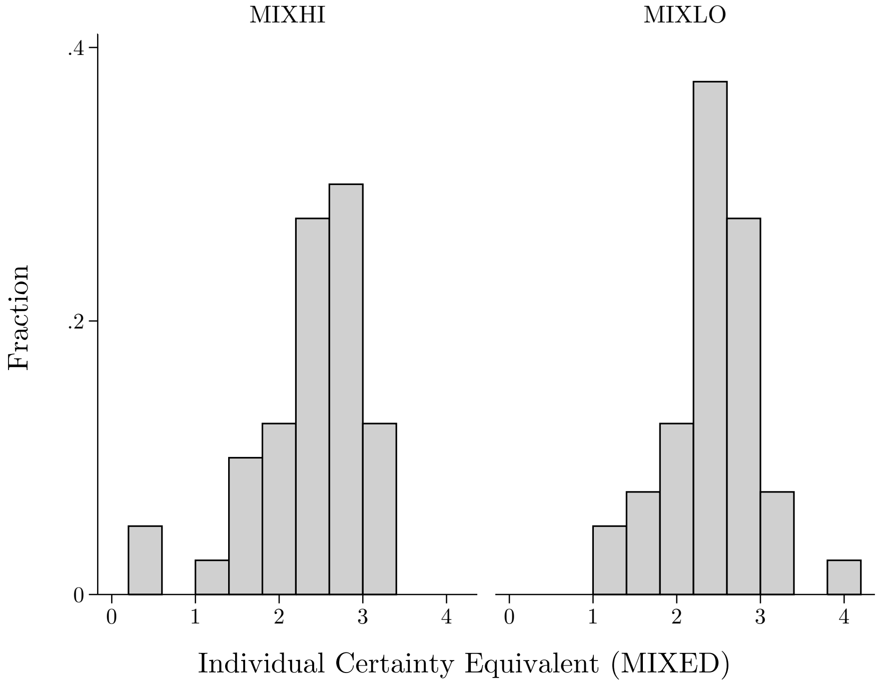


Figure B.14: Individual Certainty Equivalents in *MIXED* by Condition

Table B.15: Distribution of Individual Certainty Equivalents

Distribution of Individual Certainty Equivalents

	Mean	Standard deviation
HIGHSC	2.2	.8467361
LOWSC	2.145	.6964467
MIXED	Mean	Standard deviation
MIXHI	2.16	.6766433
MIXLO	2.24	.5494986

B.7 Instructions

Welcome to the experiment and thank you for your participation!

Please do not talk to other participants of the experiment from now on

General information on the procedure

The purpose of this experiment is to investigate economic decision making. You can earn money during the experiment, which will be paid to you individually and in cash after the experiment has ended.

The whole experiment takes about 1.5 hours and consists of 3 parts. At the beginning you will receive detailed instructions for all parts of the experiment. If you have any questions after reading the instructions or at any time during the experiment please raise your hand. One of the experimenters will then come to you and answer your question in private.

During the experiment, you and the other participants will be asked to make decisions. In some parts, you will interact with other participants. Thus both your own decisions and the decisions of other participants can determine your payoffs. Your payoffs are determined according to the rules which are explained in the following. As long as you can make your decisions, a countdown will be displayed in the upper right corner of the screen which is intended to give you an orientation for how much time you should use to make your choices. In most parts you can exceed the time limit if needed; in some parts, however, you can only act within the time limit (You will be informed about this beforehand). Information screens not requiring any decisions will disappear after the time-out.

Payment

In some parts of the experiment we will not refer points instead of Euros. Points will be converted to Euros at the end of the experiment. You will be informed about the exchange rate at the beginning of the respective part.

For your timely arrival you will receive 4 € additionally to the income earned during the experiment.

Anonymity

We evaluate the data from the experiment only in aggregate and never connect personal information to data from the experiment. At the end of the experiment you have to sign a receipt, which we need for our sponsor. The sponsor does not receive any further data from the experiment.

Aid

On your desk you will find a pen. Please leave it on there after the experiment.

Part I

Task

The first part of the experiment consists of a task that will last 5 minutes. You will see a black screen on which words in different colors will appear. Here you can see an example:

You will be asked to click one of the buttons at the bottom of the screen. You will be asked to choose the button corresponding to the color the word is written in (**not** the word itself). In the example you should click on "yellow".

After clicked a button, the screen disappears and **another word in another color** appears. Please try to solve **as many word/color combinations** as possible within 5 minutes.

After 5 minutes the first part ends automatically and the second part begins.

Payment

You receive 3 € for part I.

Part II

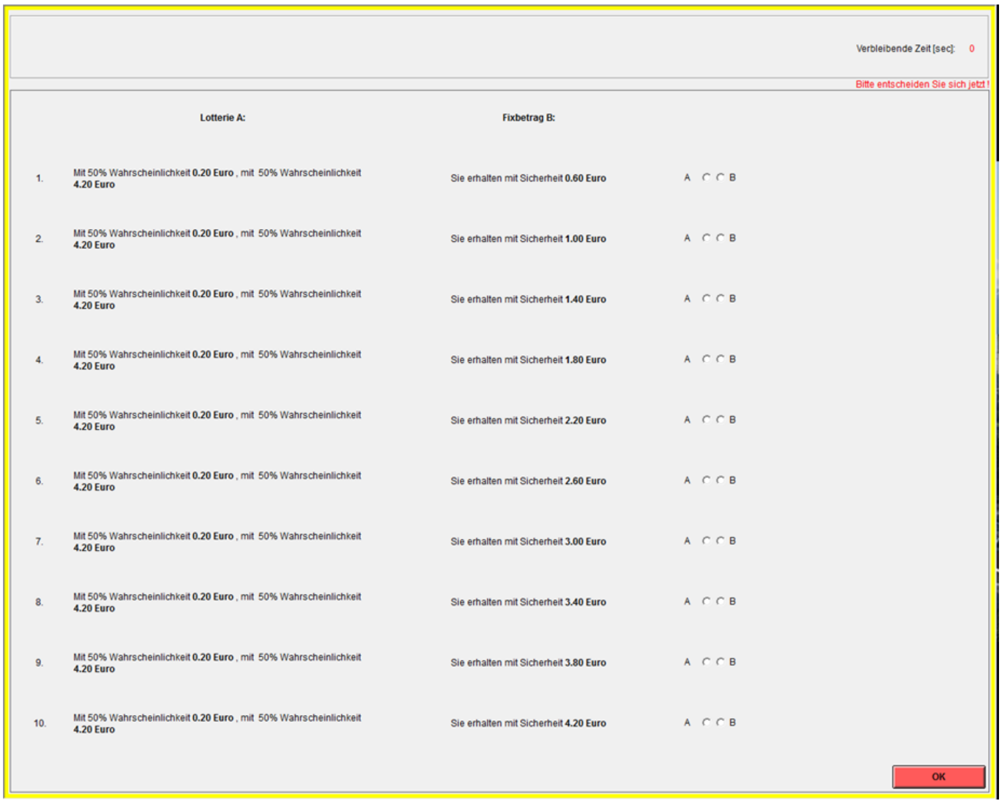
Task

In the second part you first have to answer three questions. For each question answered correctly you receive 0.5 € = 50 Cents.

Afterwards, you will be shown **10 decision problems**. In each of these problems you can choose between **a lottery and a safe amount of money**. The lottery remains unchanged within a period, whereas the safe amount of money increases with every additional decision problem. As the safe amount of money is strictly increasing from row to row, you should stay with the safe amount of money after you have switched to it once.

Your decision is only valid after you have made a choice for each problem and then confirmed it by clicking the OK-button on the bottom right of the screen. Take enough time for your decisions, as your choice – as described in the following – will determine your payoff from this part.

Here you can see what your screen will look like:



Your profit will be determined according to the following rules: First, **the computer chooses randomly and with equal probability one of the ten decision problems for payment**. If you selected the lottery in the relevant problem, the computer will simulate the outcome and you will receive it as payment. If you selected the safe amount in the relevant problem, you will receive it for sure.

For example: Assume the computer randomly chooses the first decision problem and you chose the lottery. Then the computer will simulate the outcomes of this lottery and you either receive 0.2 € (50% probability) or 4.2 € (50% probability).

Payment

The sum of your payoffs from the questions answered correctly at the beginning and your payoff from the decision problem chosen by the computer are your payment for part II of the experiment.

Please note: The computer will directly calculate the result. However, you will only learn about this at the end of the experiments, i.e. how many questions you answered correctly and which decision problem with which outcome the computer selected for you. That information will be presented to you on a separate screen at the end of the experiment.

After the end of part II, part III begins automatically.

Part III

Payment

In the third part of the experiment we refer to points rather than Euros. Points are converted to Euros at the end of the experiment according to the following exchange rate

500 points = 1 Euro (1 point = 0.002 Euros = 0.2 Cents)

Short Description

The third part of the experiment consists of a simulated stock market. The stock market lasts for 10 consecutive periods. Within these periods you can buy or sell shares of a single firm.

At the end of each period for every share that you own you receive either a dividend of 10 points (probability 50%) or 0 points (probability 50%).

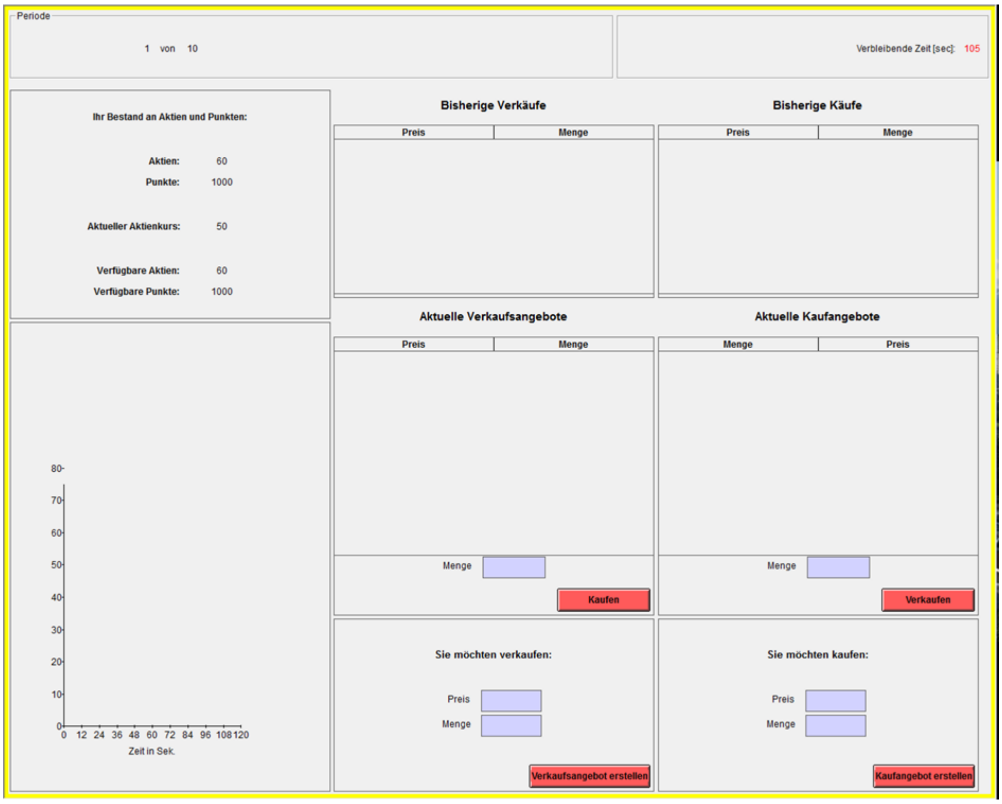
During the 2 minutes trading period you can either offer to sell or buy shares or accept existing buying or selling offers by other participants.

Detailed description: Trading Period

At the beginning of the first trading period you will receive an endowment of shares and points. Every participant receives either 20 shares and 3000 points or 60 shares and 1000 points. The distribution of endowments is random with a 50% probability of receiving each endowment.

Each period lasts exactly 120 seconds (= 2 minutes) and all screens disappear after the time out. You cannot make any trades or offers until he next trading period starts. During a trading period neither your amount of shares nor your amount of points can fall below zero.

During a trading period your screen will look like the following.



In the upper box you see the current period and how much time you have left in the current period. Below it to the left the box displays how many shares you currently own and how large your current wealth is expressed in points. Additionally the current share price and the amount of available shares and points are displayed.

Available shares are those of your shares that you have not offered for sale yet. If you offer to sell shares, you still own them, but they will be subtracted from your account as soon as someone else accepts your offer. Hence, you can only make sale offers that do not exceed your current amount of available shares.

Available points are those of your points that you have not used for buying offers yet. If you make an offer to buy shares, you still own the points, but they will be subtracted from your account as soon as someone else accepts your offer. Hence, you can only make buying offers that do not exceed your current amount of available points.

On the bottom left you can see a graph that shows the evolution of share prices in the current period. On the horizontal axis (the x-axis) you can see the time in seconds at which a trade was made. On the vertical axis (the y-axis) you can see the corresponding price.

In the upper part of the screen you see two lists that have the headlines "Previous Sales" and "Previous Purchases". Here, every trade that you made is listed. For each trade where you bought shares, price and quantity will be listed in "Previous Purchases". For each trade where you sold shares, price and quantity will be listed in "Previous Sales".

Below you find two lists with the headlines "Current Selling Offers" and "Current Buying Offers".

Accepting Selling Offers

In the list "Current Selling Offers" you find price and quantity of each offer, in which a participant offers to sell shares. Your own selling offers will also appear in this list. You can accept every offer in this list (except for your own offers) by marking the corresponding entry in the list, entering the quantity you want to buy into the field "quantity", and then confirming by clicking on the button "Buy". If you accept a selling offer, you will receive the number of shares that you have entered from the seller and the seller receives the corresponding price for each share he sold to you.

Please note: You can also buy less than the number of shares stated in the offer. In that case the offer of the seller will remain on display in the list after the trade, but the number of shares on offer will be reduced by your purchase. Example: A seller makes an offer to sell 10 shares at the price of 60 points each. A buyer buys 6 of those shares. Then an offer to buy 4 shares at the price of 60 points each will continue to be available to all other participants.

Please note that the computer automatically marks the best selling offer (i.e. the one with the lowest price) with a blue bar. You can recognize your own offers, as they are not displayed in black but in blue font.

Accepting offers to buy

In the list “Current Buying Offers” you find price and quantity of each offer, in which a participant offers to buy shares. Your own buying offers will also appear in this list. You can accept every offer in this list (except for you own offers) by marking the corresponding entry in the list, entering the quantity you want to sell into the field “quantity”, and then confirming by clicking on the button “Sell”. If you accept a buying offer, the other participant will receive the number of shares that you entered and you receive the corresponding price for each share you sold.

Please note: You can also sell less than the number of shares the buyer offers to buy. In that case the offer of the buyer will remain on display in the list after the trade, but the number of shares demanded will be reduced by your sale.

Please note that the computer automatically marks the best buying offer (i.e. the one with the highest price) with a blue bar. You can recognize your own offers according to their blue font.

Creating Selling or Buying Offers

In the bottom part of the screen you have the possibility to create your own selling or buying offers. If you want to create an offer to sell, enter the quantity of shares that you want to sell and the price per share which you demand for each unit in the field below “You Want to Sell” . After clicking the button “Create Selling Offer”, your selling offer will show up in the list “Current offers to sell”. Example: You want to sell 10 shares at a price of 55 points per share. Then you enter 10 into the field “Quantity” and 55 into the field “Price”.

If you want to create a buying offer, enter the quantity that you want to buy in the field below “You Want to Buy” and the price per share for which you are willing to buy that quantity. After clicking the button “Make Buying Offer” your offer will show up in the list “Current Buying Offers”. Example: You want to buy 20 shares at a price of 45 points per share. Then you enter 20 into the field “amount” and 45 into the field “price”.

Please note: An offer to buy or to sell that has been made cannot be cancelled. Only if no one accepts an offer during the course of a trading period, it will not be displayed in the next period of trade.

Dividends

After the end of a trading period the following screen displays a summary of the previous period showing you how many shares and points you own, whether a dividend has been paid and if so, how large your overall dividend payments were.

In each period the dividend per share either amount to 10 points (with a probability of 50%) or to 0 points (with a probability of 50%) and is the same for all shares. After the end of period 10, all shares are worthless. All participants learn the realization of the dividend simultaneously on a separate screen at the end of the corresponding period.

The following table displays the value pattern of a share, i.e. the expected value of the remaining dividends. The first column indicates the current period, in the second column you find the number of remaining dividend payments. The third column shows the average expected dividend per share and period. The last column shows the average of remaining dividends per share in the corresponding period.

Current period	Remaining dividend payments	x	Average dividend value per period (0 or 10 with equal probability)	=	Average remaining dividends per share that you own
1	10		5		50
2	9		5		45
3	8		5		40
4	7		5		35
5	6		5		30
6	5		5		25
7	4		5		20
8	3		5		15
9	2		5		10
10	1		5		5

Assume for example that four trading periods remain. As the dividend per share is either 0 or 10 points with a probability of 50% each, this yields an expected dividend of 5 points per share and period. Assume you only own one single share which you intend to hold until the market closes. Then you can expect a total dividend payment for the four remaining periods of '4 remaining periods' x '5 points' = '20 points'.

Payoff

At the end of part III the shares no remaining value. Only your amount of points will be converted to Euros according to the exchange rate stated above of 1 point = 0.002 Euros = 0.2 Cents.

Afterwards, you will see a screen displaying your payoffs from the second part.

In the following, we will ask you to completely and honestly answer some questions concerning your person. On leaving the laboratory, we will pay you your profit privately and in cash. Please remain seated until we call you up in a random order. Please leave the instructions and the pen at your desk and take your numbered seat card with you.

Practice Period

Before you start today's experiment with part I, you will first play a practice period of part III to become familiar with the stock market. The payoff from this practice period will not influence your final payoff. Please note that the realization of the dividend and your endowment are not necessarily identical to the first period of part III as the realization is random and endowments will be randomly assigned.

After completion of the practicing period part I of the experiment begins.

Appendix C

High-Frequency Trading, Maker-Taker Pricing and Behavioral Adjustments: An Experimental Study of Pricing Structures in Fast-Paced Markets

C.1 Instructions

This experiment is used for the investigation of economic decision-making. You can thereby earn money. The money will be paid out privately and in cash after the experiment.

The whole experiment approximately lasts 1 hour and consists of three independent parts. At the beginning of each part you receive detailed instructions. Please raise your hand if you have any questions after reading through the instructions or during the experiment. One of the experimenters will then come to you and answer your questions in private. For reasons of linguistic simplification we will only use male expressions in the instructions, which also apply for female participants.

In the experiment you and other participants of the experiment are asked to make decisions. In some cases you are going to interact with other participants. Your decisions as well as the decisions of other participants determine your earnings from the experiment according to the rules explained below.

While making your decisions, there is a clock counting down in the right upper corner of your computer screen. This clock serves as a guide for how much time it should take you. In most of the cases you may exceed this time frame in case you need additional time for making your decisions. However, in some parts you have to make your decisions within this time limit. Screens, which do not require any input, will be hidden after the time has expired.

Payment

In some parts of the experiment your earnings will not be counted in Euros but in points. At the end of the experiment, these points are converted to Euros at a given exchange rate. The corresponding exchange rate will be given at the beginning of the respective part.

In addition to the money earned during the experiment, you receive 4 Euros for arriving on time.

Anonymity

The data from the experiment are only evaluated on an aggregate level and names are never related to data. At the end of the experiment you have to sign a receipt, which only serves for billing with our sponsor. This sponsor does not receive any further information from the experiment.

Tools

You can find a pen on your table. Please leave it there after the experiment.

Part I

Payment

In this part of the experiment your earnings will not be counted in Euros but in points. At the end of the experiment, these points are converted to Euros at the exchange rate of

$$1000 \text{ Points} = € 1$$

$$(1 \text{ Point} = € 0.001 = 0.1 \text{ Cent})$$

Short description

The first part of the experiment refers to the simulation of a stock market. The stock market exists for 10 consecutive periods, in which you can buy and sell shares of a single company.

After the tenth period you receive the surrender value for each share in your possession. The redemption value may change in every period and will be announced after each trading period.

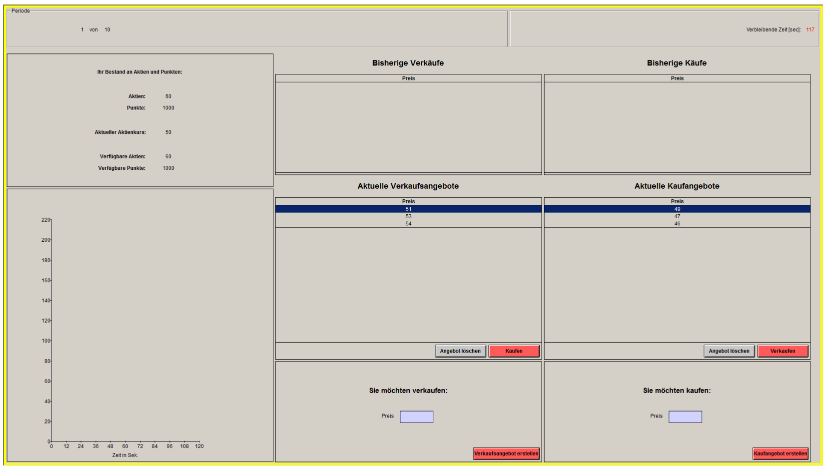
During the trading period of 120 seconds (2 minutes), you can create purchase or sales offers for shares, as well as accept the offers of other participants.

Detailed description: Trading Period

At the beginning of the first period you receive you get a basic set of shares and points. Each participant receives gets either 20 shares and 3000 points or 60 shares and 1000 points. Which set you will receive is a random choice with equal probabilities (50%).

Each period lasts for 120 seconds and all screens will be blanked out after this time has expired. No trade is possible until the beginning of the next period. In none of the trading periods your stock of shares and points can drop below zero.

During trading periods your screen looks like follows:



In the top row you can see in which period you are and how much time is left in the current period. The left-sided block below displays your current amount of shares and points. Additionally, the current stock price and the amount of available shares and points are shown.

Available shares are shares, which you did not offer for sale. If you offer shares for sale, you still own them, but as soon as somebody accepts your offer, they do not longer belong to your account of shares. Therefore, you cannot offer more than your available shares.

Available points are points, which you did not use for purchase offers. If you make a purchase offer, you still own the amount of points you need for the transaction, but as soon as somebody accepts your offer, these are no longer yours. Therefore, you can only create purchase offers that do not exceed your available points.

In the lower left part of the screen, you can see a graph, which displays the course of the share price. On the horizontal axis (x-axis) you see the time at which a trade was completed. On the vertical axis (y-axis) you see the corresponding price.

In the upper part of the screen you see two lists, which are labeled “Bisherige Verkaeufe” (prior sales) and “Bisherige Kaeufe” (prior purchases). These lists display all your completed trades. Each purchase will be listed with the corresponding price in the “Bisherige Kaeufe” list and each sale will be listed with the corresponding price in the “Bisherige Verkaeufe” list.

Below that, you can find two lists named “Aktuelle Verkaufsangebote” (current sales offers) and “Aktuelle Kaufangebote” (current purchase offers).

Accepting Sales Offers

The list “Aktuelle Verkaufsangebote” (current sales offers) contains the prices of all sale offers of all participants. Your own sales offers will also appear in this list. You can accept an offer from that list (except your own offers) by clicking the “Kaufen” (purchase) button. You then receive a share from the seller and the seller receives the corresponding price from you.

Please note that the computer automatically marks the best offer (the offer with the lowest corresponding price) with a blue bar. You can identify your own offers by the blue font color (instead of black).

Accepting Purchase Offers

The list “Aktuelle Kaufangebote” (current purchase offers) contains the prices of all purchase offers of all participants. Your own purchase offers will also appear in this list. You can accept an offer from that list (except your own offers) by clicking the “Verkaufen” (sale) button. The buyer then receives a share from you and you receive the corresponding price from the buyer.

Please note that the computer here again automatically marks the best offer (the offer with the highest corresponding price) with a blue bar. You can identify your own offers by the blue font color (instead of black).

Creating Sales and Purchase Offers

You have the option to create sales and purchase offers in the lower part of the screen.

If you want to create a sales offer, enter the price per share into the box below the "Sie moechten verkaufen" (you want to sale) section. After clicking the "Verkaufsangebot erstellen" (creating a sales offer) button, your offer appears in the "Aktuelle Verkaufsangebote" (current sales offers) list. Example: You want to sell a share for the price of 50. Then you have to enter 50 into the "Preis" (price) box.

If you want to create a purchase offer, enter the price per share into the box below the "Sie moechten kaufen" (you want to purchase) section. After clicking the "Kaufangebot erstellen" (creating a purchase offer) button, your offer appears in the "Aktuelle Kaufangebote" (current purchase offers) list. Example: You want to buy a share for the price of 80. Then you have to enter 80 into the "Preis" (price) box.

Please note: If nobody accepts your offer during a trading period, this offer will not be displayed in the next trading period.

Deleting Offers

You can withdraw an offer as long as nobody accepts it. Therefore, you have to mark your offer (in the "Aktuelle Verkaufsangebote" (current sales offers) or the "Aktuelle Kaufangebote" (current purchase offers) list) and click the "Angebot loeschen" (delete the offer) button directly below the list.

Further characteristics of the stock market

It is possible, that your stock market possesses one (or both) of the following characteristics. First of all, we describe these characteristics. At the end of the instructions you are informed whether these characteristics apply to your stock market or not.

Computerized Traders

In some of the markets, a computer will have the role of an additional trader. The computer will always create sales and purchase offers, but never accept any open offer. The offers of the computer are not marked differently than those of other traders. The computer is programmed as follows: at any time, it will have 3 purchase offers **below** the current stock price and 3 sales offers **above** the current stock price in the offer lists. In doing so, the computer will never offer more than 100 points or less than 0 points for a share. The computer is programmed to decide very quickly (within fractions of a second). Possible discounts and fees (see below) also apply to the computerized trader.

Discounts and Fees In some of the markets, there exists a discount and fee structure. By conclusion of a trade, the trader who created the sales or purchase offer receives an additional point per transaction, whereas the trader who accepted the offer has to pay an additional point per transaction.

Example: Trader A creates a sales offer for 50 points. Trader B accepts the offer. Therefore, trader A has to pay 49 points (50 points − 1 point discount), because he created the offer, and trader B has to pay 51 points (50 points + 1 point fee), because he accepted the open offer.

Estimation Questions After each trading period we will ask you to answer some estimation questions. Please answer these questions as precisely as possible. At the end, the computer randomly and with equal probabilities chooses one period and one question of the corresponding period. Your payment is higher the closer your estimation is to the real value. For a correct estimation you receive 2.00 Euros. The further away from the real value, the lower is your payment. You cannot receive less than 0 Euro for the estimation question.

Redemption Value

Initially, the redemption value is 50 points and is the same for all shares. After each trading period, the redemption value changes. It either increases by 20 points (with a probability of 50%) or decreases by 20 points (with a probability of 50%). However, the stock value cannot drop below 0 points or rise above 100 points. The change of the redemption value is the same for all participants. After the conclusion of the tenth period, you are informed about the final redemption value of the share. You receive this redemption value for each share you own at this point.

Earnings After repurchase, the shares do not have any further value. Your amount of points (consisting of your points account and the points obtained by the repurchase of shares) will be converted into Euros following the exchange rate of 1 Point = 0.001 Euros = 0.1 Cent as stated above. Additionally, you receive the amount from the estimation questions.

Trial Periods Before we are starting with today's experiment, you have to answer some comprehension questions before you have to complete two trial periods to get to know the stock market. The income from the trial periods is not part of your earnings. Please note that the basic set of the trial periods does not have to be identical to the one of the first period of Part I because the realization is random and the initial sets are randomly assigned. The trial periods are identical to Part I with respect to the characteristics of the stock market.

After the conclusion of the trial periods, Part I of the experiment begins.

Characteristics of the stock market for all periods

There are [no] computerized traders in your stock market.

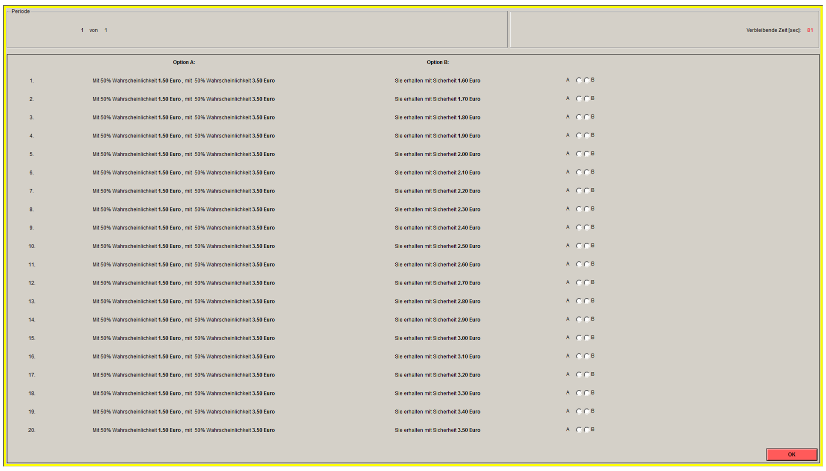
There are [no] discounts and fees in your stock market.

Part II

Exercise

The second part of the experiment consists of 20 decision-making problems. For each of these problems you can choose between a lottery (option A) and a certain amount of money (option B). The lottery is the same for all 20 decision-making problems, but the certain amount of money rises with each decision-making problem.

Specifically, your screen looks like follows:



Option A always pays you 1.50 Euros with a probability of 50% and 3.50 Euros with a probability of 50%. The secured amount of money of option B starts with 1.60 Euros and increases at intervals of 0.10 Euros (= 10 Cent) until it reaches 3.50 Euros for the twentieth decision.

Please choose your preferred option, either option A or option B, for each of the 20 decisions. Please note: since the fixed amount constantly increases, once you have chosen the certain amount of money, you should remain at your decision.

Payment

Your earnings from Part II are determined as follows: first, the computer chooses randomly and with equal probabilities one out of these 20 decision-making problems. If you have chosen the lottery for the corresponding decision, the computer will simulate this lottery and you will receive the result from the lottery. If you have chosen the certain amount of money, you will receive this amount.

Example: Assume that the computer randomly chooses the first decision-making problem and you have chosen the lottery there. Then the computer simulates the lottery and you receive either 1.50 Euros (with a probability of 50%) or 3.50 Euros (with a probability of 50%).

Please note: the computer calculates your earnings from this part of the experiment instantaneously, but you will not be informed about your earnings at this point. You will receive this information on a separate screen at the end of the experiment.

Part III

Payment

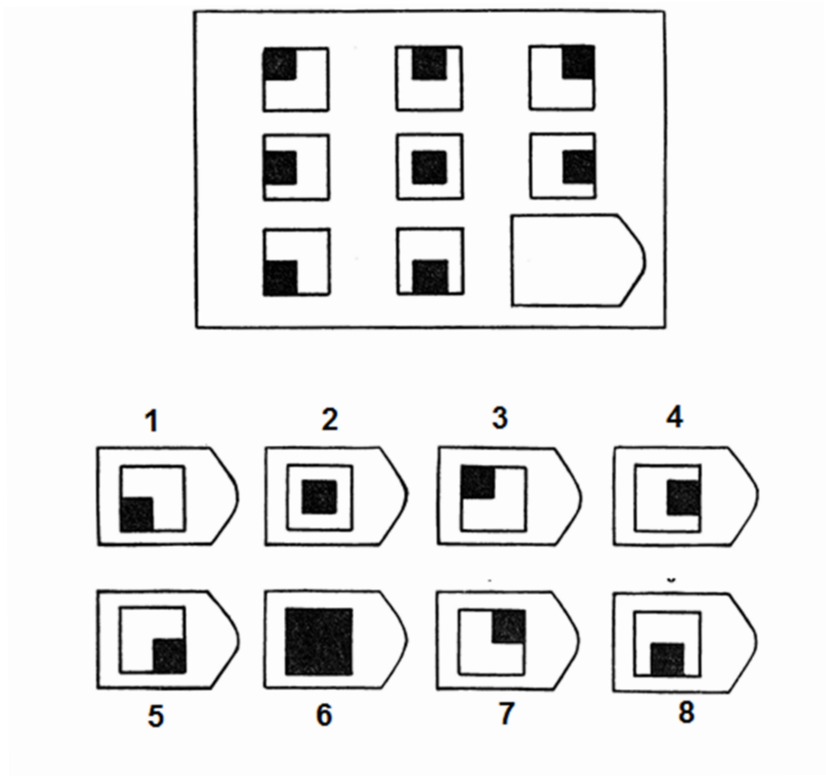
In this part of the experiment your earnings will not be counted in Euros but in points. At the end of the experiment, these points are converted to Euros at the exchange rate of

$$20 \text{ Points} = €\,1$$

$$(1 \text{ Point} = €\,0.05 = 5 \text{ Cent})$$

Description

A framed matrix is displayed on your screen, which is a rectangular arrangement of various symbols. The matrix consists of 3 columns and 3 rows. The symbol in the right lower corners is missing. Below the matrix, there are 8 symbols available for selection, where only one schematically fits into the lower right corner of the matrix. Enclosed you will find an example:



In this example, symbol 5 would be the correct answer.

Your task is to choose the correct symbol. As soon as you confirm your choice, a new matrix appears including 8 new symbols, where again only one fits into the lower right corner of the matrix.

You have to select a symbol for each matrix. Without making a choice you cannot move on to the next matrix. Furthermore, there is no chance to get back to the previous matrix and change your choice once you have confirmed your choice of a symbol.

Overall, you have 5 minutes (300 seconds) to solve as many matrices as possible. The remaining time is presented in the upper right corner of your screen. In the lower left corner of your screen, you can see how many matrices you have solved correctly and wrongly, respectively, so far and whether your last choice was correct or wrong. The matrices' level of difficulty increases over time.

Earnings

You receive 5 points for each correctly solved matrix. For each wrongly solved matrix, 5 points are deducted. Therefore, your total score is:

Total score = 5 x (# correctly solved matrices − # wrongly solved matrices)

If you have solved more matrices wrong than correct, you receive 0 points for this part. Your amount of points will be converted to Euros following the exchange rate stated above.

Following Part III, we ask you to completely and truthfully answer a few questions about personal details. After that, you will be informed about your earnings, which will be paid out in cash afterwards. Please remain seated until we call you for payment.

Appendix D

Shocking Racial Attitudes: The Cultural Legacy of Black G.I.s in Europe

D.1 Alternative Treatment Geography

As an alternative to our main specification, we define a military base as being close to a neighbourhood if the base is within k kilometers of that neighbourhood's population weighted centroid. Using this definition of 'close', our main treatment measure becomes:

$$\text{Black Unit Months}_i = \sum_t \sum_b \mathbb{1}[dist(i, b) < k] \cdot \text{BlackUnits}_{b,t}$$

where $dist(i, b)$ is the euclidean distance (in meters) between the populated weighted centroid of neighbourhood i and the given coordinates of base b. Table D.1 reports regression results with k set to 2 kilometers. The columns follow those of Table 4.2. The results are very similar to that specification. Figure D.1 shows the effect of varying k on the estimate of the coefficient on *BlackUnits*. We run 100 regressions and plot the estimated coefficient on BlackUnitMonths along with a 95% confidence interval. The point estimate decreases with distance – individuals living in areas closer to bases were more likely to have contact with them.

Table D.1: Baseline Regressions using 2 km Radius

	(1)	(2)	(3)	(4)	(5)
Within 2km radius:					
Black unit-months	-0.0113*	-0.0143**	-0.0157***	-0.0153**	-0.0146**
	(-1.95)	(-2.31)	(-2.63)	(-2.28)	(-2.14)
Unit-mnths (100s)	-0.00363	0.0375*	0.0537***	0.0707***	0.0694***
	(-0.14)	(1.89)	(2.68)	(4.10)	(3.91)
F-stat joint significance tests on:					
Grid cells		1.81969e+10	76626.0	502.0	472.7
Industrial Sectors (1931)			0.998	2.006	1.669
Population Density (1931)			6.527	2.117	2.353
Urban Rural status (1931)			0.434	1.299	1.194
Distances			1.768	1.356	1.258
Qualifications (2011)				3.708	3.895
Unemployment (2011)				0.445	0.499
Housing Tenure (2011)				1.267	1.115
Ethnic Backgrounds (2011)					0.470
Clusters	271	271	271	271	271
Observations	42181	42181	42135	42135	42135

Coefficients from OLS regressions using troops within a 2 km radius of the neighbourhood as dependent variable. The unit of observation is the neighbourhood (2011 census output area). Outcome is BNP members per 10,000 white inhabitants. Independent variables are our measure for contact with black troops 'Black unit-months' and any troops 'unit-months' (divided by one hundred) respectively. Standard errors are clustered at the local authority district level. T statistics in brackets. One, two and three stars indicate significance at the 10%, 5% and 1% levels respectively.

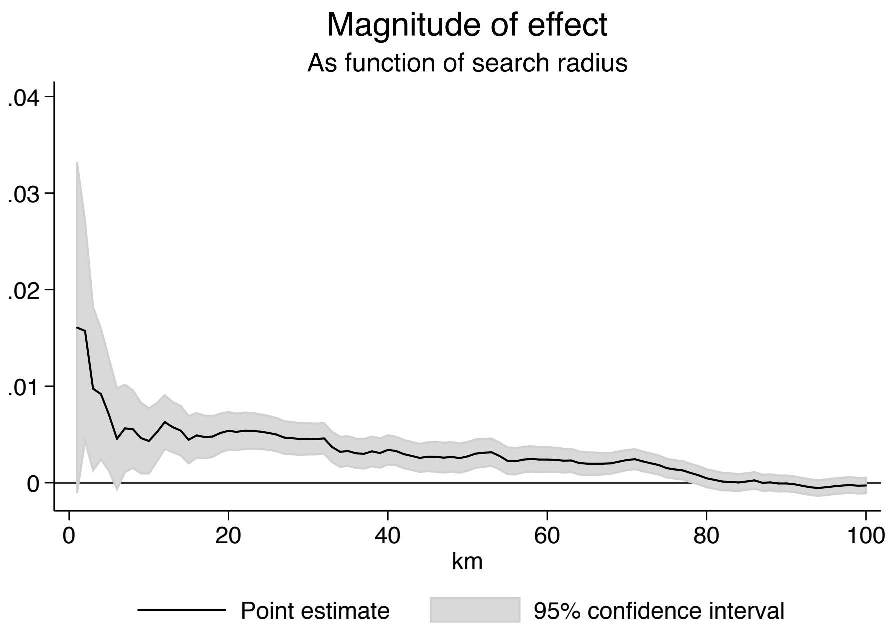


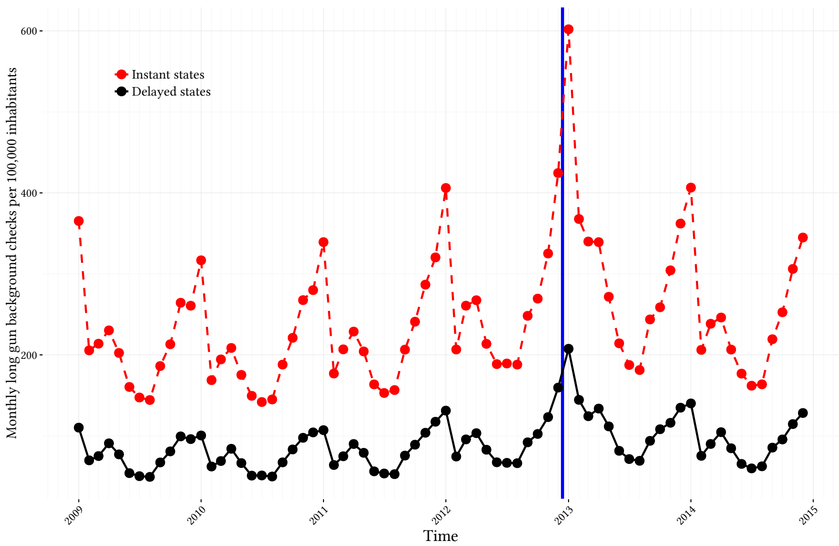
Figure D.1: Varying Search Radius

This figure shows an estimate of our treatment effect varying the radius k in which troops are considered to be close to a neighbourhood.

Appendix E

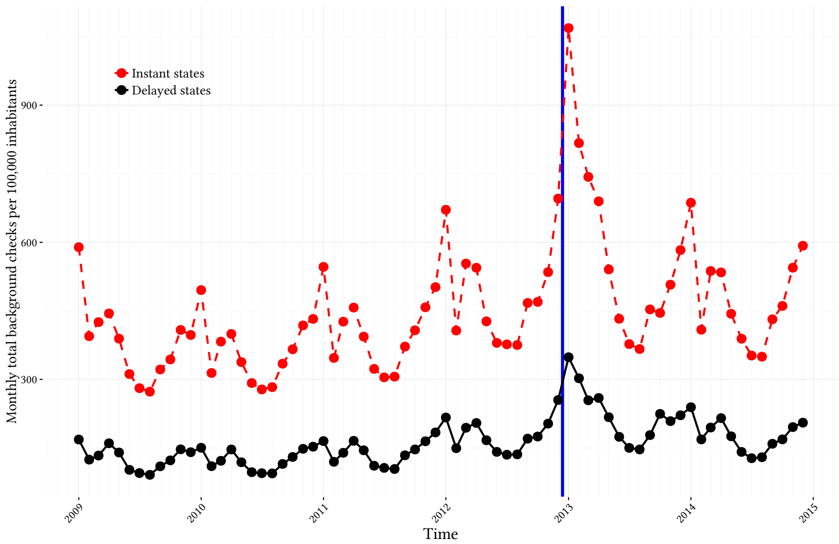
Dynamics in Gun Ownership and Crime – Evidence from the Aftermath of Sandy Hook

E.1 Figures



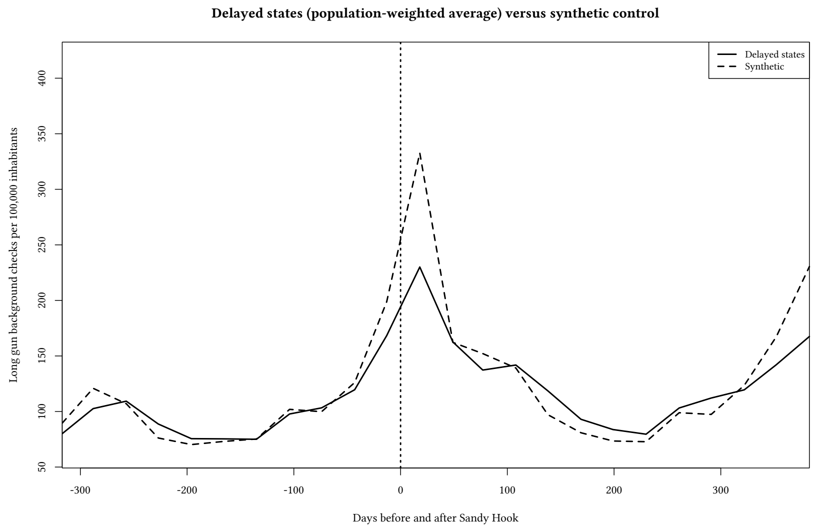
Monthly NICS background checks per 100,000 inhabitants for long guns in delayed states (black) and instant states (red) between 2009 and 2014. The blue vertical line marks the date of the shooting at Sandy Hook Elementary School.

Figure E.1: Background checks for long guns in delayed vs instant states



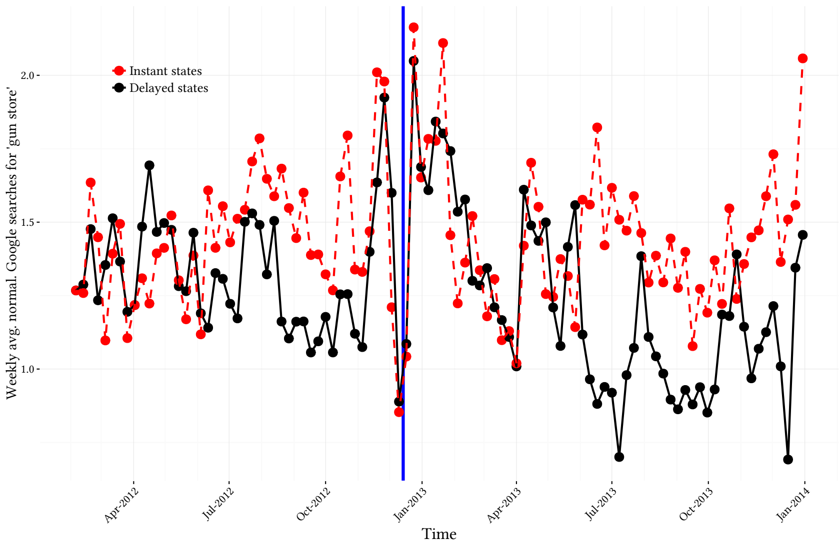
Monthly NICS background checks per 100,000 inhabitants for total guns in delayed states (black) and instant states (red) between 2009 and 2014. The blue vertical line marks the date of the shooting at Sandy Hook Elementary School.

Figure E.2: Background checks for total guns in delayed vs instant states



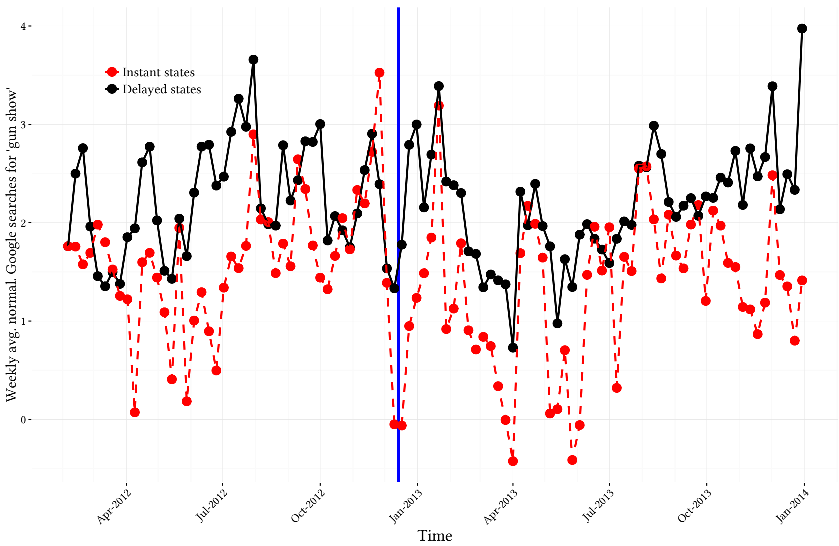
Long gun sales per 100,000 inhabitants in delayed states and a synthetic control of instant states in the months before and after the shooting at Sandy Hook Elementary School. The vertical line indicates the date of the shooting.

Figure E.3: Long gun sales in delayed states and synthetic control states



Weekly averages of daily normalized Google searches for the expression 'gun store' in delayed states (black) and instant states (red) between February 2012 and January 2014. The blue vertical line marks the date of the shooting at Sandy Hook Elementary School.

Figure E.4: Google searches for 'gun store' in delayed vs instant states



Weekly averages of daily normalized Google searches for the expression 'gun show' in delayed states (black) and instant states (red) between February 2012 and January 2014. The blue vertical line marks the date of the shooting at Sandy Hook Elementary School.

Figure E.5: Google searches for 'gun show' in delayed vs instant states

E.2 Tables

Table E.1: Handgun background checks (including CT)

	Monthly handgun sale background checks per 100,000 inhabitants					
	Handgun			Long gun		
	(1)	(2)	(3)	(4)	(5)	(6)
Instant × Post	26.199***	25.485***	101.699***	21.013**	14.515	44.114**
	(8.016)	(9.283)	(28.357)	(8.467)	(10.086)	(21.056)
State FE	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	Y	N	Y	Y
State FE×t	N	N	Y	N	N	Y
States	51	51	51	51	51	51
Observations	1,224	1,224	1,224	1,224	1,224	1,224
Mean DV	162.52	162.52	162.52	188.79	188.79	188.79
R^2	0.882	0.904	0.919	0.890	0.924	0.937

Notes: Coefficients from linear panel regressions of monthly handgun sale background checks per 100,000 inhabitants on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes months after the shooting at Sandy Hook. The level of observation is the state level. We include data from the years 2012 and 2013. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by state population.

Table E.2: Violent crimes (including CT)

	Monthly incidents per 100,000 inhabitants					
	All	Murder	Mansl'ter	Rape	Robbery	Agg. Assault
	(1)	(2)	(3)	(4)	(5)	(6)
Instant × Post	0.979	0.058***	0.004	0.024	0.461	0.436
	(0.787)	(0.020)	(0.005)	(0.169)	(0.518)	(0.406)
County FE	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
County FE×t	Y	Y	Y	Y	Y	Y
Counties	2090	2090	2090	2090	2090	2090
Observations	50,160	50,160	50,160	50,160	50,160	50,160
Mean DV	30.05	0.37	0.01	2.33	9.04	18.3
R^2	0.917	0.363	0.095	0.463	0.945	0.859

Notes: Coefficients from linear panel regressions of crime incidents per 100,000 inhabitants on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes months after the shooting at Sandy Hook. The level of observation is the county level. We include data from the years 2012 and 2013. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by county population.

Table E.3: Background checks with wider time window

	Monthly gun sale background checks per 100,000 inhabitants					
	Handgun			Long gun		
	(1)	(2)	(3)	(4)	(5)	(6)
Instant × Post	37.286***	38.433***	32.789***	32.707***	26.629***	9.703
	(10.356)	(12.700)	(7.127)	(7.758)	(8.193)	(9.183)
State FE	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	Y	N	Y	Y
State FE×t	N	N	Y	N	N	Y
States	50	50	50	50	50	50
Observations	3,600	3,600	3,600	3,000	3,000	3,000
Mean DV	134.34	134.34	134.34	158.47	158.47	158.47
R^2	0.834	0.854	0.876	0.889	0.924	0.931

Notes: Coefficients from linear panel regressions of monthly gun sale background checks per 100,000 inhabitants on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes months after the shooting at Sandy Hook. The level of observation is the state level. We include data from the years 2009 to 2014. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by state population

Table E.4: Violent crimes (varying trend levels)

	Monthly incidents per 100,000 inhabitants								
	Murder			Mansl'ter			Agg. Assault		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Instant × Post	0.055***	0.055***	0.010	0.458	0.458	0.040	0.438	0.438	0.189
	(0.021)	(0.020)	(0.014)	(0.534)	(0.522)	(0.305)	(0.414)	(0.405)	(0.234)
County FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y	Y	Y	Y
County FE×t	Y	N	N	Y	N	N	Y	N	N
State FE×t	N	Y	N	N	Y	N	N	Y	N
Region FE×t	N	N	Y	N	N	Y	N	N	Y
Counties	2084	2084	2084	2084	2084	2084	2084	2084	2084
Observations	50,016	50,016	50,016	50,016	50,016	50,016	50,016	50,016	50,016
Mean DV	0.38	0.38	0.38	9.05	9.05	9.05	18.39	18.39	18.39
R^2	0.363	0.332	0.331	0.945	0.938	0.937	0.859	0.843	0.843

Notes: Coefficients from linear panel regressions of crime incidents per 100,000 inhabitants on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes months after the shooting at Sandy Hook. The level of observation is the county level. We include data from the years 2012 and 2013. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by county population.

Table E.5: Violent crimes (include all agencies)

	Monthly incidents per 100,000 inhabitants					
	Murder		Mansl'ter		Agg. Assault	
	(1)	(2)	(3)	(4)	(5)	(6)
Instant × Post	0.055***	0.040*	0.458	0.401	0.438	0.378
	(0.021)	(0.023)	(0.534)	(0.467)	(0.414)	(0.379)
County FE	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
County FE×t	Y	Y	Y	Y	Y	Y
50% coverage rule	Y	N	Y	N	Y	N
Counties	2084	2453	2084	2453	2084	2453
Observations	50,016	58,872	50,016	58,872	50,016	58,872
Mean DV	0.38	0.36	9.05	8.66	18.39	18.23
R^2	0.363	0.303	0.945	0.903	0.859	0.796

Notes: Coefficients from linear panel regressions of crime incidents per 100,000 inhabitants on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes months after the shooting at Sandy Hook. The level of observation is the county level. We include data from the years 2012 and 2013. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by county population.

E.3 Additional Analyses

E.3.1 Theoretical Motivation: Competing Theories

This section intends to deliver additional arguments for why only present-bias and transaction costs potentially cause our findings. We will thus provide other theories that under some assumptions can explain the effect and argue why these assumptions seem too restrictive.

One candidate to explain our findings would be inattention to transaction costs. If consumers do not pay attention to the transaction costs when making the Google search for 'gun store', but only realize it when they actually buy the gun, we would observe a similar effect. Note however that once consumers arrive at a gun store to make a purchase, the transaction costs of the first trip are sunk and therefore those that intend to buy a gun should also do so. Additionally, when consumers search for 'gun store' in order to buy a gun, they presumably have spent time to deliberate the purchasing decision and therefore also thought about potential transaction costs. We therefore discount inattention to transaction costs as a plausible reason.

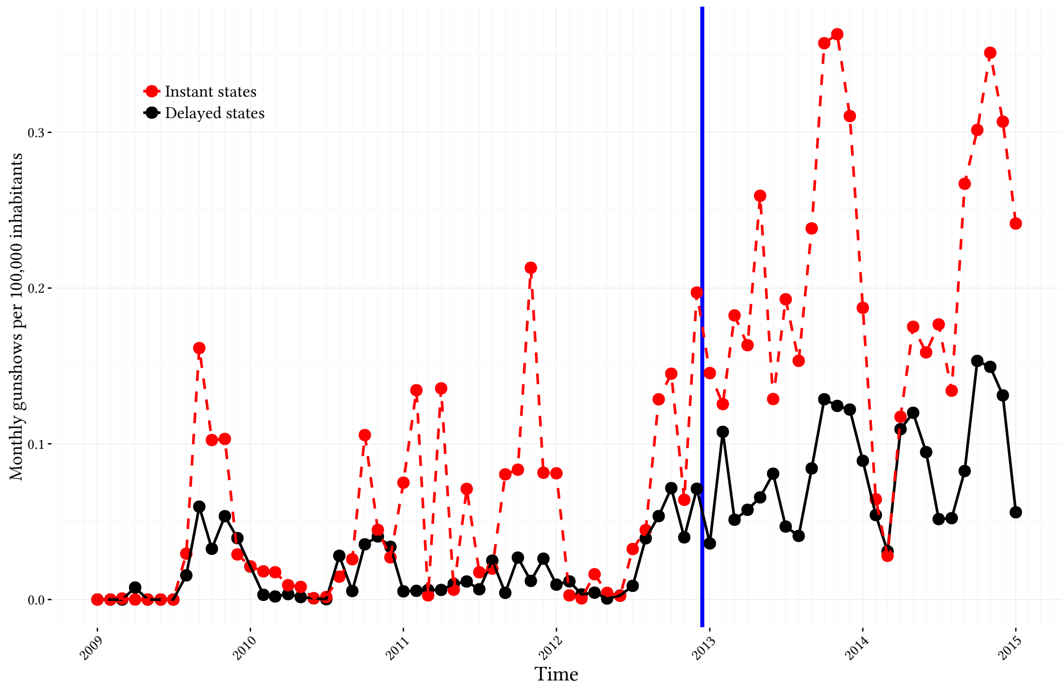
Other theoretical explanations from behavioral economics can also be ruled out to cause our findings. As already stated in the main text, projection bias or different emotional states will not cause the disparity between searches and actual sales, as long as emotions or preferences don't change between the search and the planned visit to the gun store. The reason is that these behavioral biases do not interact with the waiting periods. The decision to buy is made when the background check is conducted at the first visit at the gun store, and not when the firearm is picked up (after the waiting period expired). Our findings also indicate that preference changes between the search and gun store visit should play no role. We expect the shock to affect preferences immediately after the shock (and indeed the largest spike in gun sales can be seen in the month immediately after the shooting). Since there is not much temporal space between the Google searches and the purchases, there is little room for changes in emotional states. We also do not see any differential effect further away from the shooting, which could indicate more thoughtful and less impulsive decision making about purchasing a gun.

E.3.2 Gun Shows: Extensive Margins

One could be concerned that lower demand for firearms in delayed states arises because buyers flock to unregulated gun shows to circumvent the tedious and time-consuming process of purchasing through a federally licensed dealer. As previously noted, the majority of transactions at gun shows is presumably represented in our sample, since many exhibitors are federally licensed (and therefore mandated to perform background checks), many states have regulations for sales at gun shows and the number of guns acquired at gun shows is estimated to be comparatively small. Additionally, we have demonstrated that the demand for gun shows did not tilt towards delayed states.

To show that also the supply of gun shows did not increase stronger in delayed than in instant states, Figure E.6 shows the evolution of gun shows in both types of states graphically.

Table E.6 reports regression results of the number of monthly gun shows per 100,000 inhabitants on the joint effect of being in an instant state after the Sandy Hook shooting, again utilizing our set of controls from previous regressions interacted with month fixed effects. Standard errors are clustered on the state level, we employ month and county fixed effects and a time trend that is allowed to be different in each county. Column 1 reports the results for the years immediately before and after the shooting at Sandy Hook. We expand the time dimension in column 2 by the years 2011 and 2014 respectively, and column 3 looks at all years of the Obama administration. Overall, there does not seem to be an effect of the shooting on the number of gun shows in



Monthly number of gun shows per 100,000 inhabitants in delayed states (black) and instant states (red) between 2009 and 2014. The blue vertical line marks the date of the shooting at Sandy Hook Elementary School.

Figure E.6: Gun shows in delayed vs instant states

delayed vs instant states, as the coefficient is very close to zero and not statistically significant at any conventional level. Only column 3 shows a significant effect, but it suggests that the supply of gun shows increased stronger in instant states after the shooting. Thus we can conclude that the supply of gun shows in delayed states very likely did not increase over instant states.

Table E.6: Regressions of gun shows before and after Sandy Hook

	Monthly gun shows per 100,000 inhabitants		
	2012-2013	2011-2014	2009-2014
	(1)	(2)	(3)
Instant × Post	−0.009	0.008	0.014*
	(0.007)	(0.008)	(0.008)
County FE	Y	Y	Y
Month FE	Y	Y	Y
Controls	Y	Y	Y
County FE×t	Y	Y	Y
Counties	3134	3134	3134
Observations	75,216	150,432	225,648
Mean DV	0.04	0.04	0.03
R^2	0.537	0.408	0.376

Notes: Coefficients from linear panel regressions of gun shows per 100,000 inhabitants on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes months after the shooting at Sandy Hook. The level of observation is the county level. Columns indicate the time frame for which we include data. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, *p<0.1; **p<0.05; ***p<0.01; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by county population.

E.3.3 The 2008 Presidential Election

As an additional robustness check, and because we cannot separate the 2012 Presidential election from the shooting at Sandy Hook, we can re-run our regression of murders using the 2008 Presidential Election as the event that triggered another gun demand shift, without being violent in nature. According to anecdotal evidence, the 2008 election similarly drove up gun sales (CNN, 2008) and this is also confirmed by recent research (Depetris-Chauvin, 2015). We therefore repeat our regression from table 5.2 to see if gun sales were also differentially affected by the 2008 election. The results can be found in Table E.7.

Clearly, the results for both long- and handguns are significant across all our specifications and go in the same direction as in our main specification. Table E.8 then repeats our regression of violent crimes from Table 5.6, but now looks at the years of 2008 and 2009, and takes the 2008 Presidential election as the event date.[1] We observe an effect on murder rates that goes in the same direction as in Table 5.6 and is of similar magnitude, but that is not statistically significant due to larger standard errors. A Z-test (Clogg et al., 1995) comparing our murder estimate in Table 5.6 with our estimate in Table E.8 cannot find statistically significant differences between the two estimates ($p = .9101$). We view this exercise as suggestive that the 2008 Presidential election might have had a similar effect, but that we are unable to correctly identify the effect.

[1] For these years, we use the data by (USDOJ: FBI, 2010; USDOJ: FBI, 2011).

Table E.7: Handgun background checks (2008 election)

	Monthly handgun sale background checks per 100,000 inhabitants					
	Handgun			Long gun		
	(1)	(2)	(3)	(4)	(5)	(6)
Instant × Post	19.515***	18.651***	56.765***	29.780***	27.197***	52.925***
	(5.326)	(6.418)	(15.227)	(5.611)	(6.865)	(16.979)
State FE	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y
Controls	N	Y	Y	N	Y	Y
State FE×t	N	N	Y	N	N	Y
States	50	50	50	50	50	50
Observations	1,200	1,200	1,200	1,200	1,200	1,200
Mean DV	94.55	94.55	94.55	133.88	133.88	133.88
R^2	0.936	0.946	0.955	0.892	0.928	0.932

Notes: Coefficients from linear panel regressions of monthly handgun sale background checks per 100,000 inhabitants on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes months after the 2008 Presidential election. The level of observation is the state level. We include data from the years 2008 and 2009. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by state population.

Table E.8: Violent crimes after 2008 Presidential election

	2008 election: Monthly incidents per 100,000 inhabitants					
	All	Murder	Mansl'ter	Rape	Robbery	Assault
	(1)	(2)	(3)	(4)	(5)	(6)
Instant × Post	1.584	0.050	0.004	−0.188	0.883*	0.839
	(1.457)	(0.039)	(0.004)	(0.139)	(0.529)	(0.991)
County FE	Y	Y	Y	Y	Y	Y
Month FE	Y	Y	Y	Y	Y	Y
Controls	Y	Y	Y	Y	Y	Y
County FE×t	Y	Y	Y	Y	Y	Y
Counties	1975	1975	1975	1975	1975	1975
Observations	47,400	47,400	47,400	47,400	47,400	47,400
Mean DV	35.58	0.42	0.01	2.32	11.45	21.39
R^2	0.931	0.382	0.092	0.456	0.953	0.878

Notes: Coefficients from linear panel regressions of crime incidents per 100,000 inhabitants on our treatment indicator and controls. Instant denotes a dummy for instant states, Post denotes months after the 2008 Presidential election. The level of observation is the county level. We include data from the years 2008 and 2009. Mean DV shows the (weighted) mean of the dependent variable. Standard errors clustered at the state-level in parentheses, $^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$; Controls: Log(population); Area; % Rural; % Blacks; % Hispanics; % Below poverty line (all interacted with Month FE); Regressions are weighted by county population.

Bibliography

Abadie, A., Diamond, A., and Hainmueller, J. (2012). Synthetic control methods for comparative case studies: Estimating the effect of California's tobacco control program. *Journal of the American Statistical Association*, 105(490):493–505.

Abadie, A. and Gardeazabal, J. (2003). The economic costs of conflict: A case study of the Basque country. *American Economic Review*, 93(1):113–132.

Acemoglu, D., Johnson, S., Kermani, A., Kwak, J., and Mitton, T. (2016). The value of connections in turbulent times: Evidence from the United States. *Journal of Financial Economics*, forthcoming:1–24.

Achtziger, A., Alós-Ferrer, C., and Wagner, A. K. (2015). Money, depletion, and prosociality in the dictator game. *Journal of Neuroscience, Psychology, and Economics*, 8(1):1–14.

Achtziger, A., Alós-Ferrer, C., and Wagner, A. K. (2016). The impact of self-control depletion on social preferences in the ultimatum game. *Journal of Economic Psychology*, 53:1–16.

Agerström, J. and Rooth, D.-O. (2011). The role of automatic obesity stereotypes in real hiring discrimination. *Journal of Applied Psychology*, 96(4):790–805.

Akiyama, E., Hanaki, N., and Ishikawa, R. (2013). It is not just confusion! Strategic uncertainty in an experimental asset market. Working Paper.

Al-Ubaydli, O. and List, J. A. (2015). Do natural field experiments afford researchers more or less control than laboratory experiments? *American Economic Review*, 105(5):462–66.

Allais, M. (1953). L'extension des théories de l'équilibre économique général et du rendement social au cas du risque. *Econometrica, Journal of the Econometric Society*, pages 269–290.

Allport, G. (1954). *The Nature of Prejudice*. Addison-Wesley.

Ameriks, J., Caplin, A., and Leahy, J. (2003). Wealth accumulation and the propensity to plan. *Quarterly Journal of Economics*, 118(3):1007–1047.

Ameriks, J., Caplin, A., Leahy, J., and Tyler, T. (2007). Measuring self-control problems. *American Economic Review*, 97(3):966–972.

Andersson, O., Tyran, J.-R., Wengström, E., and Holm, H. J. (2014). Risk aversion relates to cognitive ability: Fact or fiction? Working Paper.

Andrade, E. B., Odean, T., and Lin, S. (2016). Bubbling with excitement: An experiment. *Review of Finance*, 20(2):447–466.

Andrés, A. R. and Hempstead, K. (2011). Gun control and suicide: The impact of state firearm regulations in the United States, 1995–2004. *Health Policy*, 101(1):95–103.

Angrist, J. and Pischke, J. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Army Air Force (1945). *Statistical Digest World War II*. Office of Statistical Control, 2 edition.

Arthur, W. and Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological measurement*, 54(2):394–403.

Ashraf, N., Camerer, C. F., and Loewenstein, G. (2005). Adam Smith, behavioral economist. *Journal of Economic Perspectives*, 19(3):131–145.

Augenblick, N., Niederle, M., and Sprenger, C. (2015). Working over time: Dynamic inconsistency in real effort tasks. *Quarterly Journal of Economics*, 130(3):10670–1115.

Azrael, D., Cook, P. J., and Miller, M. (2004). State and local prevalence of firearms ownership measurement, structure, and trends. *Journal of Quantitative Criminology*, 20(1):43–62.

Baillon, A., Huang, Z., Selim, A., and Wakker, P. (2013). Ambiguity attitude under time pressure. Working Paper.

Baron, M., Brogaard, J., and Kirilenko, A. (2012). The trading profits of high frequency traders. Working Paper.

Baumeister, R. F., Bratslavsky, E., Muraven, M., and Tice, D. M. (1998). Ego depletion: Is the active self a limited resource? *Journal of Personality and Social Psychology*, 74(5):1252–1265.

BBC (2001). BNP: under the skin. http://news.bbc.co.uk/hi/english/static/in_depth/programmes/2001/bnp_special/roots/1984.stm.

BBC (2005). Are you racist? The test that claims to know. http://news.bbc.co.uk/2/hi/uk_news/magazine/4447471.stm.

Benartzi, S., Thaler, R. H., et al. (1995). Myopic loss aversion and the equity premium puzzle. *Quarterly Journal of Economics*, 110(1):73–92.

Benjamin, D. J., Brown, S. A., Shapiro, J. M., et al. (2013). Who is "behavioral"? Cognitive ability and anomalous preferences. *Journal of the European Economic Association*, 11(6):1231–1255.

Beshears, J., Choi, J. J., Harris, C., Laibson, D., C., M. B., and Sakong, J. (2015). Self control and commitment: Can decreasing the liquidity of a savings account increase deposits? Working Paper.

Biais, B. and Foucault, T. (2014). HFT and market quality. *Bankers, Markets & Investors*, 128:5–19.

Biggs, M. and Knauss, S. (2011). Explaining membership in the British National Party: A multilevel analysis of contact and threat. *European Sociological Review*, 28(5):633–646.

Bisin, A. and Verdier, T. (2001). The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298–319.

Bonn, C., Lindner, F., and Schindler, D. (2016). Trading and beliefs in markets with information flows: Does market micro-structure matter? mimeo.

Bors, D. A. and Stokes, T. L. (1998). Raven's advanced progressive matrices: Norms for first-year university students and the development of a short form. *Educational and Psychological Measurement*, 58(3):382–398.

Bosch-Rosa, C., Meissner, T., and Bosch-Domènech, A. (2015). Cognitive bubbles. Working Paper.

Bourke, V. and Porter, D. (2015). The effects of make and take fees in experimental markets. Working Paper.

Brady Campaign (2016). About gun violence. http://www.bradycampaign.org/about-gun-violence.

Brañas-Garza, P., García-Muñoz, T., and González, R. H. (2012). Cognitive effort in the beauty contest game. *Journal of Economic Behavior & Organization*, 83(2):254–260.

Breaban, A. and Noussair, C. N. (2013). Emotional state and market behavior. Working Paper.

Brogaard, J. (2010). High frequency trading and its impact on market quality. Working Paper.

Bruyneel, S., Dewitte, S., Vohs, K. D., and Warlop, L. (2006). Repeated choosing increases susceptibility to affective product features. *International Journal of Research in Marketing*, 23(2):215–225.

Bruyneel, S. D., Dewitte, S., Franses, P. H., and Dekimpe, M. G. (2009). I felt low and my purse feels light: Depleting mood regulation attempts affect risk decision making. *Journal of Behavioral Decision Making*, 22(2):153–170.

Bucciol, A., Houser, D., and Piovesan, M. (2011). Temptation and productivity: A field experiment with children. *Journal of Economic Behavior & Organization*, 78(1):126–136.

Bucciol, A., Houser, D., and Piovesan, M. (2013). Temptation at work. *PloS one*, 8(1):e53713.

Buckert, M., Oechssler, J., and Schwieren, C. (2015). Imitation under stress. Working Paper.

Buckert, M., Schwieren, C., Kudielka, B. M., and Fiebach, C. J. (2014). Acute stress affects risk taking but not ambiguity aversion. *Frontiers in Neuroscience*, 8:82.

Buckton, H. (2006). *Friendly Invasion: Memories of Operation Bolero: the American Occupation of Britain 1942-1945*. Phillimore.

Bureau of Alcohol, Tobacco and Firearms (1999). Gun shows: Brady checks and crime gun traces. Report.

Cappelen, A. W., Nielsen, U. H., Tungodden, B., Tyran, J.-R., and Wengström, E. (2015). Fairness is intuitive. *Experimental Economics*, forthcoming.

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

Cardella, L., Hao, J., and Kalcheva, I. (2015). Make and take fees in the US equity market. Working Paper.

Carter, E. C. and McCullough, M. E. (2013). Is ego depletion too incredible? evidence for the overestimation of the depletion effect. *Behavioral and Brain Sciences*, 36(06):683–684.

Casari, M., Ham, J. C., and Kagel, J. H. (2007). Selection bias, demographic effects, and ability effects in common value auction experiments. *American Economic Review*, 97(4):1278–1304.

Cavallo, E., Galiani, S., Noy, I., and Pantano, J. (2013). Catastrophic natural disasters and economic growth. *Review of Economics and Statistics*, 95(5):1549–1561.

Chetty, R. (2015). Behavioral economics and public policy: A pragmatic perspective. *American Economic Review*, 105(5):1–33.

Chetty, R., Looney, A., and Kroft, K. (2009). Salience and taxation: Theory and evidence. *American Economic Review*, 99(4):1145–1177.

Chicoine, L. E. (2016). Homicides in Mexico and the expiration of the US federal assault weapons ban: a difference-in-discontinuities approach. Working Paper.

Choi, J. J., Laibson, D., and Madrian, B. C. (2011). $100 bills on the sidewalk: Suboptimal investment in 401(k) plans. *Review of Economics and Statistics*, 93(3):748–763.

Claessens, B. J., Van Eerde, W., Rutte, C. G., and Roe, R. A. (2007). A review of the time management literature. *Personnel Review*, 36(2):255–276.

Clogg, C. C., Petkova, E., and Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology*, 100(5):1261–1293.

CNBC (2012). The Sandy Hook effect: Gun sales rise as stocks fall. `http://www.cnbc.com/id/100325110`.

CNN (2008). Gun sales surge after Obama's election. `http://edition.cnn.com/2008/CRIME/11/11/obama.gun.sales/`.

CNN Money (2012). Obama's re-election drives gun sales. `http://money.cnn.com/2012/11/09/news/economy/gun-control-obama/`.

Conte, A., Scarsini, M., and Sürücü, O. (2016). The impact of time limitation: Insights from a queueing experiment. *Judgment and Decision Making*, 11(3):260.

Cook, P. (1978). *The Effect of Gun Availability on Robbery and Robbery Murder: A Cross-Section Study of 50 Cities*. Center for the Study of Justice Policy, Institute of Policy Sciences and Public Affairs, Duke University.

Cook, P. J. and Ludwig, J. (2006). The social costs of gun ownership. *Journal of Public Economics*, 90(1):379–391.

Corgnet, B., Hernán-González, R., Kujal, P., and Porter, D. (2014). The effect of earned versus house money on price bubble formation in experimental asset markets. *Review of Finance*, 19(4):1–34.

Dahl, G. and DellaVigna, S. (2009). Does movie violence increase violent crime? *Quarterly Journal of Economics*, 124(2):677–734.

Das, R., Hanson, J. E., Kephart, J. O., and Tesauro, G. (2001). Agent-human interactions in the continuous double auction. In *International Joint Conference on Artificial Intelligence*, volume 17, pages 1169–1178.

De Haan, T. and Van Veldhuizen, R. (2015). Willpower depletion and framing effects. *Journal of Economic Behavior & Organization*, 117:47–61.

De Luca, M. and Cliff, D. (2011). Human-agent auction interactions: Adaptive-aggressive agents dominate. In *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*, volume 22, page 178.

DellaVigna, S. (2009). Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2):315–372.

Depetris-Chauvin, E. (2015). Fear of Obama: An empirical study of the demand for guns and the US 2008 presidential election. *Journal of Public Economics*, 130:66–79.

Dohmen, T., Falk, A., Huffman, D., and Sunde, U. (2010). Are risk aversion and impatience related to cognitive ability? *American Economic Review*, 100(3):1238 – 1260.

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., and Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the European Economic Association*, 9(3):522–550.

Dube, A., Dube, O., and García-Ponce, O. (2013). Cross-border spillover: US gun laws and violence in Mexico. *American Political Science Review*, 107(03):397–417.

Duflo, E., Kremer, M., and Robinson, J. (2011). Nudging farmers to use fertilizer: Theory and experimental evidence from Kenya. *American Economic Review*, 101(6):2350–2390.

Dufwenberg, M., Lindqvist, T., and Moore, E. (2005). Bubbles and experience: An experiment. *American Economic Review*, 95(5):1731–1737.

Duggan, M. (2001). More guns, more crime. *Journal of Political Economy*, 109(5):1086–1114.

Duggan, M., Hjalmarsson, R., and Jacob, B. A. (2011). The short-term and localized effect of gun shows: Evidence from California and Texas. *Review of Economics and Statistics*, 93(3):786–799.

Dustmann, C. and Preston, I. (2001). Attitudes to ethnic minorities, ethnic context and location decisions. *Economic Journal*, 111(470):353–373.

Eckel, C. C. and Füllbrunn, S. C. (2015). Thar “she” blows? Gender, competition, and bubbles in experimental asset markets. *American Economic Review*, 105(2):906–920.

El Haji, A., Krawczyk, M., Sylwestrzak, M., Zawojska, E., et al. (2016). Time pressure and risk taking in auctions: A field experiment. Working Paper.

Fabio, A., Duell, J., Creppage, K., O’Donnell, K., and Laporte, R. (2016). Gaps continue in firearm surveillance: Evidence from a large US city bureau of police. *Social Medicine*, 10(1):13–21.

Farjam, M. and Kirchkamp, O. (2016). Bubbles in hybrid markets-how expectations about algorithmic trading affect human trading. Working Paper.

Fellner, G. and Maciejovsky, B. (2007). Risk attitude and market behavior: Evidence from experimental asset markets. *Journal of Economic Psychology*, 28(3):338–350.

Fenton-O’Creevy, M., Soane, E., Nicholson, N., and Willman, P. (2011). Thinking, feeling and deciding: The influence of emotions on the decision making and performance of traders. *Journal of Organizational Behavior*, 32(8):1044–1061.

Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, 10(2):171–178.

Fleegler, E. W., Lee, L. K., Monuteaux, M. C., Hemenway, D., and Mannix, R. (2013). Firearm legislation and firearm-related fatalities in the United States. *JAMA Internal Medicine*, 173(9):732–740.

Foucault, T., Kadan, O., and Kandel, E. (2005). Limit order book as a market for liquidity. *Review of Financial Studies*, 18(4):1171–1217.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4):25–42.

Frederick, S., Loewenstein, G., and O’Donoghue, T. (2002). Time discounting and time preference: A critical review. *Journal of Economic Literature*, 40(2):351–401.

Freeman, N. and Muraven, M. (2010). Self-control depletion leads to increased risk taking. *Social Psychological and Personality Science*, 1(2):175–181.

Friehe, T. and Schildberg-Hörisch, H. (2014). Crime and self-control revisited: Disentangling the effect of self-control on risk and social preferences. Working Paper.

Fudenberg, D. and Levine, D. K. (2006). A dual-self model of impulse control. *American Economic Review*, 96(5):1449–1476.

Gailliot, M. T., Gitter, S. A., Baker, M. D., Baumeister, R. F., et al. (2012). Breaking the rules: Low trait or state self-control increases social norm violations. *Psychology*, 3(12):1074–1083.

Gathergood, J. (2012). Self-control, financial literacy and consumer over-indebtedness. *Journal of Economic Psychology*, 33(3):590–602.

Gerhardt, H., Schildberg-Hörisch, H., and Willrodt, J. (2015). Does self-control depletion affect risk attitudes? Working Paper.

Glaeser, E. L. and Glendon, S. (1998). Who owns guns? criminals, victims, and the culture of violence. *American Economic Review*, 88(2):458–462.

Gosling, S. D., Rentfrow, P. J., and Swann, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528.

Greenwald, A. G. and Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological review*, 102(1):4–27.

Greenwald, A. G., McGhee, D. E., and Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. *Journal of Personality and Social Psychology*, 74(6):1464.

Greiner, B. (2015). Subject pool recruitment procedures: Organizing experiments with ORSEE. *Journal of the Economic Science Association*, 1(1):114–125.

Grossklags, J. and Schmidt, C. (2006). Software agents and market (in)efficiency: a human trader experiment. *IEEE transactions on systems, man, and cybernetics, part C (applications and reviews)* , 36(1):56–67.

Hagger, M. S., Wood, C., Stiff, C., and Chatzisarantis, N. L. (2010). Ego depletion and the strength model of self-control: A meta-analysis. *Psychological Bulletin*, 136(4):495–525.

Hahn, R. A., Bilukha, O. O., Crosby, A., Fullilove, M. T., Liberman, A., Moscicki, E. K., and Briss, P. (2003). First reports evaluating the effectiveness of strategies for preventing violence: firearms laws. *Morbidity and Mortality Weekly Report*, 52(RR-14):11–20.

Handa, P. and Schwartz, R. A. (1996). Limit order trading. *Journal of Finance*, 51(5):1835–1861.

Hargreaves Heap, S. and Zizzo, D. (2011). Emotions and chat in a financial markets experiment. Working Paper.

Hefti, A. M., Heinke, S., and Schneider, F. (2016). Mental capabilities, trading styles, and asset market bubbles: theory and experiment. Working Paper.

Hemenway, D. and Miller, M. (2000). Firearm availability and homicide rates across 26 high-income countries. *Journal of Trauma and Acute Care Surgery*, 49(6):985–988.

Hendershott, T., Jones, C. M., and Menkveld, A. J. (2011). Does algorithmic trading improve liquidity? *Journal of Finance*, 66(1):1–33.

Hepburn, L. M. and Hemenway, D. (2004). Firearm availability and homicide: A review of the literature. *Aggression and Violent Behavior*, 9(4):417–440.

Hirshleifer, D. and Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics*, 36(1):337–386.

Hofmann, W., Friese, M., and Strack, F. (2009). Impulse and self-control from a dual-systems perspective. *Perspectives on Psychological Science*, 4(2):162–176.

Huber, J. and Kirchler, M. (2012). The impact of instructions and procedure on reducing confusion and bubbles in experimental asset markets. *Experimental Economics*, 15(1):89–105.

Huffington Post (2013). Gun sales exploded in the year after Newtown shooting. `http://www.huffingtonpost.com/2013/12/06/gun-sales-newtown_n_4394185.html`.

Inzlicht, M. and Schmeichel, B. J. (2012). What is ego depletion? Toward a mechanistic revision of the resource model of self-control. *Perspectives on Psychological Science*, 7(5):450–463.

Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–291.

Kandasamy, N., Hardy, B., Page, L., Schaffner, M., Graggaber, J., Powlson, A. S., Fletcher, P. C., Gurnell, M., and Coates, J. (2014). Cortisol shifts financial risk preferences. *Proceedings of the National Academy of Sciences*, 111(9):3608–3613.

Kaplan, M. S. and Geling, O. (1998). Firearm suicides and homicides in the United States: regional variations and patterns of gun ownership. *Social Science & Medicine*, 46(9):1227–1233.

Kates, D. B. and Polsby, D. D. (2000). Long-term nonrelationship of widespread and increasing firearm availability to homicide in the United States. *Homicide Studies*, 4(2):185–201.

Keynes, J. M. (1936). *The General Theory of Interest, Employment and Money.* Macmillan.

Kirchler, M., Andersson, D., Bonn, C., Johannesson, M., Sørensen, E. Ø., Stefan, M., Tinghög, G., and Västfjäll, D. (2014). The effect of fast and slow decisions on financial risk taking. Working Paper.

Kirchler, M., Bonn, C., Huber, J., and Razen, M. (2015). The "inflow-effect"–trader inflow and price efficiency. *European Economic Review*, 77:1–19.

Kirchler, M., Huber, J., and Stöckl, T. (2012). Thar she bursts: Reducing confusion reduces bubbles. *American Economic Review*, 102(2):865–883.

Kleck, G. (2004). Measures of gun ownership levels for macro-level crime and violence research. *Journal of Research in Crime and Delinquency*, 41(1):3–36.

Kleck, G. (2015). The impact of gun ownership rates on crime rates: A methodological review of the evidence. *Journal of Criminal Justice*, 43(1):40–48.

Kleck, G. and Patterson, E. B. (1993). The impact of gun control and gun ownership levels on violence rates. *Journal of Quantitative Criminology*, 9(3):249–287.

Kleven, H. J., Landais, C., and Saez, E. (2013). Taxation and international migration of superstars: Evidence from the European football market. *American Economic Review*, 103(5):1892–1924.

Kocher, M. G., Myrseth, K. O. R., Martinsson, P., and Wollbrant, C. E. (2016). Strong, bold, and kind: Self-control and cooperation in social dilemmas. *Experimental Economics*, Forthcoming.

Kocher, M. G., Pahlke, J., and Trautmann, S. T. (2013). Tempus fugit: time pressure in risky decisions. *Management Science*, 59(10):2380–2391.

Kocher, M. G. and Sutter, M. (2006). Time is money–time pressure, incentives, and the quality of decision-making. *Journal of Economic Behavior & Organization*, 61(3):375–392.

Kostek, J. and Ashrafioun, L. (2014). Tired winners: The effects of cognitive resources and prior winning on risky decision making. *Journal of Gambling Studies*, 30(2):423–434.

Kotabe, H. P. and Hofmann, W. (2015). On integrating the components of self-control. *Perspectives on Psychological Science*, 10(5):618–638.

Kovandzic, T., Schaffer, M. E., and Kleck, G. (2012). Gun prevalence, homicide rates and causality: A GMM approach to endogeneity bias. In Gadd, D., Karstedt, S., and Messner, S. F., editors, *The Sage Handbook of Criminological Research Methods*. Sage, Thousand Oaks, CA.

Kovandzic, T., Schaffer, M. E., and Kleck, G. (2013). Estimating the causal effect of gun prevalence on homicide rates: A local average treatment effect approach. *Journal of Quantitative Criminology*, 29(4):477–541.

Kuhnen, C. M. and Melzer, B. T. (2014). Non-cognitive abilities and loan delinquency. Working paper.

Lahav, Y. and Meer, S. (2012). The effect of induced mood on prices in asset markets – experimental evidence. Working Paper.

Laibson, D. (1997). Golden eggs and hyperbolic discounting. *Quarterly Journal of Economics*, 112(2):443–478.

Lang, M. (2016). State firearm sales and criminal activity: Evidence from firearm background checks. *Southern Economic Journal*, 83(1):45–68.

Law Center to Prevent Gun Violence (2016). Gun laws by state. `http://smartgunlaws.org/search-gun-law-by-state/`.

Lee, U. (1966). *The Employment of Negro Troops*, volume 8. Government Printing Office.

Lindner, F. and Schindler, D. (2016). Speculative overpricing in asset markets with information flows revisited. mimeo.

Lo, A. W., Repin, D. V., and Steenbarger, B. N. (2005). Fear and greed in financial markets: A clinical study of day-traders. *American Economic Review*, 95(2):352–359.

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

Lott, J. R. (2013). *More guns, less crime: Understanding crime and gun control laws*. University of Chicago Press.

Lott, Jr, J. R. and Mustard, D. B. (1997). Crime, deterrence, and right-to-carry concealed handguns. *Journal of Legal Studies*, 26(1):1–68.

Ludwig, J. and Cook, P. J. (2000). Homicide and suicide rates associated with implementation of the Brady Handgun Violence Prevention Act. *Journal of the American Medical Association*, 284(5):585–591.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: An integrative review. *Psychological Bulletin*, 109(2):163–203.

Madhavan, A. (2012). Exchange-traded funds, market structure, and the flash crash. *Financial Analysts Journal*, 68(4):20–35.

Malinova, K. and Park, A. (2015). Subsidizing liquidity: The impact of make/take fees on market quality. *Journal of Finance*, 70(2):509–536.

Manski, C. F. (2004). Measuring expectations. *Econometrica*, 72(5):1329–1376.

Martinsson, P., Myrseth, K. O. R., and Wollbrant, C. (2014). Social dilemmas: When self-control benefits cooperation. *Journal of Economic Psychology*, 45:213–236.

Maruping, L. M., Venkatesh, V., Thatcher, S. M., and Patel, P. C. (2015). Folding under pressure or rising to the occasion? Perceived time pressure and the moderating role of team temporal leadership. *Academy of Management Journal*, 58(5):1313–1333.

Mas, A. (2006). Pay, reference points, and police performance. *Quarterly Journal of Economics*, 121(3):783–821.

Masatlioglu, Y., Nakajima, D., and Ozdenoren, E. (2011). Revealed willpower. mimeo.

Meier, S. and Sprenger, C. (2010). Present-biased preferences and credit card borrowing. *American Economic Journal: Applied Economics*, 2(1):193–210.

Michailova, J. and Schmidt, U. (2016). Overconfidence and bubbles in experimental asset markets. *Journal of Behavioral Finance*, forthcoming. Working Paper.

Miller, M., Azrael, D., and Hemenway, D. (2002). Rates of household firearm ownership and homicide across US regions and states, 1988-1997. *American Journal of Public Health*, 92(12):1988–1993.

Miller, M., Hemenway, D., and Azrael, D. (2007). State-level homicide victimization rates in the US in relation to survey measures of household firearm ownership, 2001–2003. *Social Science & Medicine*, 64(3):656–664.

Miller, R. M. (2002). Can markets learn to avoid bubbles? *Journal of Psychology and Financial Markets*, 3(1):44–52.

Moody, C. E. (2001). Testing for the effects of concealed weapons laws: specification errors and robustness. *Journal of Law and Economics*, 44(S2):799–813.

Moody, C. E. and Marvell, T. B. (2005). Guns and crime. *Southern Economic Journal*, 71(4):720–736.

Moore, B. (2013). African Americans in the military. In Harris, R. and Terborg-Penn, R., editors, *The Columbia Guide to African American History Since 1939*, Columbia guides to American history and cultures. Columbia University Press.

Muraven, M. (2010). Building self-control strength: Practicing self-control leads to improved self-control performance. *Journal of Experimental Social Psychology*, 46(2):465–468.

Muraven, M. and Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin*, 126(2):247–259.

Muraven, M., Baumeister, R. F., and Tice, D. M. (1999). Longitudinal improvement of self-regulation through practice: Building self-control strength through repeated exercise. *Journal of Social Psychology*, 139(4):446–457.

Muraven, M. and Slessareva, E. (2003). Mechanisms of self-control failure: Motivation and limited resources. *Personality and Social Psychology Bulletin*, 29(7):894–906.

Nadkarni, S. and Herrmann, P. (2010). CEO personality, strategic flexibility, and firm performance: The case of the Indian business process outsourcing industry. *Academy of Management Journal*, 53(5):1050–1073.

Nadler, A., Jiao, P., Alexander, V., Johnson, C., and Zak, P. (2015). Testosterone and trading: A biological driver of asset mispricing. Working Paper.

New York Times (2012). N.R.A. envisions 'a good guy with a gun' in every school. `http://www.nytimes.com/2012/12/22/us/nra-calls-for-armed-guards-at-schools.html`.

New Yorker (2014). The reckoning. `http://www.newyorker.com/magazine/2014/03/17/the-reckoning`.

Noussair, C. N. and Tucker, S. (2013). Experimental research on asset pricing. *Journal of Economic Surveys*, 27(3):554–569.

Noussair, C. N., Tucker, S. J., and Xu, Y. (2014). A futures market reduces bubbles but allows greater profit for more sophisticated traders. Working Paper.

NRA (2016). Institute for Legislative Action. `https://www.nraila.org/`.

Nunn, N. and Wantchekon, L. (2011). The slave trade and the origins of mistrust in Africa. *American Economic Review*, 101(7):3221–3252.

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

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics*, 116(2):257–270.

Olson, L. (2010). *Citizens of London: The Americans Who Stood with Britain in Its Darkest, Finest Hour*. Random House Publishing Group.

Ozdenoren, E., Salant, S. W., and Silverman, D. (2012). Willpower and the optimal control of visceral urges. *Journal of the European Economic Association*, 10(2):342–368.

Palan, S. (2013). A review of bubbles and crashes in experimental asset markets. *Journal of Economic Surveys*, 27(3):570–588.

Palfrey, T. R. and Wang, S. W. (2012). Speculative overpricing in asset markets with information flows. *Econometrica*, 80(5):1937–1976.

Petersen, G.-K., Spickers, T., Glaser, M., and Brodbeck, F. C. (2015). How private investors' stress influences investor behavior and financial markets. Presentation at Experimental Finance Conference.

Pettigrew, T. F. (1998). Intergroup contact theory. *Annual Review of Psychology*, 49(1):65–85.

Pettigrew, T. F. and Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5):751–783.

Pogue, F. (1954). *The Supreme Command*. United States Army in World War II: European Theater of Operations. Office of the Chief of Military History, Department of the Army.

Popov, I. (2016). The real effects of policy deliberation: Evidence from gun parts during gun control debate. mimeo.

Porter, D. P. and Smith, V. L. (1995). Futures contracting and dividend uncertainty in experimental asset markets. *Journal of Business*, 68(4):509–541.

Rash, F. (1942). Internal Memo, Headquarters, Services of Supply, European Theatre of Operations. National Archives reference: NAII; 290; 57/17/087; Box 43, File 218, “NEGRO TROOPS”.

Raven, J. C. and Court, J. H. (1998). *Manual for Raven progressive matrices and vocabulary scales*. Psychological Corporation.

Reynolds, D. (1995). *Rich Relations: the American Occupation of Britain, 1942-1945*. Random House.

Reynolds, D. (2006). *From World War to Cold War: Churchill, Roosevelt, and the International History of the 1940s*. OUP Oxford.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological monographs: General and applied*, 80(1):1.

Rudolph, K. E., Stuart, E. A., Vernick, J. S., and Webster, D. W. (2015). Association between Connecticut’s permit-to-purchase handgun law and homicides. *American Journal of Public Health*, 105(8):e49–e54.

Ruppenthal, R. (1978). *Logistical Support of the Armies: May 1941 – September 1944*. Number 1 in The European Theater of Operations: United States Army in World War 2. Department of the Army.

Sasser, C. (2014). *Two Fronts, One War*. Frontline Books.

Schmeichel, B. J., Vohs, K. D., and Baumeister, R. F. (2003). Intellectual performance and ego depletion: Role of the self in logical reasoning and other information processing. *Journal of Personality and Social Psychology*, 85(1):33–46.

Schmeichel, B. J. and Zell, A. (2007). Trait self-control predicts performance on behavioral tests of self-control. *Journal of Personality*, 75(4):743–756.

Schwarzer, R. and Jerusalem, M. (1995). Generalized self-efficacy scale. In Weinman, J., Wright, S., and Johnston, M., editors, *Measures in health psychology: A user’s portfolio. Causal and control beliefs*, pages 35–37. NFER-NELSON.

Scott, S. L. and Varian, H. R. (2014). Bayesian variable selection for nowcasting economic time series. In *Economic Analysis of the Digital Economy*, NBER Chapters, pages 119–135. National Bureau of Economic Research, Inc.

Securities and Exchange Commission (SEC) (2015). Maker-taker fees on equities exchanges. Memorandum.

Siegel, M., Ross, C. S., and King III, C. (2013). The relationship between gun ownership and firearm homicide rates in the United States, 1981–2010. *American Journal of Public Health*, 103(11):2098–2105.

Smith, G. (1987). *When Jim Crow Met John Bull: Black American Soldiers in World War II Britain*. I.B. Tauris.

Smith, V. L., Suchanek, G. L., and Williams, A. W. (1988). Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica*, 56(5):1119–1151.

Sorenson, S. B. and Berk, R. A. (2001). Handgun sales, beer sales, and youth homicide, California, 1972-1993. *Journal of Public Health Policy*, 22(2):182–197.

Spiliopoulos, L. and Ortmann, A. (2015). The BCD of response time analysis in experimental economics. Working Paper.

Sprenger, C. (2015). Judging experimental evidence on dynamic inconsistency. *American Economic Review*, 105(5):280–285.

Stanley, D. A., Sokol-Hessner, P., Banaji, M. R., and Phelps, E. A. (2011). Implicit race attitudes predict trustworthiness judgments and economic trust decisions. *Proceedings of the National Academy of Sciences*, 108(19):7710–7715.

Stöckl, T., Huber, J., and Kirchler, M. (2010). Bubble measures in experimental asset markets. *Experimental Economics*, 13(3):284–298.

Stojić, H., Anreiter, M. R., and Martinez, J. A. C. (2013). An experimental test of the dual self model. Working Paper.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6):643–662.

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

Sugden, R., Wang, M., and Zizzo, D. J. (2015). Take it or leave it: Experimental evidence on the effect of time-limited offers on consumer behaviour. Working Paper.

Sutter, M., Kocher, M., and Strauß, S. (2003). Bargaining under time pressure in an experimental ultimatum game. *Economics Letters*, 81(3):341–347.

Tangney, J. P., Baumeister, R. F., and Boone, A. L. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality*, 72(2):271–324.

Targonski, J. R. (2011). *A comparison of imputation methodologies in the offenses-known Uniform Crime Reports*. PhD thesis, University of Illinois at Chicago.

Thaler, R. H. and Shefrin, H. M. (1981). An economic theory of self-control. *Journal of Political Economy*, 89(2):392–406.

The Brady Campaign (2013). 2013 State Scorecard – why gun laws matter. `http://www.bradycampaign.org/sites/default/files/SCGLM-Final10-spreads-points.pdf`.

The Guardian (2009). Racism inquiry may bring ban on BNP membership for teachers. `http://www.theguardian.com/education/2009/sep/30/ed-balls-orders-racism-inquiry`.

The Intercept (2015). Gun industry executives say mass shootings are good for business. `https://theintercept.com/2015/12/03/mass-shooting-wall-st/`.

The Spectator (2009). The rise of British racism may be horribly close. `http://www.spectator.co.uk/features/3648608/the-rise-of-british-racism-may-be-horribly-close/`.

Tice, D. M. and Bratslavsky, E. (2000). Giving in to feel good: The place of emotion regulation in the context of general self-control. *Psychological Inquiry*, 11(3):149–159.

Tinghög, G., Andersson, D., Bonn, C., Böttiger, H., Josephson, C., Lundgren, G., Västfjäll, D., Kirchler, M., and Johannesson, M. (2013). Intuition and cooperation reconsidered. *Nature*, 498(7452):E1–E2.

Toplak, M. E., West, R. F., and Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39(7):1275–1289.

Trautmann, S. T. (2014). Risk taking under stress: the role(s) of self-selection. a comment on buckert et al.(2014). *Frontiers in Neuroscience*, 8.

Trautmann, S. T. and Kuilen, G. v. d. (2015). Belief elicitation: A horse race among truth serums. *Economic Journal*, 125(589):2116–2135.

Tversky, A. and Kahneman, D. (1992). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323.

Uhlmann, E. L., Greenwald, A., Poehlmann, A., and Banaji, M. (2009). Understanding and using the implicit association test: III. meta-analysis of predictive validity. *Journal of Personality and Social Psychology*, 97(1):17–41.

Unger, A. and Stahlberg, D. (2011). Ego-depletion and risk behavior: Too exhausted to take a risk. *Social Psychology*, 42(1):28–38.

United States Department of Justice. Federal Bureau of Investigation (USDOJ: FBI) (2010). Uniform crime reporting program data: Offenses known and clearances by arrest, 2008. ICPSR27648-v1. Inter-University Consortium for Political and Social Research. `http://doi.org/10.3886/ICPSR27648.v1`.

United States Department of Justice. Federal Bureau of Investigation (USDOJ: FBI) (2011). Uniform crime reporting program data: Offenses known and clearances by arrest, 2009. ICPSR30766-v1. Inter-University Consortium for Political and Social Research. `http://doi.org/10.3886/ICPSR30766.v1`.

United States Department of Justice. Federal Bureau of Investigation (USDOJ: FBI) (2012). Uniform crime reporting program data: Offenses known and clearances by arrest, 2010. ICPSR33526-v1. Inter-University Consortium for Political and Social Research. `http://doi.org/10.3886/ICPSR33526.v1`.

United States Department of Justice. Federal Bureau of Investigation (USDOJ: FBI) (2013). Uniform crime reporting program data: Offenses known and clearances by arrest, 2011. ICPSR34586-v1. Inter-University Consortium for Political and Social Research. `http://doi.org/10.3886/ICPSR34586.v1`.

United States Department of Justice. Federal Bureau of Investigation (USDOJ: FBI) (2014a). Uniform crime reporting program data: Offenses known and clearances by arrest, 2012. ICPSR35021-v1. Inter-University Consortium for Political and Social Research. `http://doi.org/10.3886/ICPSR35021.v1`.

United States Department of Justice. Federal Bureau of Investigation (USDOJ: FBI) (2014b). Uniform crime reporting program data: Supplementary homicide reports, 2012. ICPSR35023-v1. Inter-University Consortium for Political and Social Research. `http://doi.org/10.3886/ICPSR35023.v1`.

United States Department of Justice. Federal Bureau of Investigation (USDOJ: FBI) (2015a). Uniform crime reporting program data: Offenses known and clearances by arrest, 2013. ICPSR36122-v1. Inter-University Consortium for Political and Social Research. `http://doi.org/10.3886/ICPSR36122.v1`.

United States Department of Justice. Federal Bureau of Investigation (USDOJ: FBI) (2015b). Uniform crime reporting program data: Supplementary homicide reports, 2013. ICPSR36124-v1. Inter-University Consortium for Political and Social Research. `http://doi.org/10.3886/ICPSR36124.v1`.

Vohs, K. D., Baumeister, R. F., and Schmeichel, B. J. (2012). Motivation, personal beliefs, and limited resources all contribute to self-control. *Journal of Experimental Social Psychology*, 48(4):943–947.

Voigtländer, N. and Voth, H.-J. (2012). Persecution perpetuated: The medieval origins of anti-semitic violence in Nazi Germany. *Quarterly Journal of Economics*, 127(3):1339–1392.

Vox (2016). What happens after a mass shooting? Americans buy more guns. `http://www.vox.com/2016/6/15/11936494/after-mass-shooting-americans-buy-more-guns`.

Waddell, S. (2010). *United States Army Logistics: From the American Revolution to 9/11.* PSI reports. Praeger Security International.

Wall Street Journal (2010). How a trading algorithm went awry. `http://www.wsj.com/articles/SB10001424052748704029304575526390131916792`.

Washington Post (2015). Clinton's claim that 40 percent of guns are sold at gun shows and over the internet. `https://www.washingtonpost.com/news/fact-checker/wp/2015/10/16/clintons-claim-that-40-percent-of-guns-are-sold-at-gun-shows-and-over-the-internet/`.

Xu, J. (2015). Optimal strategies of high frequency traders. Working Paper.

Xu, K., Nosek, B., and Greenwald, A. G. (2014). Psychology data from the race implicit association test on the project implicit demo website. *Journal of Open Psychology Data*, 2(1):e3.